AVALIAÇÃO DE CRITÉRIOS DE SELEÇÃO DE MODELOS PARA O MODELO DE REGRESSÃO BETA

SILVIA TERESA FREIRE TORRES

Orientador: Prof[a]. Dr[a]. Viviana Giampaoli
Área de concentração: Estatística Aplicada

Dissertação submetida como requerimento parcial para obtenção do grau de Mestre em Estatística pela Universidade Federal de Pernambuco

Recife, fevereiro de 2005

Universidade Federal de Pernambuco
Mestrado em Estatística

28 de fevereiro de 2005
(data)

Nós recomendamos que a dissertação de mestrado de autoria de

Silvia Teresa Freire Torres

intitulada

Avaliação de criterios de seleção de modelos para o modelo de Regressão Beta

seja aceita como cumprimento parcial dos requerimentos para o grau de Mestre em Estatística.

Klaus Leite Pinto Vasconcellos
Coordenador da Pós-Graduação em Estatística

Prof. Klaus L. P. Vasconcellos
Coordenador da Pós-graduação em Estatística, UFPE

Banca Examinadora:

Viviana Giampaoli
Viviana Giampaoli — orientador

Manoel Jose Machado Soares Lemos
Manoel Jose Machado Soares Lemos

Maria Cristina Falcão Raposo
Maria Cristina Falcão Raposo

Este documento será anexado à versão final da dissertação.

À minha filha, Marcela e
à minha sobrinha, Carol.

Agradecimentos

A Deus e aos espíritos de luz que me deram forças para vencer mais uma etapa da vida.

À minha mãe, Maria José, pelo amor incondicional e pelo esforço para me educar.

À Marcela, minha filha, por todo amor, ajuda, incentivo, paciência e compreensão.

À minha sobrinha, Carol, pela ajuda e pelo seu amor por mim.

Ao meu esposo, Marcos Torres, por seu amor, estímulo, compreensão e paciência.

À minha orientadora, Viviana Giampaoli, pelo estímulo, orientação, dedicação, amizade e paciência.

À professora Cristina Raposo, pelas idéias e sugestões que contribuíram para o desenvolvimento deste trabalho e amizade.

Ao professor Francisco Cribari-Neto pelo incentivo e confiança.

Aos professores do Programa de Mestrado em Estatística, por suas contribuições e por sua ajuda.

À Valéria Bittencourt, pela competência, eficiência, carinho e apoio ao longo deste curso.

Aos meus amigos Fernando César e Cristina Morais, pelo incentivo, companheirismo e participação.

Aos colegas do mestrado, que de alguma forma contribuíram para o término deste trabalho.

A todos que direta ou indiretamente contribuíram para a conclusão deste trabalho.

À banca examinadora, pelas críticas e sugestões.

Resumo

O modelo de regressão Beta possui grande aplicabilidade prática, em particular, na modelagem de taxas e proporções e, tal como nos demais modelos de regressão, também são requeridos métodos que determine qual o melhor modelo.

A presente dissertação tem como objetivo principal implementar e avaliar o desempenho de diferentes critérios de seleção de modelos para o modelo de regressão Beta. Para tal, mediante diferentes estudos de simulações de Monte Carlo, analisamos alguns critérios selecionados levando em consideração suas propriedades assintóticas, os quais foram obtidos por meio da função de máxima verossimilhança. Os resultados das simulações revelaram que os desempenhos dos referidos critérios dependem da especificação do modelo e também do tamanho da amostra. Apresentamos ainda uma aplicação relacionada ao Índice de Desenvolvimento Humano, que é uma variável adequada à modelagem em estudo, visto que seus valores variam no intervalo (0,1). Nesta aplicação, observamos que com a amostra de todos os municípios da região Nordeste os diferentes critérios utilizados selecionaram o mesmo modelo.

Abstract

The Beta regression model holds a great practical applicability, in particular, for modelling rates and proportions and such as the others regression models, it also requires methods to determinate which is the best model.

The main objetive of this work is to implement and evaluate the performance of differents model selection criteria in the Beta regression model. For such, by using differents studies of Monte Carlo simulations, we have analised some criteria selected by taking into consideration its asymptotic properties, which were obtained by maximum likelihood function. The simulations results show that the performances of those criteria depend on the model specification as well as on the sample size. We have also presented an application related to the Human Development Index from United Nations Development Programme (UNDP), which is a right variable for the modelling in study, since its values vary in the interval (0,1). In this application, we have observed that with the sample of all Northeast's cities the differents used criteria selected the same model.

Índice

Capítulo 1

Introdução

Em análises estatísticas, em particular na análise de regressão, surge sempre uma pergunta importante: qual é o melhor modelo?

Quando uma distribuição associada à variável resposta é determinada, ainda é preciso definir as variáveis mais importantes. Logicamente, a medida em que mais variáveis são acrescentadas melhor será o ajuste, porém o modelo será mais complexo. Assim, um dos objetivos principais da seleção de modelos é atingir um equilíbrio entre uma melhora no ajuste e a complexidade do modelo.

Para decidir qual é o modelo mais apropriado dentre um conjunto de modelos candidatos foram criados os denominados critérios de seleção de modelos. O critério de seleção de modelos pseudo R^2 (R_p^2), o qual aparece em muitos textos de regressão hoje. Desde o início da década de 70 surgiram muitos destes critérios, como o Cp de Mallows (Mallows, 1973), o critério de informação de Akaike (AIC, Akaike, 1974) e o critério de informação bayesiano (BIC, Akaike, 1978) entre outros, até os artigos mais recentes como o de Kuha (2004), o qual analisa o desempenho dos dois critérios de seleção mais frequentemente utilizados, quais sejam, o AIC e o BIC, ambos utilizando a função de penalidade. Segundo Kuha os mencionados critérios não obstante terem fundamentos distintos, apresentam algumas similaridades, a exemplo de interpretações análogas da função de penalidade. Kuha realizou simulações e utilizou um conjunto de dados reais para demonstrar que se pode obter informações úteis para seleção de modelos utilizando os dois critérios juntos, especialmente quando se busca o máximo possível achar modelos favoráveis por ambos os critérios.

Um outro artigo de Rao e Wu (2005), por sua vez, considera o problema

de seleção de modelo na regressão clássica baseado na validação cruzada com a adição de uma função de penalidade para penalizar o sobreajuste. Rao e Wu observaram que, sob algumas condições fracas, o novo critério por eles desenvolvido mostra-se fortemente consistente, no sentido de que com probabilidade igual a 1, para todo n grande, o critério escolhe o modelo verdadeiro que apresenta o menor número de regressores. A função de penalidade depende do tamanho da amostra e é escolhida para assegurar a consistência na seleção do modelo verdadeiro. Desta forma, os autores referidos concluíram que uma função de penalidade baseada em dados observados, que a torne aleatória, preserva a consistência e fornece melhor desempenho do que uma função de penalidade fixa.

Uma excelente referência deste asunto para modelos de regressão normal e de séries temporais é o livro de McQuarrie e Tsai (1998). Estes autores destacam que um critério nem sempre é melhor que outro. O fato de que certos critérios têm um desempenho melhor que outros para um modelo específico, foi o que motivou a nossa pesquisa.

Nesta dissertação apresentamos uma análise do desempenho de vários critérios para um modelo particular proposto por Ferrari e Cribari-Neto (2004) o chamado modelo Beta. A dissertação está organizada em cinco capítulos: neste primeiro, destacamos alguns critérios de seleção de modelos tomando por base a constante preocupação dos pesquisadores na escolha do melhor modelo; no segundo capítulo apresentamos o modelo de regressão Beta; no terceiro capítulo definimos os diferentes critérios para o modelos Beta; no capítulo quatro, discutimos os resultados das simulações de Monte Carlo para dois modelos específicos, apresentando o desempenho de cada um deles em relação as eficiências observadas segundo a distância de L_2. E, finalmente, no quinto capítulo, são apresentados os resultados de uma aplicação usando como variável a ser explicada o Índice de Desenvolvimento Humano Municipal e algumas variáveis explicativas do mesmo, incluindo uma análise exploratória dos dados e aplicando a teoria apresentada no capítulo 3 para obtenção do melhor modelo através dos critérios analisados e, por último, um capítulo com as principais conclusões.

Nesta dissertação para realizar os ajustes dos correspondentes modelos, os cálculos dos critérios, bem como as simulações utilizamos os programas computacionais através da linguagem matricial de programação Ox em sua versão 3.40 para plataformas computacionais Windows, desenvolvida por Jurden A. Doornik. Ox pode ser utilizada gratuitamente para uso acadêmico e está disponível em `http//www.doornik.com`. Para a análise dos dados e a

construção dos gráficos, por sua vez, foi utilizado o pacote estatístico `R` versão 1.9.0, também para sistemas operacionais Windows, disponível gratuitamente no endereço `http//www.r-project.org`.

Capítulo 2

O Modelo Beta

Para analisar a relação existente entre variáveis aleatórias é bastante utilizado o modelo de regressão linear normal padrão, no entanto, este modelo requer algumas suposições que em muitas aplicações não são satisfeitas. Se a variável resposta estiver restrita ao intervalo (0,1), o mesmo não pode ser aplicado, visto que podemos obter valores ajustados fora do intervalo. Uma solução é transformar a variável resposta. Entretanto, defrontamo-nos com algumas desvantagens. Uma delas é que os parâmetros da resposta tornam-se de difícil interpretação em termos da variável dependente original, a outra é que, em geral, medidas de razão têm comportamento assimétrico, não satisfazendo a suposição de normalidade. A fim de reverter estas desvantagens, Ferrari e Cribari-Neto (2004) propuseram um modelo, chamado modelo Beta, baseado na suposição que a variável resposta tem distribuição Beta, esta distribuição será apresentada na próxima seção. Neste capítulo apresentamos também o modelo Beta.

2.1 A Distribuição Beta

A densidade Beta é dada por

$$f(y;p,q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1}(1-y)^{q-1} I_{(0,1)}(y), \tag{2.2.1}$$

onde $p > 0$ e $q > 0$ são parâmetros que indexam a distribuição Beta e

$\Gamma(\cdot)$ é a função gama. Sendo a função gama dada por

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

e

$$I_{(0,1)}(y) = \begin{cases} 1 & se \quad y \in (0,1) \\ 0 & se \quad y \notin (0,1). \end{cases}$$

A média e a variância da variável y são, respectivamente,

$$E(y) = \frac{p}{p+q}, \tag{2.2.2}$$

e

$$var(y) = \frac{pq}{(p+q)^2(p+q+1)}. \tag{2.2.3}$$

Uma outra medida que vale destacar é a moda da distribuição, existente quando os parâmetros p e q são maiores que 1, sendo a mesma dada por

$$moda(y) = \frac{p-1}{p+q-2}.$$

A densidade Beta apresenta uma grande diversidade de formas , devido aos valores dos parâmetros que a indexam. Por exemplo, quando $p = q = 1$ a densidade será a "uniforme padrão", porém se $p = q = 1/2$ será a "arco seno" que é utilizada para análise de passeios aleatórios, já quando $p + q = 1$ com $p \neq 1/2$ teremos a densidade "arco seno generalizada". Destacando que, quando $p = q$ as densidades terão formas simétricas, caso contrário serão assimétricas.

A família de densidades Beta pertence a família de densidades exponencial, uma vez que pode ser escrita da forma

$$\begin{aligned} f(y;p,q) &= \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1}(1-y)^{q-1} I_{(0,1)}(y) \\ &= exp\{\phi[t(y)\theta - b(\theta) + c(y,\theta)]\}. \end{aligned}$$

2.2 O Modelo Beta

Os modelos de regressão propostos por Kieschnick e McCullough (2003) ou Ferrari e Cribari-Neto (2004) devem ser utilizados quando a variável resposta

tiver distribuição Beta, pois sitando Johnson, Kots e Balakrishnan (1995, p. 235), "Beta distributions are very versatile and a variety of uncertanties can be usefully modeled by them. This flexibility encourages its empirical use in a wide range of applications".

Como na análise de regressão é normalmente útil modelar a média da variável resposta, bem como definir o modelo de forma que contenha um parâmetro de precisão, Ferrari e Cribari-Neto (2004) propõem uma parametrização diferente da densidade Beta com a finalidade de conseguir uma estrutura de regressão associada a um parâmetro de precisão. A densidade Beta da equação (2.2.1), que é indexada por p e q, será escrita da seguinte forma

$$f(y;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1} I_{(0,1)}(y), \qquad (2.3.1)$$

onde $\mu > 0$ e $\phi > 0$ quando fazemos $\mu = p/(p+q)$ e $\phi = p+q$, isto é, $p = \mu\phi$ e $q = (1-\mu)\phi$. Por conseguinte, as equações (2.2.2) e (2.2.3) torna-se-ão

$$E(y) = \mu,$$

e

$$var(y) = \frac{V(\mu)}{1+\phi},$$

onde $V(\mu) = \mu(1-\mu)$, tal que μ é a média da resposta e ϕ pode ser interpretado como um parâmetro de precisão. Percebe-se que quanto maior for o valor de ϕ tanto menor será a variância de y, fixando-se μ. Vale a pena salientar que quando $\mu = 1/2$ a distribuição é simétrica e quando $\mu \neq 1/2$ a mesma é assimétrica. Em particular para $\mu = 1/2$ e $\phi = 2$ a densidade se reduz a uniforme padrão.

Como já foi citado, trabalharemos aqui com variável resposta restrita ao intervalo (0,1), porém o modelo ainda é adequado quando a resposta está restrita ao intervalo (a,b), onde $a < b$ são escalares desconhecidos. Então, ao invés de modelarmos y, utilizaremos $(y-a)/(b-a)$ que ficará portanto definido no intervalo $(0,1)$. Ferrari e Cribari-Neto (2004) propõem o modelo descrito a seguir.

Sejam $y_1, \ldots, y_n$ variáveis aleatórias independentes, as quais seguem densidades em (2.3.1) com média $\mu_t, t = 1, \ldots, n$ e precisão ϕ desconhecidas. Assuma que a média de y_t no modelo pode ser escrita como

$$g(\mu_t) = \sum_{i=1}^{k} x_{ti}\beta_i = \eta_t, \tag{2.3.2}$$

onde $\beta = (\beta_1, \ldots, \beta_k)^T$, $k < n$ é um vetor de parâmetros de regressão desconhecidos ($\beta \in R^k$), $x_{t1}, \ldots, x_{tk}$ são observações das k covariáveis conhecidas e fixadas e $g(\cdot)$ é uma função monótona e duas vezes diferenciável, restrita ao intervalo (0,1), denominada função de ligação. Nota-se que a variância de y_t é uma função de μ_t e, por conseguinte, dos valores das covariáveis. Logo, o modelo aqui definido admite variável resposta com variância não constante.

McCullagh e Nelder (1989, §4.3.1) comparam várias funções de ligação associadas aos modelos lineares generalizados, tais como, a logit ou logística em que $g(\mu) = log\{\mu/(1-\mu)\}$, a função probit ou Normal inversa, isto é, $g(\mu) = \Phi^{-1}(\mu)$, onde $\Phi(\cdot)$ é a função de distribuição acumulada da variável aleatória normal padrão, a função de ligação log-log $g(\mu) = -log\{-log(\mu)\}$, a ligação log-log complementar $g(\mu) = log\{-log(1-\mu)\}$, entre outras. Atkinson (1985, cap.7) apresenta outras transformações da função de ligação.

No modelo linear generalizado proposto por McCullagh e Nelder (1989), o vetor de observações y que contém n componentes, é considerado uma realização da variável aleatória Y cujas componentes são distribuídas independentemente. Assume-se que cada componente de Y pertence a família de densidades exponencial, como segue

$$\pi(y; \theta_t, \phi) = \exp\{\phi[y\theta_t - b(\theta_t) + c(y, \phi)]\},$$

onde $b(\cdot)$ e $c(\cdot, \cdot)$ são funções específicas. Se $\phi > 0$, parâmetro de escala, é conhecido, configurará um modelo da família exponencial uniparamétrica, ou seja, com um parâmetro natural θ_t. No entanto, se ϕ for desconhecido, poderá ou não ser da família exponencial bi-paramétrica. Verifica-se facilmente que

$$E(y_t) = \mu_t = b'(\theta_t)$$

e

$$var(y_t) = \frac{1}{\phi} b''(\theta_t),$$

desta feita $var(y_t) = \phi^{-1} V_t$, onde $V_t = \frac{d\mu_t}{d\theta_t}$ é denominada *função de variância*, dependendo unicamente da média e, sendo assim, o parâmetro natural pode ser expresso através de uma relação funcional unívoca da média $\theta = \int V^{-1} d\mu$ $= q(\mu)$. Desta forma, pode-se demonstrar que especificando o parâmetro $\mu =$

$q^{-1}(\theta)$, a distribuição $\pi(y;\theta_t,\phi)$ é univocamente determinada (Patil e Shorrock, 1965). Portanto, uma relação funcional variância-média caracteriza a distribuição na família exponencial $\pi(y;\theta_t,\phi)$. Todavia, esta relação não caracteriza a distribuição na família exponencial não-linear

$$\begin{aligned}\pi(y;\theta,\phi) &= exp\{\phi[t(y)\theta - b(\theta) + c(y,\phi)]\} \\ &= \frac{\Gamma(\theta)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1} I_{(0,1)}(y),\end{aligned}$$

onde $t(y) = \log(y/(1-y))$, $\theta = \mu$, $b(\theta) = \phi^{-1}\log[\Gamma(\mu\phi)\Gamma((1-\mu)\phi)\Gamma(\theta)^{-1}]$ e $c(y,\phi) = \phi^{-1}[(\phi-1)\log(1-y) - \log(y)]$. Assim, esta relação não caracteriza a distribuição Beta na família exponencial supra citada. Isto é comprovado pois se Y tem distribuição Beta com parâmetros $p = \phi\mu$ e $q = \phi(1-\mu)$ com ϕ conhecido, obtém-se a função de variância do modelo binomial, negando a relação unívoca entre a variância e a distribuição $\pi(y;\theta,\phi)$.

