VIII ERMAC-R3

5º Encontro Regional de Matemática Aplicada e Computacional
20 a 22-Novembro-2008
Universidade Federal do Rio Grande do Norte - Natal/RN

Modelagem da Taxa de Mortalidade Infantil nos Estados brasileiros utilizando a teoria dos Modelos Lineares Generalizados

Carlos Sérgio Araújo dos Santos
Programa de Pós-graduação em Biometria e Estatística Aplicada, UFRPE,
52171-900, Recife, PE
E-mail: carlossergioaraujo@gmail.com,

Joseilme Fernandes Gouveia
Programa de Pós-graduação em Biometria e Estatística Aplicada, UFRPE,
52171-900, Recife, PE
E-mail: juniorfg99@hotmail.com.

Resumo

A Taxa de Mortalidade Infantil (TMI), reflete, de maneira geral, os níveis de saúde e de desenvolvimento socioeconômico de uma determinada região. Observa-se uma disparidade expressiva nas TMI entre os estados brasileiros. Esse fato tem provocado o interesse de diversos pesquisadores na busca dos determinantes dessas desigualdades. Assim, a utilização da teoria dos Modelos Lineares Generalizados (MLG), torna-se um ferramental importante para modelagem dessas taxas. Os MLG's são uma classe de modelos de regressão em que a regressão linear com a variável resposta normalmente distribuída é um caso particular. São formados por um componente aleatório, que possui distribuição pertencente à família exponencial, um componente sistemático, linearizados por uma função de ligação. Serão utilizadas para modelar a TMI, variáveis de cunho epidemiológico, demográfico e socioeconômico. Os programas utilizados para o ajuste do banco de dados foram desenvolvidos no software R na versão 2.7.0. O modelo Gama com função de ligação identidade foi o modelo que melhor ajustou-se aos dados. Pretende-se com esse estudo realizar previsões dessas taxas e, consequentemente, subsidiar o planejamento dos gestores em saúde a adotar políticas públicas nas áreas de saúde materna e infantil nos estados brasileiros.

Palavras-chave

Taxa de Mortalidade Infantil, Modelos Lineares Generalizados, Modelos de Regressão, Modelo Gama.

Introdução

A Taxa de Mortalidade Infantil (TMI) é considerada um dos mais importantes indicadores epidemiológicos utilizados internacionalmente, pois esse indicador contribui na avaliação dos níveis de saúde e de desenvolvimento socioeconômico de uma determinada região [11].

Ao longo das ultimas décadas, observa-se uma redução da TMI, em consequência da melhoria das condições gerais de vida e do acesso aos serviços de saúde. Tendo em vista que a redução dessas taxas configura-se como uma das principais metas na área de saúde em diversos países [15].

No Brasil, percebe-se uma tendência de queda na mortalidade infantil, entretanto, essa taxa ainda situa-se em um patamar elevado. Observa-se também uma expressiva desigualdade entre suas unidades federativas. Essas disparidades são consequências dos diferentes níveis de saúde, localização geográfica e social existente entre os estados.

Assim, buscar padrões de relacionamento que explique esse indicador, constitui uma importante fonte na busca dos determinantes dessas desigualdades.

Baseando-se nesse contexto, realizar uma modelagem dessa taxa considerando variáveis de cunho epidemiológico, demográfico e socioeconômico, é de suma importância para a realização de previsões que servirão de subsídios em processos de planejamento, gestão e avaliação de políticas e ações de saúde voltadas para atenção pré-natal e ao parto, bem como para a proteção da saúde infantil.

Utiliza-se a teoria dos Modelos Lineares Generalizados (MLG), que considera modelos probabilísticos pertencente à família exponencial. Tornando-se assim, um ferramental importante para modelar a TMI nos estados brasileiros.

Objetivo

Objetiva-se assim, realizar uma modelagem das Taxas de Mortalidade Infantil para os Estados brasileiros com referência ao ano de 2004, utilizando variáveis de cunho epidemiológico, socioeconômico e demográfico.

Material e Métodos

Os indicadores básicos para realização da modelagem foram coletados das fontes oficiais de dados contínuos do DATASUS do Ministério da Saúde [10] (Taxa de Mortalidade Infantil e Taxa de Fecundidade Total), IBGE [7](Taxa de Analfabetismo, Cobertura de esgotamento sanitário e Esperança de vida ao nascer) e IPEA [8] (Produto Interno Bruto per capita).