Dentre as várias funções de ligação existentes, uma particularmente usada é a ligação logit, que é descrita da seguinte forma

$$\mu_t = \frac{e^{x_t^T\beta}}{1 + e^{x_t^T\beta}},$$

onde $x_t^T = (x_{t1},\ldots,x_{tk}), t = 1,\ldots,n$. Neste caso, o parâmetro de regressão tem uma importante interpretação. Admita que o valor do $i-$ésimo regressor seja acrescido por c unidades e que as demais variáveis independentes permaneçam inalteradas. Seja $\mu^\dagger$ a média de y sob os novos valores das covariáveis, enquanto μ denota a média de y sob os valores das covariáveis originais. Então, demonstra-se facilmente que

$$e^{c\beta_i} = \frac{\mu^\dagger/(1-\mu^\dagger)}{\mu(1-\mu)},$$

isto é, $e^{c\beta_i}$ é a razão de chances.

Por ser o método da máxima verossimilhança selecionado para a estimação dos parâmetros do modelo proposto, consideremos a função de log-verossimilhança baseada na amostra de n observações independentes, isto é,

$$\ell(\beta,\phi) = \sum_{t=1}^{n} \ell_t(\mu_t,\phi), \tag{2.3.3}$$

onde

$$\begin{aligned}\ell_t(\mu_t,\phi) &= \log\Gamma(\phi)-\log\Gamma(\mu_t\phi)-\log\Gamma((1-\mu_t)\phi)\\ &\quad+(\mu_t\phi-1)\log y_t+\{(1-\mu_t)\phi-1\}\log(1-y_t),\end{aligned}$$

com μ_t definida de tal forma que satisfaz (2.3.2), onde $\mu_t = g^{-1}(\eta_t)$, é uma função de β. A função escore é obtida pela diferenciação da função de log-verossimilhança em relação aos parâmetros desconhecidos. A isto se segue que, para $i = 1,\ldots,k$,

$$\frac{\partial\ell(\beta,\phi)}{\partial\beta_i}=\sum_{t=1}^{n}\frac{\partial\ell_t(\mu_t,\phi)}{\partial\mu_t}\frac{d\mu_t}{d\eta_t}\frac{\partial\eta_t}{\partial\beta_i}. \tag{2.3.4}$$

Note que $d\mu_t/d\eta_t = 1/g'(\mu_t)$. Também,

$$\frac{\partial\ell_t(\mu_t,\phi)}{\partial\mu_t}=\phi\left[\log\frac{y_t}{1-y_t}-\{\psi(\mu_t\phi)-\psi(1-\mu_t)\phi\}\right], \tag{2.3.5}$$

onde $\psi(\cdot)$ é a função digamma, isto é, $\psi(z) = d\log\Gamma(z)/dz$, $z > 0$. Seja $y_t^* = \log\{\frac{y_t}{1-y_t}\}$ e $\mu_t^* = \{\psi(\mu_t\phi)-\psi(1-\mu_t)\phi\}$. Sob certas condições de regularidade, (ver, Martínez 2004) sabe-se que o valor esperado da derivada em (2.3.5) iguala-se a zero, de forma que $\mu_t^* = E(y_t^*)$. Por conseguinte,

$$\frac{\partial\ell(\beta,\phi)}{\partial\beta_i}=\phi\sum_{t=1}^{n}(y_t^*-\mu_t^*)\frac{1}{g'(\mu_t)}x_{ti}. \tag{2.3.6}$$

A função escore para β pode ser definida de forma matricial como segue

$$U_\beta(\beta,\phi)=\phi X^T T(y^*-\mu^*), \tag{2.3.7}$$

onde X é uma matriz $n\times k$ cuja $t-$ésima linha é x_t^T, $T = diag\{1/g'(\mu_1),\ldots,1/g'(\mu_n)\}$, $y^* = (y_1^*,\ldots,y_n^*)^T$ e $\mu^* = (\mu_1^*,\ldots,\mu_n^*)^T$. De forma semelhante, temos que para o parâmetro de precisão a função escore pode ser escrita como

$$U_\phi(\beta,\phi)=\sum_{t=1}^{n}\{\mu_t(y_t^*-\mu_t^*)+\log(1-y_t)-\psi((1-\mu_t)\phi)+\psi(\phi)\}. \tag{2.3.8}$$

Obteremos a seguir a expressão da matriz de informação de Fisher. Seja $W = diag\{w_1, \ldots, w_n\}$, com

$$w_t = \phi\{\psi'(\mu_t\phi) + \psi'((1-\mu_t)\phi)\}\frac{1}{\{g'(\mu_t)\}^2},$$

$c = (c_1, \ldots, c_n)^T$, com $c_t = \phi\{\psi'(\mu_t\phi)\mu_t - \psi'((1-\mu_t)\phi)(1-\mu_t)\}$, onde $\psi'(\cdot)$ é a função trigamma. Também, seja $D = diag\{d_1, \ldots, d_n\}$, com $d_t = \psi'(\mu_t\phi)\mu_t^2 + \psi'(1-\mu_t)\phi)(1-\mu_t)^2 - \psi'(\phi)$. Pode-se provar que a matriz de informação de Fisher é dada por

$$K = K(\beta, \phi) = \begin{pmatrix} K_{\beta\beta} & K_{\beta\phi} \\ K_{\phi\beta} & K_{\phi\phi} \end{pmatrix}, \tag{2.3.9}$$

onde $K_{\beta\beta} = \phi X^T W X$, $K_{\beta\phi} = K_{\phi\beta}^T = X^T T c$ e $K_{\phi\phi} = tr(D)$. Observe que $K_{\beta\phi} = K_{\phi\beta}^T \neq 0$, o que indica que os parâmetros β e ϕ não são ortogonais, diferentemente do que é verificado na classe de modelos lineares generalizados (McCullagh e Nelder, 1989).

Sob condições de regularidade usuais para estimação de máxima verossimilhança, quando o tamanho da amostra é grande, temos que

$$\begin{pmatrix} \hat{\beta} \\ \hat{\phi} \end{pmatrix} \overset{A}{\sim} \mathcal{N}_{k+1}\left(\begin{pmatrix} \beta \\ \phi \end{pmatrix}, K^{-1}\right),$$

onde $\hat{\beta}$ e $\hat{\phi}$ são os estimadores de máxima verossimilhança de β e ϕ, respectivamente, e $\mathcal{N}_{k+1}$ uma distribuição normal $(k+1)$-variada. Por esta razão é útil obter uma expressão para K^{-1}, a qual pode ser usada para obtenção dos erros padrões assintóticos das estimativas de máxima verossimilhança. Utilizando a expressão padrão para a inversa de matrizes particionadas (ver, por exemplo, Rao, 1973, p. 33), obtem-se a inversa da matriz de informação de Fisher (2.3.8) como segue

$$K^{-1} = K^{-1}(\beta, \phi) = \begin{pmatrix} K^{\beta\beta} & K^{\beta\phi} \\ K^{\phi\beta} & K^{\phi\phi} \end{pmatrix}, \tag{2.3.10}$$

onde

$$K^{\beta\beta} = \frac{1}{\phi}(X^T W X)^{-1}\left\{I_k + \frac{X^T T c c^T T^T X (X^T W X)^{-1}}{\gamma\phi}\right\},$$

com $\gamma = tr(D) - \phi^{-1} c^T X (X^T W X)^{-1} X^T T c$,

$$K^{\beta\phi} = (K^{\phi\beta})^T = -\frac{1}{\gamma\phi}(X^T W X)^{-1} X^T T c,$$

e $K^{\phi\phi} = \gamma^{-1}$. Aqui, I_k é a matriz identidade de ordem $k \times k$.

Os estimadores de máxima verossimilhança de β e ϕ são obtidos pela solução do sistema,

$$\begin{cases} U_\beta(\beta, \phi) = 0 \\ U_\phi(\beta, \phi) = 0 \end{cases}$$

e observamos que os mesmos não possuem forma fechada. Assim, eles necessitam ser obtidos numericamente pela maximização da função de log-verossimilhança através de um algoritmo de otimização não-linear, tal como um algoritmo de Newton (Newton-Rapson, Fisher's scoring, BHHH, etc.) ou o algoritmo quasi-Newton (BFGS); para detalhes, ver Nocedal e Wright (1999). Neste trabalho utilizamos o algoritmo BFGS. Ferrari e Cribari-Neto (2004) sugerem usar como uma estimativa pontual inicial para β a estimativa de mínimos quadrados ordinários do vetor de parâmetros obtida a partir de uma regressão linear da resposta transformada $g(y_1), \ldots, g(y_n)$ em X, isto é, $(X^TX)^{-1}X^Tz$, onde $z = (g(y_1), \ldots, g(y_n))^T$. Também necessitamos de um valor inicial para o parâmetro de escala ϕ, $\phi = \{\mu_t(1-\mu_t)/var(y_t)\} - 1$, que é obtido facilmente de $var(y_t) = \mu_t(1-\mu_t)/(1+\phi)$. Note que ao expandir até primeira ordem a função $g(y_t)$ em série de Taylor em torno do ponto μ_t e tomando a variância, deduzimos que

$$var(g(y_t)) \approx var\{g(\mu_t) + (y_t - \mu_t)g'(\mu_t)\} = var(y_t)\{g'(\mu_t)\}^2,$$

isto é,

$$var(y_t) \approx var(g(y_t))/\{g'(\mu_t)\}^2.$$

Então, a sugestão dada por Ferrari e Cribari-Neto (2004) é considerar como estimativa inicial para ϕ

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\breve{\mu}_t(1-\breve{\mu}_t)}{\breve{\sigma}_t^2} - 1,$$

onde $\breve{\mu}_t$ é obtido aplicando $g^{-1}(\cdot)$ ao t-ésimo valor ajustado da regressão linear de $g(y_1), \ldots, g(y_n)$ em X, isto é, $\breve{\mu}_t = g^{-1}(x_t^T(X^TX)^{-1}X^Tz)$, e $\breve{\sigma}_t^2 = \breve{e}^T\breve{e}/[(n-k)\{g'(\breve{\mu}_t)\}^2]$, sendo $\breve{e} = z - (X^TX)^{-1}X^Tz$ o vetor de resíduos de mínimos quadrados da regressão linear sob a variável resposta transformada.

Capítulo 3

Critérios de Seleção de Modelos

Como já foi mencionado os critérios de seleção de modelos são um guia para a escolha do melhor modelo. Existem várias maneiras de comparar os diferentes critérios existentes na literatura. Uma maneira é analisar o número de vezes que cada critério identifica o verdadeiro modelo e uma outra maneira é a utilização de uma medida de distância entre o modelo escolhido e o verdadeiro modelo. Em ambos os casos existe a necessidade de realizar simulações e também que o modelo verdadeiro pertença ao conjunto de modelos candidatos. Em qualquer conjunto de modelos existirá algum modelo mais próximo ao verdadeiro modelo. A razão que compara as distâncias entre os modelos escolhidos e o modelo mais próximo é chamada de *eficiência observada*. Todos estes conceitos e definições serão apresentados nas próximas seções.

3.1 Critérios Consistentes e Eficientes

Muitos pesquisadores assumem que o modelo verdadeiro tem dimensão finita e que o mesmo pertence ao conjunto de modelos candidatos. Sob esta suposição, o objetivo da seleção do modelo é escolher o modelo verdadeiro a partir deste conjunto. Um critério de seleção do modelo, que identifica o modelo correto com probabilidade assintoticamente igual a 1 é definido como *critério de seleção consistente*.

Outros pesquisadores não aceitam a suposição supra citada, e assumem que ou o modelo verdadeiro tem dimensão infinita ou não está incluído no conjunto de modelos candidatos. O objetivo dos critérios de seleção do mo-

delo é escolher um modelo que mais se aproxime do modelo verdadeiro a partir de um conjunto de modelos candidatos com dimensão finita. O modelo candidato que mais se aproxima do modelo verdadeiro é definido como o modelo adequado. Em grandes amostras, um critério de seleção de modelos que, elege o modelo com distribuição do erro médio quadrado mínimo, é definido *assintoticamente eficiente* (Shibata, 1980).

McQuarrie e Tsai (1998) apresentam o conceito de sinal de ruído para realizar as comparações, em termos também de sobreajustamento, entre os diferentes critérios de seleção de modelos de regressão normal. Com este intuito obtem analiticamente os momentos das diferenças num critério dado de dois modelos diferentes aninhados. Os resultados obtidos estão fortemente baseados na suposição de normalidade e não podem ser estendidos facilmente ao caso do modelo Beta. Na seção seguinte apresentaremos a definição de cada um dos critérios para o modelo Beta destacando que a obtenção de cada um deles requer cálculos númericos que foram implementados computacionalmente para a realização deste trabalho.

3.2 Critérios de Seleção do Modelo Beta

Para escolher o melhor modelo foram criados e apresentados na literatura vários critérios ao longo do tempo. O primeiro foi o coeficiente de determinação (R^2), contudo, não é uma boa estratégia, pois o mesmo sempre aumenta com a inclusão de novas covariáveis; para contornar este problema foi criado um coeficiente de determinação ajustado, denominado pseudo $R^2(R_p^2)$ que é definido como o quadrado do coeficiente de correlação amostral entre $\hat{\eta}$ e $g(y)$. Note que $0 \leq R_p^2 \leq 1$ e, quando $R_p^2 = 1$ existe uma concordância perfeita entre $\hat{\eta}$ e $g(y)$ e, por consequência, entre $\hat{\mu}$ e y.

Embora o desempenho do R_p^2 não seja muito bom sob certas circunstâncias, ele é uma ferramenta suporte em muitos estudos de modelagem com regressão. Dentre os possíveis modelos propostos, o melhor modelo é aquele que maximiza o pseudo R^2. Para o cálculo do mesmo utilizaremos a seguinte expressão:

$$R_p^2 = 1 - \frac{SSE/(n-k)}{SST/(n-1)}, \tag{3.2.1}$$

onde $SSE = \sum_{t=1}^{n}(g(y_t) - \hat{\eta}_t)^2$, $SST = \sum_{t=1}^{n}(g(y_t) - \overline{g(y_t)})^2$, n é o número de

observações e k é o número de parâmetros do modelo proposto.

Uma outra medida para verificar a discrepância do ajuste, proposta por Nelder e Wedderburn (1972) foi a *deviance* (traduzida por Cordeiro (1986) como desvio). Nesta medida o modelo saturado serve como base de medida do ajustamento para o modelo em investigação. A mesma pode ser obtida através da estatística que é o dobro da diferença entre os máximos da log-verossimilhança do modelo saturado e o modelo em investigação, denotado por $l_t(\tilde{\mu}_t, \phi)$ e $l_t(\mu_t, \phi)$ respectivamente.

$$D(y; \mu, \phi) = \sum_{t=1}^{n} (2(l_t(\tilde{\mu}_t, \phi) - l_t(\mu_t, \phi))),$$

onde $\tilde{\mu}_t$ é o valor de μ_t que é solução do sistema de equações $\partial l_t / \partial \mu_t = 0$, isto é de (2.3.4), $\phi(y_t^* - \mu_t^*) = 0$.

Se ϕ for desconhecido, poderá ser obtida uma aproximação desta quantidade que é normalmente denominada como a *deviance* para o modelo corrente e é representada por $D(y; \hat{\mu}, \hat{\phi})$. Tem-se, então, $D(y; \hat{\mu}, \hat{\phi}) = \sum_{t=1}^{n} (r_t^d)^2$, onde

$$r_t^d = sign(y_t - \hat{\mu}_t)\{2(l_t(\tilde{\mu}_t, \hat{\phi}) - l_t(\hat{\mu}_t, \hat{\phi}))\}^{1/2},$$

sendo r_t^d chamada de $t-$ésima *deviance residual*.

Vale destacar que a $t-$ésima observação contribui para a *deviance* com $(r_t^d)^2$, e se o valor absoluto de r_t^d for grande indica que a observação é discrepante.

A *deviance* é sempre maior ou igual a zero, e decresce quando ocorre a inclusão de covariáveis na componente sistemática, até atingir o valor zero para o modelo saturado ou completo. Note que, a utilização de um grande número de covariáveis visando a redução da *deviance*, resulta em um modelo com alto grau de complexidade de interpretação, e portanto, o mais indicado é escolher modelos simples com *deviance* moderada, localizados entre os modelos mais complicados e os que se ajustam mal aos dados.

Na prática, sem muito rigor, costuma-se ainda utilizar o ponto crítico $\chi^2_{(n-p)}(\alpha)$ da distribuição qui-quadrado ao nível de significância igual a α e p o posto da matriz do modelo. Portanto, se

$$D(y; \mu, \phi) \leq \chi^2_{(n-p)}(\alpha),$$

pode-se considerar que há evidências que o modelo proposto esteja bem ajustado aos dados, a um nível de $100\alpha\%$ de significância, caso contrário devemos descartar o modelo pois o mesmo pode ser considerado inadequado.

Apesar deste critério ser bastante utilizado, não o utilizaremos nas simulações por se tratar de um critério de seleção apropriado para a comparação de modelos encaixados, porém usaremos na aplicação prática.

Akaike (1974) propõe o critério AIC (Akaike Information Criterion), que foi desenvolvido a partir dos Estimadores de Máxima Verossimilhança (EMV) para decidir qual o modelo mais adequado, quando utilizamos muitos modelos com quantidades diferentes de parâmetros, isto é, selecionar um modelo que esteja bem ajustado com um número reduzido de parâmetros, sendo este critério assintoticamente eficiente, no entanto, não é assintoticamente consistente. O AIC foi o primeiro critério baseado na informação de Kullback-Leibler (K-L), ele é assintoticamente não viesado para K-L. O critério AIC supõe que o modelo verdadeiro pertence ao conjunto de modelos ajustados. Esta suposição pode ser irrealista na prática, porém permite calcular valores esperados em distribuições centrais e considerar o conceito de sobreajustamento. Em geral,

$$AIC = -2l(\hat{\mu}, \hat{\phi}) + 2(k+1). \tag{3.2.2}$$

Visto não ser o AIC adequado para pequenas amostras, Sugiura (1978) e Hurvich e Tsai (1989) derivaram o AICc estimando a discrepância esperada de Kullback-Leibler diretamente dos modelos de regressão. Hurvich e Tsai também adotaram a suposição de que o modelo verdadeiro pertence ao conjunto de modelos candidatos, e mostraram que o AICc de fato tem melhor desempenho que o AIC em pequenas amostras, pois corrige o sobreajustamento; porém é assintoticamente equivalente ao AIC e portanto, é assintoticamente eficiente. Como no caso do AIC, os parâmetros associados ao modelo candidato são estimados por Máxima Verossimilhança. Temos que,

$$AICc = -2l(\hat{\mu}, \hat{\phi}) + 2(k+1)\Big(\frac{n}{n-k-2}\Big). \tag{3.2.3}$$

Akaike (1978) e Schwarz (1978) introduziram critérios de seleção de modelos concebidos através de uma perspectiva Bayesiana. Schwarz desenvolveu o SIC (Schwarz Information Criterion) para seleção de modelos da família Koopman-Darmois, ao passo que Akaike desenvolveu o critério de seleção de modelos BIC (Bayesian Information Criterion) para seleção de modelos em regressão linear. Neste trabalho utilizaremos o critério BIC por se tratar de modelos de regressão. Ao contrário do AIC, o critério BIC assume que o modelo verdadeiro tem dimensão infinita e portanto não pertence ao conjunto

de modelos candidatos. O BIC é definido por

$$BIC = -2l(\hat{\mu}, \hat{\phi}) + (k+1)log(n). \tag{3.2.4}$$

Hannan e Quinn (1979) propuseram o critério HQ para modelos de séries temporais autoregressivos, porém o mesmo pode ser estendido para outros modelos. Ele é obtido através de

$$HQ = -2l(\hat{\mu}, \hat{\phi}) + 2(k+1)log(log(n)). \tag{3.2.5}$$

Tanto o BIC quanto o HQ são assintoticamente consistentes, no entanto, muitos autores apontam que o HQ comporta-se como o critério eficiente AIC.

Outro critério de seleção de modelos a ser utilizado é o HQc que se originou na modificação da função de penalidade do HQ, usando análise semelhante a que foi realizada entre o AIC e AICc, sendo definido por

$$HQc = -2l(\hat{\mu}, \hat{\phi}) + \frac{2(k+1)log(log(n))n}{n-k-3}. \tag{3.2.6}$$

Pode-se provar que o HQc não apenas corrige o desempenho do HQ em pequenas amostras, como também é um critério assintoticamente consistente.

3.3 Eficiência Observada L_2

A medida de distância utilizada neste trabalho é L_2, definida por

$$L_2 = ||\mu(M_v) - \mu(M_c)||^2,$$

onde $\mu(M_v)$ e $\mu(M_c)$, denotam o vetor de médias do modelo verdadeiro(M_v) e do modelo candidato (M_c). Uma vantagem de L_2 é que o mesmo só depende das médias das duas distribuições e não das densidades atuais. A medida L_2 pode ser aplicada quando os erros não são normalmente distribuídos.

Podemos definir a eficiência assintótica e a eficiência observada para pequenas amostras. Shibata (1980) sugere o uso da distância esperada, $E(L_2) = E(||\mu(M_v) - \mu(M_c)||^2)$, como medida de distância. Usando a medida de Shibata, McQuarrie e Tsai (1998) assumem que existe entre os modelos candidatos um modelo que é o mais ajustado do modelo verdadeiro em termos da esperança de L_2, $E(L_2(M_a))$. Suponha que um critério de seleção de modelos escolha o modelo M_k, o qual tenha a esperança de L_2, $E(L_2(M_k))$. Certamente, $E(L_2(M_k)) \geq E(L_2(M_a))$.

Um critério de seleção de modelos é definido assintoticamente eficiente se

$$\lim_{n\to\infty} P\Big(\frac{E(L_2(M_a))}{E(L_2(M_k))}\Big) = 1,$$

onde n representa o tamanho da amostra. Para pequenas amostras, definimos analogamente a *eficiência observada* L_2 por

$$\textit{eficiência observada } L_2 = \frac{L_2(M_a)}{L_2(M_k)},$$

onde $L_2(M_a) = ||\mu(M_v) - \hat{\mu}(M_a)||^2$, $L_2(M_k) = ||\mu(M_v) - \hat{\mu}(M_k)||^2$ e $\hat{\mu}$ é o vetor de valores preditos dos respectivos modelos.

Capítulo 4

Estudos de Simulação

Neste capítulo analisamos dois particulares modelos Beta com o objetivo de verificar qual o melhor critério de seleção nestes casos, levando em consideração seis critérios de seleção de modelos (R_p^2, AIC, AICc, BIC, HQ, HQc) definidos nas equações 3.2.1 a 3.2.6. Dentre os critérios existentes na literatura houve uma preocupação de selecionarmos critérios assintoticamente eficientes e consistentes.

Avaliamos através de simulações de Monte Carlo o desempenho dos critérios para os dois modelos, levando em consideração diferentes tamanhos de amostras ($n = 20, 40, 60, 200$) e o número de réplicas $R = 5000$. Foram utilizadas sete covariáveis com distribuição exponencial de parâmetro 3 e incluímos o intercepto, obtendo em cada simulação um conjunto de 255 subconjuntos considerados como modelos candidatos potenciais. Para ambos os modelos,

$C^* = 5$, $\phi^* = 120$ e o intercepto $\beta_0 = 1$.

Modelo 1:

$$\beta_1 = \beta_2 = \beta_3 = \beta_4 = 1.$$

Os valores de x_{tj} são gerados em cada réplica independentes e identicamente distribuídos exp(3), sendo $t = 1, 2, \ldots, n$ e $j = 1, 2, \ldots, 7$ e $x_{t0} = 1$.

Os valores de $y_t(y_1, \ldots, y_n)$ em cada réplica são gerados a partir da parametrização (2.3.1) da distribuição $Beta(\mu_t, \phi^*)$,
onde

$$\mu_t = \frac{exp(\beta_0 + \beta_1 x_{t1} + \ldots + \beta_4 x_{t4})}{1 + exp(\beta_0 + \beta_1 x_{t1} + \ldots + \beta_4 x_{t4})}.$$

De posse dos valores da variável resposta e da matriz de regressores, tomamos como estimativas iniciais dos parâmetros para obtenção da maximização da função de log-verossimilhança (2.3.3) através do método BFGS, as estimativas obtidas da regressão auxiliar

$$(X^T X)^{-1} X^T z,$$

onde z é o vetor da variável resposta transformado

$$z = (g(y_1), \ldots, g(y_n))^T.$$

Posteriormente, calculamos os valores da distância L_2 por (3.3.1) e as respectivas eficiências observadas L_2 por (3.3.2) dos modelos candidatos. Notemos que nesta situação o modelo verdadeiro é conhecido e portanto é possível calcular a distância L_2 e determinamos assim dentre todos os modelos candidatos ajustados aquele que torna mínima esta distância, destacamos também que nem sempre esta distância seleciona o verdadeiro modelo. Em cada réplica foi escolhido um modelo através dos critérios utilizados e a partir do modelo selecionado foi calculada a eficiência observada definida em (3.3.2). Posteriormente, calculamos a média, a mediana, o desvio-padrão e o coeficiente de variação para as eficiências observadas resultantes de cada critério.

Modelo 2: é idêntico ao Modelo 1 com exceção dos parâmetros β que são bem menores, definidos como

$$\beta_1 = 1, \beta_2 = 1/2, \beta_3 = 1/3, \beta_4 = 1/4.$$

A diminuição dos valores dos parâmetros β, tal como referido por McQuarrie e Tsai (1998), torna o modelo "menos identificável" visto que as variáveis associadas a estes parâmetros têm uma menor associação com a variável resposta.

Contamos o número de vezes em que o critério selecionou o modelo verdadeiro, que corresponde ao número de vezes que os modelos de ordem C foram escolhidos $C = 1, 2, \ldots, 8$, sendo C o número de colunas da matriz de regressores. Posteriormente calculamos a média, a mediana, o desvio-padrão e o coeficiente de variação das eficiências observadas L_2. O desempenho dos critérios de seleção de modelos foi baseado na média da eficiência observada L_2, onde a maior eficiência observada denota melhor desempenho. O critério com a maior eficiência observada é o de escore 1 (o melhor) ao passo que o critério com a menor eficiência observada será classificado de escore 6 (o pior).

4.1 Resultados das Simulações

Os resultados das simulações para os dois modelos estão apresentados nas Tabelas A.1 a A.8 inseridas no Apêndice A; nas quais apresentamos a freqüência do modelo selecionado e as eficiências observadas L_2.

A partir da Tabela A.1 com $n = 20$, considerando o número de vezes em que o critério escolheu o verdadeiro modelo M_v, podemos notar que L_2 nem sempre selecionou M_v, que o AICc dentre todos os critérios foi o que mais selecionou o modelo verdadeiro obtendo uma freqüência (2691) seguido pelo critério HQc (2628). Lembrando que estes critérios foram modificados para corrigir os problemas de sobreajustamento dos critérios originais para o modelo de regressão normal. Desta tabela observamos que para este modelo Beta novamente os critérios AIC e HQ tendem a sobreajustar já que selecionam modelos de dimensão maior um número considerável de vezes. Este problema é corrigido em parte pelos critérios AICc e HQc, porém passam a ter um leve problema de baixo-ajustamento (*underfitting*) levando a uma possível perda de eficiência. Contudo podemos observar que o critério com melhor desempenho foi o AICc com eficiência observada média de 0.7317, seguido do BIC (0.7135) e com uma eficiência média observada praticamente igual ao do HQc (0.7134). Podemos notar que as dispersões relativas da eficiência observada em todos os critérios estão muito próximas, porém os coeficientes de variação são consideravelmente altos, superiores a 42%. Assim, observando as medianas das eficiências observadas podemos notar mais claramente a superioridade dos critérios AICc e HQc. O critério que obteve menor desempenho foi o R_p^2 com eficiência observada média de 0.6413.

Na Tabela A.2 com $n = 40$, também observamos que o critério com melhor desempenho foi o HQc com eficiência observada média de 0.8930, seguido do BIC com 0.8771 e, o que obteve menor desempenho foi o R_p^2 com eficiência observada média de 0.7029. Aqui, os critérios AIC e R_p^2 tendem a sobreajustar.

Verificamos na Tabela A.3 com $n = 60$, que nenhum dos critérios apresentou problemas de baixo-ajustamento e que o BIC e o HQc foram os critérios com melhores desempenhos apresentando eficiência observada média de 0.9034 e de 0.8838, respectivamente, e os que se mostraram com piores desempenhos foram o AIC e R_p^2 com 0.7526 e 0.7035 de eficiência média.

Observamos na Tabela A.4 onde $n = 200$, como para $n = 40$ e para $n = 60$ não ocorreu problemas de baixo-ajustamento e os melhores critérios foram o BIC e o AICc com eficiência observada média de 0.9501 e de 0.8917,

respectivamente, enquanto que o R_p^2 e o AIC foram os que se mostraram como os piores critérios com 0.7778 e 0.7245 de eficiência observada média. Observamos ainda, que as medianas da eficiência observada de todos os critérios foram iguais a 1 exceto para o R_p^2.

Da comparação entre as Tabelas A.1 a A.4 observamos, como era esperado pelas propriedades de consistência dos critérios, que o número de vezes que o modelo escolhido é o verdadeiro, aumentou em todos os casos com o aumento de tamanho da amostra, exceto o AICc quando o n passou de 40 para 60 e de 60 para 200 e o HQc quando o n passou de 40 para 60. Também observamos que os critérios AIC e $AICc$ têm o mesmo problema de sobreajustamento. Porém o critério HQc parece corrigir este problema do HQ. Notamos ainda, que em todos os tamanhos de amostra os piores critérios foram o R_p^2 e o AIC. Verificamos que tanto a freqüência do modelo verdadeiro obtida em cada critério quanto a eficiência observada média, obedecem a mesma ordem dos critérios, que as eficiências observadas medianas dos quatro melhores critérios atingiram o limite máximo (1.000), que ocorreu uma inversão do AICc e HQc para $n = 20$ e $n = 40$ e, que quando $n = 60$ e $n = 200$ todos os critérios apresentaram os mesmos escores. Em geral, para este modelo podemos concluir que o critério BIC obteve um bom desempenho já que seu escore foi 1 ou 2 em todos os casos.

Quanto aos resultados da simulação para o Modelo 2, com os dados apresentados na Tabela A.5 ($n = 20$) destacamos que para este modelo as correções não funcionaram bem para os critérios AIC e HQ, pois estes critérios ainda escolhem modelos de dimensão menor, isto é, há um baixo-ajustamento, o que tem como conseqüência uma perda significativa em termos de eficiência observada. Quando consideramos o tamanho $n = 40$ na Tabela A.6, podemos observar que os critérios corrigidos têm o mesmo problema de sobreajustamento dos critérios originais. As simulações com $n = 20$ e $n = 40$ revelam que o critério BIC, teve um desempenho melhor se considerarmos tanto em relação ao sobreajustamento e baixo-ajustamento, quanto ao observar que suas eficiências observadas médias e medianas são bem próximas aos critérios AIC e HQ. Para $n = 60$ e $n = 200$, Tabela A.7 e A.8, percebemos que o AICc realiza correções no AIC, o mesmo não acontece com o HQc em relação ao HQ. Com as simulações realizadas para o Modelo 2, concluímos que não podemos eleger um critério como melhor, devido os mesmos assumirem escores diferentes dependendo do tamanho da amostra.

Comparando os resultados obtidos nas simulações para os dois tipos de modelos confirmamos que o Modelo 1 representa o caso onde é mais fácil

identificar como modelo candidato o modelo verdadeiro. Assim, da comparação das Tabelas do Modelo 1 (A.1 a A.3) com as do Modelo 2 (A.5 a A.7), respectivamente, podemos concluir que o número de vezes que o modelo verdadeiro é escolhido é bem menor para o Modelo 2 que para o Modelo 1. Também as eficiências médias e medianas do segundo modelo são menores, este fato, conforme já referido é devido ao fato de que o Modelo 2 é "fracamente identificável".

Vale destacar que, para $n = 200$, Tabela A.4 e A.8, os dois Modelos apresentaram resultados bem próximos em todos os critérios aqui analizados e todos os critérios obedeceram a mesma ordem de classificação.

Capítulo 5

Aplicação

A aplicação foi feita com objetivo de escolher o melhor modelo para explicar o Índice de Desenvolvimento Humano Municipal (IDHM) dos municípios nordestinos no ano 2000. O IDHM foi criado a partir do Programa das Nações Unidas para o Desenvolvimento (PNUD), que tem como mandato central o combate à pobreza.