O ano de 2004 foi escolhido como referencia para modelagem pois as estimativas dos indicadores estariam mais proximas da atualidade, apesar de que alguns indicadores já possuírem estimativas mais recentes. O ano de 2004 é o último ano que possui estimativas para todos os indicadores utilizados para essa modelagem, além do mais, ajustar dados referentes ao mesmo ano ou período específico, torna-se mais viável em termos interpretativos.

As observações foram os 26 estados da federação mais o Distrito Federal.

Variáveis do modelo

Foram utilizadas as seguintes variáveis nesse estudo:

Variável epidemiológica: Taxa de Mortalidade Infantil.

Variáveis demográficas: Taxa de Fecundidade Total e Esperança de vida ao nascer.

Variáveis socioeconômicas: Produto Interno Bruto per capita, cobertura de esgotamento sanitário e Taxa de analfabetismo.

Cada uma dessas variáveis serão descritas a seguir.

Taxa de Mortalidade Infantil: Número de óbitos menores de um ano de idade, por mil nascidos vivos, na população residente em determinado espaço geográfico, no ano considerado.

Taxa de fecundidade total (TFT): Número médio de filhos nascidos vivos, tidos por uma mulher ao final do seu período reprodutivo, na população residente em determinado espaço geográfico.

Esperança de vida ao nascer: Número médio de anos de vida esperados para um recém-nascido, mantido ao padrão de mortalidade existente, na população residente em determinado espaço geográfico, no ano considerado.

Produto Interno Bruto per capita (PIB): Valor médio agregado por indivíduo, em moeda corrente e a preços de mercado, de bens e serviços finais produzidos em determinado espaço geográfico, no ano considerado.

Cobertura de esgotamento sanitário: Percentual da população que dispõe de escoadouro de dejetos através de ligação do domicílio à rede coletora ou fossa séptica, em determinado espaço geográfico, no ano considerado.

Taxa de analfabetismo: Percentual de pessoas de 15 ou mais anos de idade que não sabem ler e escrever pelo menos um bilhete simples no idioma que conhecem, na população total residente na mesma faixa etária, em determinado espaço geográfico, no ano considerado.

Modelos Lineares Generalizados

Os Modelos Lineares Generalizados foram introduzidos por Nelder e Wedderburn (1972), cujos modelos são usados geralmente quando se tem uma única variável aleatória Y associada a um conjunto de variáveis explanatórias $x_1, \ldots, x_p$. O MLG é composto por três componentes [4]:

*i)*Componente aleatório: representado por um conjunto de variáveis aleatórias independentes $y_1, \ldots, y_n$ provinientes de uma mesma distribuição que faz parte da família exponencial de distribuições com médias $\mu_1, \ldots, \mu_n$, ou seja,

$$E(y_i) = \mu_i, \qquad i = 1, \ldots, n.$$

sendo $\phi > 0$ um parâmetro de dispersão e o parâmetro θ_i denominado parâmetro canônico. Então, a função de densidade de probabilidade de Y_i é dada por:

$$f(y_i; \theta_i, \phi) = exp\{\phi^{-1}[y_i\theta_i - b(\theta_i)] + c(y_i, \phi)\} \quad (1)$$

sendo $b(.)$ e $c(.)$ funções conhecidas.

*ii)*Componente sistemático: as variáveis explicativas entram na forma de uma soma linear de seus efeitos

$$\eta_i = \sum_{r=1}^{p} x_{ir}\beta_j = x_i^t\beta \qquad ou \qquad \eta = X\beta \qquad (2)$$

sendo $X = (x_1, \ldots, x_n)^t$ a matriz do modelo, $\beta = (\beta_1, \ldots, \beta_n)^t$ o vetor de parâmetros e $\eta = (\eta_1, \ldots, \eta_n)^t$ o preditor linear.

iii) Função de ligação: uma função que relaciona o componente aleatório ao componente sistemático, ou seja, vincula a média ao preditor linear, isto é,

$$\eta_i = g(\mu_i) \qquad (3)$$

sendo $g(.)$ uma função monótona diferenciável.

Portanto, uma decisão importante na escolha do MLG é definir os termos do trinômio: (i) distribuição da variável resposta; (ii) matriz do modelo e (iii) função de ligação. Nesses termos, um MLG é definido por uma distribuição da família (1), uma estrutura linear (2) e uma função de ligação (3) [4].

Algumas referências para o estudo dos MLG são: Cordeiro (1986) [3], McCullagh e Nelder (1989) [9] e Paula (2004) [13].