Tendo sede em Nova York, Estados Unidos, o PNUD foi criado em 1965 para estimular o desenvolvimento global por meio do intercâmbio de conhecimentos, experiências e recursos entre os países. O PNUD, hoje com escritórios em 166 países, dá suporte à elaboração e à implementação de políticas públicas e iniciativas que promovem o desenvolvimento sustentável, vendo o ser humano como agente e sujeito do próprio desenvolvimento.

O trabalho do PNUD no Brasil tem evoluído constantemente. Hoje, visa o combate à pobreza; o manejo do meio ambiente e sua utilização de forma sustentável e limpa; e a modernização do Estado brasileiro. O PNUD atua como interlocutor e articulador junto a países, organizações públicas e privadas, agências de fomento e instituições financeiras, contribuindo para a transferência de conhecimentos entre os variados atores do desenvolvimento. O relacionamento entre o PNUD e a sociedade brasileira transcende os compromissos oficiais de cooperação: dificuldades e alegrias vividas pelos brasileiros são compartilhadas pelos funcionários locais e técnicos do Programa, por meio dos diversos projetos que vêm sendo implantados conjuntamente.

Os gastos do Programa no Brasil representam cerca de 10% de seu desembolso em todo mundo. Nos últimos anos, houve uma grande diminuição de doações para o fundo de programas de desenvolvimento na América Latina, de forma que o governo brasileiro viu-se obrigado a financiar e gerenciar os

projetos que executa em parceria com o PNUD, para que os projetos do Brasil não parassem. No Brasil, atualmente, o PNUD tem cerca de 180 projetos em andamento, com prioridades nas áreas de governança democrática, redução da pobreza, tecnologia de informação, políticas de energia e meio ambiente e combate ao vírus HIV/AIDS.

Para maiores informações ver `http//www.undp.org.br.`

O conceito de Desenvolvimento Humano, base tanto do Relatório de Desenvolvimento Humano (RDH), publicado anualmente, quanto do Índice Desenvolvimento Humano (IDH), parte da premissa que para aferir o avanço de uma população devemos considerar características sociais, culturais e políticas além das dimensões econômicas.

O IDH foi criado a partir dos seguintes indicadores: o de educação (alfabetização e taxa de matrículas), o de longevidade (esperança de vida ao nascer) e o de renda (Produto Interno Bruto per capita), com o intuito de medir o nível de desenvolvimento dos países. Em função da forma como é definido, o índice varia de 0 (nenhum desenvolvimento humano) a 1 (desenvolvimento humano total). Vale destacar que conceitualmente se considera de baixo desenvolvimento humano os países com IDH igual ou inferior a 0.499, de médio desenvolvimento, os que têm IDH entre 0.500 e 0.799 e, finalmente, de alto desenvolvimento, aqueles que detêm IDH maior ou igual a 0.800.

O Índice de Desenvolvimento Humano Municipal (IDHM) com origem no Brasil, pode ser consultado no Atlas de Desenvolvimento Humano, um banco eletrônico com informações sócio-econômicas sobre os 5.507 municípios do país, os 26 Estados e o Distrito Federal.

Para aferir o nível de desenvolvimento humano de municípios, os indicadores utilizados são os mesmos do cálculo do IDH, no entanto, alguns destes são usados de forma diferente, para se tornarem mais apropriados para avaliar as condições de núcleos menores.

Na dimensão educação, considera-se dois indicadores com diferentes pesos, sendo peso dois, a taxa de alfabetização de pessoas acima de 15 anos de idade e, peso um a taxa bruta de frequência à escola; sendo o primeiro indicador o percentual de pessoas com mais de 15 anos capazes de ler e escrever um bilhete simples e o segundo o somatório de pessoas (independente da idade) que frequentam os cursos fundamental, secundário e superior, dividido pela população da localidade na faixa etária de 7 a 22 anos, exceto as classes especiais de alfabetização, que não entram neste cálculo.

Na dimensão longevidade, considera-se o mesmo indicador do IDH usado para países, a esperança de vida ao nascer, ou seja, o número médio de anos

que uma pessoa nascida naquela localidade no ano de referência deve viver, desde que a taxa de mortalidade existente se mantenha constante. Este indicador leva em conta as condições sociais, de saúde e de salubridade por considerar as taxas de mortalidade das diferentes faixas etárias daquela localidade. Como as estatísticas do registro civil nos municípios são inadequadas, recorreu-se a técnicas indiretas para se obter as estimativas de mortalidade. Tomou-se por base as perguntas do Censo do Instituto Brasileiro de Geografia e Estatística (IBGE) sobre o número de filhos nascidos vivos e o número de filhos ainda vivos na data em que o Censo foi realizado. A partir destas informações calcula-se as proporções de óbitos e posteriormente a probabilidade de morrer e finalmente transforma-se estas probabilidades em tábuas de vida, de onde é extraída a esperança de vida ao nascer. O IDHM-Longevidade é obtido através da fórmula: (valor observado do indicador - 25)/ (85 - 25), pois se usa como parâmetro máximo 85 anos e como parâmetro mínimo 25 anos.

Finalmente, na dimensão renda é usado a renda municipal per capita média, isto é, soma-se a renda de todos os habitantes do município (inclusive salários, pensões, aposentadorias e transferências governamentais, entre outros) e divide-se pelo número de pessoas que residem no município. Para transformar em índice, primeiro convertem-se os valores anuais mínimo e máximo expressos em dólar PPC (Paridade do Poder de Compra), utilizados pelo PNUD no cálculo do IDH-Renda dos países, correspondentes aos valores anuais de PIB per capita de US$ 100 PPC e US$ 40.000 PPC respectivamente, em valores mensais em reais: R$ 3,90 e R$ 1.560,17 em 2000. Estes valores são obtidos pela multiplicação da renda per capita mensal em reais pelo fator (R$ 297/US$ 7.625 PPC), que é a relação entre a renda per capita média mensal (em reais) e o PIB per capita anual (em dólares PPC) do Brasil em 2000. Por fim, o sub-índice é calculado através da fórmula: [ln(valor observado do indicador) - ln(1.560,17)]/ [ln(1.560,17) - ln(3,90)].

Após calcular os sub-índices: IDHM-E, para educação; IDHM-L, para saúde (ou longevidade); IDHM-R, para renda; obtem-se o IDHM, que é a média aritmética simples desses três sub-índices, ou seja

$$IDHM = \frac{IDHM\text{-}E + IDHM\text{-}L + IDHM\text{-}R}{3}.$$

Como foram fixados valores de referência de mínimo e de máximo em cada categoria, a esses valores corresponderão nos respectivos sub-índices os valores zero e um, respectivamente. Os sub-índices de cada município

serão valores proporcionais dentro dessa escala e portanto quanto melhor o desempenho municipal naquela dimensão, mais próximo o seu índice estará de 1, acontecendo o mesmo com o IDHM.

Para aplicar as técnicas de seleção de modelos objetivando selecionar o melhor modelo de regressão Beta para explicar o IDHM de 2000 dos municípios nordestinos foi então retirado do Atlas de Desenvolvimento Humano, tendo a preocupação de não utilizar informações que já faziam parte da composição do IDHM de 2000 (y_t), as informações municipais das seguintes variáveis:

- x_{t1}: IDHM de 1991.

- x_{t2}: índice de Gini, que mede o grau de desigualdade existente na distribuição de indivíduos segundo a renda domiciliar per capita. Seu valor varia de 0, quando não há desigualdade (a renda de todos os indivíduos tem o mesmo valor), a 1, quando a desigualdade é máxima (apenas um um indivíduo detém toda a renda da sociedade e a renda de todos os outros indivíduos é nula); que vamos representar por GINI.

- x_{t3}: proporção de pessoas com renda domiciliar per capita abaixo de R$ 37,75 (linha de indigência), equivalente a 1/4 do salário mínimo vigente em agosto de 2000. O universo de indivíduos é limitado àqueles que vivem em domicílios particulares permanentes; representado por INDIG.

- x_{t4}: proporção de pessoas que vivem em domicílios com água encanada, canalizada para um ou mais cômodos, proveniente de rede geral, de poço, de nascente ou de reservatório abastecido por água das chuvas ou carro-pipa; que vamos representar por ÁGUA.

- x_{t5}: proporção de pessoas que vivem em domicílios urbanos com serviço de coleta de lixo, ou seja, em domicílios em que a coleta de lixo é realizada diretamente por empresa pública ou privada, ou em que o lixo é depositado em caçamba, tanque ou depósito fora do domicílio, para posterior coleta pela prestadora do serviço; denominado por LIXO.

- x_{t6}: proporção de pessoas que vivem em domicílios com energia elétrica, proveniente ou não de uma rede geral, com ou sem medidor; que vamos representar por ENERGIA.

5.1 Análise Exploratória dos Dados

Antes da modelagem propriamente dita foram feitas análises descritivas/ exploratórias dos dados já referidos, cujos principais aspectos encontram-se a seguir descritos.

Inicialmente serão apresentadas as análises referentes ao IDHM calculado nos anos de 1991 e de 2000 para os municípios, em cada um dos estados, e para a região Nordeste como um todo.

Considerando a classificação já referida do desenvolvimento humano, os dados apresentados na Tabela A.9 inserida no Apêndice A, ano 1991 revelam que quatro estados: Alagoas, Maranhão, Paraíba e Piauí, foram considerados de baixo desenvolvimento humano levando-se em consideração a média do IDHM em cada estado, enquanto os demais, apesar de muito próximos do limite que os classificaria como de baixo desenvolvimento, foram classificados como de desenvolvimento humano médio, no entanto, todos os estados apresentaram desenvolvimento humano médio levando em consideração os seus respectivos IDHM, o que pode ser observado a partir do Atlas de Desenvolvimento Humano. Podemos notar ainda que os estados do Ceará, com 9.59%, e do Rio Grande do Norte, com 9.92%, apresentaram menores coeficientes de variação do IDHM, ou seja, eram os estados com municípios mais homogêneos, ao passo que Pernambuco e Piauí tiveram os maiores coeficientes de variação, 12.54% e 12.47%, respectivamente. Todos os estados apresentaram valores com assimetria positiva, isto é, a maioria dos municípios possuiu índice de desenvolvimento humano inferior ao médio dos seus respectivos estados, destacando apenas o estado do Piauí com um comportamento simétrico.

Observamos também que a região Nordeste como um todo, apresentou um IDHM médio de 0.503 quase no limite da faixa que a classificaria como de baixo desenvolvimento; no entanto ainda se classificando como médio, mantendo-se ainda como médio pela observação do IDHM da região.

Em 2000, como podemos observar também na Tabela A.9, todos os estados foram classificado com desenvolvimento humano médio, tanto quando observamos a média do IDHM quanto o IDHM de cada estado. O Ceará, por sua vez, apresentou o menor desvio-padrão, 0.037, e a menor dispersão relativa, 5.93%, indicando que o mesmo foi o estado com situação municipal mais homogênea da região Nordeste. Analisando os coeficientes de assimetria, constata-se, tal como no ano de 1991, uma assimetria positiva, exceto para o estado do Piauí com um comportamento simétrico.

Conforme pode-se constatar ainda na Tabela A.9 analisando os dados de cada um dos municípios da região Nordeste, em todos os estados da região, tanto em 1991 quanto no ano 2000 os municípios com os maiores valores do IDHM foram as capitais dos estados, exceto no estado de Pernambuco que o maior IDHM foi do distrito de Fernando de Noronha, cujos valores foram: 0.759 e 0.862 respectivamente nos anos 1991 e 2000. Vale destacar que a capital de Pernambuco, Recife obteve o segundo maior valor do IDHM no ano de 1991 (0.740) e o terceiro em 2000 (0.797), ficando em segundo o município de Paulista (0.799) que faz parte da Região Metropolitana do Recife.

Por outro lado, os menores valores do IDHM nos municípios nordestinos foram encontrados em pequenos municípios em termos de população e de área geográfica com valores bem baixos, variando de 0.323 observado em 1991 no município de Curral Novo do Piauí até 0.551 no município de Barroquinha no estado do Ceará. Analisando a variação do IDHM em cada município no período 1991/2000 constata-se que houve aumento em todos os municípios do Nordeste. Notamos ainda, no ano de 1991 que nenhum município obteve o IDHM classificado como alto e que apenas dois municípios tiveram o IDHM classificado como alto em 2000, a capital da Bahia, Salvador com 0.805 e o distrito de Fernando de Noronha em Pernambuco obtendo 0.862.

Podemos observar que o estado que tem o maior número de municípios é o da Bahia (415) e o menor é Sergipe (75), ambos tiveram desenvolvimento humano médio nas duas épocas. Vericamos que em todos os estados houve aumento do IDHM médio, sendo o Ceará o que mais aumentou e os estados do Maranhão e Sergipe os que menos aumentaram neste período. Notamos ainda que todos os desvios-padrão diminuíram, no período 1991/2000 e, consequentemente, as dispersões relativas também.

Analisando cada um dos sub-índices que compõe o IDHM, podemos observar a partir da Tabela A.10 no Apêndice A, referente ao ano de 1991, que em quase todos os estados, exceto Alagoas, as médias do sub-índice renda, foram inferiores ao do sub-índice educação, que ficaram por sua vez abaixo das do sub-índice longevidade, acontecendo o mesmo com a região Nordeste como um todo. Percebemos ainda que, com relação ao sub-índice renda todos se enquadram na classificação de índice de desenvolvimento humano baixo, enquanto que todos os sub-índices longevidade possuíam índice de desenvolvimento humano médio, inclusive na região Nordeste como um todo. A média do sub-índice educação em cinco estados apresentou um índice de desenvolvimento médio, estes sub-índices foram os que apresentaram maior dispersão

relativa, ou seja, menor homogeneidade. Podemos notar também que no tocante a renda, todos os estados possuíam assimetria positiva, levando-nos a concluir que a maioria dos municípios apresentaram índice inferior ao índice médio de seu estado.

Através da Tabela A.11 no Apêndice A, referente ao ano de 2000, percebemos que as médias do sub-índice renda foram menores que as do sub-índice de longevidade, tendo sido estas inferiores as do sub-índice educação, a exceção do estado de Alagoas. Vale salientar que apenas nos estados de Alagoas, Maranhão e Piauí, o sub-índice renda foi considerado baixo, sendo médio nos demais estados. No que se refere aos sub-índices educação e longevidade os valores podem ser classificados como médio. Em relação a assimetria da distribuição dos sub-índices, verificamos que em todos os estados no que tange ao sub-índice renda a assimetria foi positiva, enquanto no que se refere ao sub-índice longevidade tivemos distribuições simétricas, com ligeira assimetria negativa nos estados de Alagoas, da Bahia e do Ceará, e de assimetria positiva nos estados do Maranhão e da Paraíba. No sub-índice educação a distribuição foi simétrica nos estados da Bahia, do Maranhão e do Piauí, tendo uma assimetria positiva nos demais estados.

Para as demais variáveis relativas ao ano de 2000 aqui consideradas como variáveis com algum poder de explicação do IDHM no ano 2000, a Tabela A.12 também inserida no apêndice A apresenta algumas medidas descritivas para todos os municípios nordestinos, onde se pode destacar:

- que em média aproximadamente 77% das pessoas possuíam energia elétrica no Nordeste, enquanto que apenas 41,22% tinham água encanada;

- notamos que o índice de Gini foi o mais homogêneo, dentre as variáveis independentes apresentando o coeficiente de variação de 9.36%, no entanto a heterogeneidade dos municípios é bem maior quando se observam os valores de água encanada e a coleta de lixo com coeficiente de variação de 48.98% e 42.15%, respectivamente;

- as medianas do percentual de pessoas que vivem com água encanada e com energia elétrica são inferiores a 41% e 82.89%, respectivamente. Observamos também que com relação a proporção de pessoas com renda domiciliar per capita abaixo da linha de indigência, a mediana foi de 43.85%, ou seja, quase metade da população do município;

- analisando todas as medidas observamos que a distribuição de energia elétrica e a coleta de lixo são os indicadores que se apresentaram melhores.

5.2 Seleção do Modelo

Ao considerar os seis regressores mais o intercepto rotulamos os 127 modelos candidatos com a finalidade de apresentar os resultados obtidos por estado da região Nordeste, visando observar quais foram os modelos selecionados por cada um dos critérios. Se o modelo possui o mesmo rótulo então é composto das mesmas covariáveis.

Foram inicialmente apresentados e escolhidos os modelos para cada um dos estados da região Nordeste e, com os resultados apresentados na Tabela 5.2.1 adiante inserida, observamos que as correções do AICc e HQc não influenciam na escolha do modelo, exceto no estado do Maranhão onde encontramos AIC (33) e AICc (102) e, em Sergipe que possui o menor número de municípios onde encontramos HQ (102) e HQc (15). A exceção da Bahia e do Piauí os critérios BIC e HQc escolheram o mesmo modelo. Notamos também que nos estados de Alagoas, Piauí e Rio Grande do Norte cinco critérios selecionaram o mesmo modelo e ainda, que o critério mais selecionado no estado de Pernambuco não foi selecionado nos demais estados. Os resultados encontrados em cada estado da região Nordeste revelam que os diversos critérios não selecionaram o mesmo modelo e isto ocorre, sem dúvida, além de outros fatores, devido a quantidade de municípios em cada estado.

Todos os modelos selecionados nos estados e na região Nordeste contém o intercepto e a covariável IDHM de 1991, especificaremos as demais covariáveis que compõe cada modelo:

- Modelo 15: INDIG.
- Modelo 16: INDIG. e ÁGUA.
- Modelo 19: ÁGUA.
- Modelo 31: GINI e ENERGIA.
- Modelo 32: GINI, INDIG e LIXO.
- Modelo 33: GINI, INDIG, LIXO e ENERGIA.

- Modelo 34: GINI, INDIG e ENERGIA.
- Modelo 87: INDIG e ENERGIA.
- Modelo 88: INDIG, ÁGUA e ENERGIA.
- Modelo 100: Não contém outras covariáveis.
- Modelo 102: GINI e INDIG.
- Modelo 103: GINI, INDIG e ÁGUA.
- Modelo 104: GINI, INDIG, ÁGUA e LIXO.

Tabela 5.2.1: Rótulos dos Modelos Selecionados pelos Critérios Estudados por Estados, da Região Nordeste

Estados	Critérios					
	AIC	AICc	BIC	HQ	HQc	R_p^2
Alagoas	100	100	100	100	100	19
Bahia	104	104	102	31	31	104
Ceará	31	31	102	102	102	31
Maranhão	33	102	102	102	102	32
Paraíba	16	16	15	15	15	88
Pernambuco	87	87	87	87	87	33
Piauí	103	103	102	103	103	103
R. G. do Norte	103	103	103	103	103	34
Sergipe	103	103	15	102	15	103

Observamos da Tabela 5.2.1 já referida, que os critérios AIC e AICc escolheram os modelos diferentes unicamente no estado do Maranhão, o AIC escolheu uma variável a mais, mostrando talvez um problema de sobreajustamento por parte do AIC. Também poderia se suspeitar de problemas de sobreajustamento por parte de AIC e AICc nos modelos escolhidos para os estados de Bahia e Sergipe, e do HQ para o Sergipe. Para o estado de Piaui o critério BIC escolheu um modelo de dimensão menor que os outros critérios.

Considerando a região Nordeste como um todo, com seus 1787 municípios, foram ajustados os modelos e aplicados os diversos critérios de escolha do melhor modelo e, constatamos que dos 127 modelos candidatos todos os critérios escolheu o mesmo modelo (de rótulo 32) para região Nordeste, como era de se esperar, já que o número de municípios é grande (1787). As estatísticas de cada critério analisado e da *Deviance* do modelo selecionado estão apresentadas na Tabela 5.2.2 a seguir inserida.

Tabela 5.2.2: Estatísticas dos Critérios e da *Deviance*

Critérios	AIC	AICc	BIC	HQ
Estatísticas	-9336.89	-9336.83	-9248.47	-9322.70
Critérios	HQc	R_p^2	*Deviance*	
Estatísticas	-9322.56	0.84422	4.2274	

Notamos também que as estatísticas de alguns critérios estão bem próximas, pois as mesmas são mensuradas através da função de log-verossimilhança, diferenciadas apenas pela função de penalidade.

O modelo selecionado foi o seguinte

$$log\Big(\frac{\mu_t}{1-\mu_t}\Big) = -0.8962 + 2.606x_{t1} + 0.421x_{t2} - 0.517x_{t3} + 0.032x_{t5}, \quad (5.2.1)$$

e a estimativa do parâmetro de precisão $\hat{\phi}$ foi de 747.39 .

Notamos que existe uma relação negativa entre o IDHM de 2000 e a proporção de pessoas com renda domiciliar per capita abaixo da linha de indigência, ou seja, quanto maior a proporção de indigentes no município menor o Índice de Desenvolvimento Humano. Naturalmente confirmou-se que existe uma relação positiva com o IDHM de 1991 com uma estimativa do parâmetro associado de 2.606. O aumento do IDHM de 2000 é explicado pelo aumento do índice de Gini e da proporção de pessoas que vivem em domicílios urbanos com serviço de coleta de lixo, o que corresponde a dizer que no Nordeste do Brasil o aumento do Desenvolvimento Humano está associado ao aumento da desigualdade de renda e da melhoria das condições da coleta de lixo domiciliar.

Analisando a *deviance* = 4.2274 verificamos que o modelo está bem ajustado, pois como a amostra é grande comparamos com os percentis da Normal

padrão, com uma significância de 1.18e-05, ou seja, há evidências que o modelo escolhido está bem ajustado aos dados, a um nível de 1.18e-03% de significância. Verificamos que, como a estimativa do parâmetro de precisão foi grande a variância das estimativas da variável resposta é pequena.

A Tabela 5.2.3 a seguir, mostra que as covariáveis são estatisticamente significantes levando-se em consideração os níveis nominais usuais.

Tabela 5.2.3: Estimativas dos Parâmetros

Parâmetros	Estimativas	Desvio-padrão	Estatística z	Valor p
β_0	-0.8962	0.03285	-27.28	0.0000
β_1	2.6060	0.04946	52.69	0.0000
β_2	0.4214	0.03814	11.05	0.0000
β_3	-0.5173	0.02884	-17.94	0.0000
β_5	0.0326	0.00726	4.49	0.0000

Da comparação do modelo (5.2.1) com os modelos escolhidos para cada um dos estados destacamos a importância das variáveis percentual de indigentes e índice de Gini, além claro do IDHM de 1991.

Após uma análise detalhada da significância dos parâmetros associados a cada modelo e das análises de resíduos e de diagnóstico poderia-se escolher o modelo mais adequado nesta aplicação, porém os critérios indicam o caminho inicial dessa procura.

Capítulo 6

Conclusões

A partir dos resultados das simulações, constatamos que para o Modelo 1, os piores critérios de seleção de modelos foram o R_p^2 e o AIC independentemente do tamanho da amostra ($n = 20$, $n = 40$ e $n = 60$), enquanto que o AICc, HQc e BIC se apresentaram como os melhores, exceto para $n = 60$ e para $n = 200$ quando o HQ apresenta-se melhor que o AICc, no entanto para $n = 60$, as médias das eficiências observadas L_2 se apresentam muito próximas indicando situação de fronteira . Destacamos que o critério BIC sempre obteve escores 1 e 2, logo poderia ser considerado como o melhor.

Constatamos que num modelo "menos identificável" (Modelo 2) não foi possível eleger o melhor critério, confirmando os achados de McQuarrie e Tsai (1998) de que um critério nem sempre é melhor que outro, pois depende do modelo verdadeiro.

Confirmamos ainda que o número de vezes que o modelo verdadeiro é selecionado aumenta quando a amostra aumenta, independente do modelo verdadeiro.

Vale destacar que, para $n = 200$ Tabela A.4 e A.8 os dois Modelos apresentaram resultados bem próximos em todos os critérios aqui analizados e todos os critérios obedeceram a mesma ordem de classificação, o que nos leva a conclusão que quando o tamanho da amostra é suficientemente grande a ordem de classificação dos critérios independe do modelo verdadeiro.

Com relação a aplicação podemos destacar que:

- os modelos em cada estado da região Nordeste revelam que os diversos critérios não selecionaram o mesmo modelo e isto ocorre, sem dúvida, além de outros fatores, devido a quantidade de municípios em cada estado;

- para o conjunto de municípios da região Nordeste, todos os critérios selecionaram o mesmo modelo, provavelmente devido a região possuir um grande número de municípios (1787);

- para o modelo do Nordeste a relação positiva do IDHM ano 2000 com o IDHM ano 1991 e com a proporção de pessoas que vivem em domicílios urbanos com serviço de coleta de lixo, já era esperada. No entanto, com o índice de Gini esperava-se uma relação negativa, fato que não ocorreu tendo em vista que os municípios mais desenvolvidos do Nordeste apresentam uma grande desigualdade de renda. Observamos ainda uma relação negativa entre o IDHM ano 2000 e a proporção de indigentes, pois quando esta diminui o Índice de Desenvolvimento Humano aumenta.

Na comparação dos modelos escolhidos pelos diferentes critérios para a região Nordeste e para cada um dos estados concluimos da importância das variáveis: IDHM de 1991, percentual de indigentes e índice de Gini.

Nos resultados da aplicação observamos similares desempenhos nas simulações por parte de todos os critérios.

Apesar de não ser simples a tarefa de determinar qual é o critério de melhor desempenho. Vale a pena continuar a pesquisa neste assunto pois os critérios de seleção de modelos são muito importantes porque indicam o caminho inicial para escolha do modelo mais apropriado a um conjunto de dados.

Apêndice A

Tabelas

As tabelas aqui apresentadas são os resultados das simulações e das análises descritivas/exploratória das variáveis utilizadas na aplicação.

Tabela A.1: Freqüência do modelo selecionado e as eficiências observadas para o Modelo 1 com $n = 20$, para $R = 5000$.

C	Freqüência						
	AIC	AICc	BIC	HQ	HQc	R_p^2	L_2
1	0	2	2	0	5	0	0
2	1	21	9	1	44	0	0
3	29	205	100	42	329	23	0
4	256	977	532	302	1249	260	28
5	2063	3051	2642	2227	2949	1781	3681
6	1751	677	1317	1660	397	1916	1136
7	749	64	343	649	26	860	147
8	151	3	55	119	1	160	8
verdadeiro	1793	2691	2318	1939	2628	1368	3633

eficiência observada L_2

Medidas	AIC	AICc	BIC	HQ	HQc	R_p^2	L_2
Média	0.6816	0.7317	0.7135	0.6911	0.7134	0.6413	1.0000
Mediana	0.6871	0.9842	0.7734	0.6993	0.9796	0.6548	1.0000
Desvio-Padrão	0.2881	0.3152	0.3009	0.2911	0.3270	0.2992	0.0000
C.V.(%)	42.27	43.08	42.17	42.12	45.84	46.66	0.00
Escore	5	1	2	4	3	6	

Tabela A.2: Freqüência do modelo selecionado e as eficiências observadas para o Modelo 1 com $n = 40$, para $R = 5000$.

C	Freqüência						
	AIC	AICc	BIC	HQ	HQc	R_p^2	L_2
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	5	11	26	7	26	42	0
5	2599	3522	3952	3237	4102	1818	4066
6	1820	1277	905	1455	805	2132	846
7	520	180	110	276	64	882	84
8	56	10	7	25	3	126	4
verdadeiro	2593	3514	3942	3231	4092	1743	4066

eficiência observada L_2

Medidas	AIC	AICc	BIC	HQ	HQc	R_p^2	L_2
Média	0.7578	0.8375	0.8771	0.8125	0.8930	0.7029	1.0000
Mediana	0.9889	1.0000	1.0000	1.0000	1.0000	0.7526	1.0000
Desvio-Padrão	0.2850	0.2687	0.2495	0.2769	0.2373	0.2904	0.0000
C.V.(%)	37.61	32.09	28.45	34.08	26.57	41.31	0.00
Escore	5	3	2	4	1	6	

Tabela A.3: Freqüência do modelo selecionado e as eficiências observadas para o Modelo 1 com $n = 60$, para $R = 5000$.

C	Freqüência						
	AIC	AICc	BIC	HQ	HQc	R_p^2	L_2
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	0	0	0	0	0	21	0
5	2662	3227	4206	3446	4024	1893	4219
6	1844	1524	731	1332	892	2133	725
7	455	241	61	215	81	839	55
8	39	8	2	7	3	114	1
verdadeiro	2662	3227	4205	3446	4023	1850	4219

eficiência observada L_2

Medidas	AIC	AICc	BIC	HQ	HQc	R_p^2	L_2
Média	0.7526	0.8058	0.9034	0.8259	0.8838	0.7035	1.0000
Mediana	0.9955	1.0000	1.0000	1.0000	1.0000	0.7521	1.0000
Desvio-Padrão	0.2934	0.2835	0.2307	0.2774	0.2464	0.2932	0.0000
C.V.(%)	38.98	35.18	25.54	33.59	27.88	41.68	0.00
Escore	5	4	1	3	2	6	

Tabela A.4: Freqüência do modelo selecionado e as eficiências observadas para o Modelo 1 com $n = 200$, para $R = 5000$.

C	Freqüência						
	AIC	AICc	BIC	HQ	HQc	R_p^2	L_2
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	0	0	0	0	0	8	0
5	2953	3124	4633	3966	4116	2064	4519
6	1678	1574	357	964	840	2112	465
7	345	286	10	67	41	730	16
8	24	16	0	3	3	86	0
verdadeiro	2953	3124	4633	3966	4116	2052	4519

eficiência observada L_2

Medidas	AIC	AICc	BIC	HQ	HQc	R_p^2	L_2
Média	0.7778	0.7933	0.9501	0.8751	0.8917	0.7245	1.0000
Mediana	1.0000	1.0000	1.0000	1.0000	1.0000	0.7860	1.0000
Desvio-Padrão	0.2933	0.2906	0.1817	0.2568	0.2442	0.2879	0.0000
C.V.(%)	37.71	36.33	19.12	29.35	27.39	39.74	0.00
Escore	5	4	1	3	2	6	

Tabela A.5: Freqüência do modelo selecionado e as eficiências observadas para o Modelo 2 com $n = 20$, para $R = 5000$.

C	Freqüência						
	AIC	AICc	BIC	HQ	HQc	R_p^2	L_2
1	0	11	10	2	17	0	0
2	91	370	278	121	573	24	6
3	538	1508	1060	651	1831	258	120
4	1397	1980	1723	1482	1861	966	1011
5	1572	931	1264	1524	617	1720	3248
6	1003	181	507	882	90	1384	568
7	343	18	142	296	11	560	47
8	56	1	16	42	0	88	0
verdadeiro	552	394	492	552	283	547	2628

eficiência observada L_2

Medidas	AIC	AICc	BIC	HQ	HQc	R_p^2	L_2
Média	0.5682	0.5522	0.5595	0.5692	0.5355	0.5698	1.0000
Mediana	0.5436	0.5278	0.5353	0.5458	0.5076	0.5472	1.0000
Desvio-Padrão	0.2529	0.2568	0.2571	0.2549	0.2536	0.2495	0.0000
C.V.(%)	45.51	46.50	45.95	44.78	47.36	43.79	0.00
Escore	3	6	4	2	5	1	

Tabela A.6: Freqüência do modelo selecionado e as eficiências observadas para o Modelo 2 com $n = 40$, para $R = 5000$.

C	Freqüência						
	AIC	AICc	BIC	HQ	HQc	R_p^2	L_2
1	0	0	1	0	0	0	0
2	1	5	45	7	33	1	0
3	150	301	660	292	613	48	5
4	992	1569	1904	1417	2006	480	248
5	2239	2310	1929	2203	1996	1772	4173
6	1243	703	399	889	311	1824	539
7	337	107	61	176	41	750	33
8	38	5	1	16	0	125	2
verdadeiro	1538	1628	1439	1564	1481	1109	3994

eficiência observada L_2

Medidas	AIC	AICc	BIC	HQ	HQc	R_p^2	L_2
Média	0.6600	0.6606	0.6280	0.6550	0.6340	0.6393	1.0000
Mediana	0.6437	0.6415	0.5983	0.6319	0.6074	0.6343	1.0000
Desvio-Padrão	0.2851	0.2929	0.2979	0.2906	0.2986	0.2698	0.0000
C.V.(%)	43.20	44.34	47.44	44.37	47.10	42.20	0.00
Escore	2	1	6	3	5	4	

Tabela A.7: Freqüência do modelo selecionado e as eficiências observadas para o Modelo 2 com $n = 60$, para $R = 5000$.

C	Freqüência						
	AIC	AICc	BIC	HQ	HQc	R_p^2	L_2
1	0	0	0	0	0	0	0
2	0	1	10	1	2	0	0
3	34	58	277	87	165	7	1
4	568	831	1621	1023	1410	237	69
5	2561	2840	2631	2817	2825	1757	4410
6	1442	1090	436	925	560	1994	496
7	372	175	25	142	38	874	24
8	23	5	0	5	0	131	0
verdadeiro	2174	2438	2334	2421	2460	1421	4353

eficiência observada L_2

Medidas	AIC	AICc	BIC	HQ	HQc	R_p^2	L_2
Média	0.7188	0.7371	0.7073	0.7313	0.7269	0.6754	1.0000
Mediana	0.7384	0.8841	0.7707	0.8675	0.9372	0.6689	1.0000
Desvio-Padrão	0.2855	0.2904	0.3083	0.2941	0.3022	0.2683	0.0000
C.V.(%)	39.72	39.40	43.59	40.22	41.57	39.72	0.00
Escore	4	1	5	2	3	6	

Tabela A.8: Freqüência do modelo selecionado e as eficiências observadas para o Modelo 2 com $n = 200$, para $R = 5000$.