Estimação dos Parâmetros

Para estimar os parâmetros β, utilizou-se o método de máxima verossimilhança, cujo método é realizado através de processo iterativo. O algoritmo foi desenvolvido por Nelder e Wedderburn (1972) e se baseia no método Escore de Fisher.

Medidas de Diagnóstico

A partir dos gráficos das medidas de diagnóstico, podemos verificar a função de variância, a adequação do modelo aos dados, identificar pontos aberrantes e influentes e avaliar a distribuição do erro aleatório.

Ponto de Alavanca

Em MLG's, utiliza-se a matriz H generalizada. O elemento h_{ii} da matriz H representa uma medida de alavancagem. A matriz de projeção H é definida por:

$$H = W^{1/2}X(X^TWX)^{-1}X^TW^{1/2}$$

em que X é a matriz dos dados e $W = diag\{w_1, \ldots, w_n\}$ uma matriz diagonal de pesos que capta a informação sobre a distribuição e a função de ligação usadas [4].

Espera-se que observações distantes do espaço formado pelas variáveis explicativas apresentem valores apreciáveis de h_{ii}.

Se $h_{ii} > 2p/n$ temos que y_i é ponto de alavanca. Em que p é o número de parâmetros estimados.

Ponto de Influênte

Uma observação influente é aquela cuja omissão do conjunto de dados resulta em mudanças substanciais em certos aspectos do modelo [4]. Utiliza-se a Estatística Modificada de Cook (T_i). Se $T_i > 2(p/n)^{1/2}$, então y_i será influênte.

Ponto Aberrante

Um ponto aberrante é aquele que apresenta perfil diferente das demais observações em relação aos valores da variável resposta.

Graficamente, para detectarmos pontos aberrantes, utilizamos o resíduo padronizado versus o índice das observações t.

Critério de Informação de Akaike (AIC)

Uma medida utilizada para selecionar o modelo mais parcimonioso entre os ajustados, ou seja, que esteja bem ajustado e com um número reduzido de parâmetros, é o critério de informação de Akaike (AIC).

O critério de informação de Akaike é dado por:

$$AIC = -2l(\hat{\beta}) + 2p$$

sendo $l(.)$ o logarítmo da função de verossimilhança aos dados.

Gráfico de Probabilidade Meio-Normal com Envelope Simulado

Este gráfico é construído baseado nos resíduos padronizados. O envelope simulado é a banda de confiança. A ocorrência de pontos proximos ou fora da banda de confiança indicam que o modelo não está apropriado.

Suporte Computacional

Os programas utilizados para modelagem da Taxa de Mortalidade Infantil foram desenvolvidos no softwere R versão 2.7.0. Esse pacote computacional encontra-se disponível gratuitamente em http://www.r-project.org.br e é muito utilizado, pois apresenta diversas funções implementadas, que também podem ser obtidas nesse endereço. Mais detalhes sobre R podem ser encontrados em R Development Core Team (2007) [14].

Resultados

A metodologia de MLG's foi aplicada na modelagem da TMI em 2004, a partir de um conjunto de variáveis explicativas, avaliando-as por meio de medidas de diagnóstico.

O conjunto dos dados contém 27 observações referentes aos 26 estados brasileiros mais o Distrito Federal. A variável dependente (y) é a Taxa de Mortalidade Infantil. As variáveis independentes são: PIB per capita, Cobertura de esgotamento sanitário, Taxa de analfabetismo, Esperança de vida ao nascer e a Taxa de Fecundidade Total.

Escolha da distribuição e Ajuste do modelo

Inicialmente geramos o histograma dos dados (Figura 1), onde foi possível verificar pouca simetria dos dados em seu formato, a natureza contínua dos dados e o intervalo positivo de variação. Essas características sugerem que a distribuição dos dados segue o modelo Gama ou Normal inverso. No entanto, o modelo Normal inverso não ajustou-se adequadamente aos dados.

Para ajustarmos o modelo Gama com função de ligação identidade a esses dados, assumiremos que $y_1, \ldots, y_{27}$ são independentes e seguem uma distribuição Gama com média μ_t, $t = 1, \ldots, 27$, e parâmetro de dispersão ϕ desconhecido. Inicialmente, modelamos a média como:

$$g(\mu) = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + \beta_5 x_{t5}$$

sendo β_0 correspondente ao intercepto e $g(\mu)$ a função de ligação identidade, em que

x_{t1} representa o PIB per capita do estado t;

x_{t2} representa a Cobertura de esgotamento sanitário do estado t;

x_{t3} representa a Taxa de analfabetismo do estado t;

x_{t4} representa a Esperança de vida ao nascer do estado t;

x_{t5} representa a Taxa de Fecundidade Total do estado t.