C	Freqüência						
	AIC	AICc	BIC	HQ	HQc	R_p^2	L_2
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	5	6	87	21	25	2	0
5	2953	3099	4574	3999	4110	1707	4703
6	1676	1598	326	891	795	2237	288
7	340	278	13	84	70	907	9
8	26	19	0	5	0	147	0
verdadeiro	2951	3096	4567	3995	4103	1704	4703

eficiência observada L_2

Medidas	AIC	AICc	BIC	HQ	HQc	R_p^2	L_2
Média	0.7912	0.8038	0.9452	0.8852	0.8966	0.7117	1.0000
Mediana	1.0000	1.0000	1.0000	1.0000	1.0000	0.7342	1.0000
Desvio-Padrão	0.2772	0.2747	0.1831	0.2411	0.2324	0.2680	0.0000
C.V.(%)	35.04	34.18	19.37	27.24	25.92	37.66	0.00
Escore	5	4	1	3	2	6	

Tabela A.9: Medidas Estatísticas do IDHM por Estado, na Região Nordeste, 1991/2000.

Estados / Anos	Medidas				
	Municípios	Média	D.P.	C.V.(%)	Assim.
Alagoas					
1991	101	0.472	0.051	10.76	0.868
2000	101	0.583	0.045	7.65	0.525
Bahia					
1991	415	0.516	0.054	10.39	0.541
2000	415	0.626	0.046	6.86	0.548
Ceará					
1991	184	0.509	0.046	9.59	0.686
2000	184	0.631	0.037	5.93	0.555
Maranhão					
1991	217	0.483	0.054	11.27	0.709
2000	217	0.581	0.047	8.11	0.620
Paraíba					
1991	223	0.482	0.053	10.97	0.809
2000	223	0.592	0.044	7.44	0.847
Pernambuco					
1991	185	0.526	0.066	12.54	1.053
2000	185	0.626	0.057	9.14	0.901
Piauí					
1991	221	0.484	0.060	12.47	0.210
2000	221	0.587	0.046	7.80	0.353
R.G.do Norte					
1991	166	0.522	0.052	9.92	0.767
2000	166	0.637	0.043	6.75	0.859
Sergipe					
1991	75	0.524	0.054	10.33	0.730
2000	75	0.622	0.043	6.97	0.755
Nordeste					
1991	1787	0.503	0.058	11.53	0.592
2000	1787	0.610	0.050	8.17	0.455

continua

Tabela A.9: Medidas Estatísticas do IDHM por Estado, na Região Nordeste, 1991/2000.

Estados / Anos	Medidas				
	Curtose	Quartil1	Mediana	Quartil3	Amplit.
Alagoas					
1991	2.193	0.436	0.465	0.503	0.321
2000	1.046	0.558	0.578	0.609	0.260
Bahia					
1991	0.970	0.481	0.511	0.548	0.374
2000	0.813	0.598	0.621	0.651	0.284
Ceará					
1991	1.697	0.480	0.505	0.537	0.314
2000	1.120	0.606	0.631	0.652	0.235
Maranhão					
1991	1.482	0.447	0.478	0.520	0.355
2000	1.214	0.549	0.572	0.613	0.294
Paraíba					
1991	1.720	0.446	0.477	0.510	0.356
2000	1.720	0.560	0.588	0.614	0.289
Pernambuco					
1991	1.498	0.483	0.507	0.558	0.400
2000	1.771	0.587	0.617	0.654	0.395
Piauí					
1991	0.278	0.442	0.485	0.524	0.390
2000	0.542	0.558	0.583	0.618	0.287
R.G.do Norte					
1991	1.192	0.489	0.513	0.547	0.322
2000	0.885	0.606	0.631	0.655	0.244
Sergipe					
1991	1.941	0.485	0.517	0.557	0.329
2000	2.068	0.594	0.618	0.652	0.258
Nordeste					
1991	1.238	0.465	0.499	0.535	0.436
2000	0.959	0.577	0.608	0.639	0.395

continuação

Tabela A.9: Medidas Estatísticas do IDHM por Estado, na Região Nordeste, 1991/2000.

conclusão

Estados	Medidas	
Anos	Mínimo	Máximo
Alagoas		
1991	0.366 São José de Tapera	0.687 Maceió
2000	0.479 Traipu	0.739 Maceió
Bahia		
1991	0.377 Quijingue	0.751 Salvador
2000	0.521 Itapicuru	0.805 Salvador
Ceará		
1991	0.403 Barroquinha	0.717 Fortaleza
2000	0.551 Barroquinha	0.786 Fortaleza
Maranhão		
1991	0.366 Lagoa Grande do MA	0.721 São Luís
2000	0.484 Centro de Guilherme	0.778 São Luís
Paraíba		
1991	0.363 Damião	0.719 João Pessoa
2000	0.494 Cacimbas	0.783 João Pessoa
Pernambuco		
1991	0.359 Manari	0.759 Fernando de Noronha
2000	0.467 Manari	0.862 Fernando de Noronha
Piauí		
1991	0.323 Curral Novo do Piauí	0.713 Teresina
2000	0.479 Guaribas	0.766 Teresina
R. G. do Norte		
1991	0.411 Bodó	0.733 Natal
2000	0.544 Venha-ver	0.788 Natal
Sergipe		
1991	0.405 Poço Redondo	0.734 Aracajú
2000	0.536 Poço Redondo	0.794 Aracajú
Nordeste		
1991	0.323 Curral Novo do Piauí	0.759 Fernando de Noronha
2000	0.467 Manari	0.862 Fernando de Noronha

Fonte: Atlas do Desenvolvimento Humano.

Tabela A.10: Medidas Estatísticas dos Sub-índices por Estado, na Região Nordeste, 1991.

Estados / Sub-índices	Medidas			
	Média	Desvio-Padrão	C. V. (%)	Assimetria
Alagoas				
IDHM-E	0.438	0.086	19.71	0.635
IDHM-L	0.524	0.050	9.59	0.267
IDHM-R	0.455	0.049	10.75	1.154
IDHM	0.472	0.051	10.76	0.868
Bahia				
IDHM-E	0.516	0.099	19.23	0.271
IDHM-L	0.566	0.056	9.85	-0.454
IDHM-R	0.467	0.054	11.48	0.921
IDHM	0.516	0.054	10.39	0.541
Ceará				
IDHM-E	0.501	0.071	14.20	0.503
IDHM-L	0.581	0.044	7.63	-0.291
IDHM-R	0.445	0.050	11.14	0.812
IDHM	0.509	0.046	9.59	0.686
Maranhão				
IDHM-E	0.488	0.108	22.06	0.348
IDHM-L	0.532	0.045	8.47	0.215
IDHM-R	0.430	0.049	11.31	0.934
IDHM	0.483	0.054	11.27	0.709
Paraíba				
IDHM-E	0.478	0.088	18.29	0.376
IDHM-L	0.533	0.057	10.75	0.334
IDHM-R	0.434	0.048	11.11	1.186
IDHM	0.482	0.053	10.97	0.809

continua

Tabela A.10: Medidas Estatísticas dos Sub-índices por Estado, na Região Nordeste, 1991.

conclusão

Estados	Medidas			
Sub-índices	Média	Desvio-Padrão	C. V. (%)	Assimetria
Pernambuco				
IDHM-E	0.521	0.099	19.02	0.698
IDHM-L	0.576	0.065	11.31	0.472
IDHM-R	0.482	0.060	12.47	1.016
IDHM	0.526	0.066	12.54	1.053
Piauí				
IDHM-E	0.478	0.106	22.12	-0.168
IDHM-L	0.559	0.054	9.70	-0.281
IDHM-R	0.414	0.051	12.44	0.993
IDHM	0.484	0.060	12.47	0.210
R. G. do Norte				
IDHM-E	0.547	0.078	14.18	0.372
IDHM-L	0.553	0.053	9.56	0.166
IDHM-R	0.465	0.049	10.65	1.060
IDHM	0.522	0.052	9.92	0.767
Sergipe				
IDHM-E	0.536	0.096	17.97	0.317
IDHM-L	0.554	0.055	9.93	0.471
IDHM-R	0.483	0.047	9.78	2.084
IDHM	0.524	0.054	10.33	0.730
Nordeste				
IDHM-E	0.502	0.098	19.52	0.239
IDHM-L	0.555	0.057	10.27	0.066
IDHM-R	0.451	0.056	12.42	0.843
IDHM	0.503	0.058	11.53	0.592

Fonte: Atlas do Desenvolvimento Humano.

Tabela A.11: Medidas Estatísticas dos Sub-índices por Estado, na Região Nordeste, 2000.

Estados / Sub-índices	Medidas			
	Média	Desvio-Padrão	C. V. (%)	Assimetria
Alagoas				
IDHM-E	0.629	0.057	9.07	0.769
IDHM-L	0.634	0.050	7.93	-0.511
IDHM-R	0.486	0.050	10.23	1.008
IDHM	0.583	0.045	7.65	0.525
Bahia				
IDHM-E	0.720	0.060	8.33	0.292
IDHM-L	0.636	0.056	8.78	-0.344
IDHM-R	0.521	0.052	9.98	0.815
IDHM	0.626	0.046	6.86	0.548
Ceará				
IDHM-E	0.704	0.047	6.64	0.363
IDHM-L	0.687	0.047	6.86	-0.342
IDHM-R	0.503	0.045	8.97	1.007
IDHM	0.631	0.037	5.93	0.555
Maranhão				
IDHM-E	0.685	0.068	9.89	0.125
IDHM-L	0.586	0.045	7.65	0.431
IDHM-R	0.472	0.055	11.60	0.863
IDHM	0.581	0.047	8.11	0.620
Paraíba				
IDHM-E	0.669	0.057	8.58	0.507
IDHM-L	0.605	0.054	9.00	0.408
IDHM-R	0.502	0.044	8.78	1.639
IDHM	0.592	0.044	7.44	0.847

continua

Tabela A.11: Medidas Estatísticas dos Sub-índices por Estado, na Região Nordeste, 2000.

conclusão

Estados / Sub-índices	Medidas			
	Média	Desvio-Padrão	C. V. (%)	Assimetria
Pernambuco				
IDHM-E	0.682	0.068	9.93	0.962
IDHM-L	0.668	0.063	9.43	0.133
IDHM-R	0.528	0.063	11.80	1.189
IDHM	0.626	0.057	9.14	0.901
Piauí				
IDHM-E	0.662	0.060	9.15	0.102
IDHM-L	0.617	0.058	9.44	-0.008
IDHM-R	0.482	0.047	9.78	0.974
IDHM	0.587	0.046	7.80	0.353
R. G. do Norte				
IDHM-E	0.711	0.051	7.21	0.526
IDHM-L	0.676	0.052	7.74	0.180
IDHM-R	0.522	0.050	9.67	1.040
IDHM	0.637	0.043	6.75	0.859
Sergipe				
IDHM-E	0.717	0.056	7.88	0.638
IDHM-L	0.627	0.055	8.80	0.249
IDHM-R	0.520	0.045	8.66	1.618
IDHM	0.622	0.043	6.97	0.755
Nordeste				
IDHM-E	0.690	0.065	9.42	0.243
IDHM-L	0.636	0.063	9.91	0.055
IDHM-R	0.505	0.054	10.69	0.892
IDHM	0.610	0.050	8.17	0.455

Fonte: Atlas do Desenvolvimento Humano.

Tabela A.12: Medidas Descritivas das Covariáveis para todos os Municípios da Região Nordeste, 2000.

Medidas	Covariáveis				
	GINI	INDIG	AGUA	LIXO	ENERGIA
Média	0.5780	0.4358	0.4122	0.6670	0.7766
Desvio-padrão	0.054	0.112	0.202	0.281	0.180
C.V.(%)	9.36	25.80	48.98	42.15	23.11
Assimetria	0.4446	-0.2074	0.1149	-0.9927	-0.9518
Curtose	0.5157	0.0335	0.5192	0.1746	0.1596
Quartil1	0.5400	0.3638	0.2634	0.5231	0.6630
Mediana	0.5700	0.4385	0.4100	0.7601	0.8289
Quartil3	0.6100	0.5144	0.5570	0.8889	0.9225
Mínimo	0.3600	0.0002	0.0000	0.0000	0.1743
Máximo	0.8000	0.8165	0.9632	1.0000	0.9999

Fonte: Atlas do Desenvolvimento Humano.

Apêndice B

Programa de Simulação

Neste apêndice apresentamos o programa de simulação utilizado neste trabalho; o programa foi desenvolvido na linguagem de programação **Ox**. No programa encontra-se as simulações de Monte Carlo, contém as funções que permitem o cálculo das estimativas dos parâmetros, as funções referentes aos vários critérios, das eficiências observadas e suas medidas descritivas.

```
/*********************************************************************
 *
 * PROGRAM: criteria.ox
 *
 * USE: Simulates maximum likelihood estimation of the Beta
 *      distribution parameters, and also model selection criteria.
 *      The model considered is a bBta regression model with a new
 *      scheme of parameters (Ferrari e Cribari-Neto, 2004).
 *
 * AUTHOR: Silvia Teresa Freire Torres
 *
 * DATE: Jule 2, 2004.
 *
 * VERSION: 0.1.7
 *
 * LAST MODIFIED: December 10, 2004.
 *
 *********************************************************************/

 /*Arquivos principais */
 #include <oxstd.h>
 #include <oxfloat.h>
 #include <oxprob.h>
 #import <maximize>
 #import <solvenle>

 /*Variáveis globais */
 static decl s_vy;
 static decl s_mX;
 static decl s_vlogy;
 static decl s_vlog1minusy;
 static decl parametro;
```

```
 static decl g_mY;

//função de log-verossimilhança do modelo Beta

floglik(const vP, const adFunc, const avScore, const amHess)
{
    decl k = rows(vP)-1;
    decl eta = s_mX*vP[0:(k-1)];
    decl mu = exp(eta)./(1.0+exp(eta));
    decl phi = vP[k];
    decl ynew = log(s_vy./(1.0-s_vy));
    decl munew = polygamma(mu*phi,0)-polygamma((1.0-mu)*phi,0);
    decl T = diag(exp(eta) ./(1.0+exp(eta)).^2);
        adFunc[0] = double(sumc(loggamma(phi) - loggamma(mu*phi)
            -loggamma((1-mu)*phi) + (mu*phi-1) .* log(s_vy)
            + ((1-mu)*phi-1) .*log(1-s_vy)));

//Derivadas analíticas da função de log-verossimilhança
  if(avScore)
  {
    (avScore[0])[0:(k-1)] = phi*s_mX'*T*(ynew-munew);
    (avScore[0])[k] = double(sumc(polygamma(phi,0)- mu .*
    polygamma(mu*phi,0) - (1.0-mu) .* polygamma((1.0-mu)*phi,0)+
    mu.*log(s_vy)+(1.0-mu).*log(1.0-s_vy)));
  }
    if(isnan(adFunc[0]) || isdotinf(adFunc[0]))
   return 0;
   else
   return 1; //1 indica sucesso
}
fgerador_vY(const func, const v_mu, const d_phi, const cn)
{
    decl ci;
    decl vecY=zeros(cn,1);
    for (ci=0;ci<cn;ci++)
     {
      decl p=(v_mu[ci]*d_phi);
      decl q=((1-v_mu[ci])*d_phi);
      vecY[ci]=ranbeta(1,1,p,q);
     }
     func[0]=vecY;
     return 1;
}

/*Parâmetros da simulação*/
 const decl INREP = 5000; //numero de replicas Monte Carlo
 const decl INMIN = 20;  //minimo tamanho de amostra
 const decl INMAX = 60;  //maximo tamanho de amostra

/*Corpo principal*/
 main()
 {

// Declaração de vaiaveis usadas
decl ir1, cfailure, cn, vp1, dExecTime, v_eta, v_mu, cs, mmle, v_zEstim,
     v_eEstim, phi_ini, dfunc1, p, q, v_theta, mResMean, vx, vbeta,
     d_phi, v_z, v_betaini, sigmaEst, v_thetaIni, d_denom, m_inv,
```

```
     b_SomaQua, v_muVee, v_sigma2Est, a, b, c, d, mx, k, e, f, g, h,
     i, j, dfunc, npar, gerY, gerY1, b_totalQua, RQua, v_etafinal, v_erro,
     b_erroQua, eta1, v_muverdadeiro,  s_mx, r, v_total, v_zmedia, vlogver,
     v_varassin, mX, vAIC, vlogverossimilhanca, pos, AICck1, AICck2, AICck3,
     AICck4, AICck5, AICck6, AICck7, AICck8, AICreal, AIC, mAIC, AICc,
     AICcck1, AICcck2, AICcck3, AICcck4, AICcck5, AICcck6, AICcck7, AICcck8,
     AICcreal, mAICc, BIC, BICck1, BICck2, BICck3, BICck4, BICck5, BICck6,
     BICck7, BICck8, BICreal, mBIC, mHQ, HQ, HQck1, HQck2, HQck3, HQck4,
     HQck5, HQck6, HQck7, HQck8, HQreal, HQc, mHQc, HQcck1, HQcck2, HQcck3,
     HQcck4, HQcck5, HQcck6, HQcck7, HQcck8, HQcreal, mRQua, RQck1, RQck2,
     RQck3, RQck4, RQck5, RQck6, RQck7, RQck8, RQreal, vp2, v_mu1, Dck1,
     Dck2, Dck3, Dck4, Dck5, Dck6, Dck7, Dck8, Dreal, Dev, mDev, L2ck1,
     L2ck2, L2ck3, L2ck4, L2ck5, L2ck6, L2ck7, L2ck8, L2real, L2, mL2,
     ncombinacoes, posicaoreal, L2EficObservAIC, L2EficObservAICc,
     L2EficObservBIC, L2EficObservHQ, L2EficObservHQc, L2EficObservRQ,
     MediaL2EficObservAIC, DPadraoL2EficObservAIC, MedianaL2EficObservAIC,
     MediaL2EficObservAICc, DPadraoL2EficObservAICc, MedianaL2EficObservAICc,
     MediaL2EficObservBIC, DPadraoL2EficObservBIC, MedianaL2EficObservBIC,
     MediaL2EficObservHQ, DPadraoL2EficObservHQ, MedianaL2EficObservHQ,
     MediaL2EficObservHQc, DPadraoL2EficObservHQc, MedianaL2EficObservHQc,
     MediaL2EficObservRQ, DPadraoL2EficObservRQ, MedianaL2EficObservRQ,
     MediaL2EficObservDev, DPadraoL2EficObservDev, MedianaL2EficObservDev,
     L2EficObservDev, logv;

// arranque do relogio para o cálculo do tempo de execução
dExecTime = timer();

// turn off warnings
oxwarning( 0 );

// Gerador de números uniformes pseudo-aleatórios. "George Marsaglia"
ranseed("GM");

// Parametros verdaderos do modelo
    vbeta=<1;1;1/2;1/3;1/4;1/5;1/6;1/7>;
// vbeta=ones(8,1);
println("beta=",vbeta');
d_phi = 120;          //phi

ranseed({1965, 2001});

posicaoreal=223;
// inicializacao do contador de falhas da convergencia
cfailure = 0;

ranseed( -1 );

//Loop de Monte Carlo
for (cn=INMIN;cn<=INMAX;cn += 20)
{
 AICck1=AICck2=AICck3=AICck4=AICck5=AICck6=AICck7=AICck8=AICreal=AICcck1=
 AICcck2=AICcck3=AICcck4=AICcck5=AICcck6=AICcck7=AICcck8=AICcreal=BICck1=
 BICck2=BICck3=BICck4=BICck5=BICck6=BICck7=BICck8=BICreal=HQck1=HQck2=
 HQck3=HQck4=HQck5=HQck6=HQck7=HQck8=HQreal=HQcck1=HQcck2=HQcck3=HQcck4=
 HQcck5=HQcck6=HQcck7=HQcck8=HQcreal=RQck1=RQck2=RQck3=RQck4=RQck5=RQck6=
 RQck7=RQck8=RQreal=Dck1=Dck2=Dck3=Dck4=Dck5=Dck6=Dck7=Dck8=Dreal=L2ck1=
 L2ck2=L2ck3=L2ck4=L2ck5=L2ck6=L2ck7=L2ck8=L2real=0;
```

```
L2EficObservAIC=L2EficObservAICc=L2EficObservBIC=L2EficObservHQ=
L2EficObservHQc=L2EficObservRQ=L2EficObservDev=zeros(1,INREP);

   for (cs=0;cs<INREP;++cs)
   {
   // matriz de dados explicativos
   vx =ones(cn,1)~ranexp(cn,7,3.0);//gerador dos regressores X
   mX=vx[][:4];
   eta1=mX*vbeta[0:4];
   v_muverdadeiro=exp(eta1)./(1.0+exp(eta1));

   fgerador_vY(&gerY1,v_muverdadeiro, d_phi, cn);//gerador de Y
   s_vy=gerY1;

   s_mX=mX;
   npar=columns(s_mX); // numero de parametros do modelo
   v_eta=s_mX*vbeta[0:(npar-1)];
   v_mu1=exp(v_eta)./(1.0+exp(v_eta));
   v_z = log(s_vy ./ (1.0 - s_vy));
   v_zmedia = meanc(v_z);
   v_total = v_z-meanc(v_z);
   b_totalQua = v_total'*v_total;

   s_mx = vx;  //matriz de regressores X
   ncombinacoes=255;
   mmle=zeros(9,ncombinacoes);
   AIC=zeros(1,ncombinacoes);
   AICc=zeros(1,ncombinacoes);
   BIC=zeros(1,ncombinacoes);
   HQ=zeros(1,ncombinacoes);
   L2=zeros(1,ncombinacoes);
   HQc=zeros(1,ncombinacoes);
   vlogver=zeros(1,ncombinacoes);
   k=zeros(1,ncombinacoes);
   RQua=zeros(1,ncombinacoes);
   Dev=zeros(1,ncombinacoes);
   r=0;
   for (a=0;a<columns(s_mx);++a)
   {
     for (b=a;b<columns(s_mx);++b)
     {
       for (c=b+2;c<columns(s_mx);++c)
       {
        for (d=c;d<columns(s_mx);++d)
        {
         s_mX=s_mx[][a:b]~s_mx[][c:d];
         p=columns(s_mX);
         k[][r]=columns(s_mX);;
         npar=columns(s_mX); // numero de parametros do modelo
         v_eta=s_mX*vbeta[0:(npar-1)];
         v_mu=exp(v_eta)./(1.0+exp(v_eta));
         d_denom = cn-npar;
         m_inv = invertsym(s_mX'*s_mX)*s_mX';  //inversao auxiliar

  //Regressão auxiliar para determinar pontos inicias no momento
 //da maximizacao por BFGS
       v_betaini = m_inv*v_z; //estimacao inicial dos betas
```

```
      v_zEstim = s_mX*v_betaini; // vetor v_z estimado
      v_eEstim = v_z-v_zEstim; // vetor de residuos estimados
      b_SomaQua = v_eEstim'*v_eEstim; // soma de quadrados
      v_muVee = exp(v_zEstim) ./ (1.0+exp(v_zEstim));  //vetor mu transformado
      v_sigma2Est = b_SomaQua/(d_denom); //*v_invgDerPrima2;
      phi_ini = double(meanc( 1.0 ./ (v_sigma2Est*
               (v_muVee .* (1.0-v_muVee)))) - 1.0); //phi inicial
      v_thetaIni = v_betaini|phi_ini; //vetor theta inicial para a maximizacao

 // valores de arranque
     vp1 = v_thetaIni;

 //Maximizacao da verossimilhanca por o Metodo BFGS
//parametros da maximizacao por BFGS
     MaxControl( 50, -1 );

 // Estimacao por maxima verossimilhanca
     ir1 = MaxBFGS(floglik, &vp1, &dfunc1, 0, FALSE);
     if( ir1 == MAX_CONV || ir1 == MAX_WEAK_CONV )
     {
       v_etafinal =s_mX*vp1[0:(npar-1)];
       v_erro = v_z-v_etafinal;
       b_erroQua = v_erro'*v_erro;
       v_mu=exp(v_etafinal)./(1.0+exp(v_etafinal));
       L2[][r] = (v_mu1-v_mu)'*(v_mu1-v_mu);
       mmle[][r] = vp1;
       vlogver[][r]=dfunc1;
       AIC[][r] = -2*dfunc1+2*(npar+1);
       AICc[][r] = -2*dfunc1+(2*(npar+1)*cn)/(cn-npar-2);
       BIC[][r] = -2*dfunc1+(npar+1)*log(cn);
       HQ[][r] = -2*dfunc1+2*(npar+1)*log(log(cn));
       HQc[][r] = -2*dfunc1+2*cn*(npar+1)*log(log(cn))/(cn-npar-3);
       RQua[][r] = 1-((b_erroQua/b_totalQua)*((cn-1)/(cn-npar)));
       ++r;
       }
      }
     }
    }
   }
for (a=0;a<columns(s_mx);++a)
{
 for (b=a;b<columns(s_mx);++b)
 {
  for (c=b+2;c<columns(s_mx);++c)
  {
   for (d=c;d<columns(s_mx);++d)
   {
    for (e=d+2;e<columns(s_mx);++e)
    {
      for (f=e;f<columns(s_mx);++f)
      {
       s_mX=s_mx[][a:b]~s_mx[][c:d]~s_mx[][e:f];
       p=columns(s_mX);
       k[][r]=columns(s_mX);;
       npar=columns(s_mX); // numero de parametros do modelo
       v_eta=s_mX*vbeta[0:(npar-1)];
       v_mu=exp(v_eta)./(1.0+exp(v_eta));
```

```
      d_denom = cn-npar;
      m_inv = invertsym(s_mX'*s_mX)*s_mX';  //inversao auxiliar

 //Regressão auxiliar para determinar pontos inicias no momento
//da maximizacao por BFGS
      v_betaini = m_inv*v_z; //estimacao inicial dos betas
      v_zEstim = s_mX*v_betaini; // vetor v_z estimado
      v_eEstim = v_z-v_zEstim; // vetor de residuos estimados
      b_SomaQua = v_eEstim'*v_eEstim; // soma de quadrados
      v_muVee = exp(v_zEstim) ./ (1.0+exp(v_zEstim));  //vetor mu transformado
      v_sigma2Est = b_SomaQua/(d_denom); //*v_invgDerPrima2;
      phi_ini = double(meanc( 1.0 ./ (v_sigma2Est*
                (v_muVee .* (1.0-v_muVee)))) - 1.0); //phi inicial
      v_thetaIni = v_betaini|phi_ini; //vetor theta inicial para a maximizacao

 // valores de arranque
     vp1 = v_thetaIni;

 //Maximizacao da verossimilhanca por o Metodo BFGS
//parametros da maximizacao por BFGS
     MaxControl( 50, -1 );

 // Estimacao por maxima verossimilhanca
     ir1 = MaxBFGS(floglik, &vp1, &dfunc1, 0, FALSE);
     if( ir1 == MAX_CONV || ir1 == MAX_WEAK_CONV )
     {
       v_etafinal =s_mX*vp1[0:(npar-1)];
       v_erro = v_z-v_etafinal;
       b_erroQua = v_erro'*v_erro;
       v_mu=exp(v_etafinal)./(1.0+exp(v_etafinal));
       L2[][r] = (v_mu1-v_mu)'*(v_mu1-v_mu);
       mmle[][r] = vp1;
       vlogver[][r]=dfunc1;
       AIC[][r] = -2*dfunc1+2*(npar+1);
       AICc[][r] = -2*dfunc1+(2*(npar+1)*cn)/(cn-npar-2);
       BIC[][r] = -2*dfunc1+(npar+1)*log(cn);
       HQ[][r] = -2*dfunc1+2*(npar+1)*log(log(cn));
       HQc[][r] = -2*dfunc1+2*cn*(npar+1)*log(log(cn))/(cn-npar-3);
       RQua[][r] = 1-((b_erroQua/b_totalQua)*((cn-1)/(cn-npar)));
       ++r;
       }
      }
     }
    }
   }
  }
 }
 for (a=0;a<columns(s_mx);++a)
 {
  for (b=a;b<columns(s_mx);++b)
  {
   for (c=b+2;c<columns(s_mx);++c)
   {
    for (d=c;d<columns(s_mx);++d)
    {
     for (e=d+2;e<columns(s_mx);++e)
     {
```

```
      for (f=e;f<columns(s_mx);++f)
      {
        for (g=f+2;g<columns(s_mx);++g)
        {
         for (h=g;h<columns(s_mx);++h)
         {
         s_mX=s_mx[][a:b]~s_mx[][c:d]~s_mx[][e:f]~s_mx[][g:h];
       p=columns(s_mX);
       k[][r]=columns(s_mX);;
       npar=columns(s_mX); // numero de parametros do modelo
       v_eta=s_mX*vbeta[0:(npar-1)];
       v_mu=exp(v_eta)./(1.0+exp(v_eta));
       d_denom = cn-npar;
       m_inv = invertsym(s_mX'*s_mX)*s_mX';  //inversao auxiliar

 //Regressão auxiliar para determinar pontos inicias no momento
//da maximizacao por BFGS
      v_betaini = m_inv*v_z; //estimacao inicial dos betas
      v_zEstim = s_mX*v_betaini; // vetor v_z estimado
      v_eEstim = v_z-v_zEstim; // vetor de residuos estimados
      b_SomaQua = v_eEstim'*v_eEstim; // soma de quadrados
      v_muVee = exp(v_zEstim) ./ (1.0+exp(v_zEstim));  //vetor mu transformado
      v_sigma2Est = b_SomaQua/(d_denom); //*v_invgDerPrima2;
      phi_ini = double(meanc( 1.0 ./ (v_sigma2Est*
               (v_muVee .* (1.0-v_muVee)))) - 1.0); //phi inicial
      v_thetaIni = v_betaini|phi_ini; //vetor theta inicial para a maximizacao

 // valores de arranque
     vp1 = v_thetaIni;

 //Maximizacao da verossimilhanca por o Metodo BFGS
//parametros da maximizacao por BFGS
     MaxControl( 50, -1 );

 // Estimacao por maxima verossimilhanca
     ir1 = MaxBFGS(floglik, &vp1, &dfunc1, 0, FALSE);
     if( ir1 == MAX_CONV || ir1 == MAX_WEAK_CONV )
     {
       v_etafinal =s_mX*vp1[0:(npar-1)];
       v_erro = v_z-v_etafinal;
       b_erroQua = v_erro'*v_erro;
       v_mu=exp(v_etafinal)./(1.0+exp(v_etafinal));
       L2[][r] = (v_mu1-v_mu)'*(v_mu1-v_mu);
       mmle[][r] = vp1;
       vlogver[][r]=dfunc1;
       AIC[][r] = -2*dfunc1+2*(npar+1);
       AICc[][r] = -2*dfunc1+(2*(npar+1)*cn)/(cn-npar-2);
       BIC[][r] = -2*dfunc1+(npar+1)*log(cn);
       HQ[][r] = -2*dfunc1+2*(npar+1)*log(log(cn));
       HQc[][r] = -2*dfunc1+2*cn*(npar+1)*log(log(cn))/(cn-npar-3);
       RQua[][r] = 1-((b_erroQua/b_totalQua)*((cn-1)/(cn-npar)));
       ++r;
       }
      }
     }
    }
   }
```

```
   }
  }
 }
}
for (a=0;a<columns(s_mx);++a)
{
 for (b=a;b<columns(s_mx);++b)
 {
  for (c=b+2;c<columns(s_mx);++c)
  {
   for (d=c;d<columns(s_mx);++d)
   {
    for (e=d+2;e<columns(s_mx);++e)
    {
     for (f=e;f<columns(s_mx);++f)
     {
      for (g=f+2;g<columns(s_mx);++g)
      {
       for (h=g;h<columns(s_mx);++h)
       {
        for (i=h+2;i<columns(s_mx);++i)
        {
         for (j=i;j<columns(s_mx);++j)
         {
         s_mX=s_mx[][a:b]~s_mx[][c:d]~s_mx[][e:f]~s_mx[][g:h]~s_mx[][i:j];
        p=columns(s_mX);
        k[][r]=columns(s_mX);
        npar=columns(s_mX); // numero de parametros do modelo
        v_eta=s_mX*vbeta[0:(npar-1)];
        v_mu=exp(v_eta)./(1.0+exp(v_eta));
        d_denom = cn-npar;
        m_inv = invertsym(s_mX'*s_mX)*s_mX';  //inversao auxiliar

 //Regressão auxiliar para determinar pontos inicias no momento
//da maximizacao por BFGS
      v_betaini = m_inv*v_z; //estimacao inicial dos betas
      v_zEstim = s_mX*v_betaini; // vetor v_z estimado
      v_eEstim = v_z-v_zEstim; // vetor de residuos estimados
      b_SomaQua = v_eEstim'*v_eEstim; // soma de quadrados
      v_muVee = exp(v_zEstim) ./ (1.0+exp(v_zEstim));  //vetor mu transformado
      v_sigma2Est = b_SomaQua/(d_denom); //*v_invgDerPrima2;
      phi_ini = double(meanc( 1.0 ./ (v_sigma2Est*
                (v_muVee .* (1.0-v_muVee)))) - 1.0); //phi inicial
      v_thetaIni = v_betaini|phi_ini; //vetor theta inicial para a maximizacao

 // valores de arranque
     vp1 = v_thetaIni;

 //Maximizacao da verossimilhanca por o Metodo BFGS
//parametros da maximizacao por BFGS
     MaxControl( 50, -1 );

 // Estimacao por maxima verossimilhanca
     ir1 = MaxBFGS(floglik, &vp1, &dfunc1, 0, FALSE);
     if( ir1 == MAX_CONV || ir1 == MAX_WEAK_CONV )
     {
       v_etafinal =s_mX*vp1[0:(npar-1)];
```

```
          v_erro = v_z-v_etafinal;
          b_erroQua = v_erro'*v_erro;
          v_mu=exp(v_etafinal)./(1.0+exp(v_etafinal));
          L2[][r] = (v_mu1-v_mu)'*(v_mu1-v_mu);
          mmle[][r] = vp1;
          vlogver[][r]=dfunc1;
          AIC[][r] = -2*dfunc1+2*(npar+1);
          AICc[][r] = -2*dfunc1+(2*(npar+1)*cn)/(cn-npar-2);
          BIC[][r] = -2*dfunc1+(npar+1)*log(cn);
          HQ[][r] = -2*dfunc1+2*(npar+1)*log(log(cn));
          HQc[][r] = -2*dfunc1+2*cn*(npar+1)*log(log(cn))/(cn-npar-3);
          RQua[][r] = 1-((b_erroQua/b_totalQua)*((cn-1)/(cn-npar)));
          ++r;
         }
        }
       }
      }
     }
    }
   }
  }
 }
}
  for (a=0;a<columns(s_mx);++a)
  {
   for (b=a;b<columns(s_mx);++b)
   {
          s_mX=s_mx[][a:b];
          p=columns(s_mX);
          k[][r]=columns(s_mX);;
          npar=columns(s_mX); // numero de parametros do modelo
          v_eta=s_mX*vbeta[0:(npar-1)];
          v_mu=exp(v_eta)./(1.0+exp(v_eta));
          d_denom = cn-npar;
          m_inv = invertsym(s_mX'*s_mX)*s_mX';  //inversao auxiliar


    //Regressão auxiliar para determinar pontos inicias no momento
   //da maximizacao por BFGS
         v_betaini = m_inv*v_z; //estimacao inicial dos betas
         v_zEstim = s_mX*v_betaini; // vetor v_z estimado
         v_eEstim = v_z-v_zEstim; // vetor de residuos estimados
         b_SomaQua = v_eEstim'*v_eEstim; // soma de quadrados
         v_muVee = exp(v_zEstim) ./ (1.0+exp(v_zEstim));  //vetor mu transformado
         v_sigma2Est = b_SomaQua/(d_denom); //*v_invgDerPrima2;
         phi_ini = double(meanc( 1.0 ./ (v_sigma2Est*
                   (v_muVee .* (1.0-v_muVee)))) - 1.0); //phi inicial
         v_thetaIni = v_betaini|phi_ini; //vetor theta inicial para a maximizacao

    // valores de arranque
        vp1 = v_thetaIni;

    //Maximizacao da verossimilhanca por o Metodo BFGS
   //parametros da maximizacao por BFGS
        MaxControl( 50, -1 );
```

```
   // Estimacao por maxima verossimilhanca
       ir1 = MaxBFGS(floglik, &vp1, &dfunc1, 0, FALSE);
       if( ir1 == MAX_CONV || ir1 == MAX_WEAK_CONV )
       {
         v_etafinal =s_mX*vp1[0:(npar-1)];
         v_erro = v_z-v_etafinal;
         b_erroQua = v_erro'*v_erro;
         v_mu=exp(v_etafinal)./(1.0+exp(v_etafinal));
         L2[][r] = (v_mu1-v_mu)'*(v_mu1-v_mu);
         mmle[][r] = vp1;
         vlogver[][r]=dfunc1;
         AIC[][r] = -2*dfunc1+2*(npar+1);
         AICc[][r] = -2*dfunc1+(2*(npar+1)*cn)/(cn-npar-2);
         BIC[][r] = -2*dfunc1+(npar+1)*log(cn);
         HQ[][r] = -2*dfunc1+2*(npar+1)*log(log(cn));
         HQc[][r] = -2*dfunc1+2*cn*(npar+1)*log(log(cn))/(cn-npar-3);
         RQua[][r] = 1-((b_erroQua/b_totalQua)*((cn-1)/(cn-npar)));
         ++r;
        }
       }
      }
if (r==ncombinacoes)
{
  mL2 = min(L2);
  pos=vecrindex(L2,mL2);
  if (k[pos]==1)
  ++L2ck1;
  if ( k[pos]==2)
  ++L2ck2;
  if ( k[pos]==3)
  ++L2ck3;
  if ( k[pos]==4)
  ++L2ck4;
  if ( k[pos]==5)
  ++L2ck5;
  if ( k[pos]==6)
  ++L2ck6;
  if ( k[pos]==7)
  ++L2ck7;
  if ( k[pos]==8)
  ++L2ck8;
  if (pos==posicaoreal)
  ++L2real;
  mAIC=min(AIC);
  pos=vecrindex(AIC,mAIC);
  L2EficObservAIC[][cs]=mL2/L2[pos];
  if ( k[pos]==1)
  ++AICck1;
  if ( k[pos]==2)
  ++AICck2;
  if ( k[pos]==3)
  ++AICck3;
  if ( k[pos]==4)
  ++AICck4;
  if ( k[pos]==5)
  ++AICck5;
  if ( k[pos]==6)
```

```
++AICck6;
if ( k[pos]==7)
++AICck7;
if ( k[pos]==8)
++AICck8;
if (pos==posicaoreal)
++AICreal;
mAICc=min(AICc);
pos=vecrindex(AICc,mAICc);
L2EficObservAICc[][cs]=mL2/L2[pos];
if ( k[pos]==1)
++AICcck1;
if ( k[pos]==2)
++AICcck2;
if ( k[pos]==3)
++AICcck3;
if ( k[pos]==4)
++AICcck4;
if ( k[pos]==5)
++AICcck5;
if ( k[pos]==6)
++AICcck6;
if ( k[pos]==7)
++AICcck7;
if ( k[pos]==8)
++AICcck8;
if (pos==posicaoreal)
++AICcreal;
mBIC=min(BIC);
pos=vecrindex(BIC,mBIC);
L2EficObservBIC[][cs]=mL2/L2[pos];
if ( k[pos]==1)
++BICck1;
if ( k[pos]==2)
++BICck2;
if ( k[pos]==3)
++BICck3;
if ( k[pos]==4)
++BICck4;
if ( k[pos]==5)
++BICck5;
if ( k[pos]==6)
++BICck6;
if ( k[pos]==7)
++BICck7;
if ( k[pos]==8)
++BICck8;
if (pos==posicaoreal)
++BICreal;
mHQ=min(HQ);
pos=vecrindex(HQ,mHQ);
L2EficObservHQ[][cs]=mL2/L2[pos];
if ( k[pos]==1)
++HQck1;
if ( k[pos]==2)
++HQck2;
if ( k[pos]==3)
```

```
++HQck3;
if ( k[pos]==4)
++HQck4;
if ( k[pos]==5)
++HQck5;
if ( k[pos]==6)
++HQck6;
if ( k[pos]==7)
++HQck7;
if ( k[pos]==8)
++HQck8;
if (pos==posicaoreal)
++HQreal;
mHQc=min(HQc);
pos=vecrindex(HQc,mHQc);
L2EficObservHQc[][cs]=mL2/L2[pos];
if ( k[pos]==1)
++HQcck1;
if ( k[pos]==2)
++HQcck2;
if ( k[pos]==3)
++HQcck3;
if ( k[pos]==4)
++HQcck4;
if ( k[pos]==5)
++HQcck5;
if ( k[pos]==6)
++HQcck6;
if ( k[pos]==7)
++HQcck7;
if ( k[pos]==8)
++HQcck8;
if (pos==posicaoreal)
++HQcreal;
mRQua=max(RQua);
L2EficObservRQ[][cs]=mL2/L2[pos];
if ( k[pos]==1)
++RQck1;
if ( k[pos]==2)
++RQck2;
if ( k[pos]==3)
++RQck3;
if ( k[pos]==4)
++RQck4;
if ( k[pos]==5)
++RQck5;
if ( k[pos]==6)
++RQck6;
if ( k[pos]==7)
++RQck7;
if ( k[pos]==8)
++RQck8;
if (pos==posicaoreal)
++RQreal;
}
else
{
```

```
    ++cfailure;
     --cs;
  }
 }

 MediaL2EficObservAIC = meanr(L2EficObservAIC);
 DPadraoL2EficObservAIC = sqrt(cs*varr(L2EficObservAIC)/(cs-1));
 MedianaL2EficObservAIC=quantiler(L2EficObservAIC,0.50);
 MediaL2EficObservAICc = meanr(L2EficObservAICc);
 DPadraoL2EficObservAICc = sqrt(cs*varr(L2EficObservAICc)/(cs-1));
 MedianaL2EficObservAICc=quantiler(L2EficObservAICc,0.50);
 MediaL2EficObservBIC = meanr(L2EficObservBIC);
 DPadraoL2EficObservBIC = sqrt(cs*varr(L2EficObservBIC)/(cs-1));
 MedianaL2EficObservBIC=quantiler(L2EficObservBIC,0.50);
 MediaL2EficObservHQ = meanr(L2EficObservHQ);
 DPadraoL2EficObservHQ = sqrt(cs*varr(L2EficObservHQ)/(cs-1));
 MedianaL2EficObservHQ=quantiler(L2EficObservHQ,0.50);
 MediaL2EficObservHQc = meanr(L2EficObservHQc);
 DPadraoL2EficObservHQc = sqrt(cs*varr(L2EficObservHQc)/(cs-1));
 MedianaL2EficObservHQc=quantiler(L2EficObservHQc,0.50);
 MediaL2EficObservRQ = meanr(L2EficObservRQ);
 DPadraoL2EficObservRQ = sqrt(cs*varr(L2EficObservRQ)/(cs-1));
 MedianaL2EficObservRQ=quantiler(L2EficObservRQ,0.50);
 MediaL2EficObservDev = meanr(L2EficObservDev);
 DPadraoL2EficObservDev = sqrt(cs*varr(L2EficObservDev)/(cs-1));
 MedianaL2EficObservDev=quantiler(L2EficObservDev,0.50);

/*Impressão dos resultados*/
 println("MediaL2EficObservAIC",MediaL2EficObservAIC);
 println("DesvioPadraoL2EficObservAIC=",DPadraoL2EficObservAIC);
 println("MedianaL2EficObservAIC=",MedianaL2EficObservAIC);
 println("MediaL2EficObservAICc=",MediaL2EficObservAICc);
 println("DesvioPadraoL2EficObservAICc=",DPadraoL2EficObservAICc);
 println("MedianaL2EficObservAICc=",MedianaL2EficObservAICc);
 println("MediaL2EficObservBIC",MediaL2EficObservBIC);
 println("DesvioPadraoL2EficObservBIC=",DPadraoL2EficObservBIC);
 println("MedianaL2EficObservBIC=",MedianaL2EficObservBIC);
 println("MediaL2EficObservHQ",MediaL2EficObservHQ);
 println("DesvioPadraoL2EficObservHQ=",DPadraoL2EficObservHQ);
 println("MedianaL2EficObservHQ=",MedianaL2EficObservHQ);
 println("MediaL2EficObservHQc",MediaL2EficObservHQc);
 println("DesvioPadraoL2EficObservHQc=",DPadraoL2EficObservHQc);
 println("MedianaL2EficObservHQc=",MedianaL2EficObservHQc);
 println("MediaL2EficObservRQ=",MediaL2EficObservRQ);
 println("DesvioPadraoL2EficObservRQ=",DPadraoL2EficObservRQ);
 println("MedianaL2EficObservRQ=",MedianaL2EficObservRQ);
 println("contL2","\n k1=",L2ck1,"\n k2=",L2ck2,
 "\n k3=",L2ck3,"\n k4=",L2ck4,"\n k5=",L2ck5,"\n k6=",
 L2ck6,"\n k7=",L2ck7,"\n k8=",L2ck8,"\n kreal=",L2real);
 println("contAIC","\n k1=",AICck1,"\n k2=",AICck2,
 "\n k3=",AICck3,"\n k4=",AICck4,"\n k5=",AICck5,"\n k6=",
 AICck6,"\n k7=",AICck7,"\n k8=",AICck8,"\n kreal=",AICreal);
 println("contAICc","\n k1=",AICcck1,"\n k2=",AICcck2,
 "\n k3=",AICcck3,"\n k4=",AICcck4,"\n k5=",AICcck5,
 "\n k6=",AICcck6,"\n k7=",AICcck7,"\n k8=",AICcck8,
 "\n kreal=",AICcreal);
 println("contBIC","\n k1=",BICck1,"\n k2=",BICck2,"\n k3=",
```

```
  BICck3,"\n k4=",BICck4,"\n k5=",BICck5,"\n k6=",BICck6,
  "\n k7=",BICck7,"\n k8=",BICck8,"\n kreal=",BICreal);
  println("contHQ","\n k1=",HQck1,"\n k2=",HQck2,"\n k3=",
  HQck3,"\n k4=",HQck4,"\n k5=",HQck5,"\n k6=",HQck6,"\n k7=",
  HQck7,"\n k8=",HQck8,"\n kreal=",HQreal);
  println("contHQc","\n k1=",HQcck1,"\n k2=",HQcck2,"\n k3=",
  HQcck3,"\n k4=",HQcck4,"\n k5=",HQcck5,"\n k6=",HQcck6,
  "\n k7=",HQcck7,"\n k8=",HQcck8,"\n kreal=",HQcreal);
  println("contRQ","\n k1=",RQck1,"\n k2=",RQck2,"\n k3=",
  RQck3,"\n k4=",RQck4,"\n k5=",RQck5,"\n k6=",RQck6,
  "\n k7=",RQck7,"\n k8=",RQck8,"\n kreal=",RQreal);
  print( "\n\nNUM. OBSERVA. DA AMOSTRA: ", cn );
  print( "\nNUM. DE REPLICAS COM FALHAS ", cfailure );
  println("Tempo=",dExecTime);
  println("r=",r);
  }
}
```

Bibliografia

[1] Akaike, H. (1974). A new look at statistical model identification. *IEEE Transactions on Automatic Control AU*, **19**, 716-722.

[2] Akaike, H. (1978). A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mathematics A*, **30**, 9-14.

[3] Athinson, A. C. (1985). *Plots, Transformation and Regrression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*. New York: Oxford University Press.

[4] Cordeiro, G. M. (1986). *Modelos Lineares Generalizados*. Campinas, VII SINAPE. 286p.

[5] Ferrari, S. L. P. e Cribari-Neto, F. (2004). Beta regression for modeling rates and proportions. *Journal of Applied Statistics*, **31**, 7, 799-816.

[6] Hannan, E. J. e Quinn, B. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society B*, **41**, 190-191.

[7] Huvich, C. M. e Tsai, C-L. (1989). Regression and time series model selection in small samples. *Biometrics*, **76**, 297-307.

[8] Huvich, C. M. e Tsai, C-L. (1995). Model Selection for Extended Quasi-likelihood Models in Small Samples. *Biometrics*, **51**, 1077-1084.

[9] Johnson, N., Kotz, S. e Balakrishnan, N. (1995). *Continuous Univariate Distributions*. New York: John Wiley and Sons. 2ª ed.

[10] Kieschnick, R. L. e McCullough, B. D. (2003). Regression analysis of variates observed (0,1): percentages, proportions and fractions. *Statistical Modeling*, **3**, 193-213.

[11] Kuha, J. (2004). AIC and BIC - Comparisons of assumptions and performance. *Sociological Methods & Research*, **33**, 2, 188-229.

[12] Mallows, C. L. (1973). Some comments on Cp. *Technometrics*, **37**, 661-675.

[13] Martínez, R. O. (2004). *Estimação Pontual e Intervalar em um Modelo de Regressão Beta*. Dissertação de Mestrado, UFPE.

[14] McCullagh, P. e Nelder, J. (1989). *Generalized Linear Models*. London: Chapman and Hall.

[15] McQuarrie, A. D. R. e Tsai, C-L. (1998). *Regression and Time Series Model Selection*. Singapore: World Scientific Publishing Co. Pte. Ltd.

[16] Nações Unidas (2003). PNUD - Programa das Nações Unidas para o Desenvolvimento: Atlas do Desenvolvimento Humano no Brasil. `http//www.undp.org.br.`

[17] Nelder, J. A. e Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society A*, **135**, 3, 370-84.

[18] Nocedal, J. e Wright, S. J. (1999). *Numerical Optimization*. New York: Springer-Verlag.

[19] Patil, G. P. e Shorrock, R. (1965). On certain properties of the exponential-type distribuition. *Journal of the Royal Statistical Society B*, **27**, 94-99.

[20] Rao, C. R. (1973). *Linear Statistical Inference and its Applications*, 2ª ed. New York: Wiley.

[21] Rao, C. R. e Wu, Y. (2005). Linear Model Selection by Cross-validation. *Journal of Statistical Planning and Inference*, **128**, 1, 231-240.

[22] Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461-464.

[23] Shibata, R. (1980). An optimal selection of regression variables. *Biometrika*, **68**, 45-54.

[24] Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections. *Communicationin Statistics - Theory and Methods*, **7**, 13-26.