Testamos então, o modelo Gama para os dados. A função de ligação identidade foi a escolhida pelo menor desvio, comparando-se com as demais funções de ligação, além do mais, o processo de estimação dos parâmetros não convergiu com as ligações logarítmicas e inversa. A maioria das variáveis foram significativas e, entre os modelos testados, foi o modelo que produziu o menor AIC.

Conclui-se que o modelo Gama com função de ligação identidade é um bom modelo para ajustar os dados de TMI.

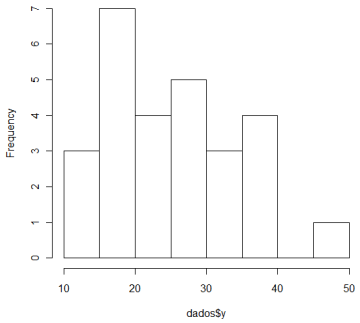

Figura 1: Histograma dos dados de TMI.

Os resultados da estimação dos parâmetros são dados na Tabela 1. Observa-se que as variáveis Cobertura de esgotamento sanitário e TFT não são significantes para o modelo. O parâmetro da variável TFT apresenta um p-valor alto (0,82030). Por conta disto, retirou-se essa variável do modelo e ajustou-se os dados novamente.

A retirada da variável TFT, melhorou significativamente o ajuste do modelo, reduzindo o Erro Padrão e o P-valor relacionado ao parâmetro das variáveis e aumentando a significância da variável Cobertura de esgotamento sanitário (0,09660) (Tabela 2).

Tabela 1: Estimativas dos parâmetros do modelo Gama com função de ligação identidade ajustado para os dados de TMI.

Parâmetro	Estimativa	Erro Padrão	P-valor
β_0	78,35191	40,48101	0,06650
β_1	0,00045	0,00024	0,07590
β_2	-0,04626	0,02725	0,10430
β_3	1,34450	0,24649	0,00002
β_4	-0,95803	0,51358	0,07620
β_5	0,41352	1,79751	0,82030
ϕ	0,01291	0,00652	-

Tabela 2: Estimativas dos parâmetros do modelo Gama com função de ligação identidade ajustado para os dados de TMI sem a variável TFT.

Parâmetro	Estimativa	Erro Padrão	P-valor
β_0	84,09522	29,50490	0,00930
β_1	0,00043	0,00018	0,06600
β_2	-0,04610	0,02205	0,09660
β_3	1,32883	0,22793	0,00010
β_4	-1,02196	0,40368	0,01900
ϕ	0,01236	0,00601	-

Em seguida, calculou-se as medidas de diagnóstico para verificar se há algum ponto com comportamento atípico. A Figura 2 indica os pontos 4 e 27 como pontos de alavanca. As observações 3, 4, 9, 17 e 27 aparecem como possíveis pontos influentes. Os pontos 3, 9 e 18 aparecem como pontos aberrantes. O gráfico de probabilidade meio-normal com envelope simulado, mostra que todos os pontos se encontram dentro da banda de confiança indicando um bom ajuste dos dados.

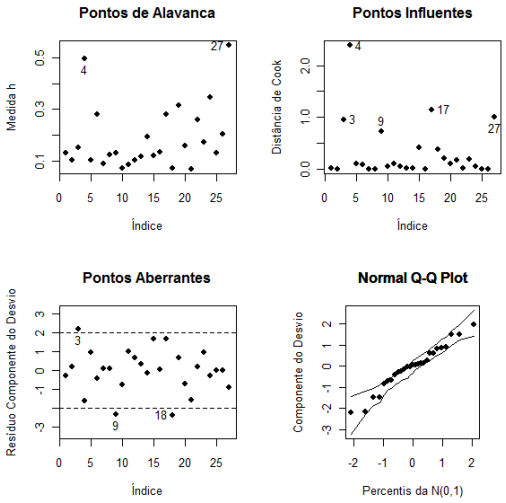

Figura 2: Gráficos de diagnóstico do modelo Gama com função de ligação identidade ajustado para os dados de TMI sem a variável TFT..

A retirada dessas observações não causou grandes

mudanças na variação percentual das estimativas dos parâmetros. Porém, o ponto 18 (referente ao estado do Espírito Santo), não é ponto influente nem alavanca, por isso merece um pouco mais de atenção. Portanto, retirou-se esse ponto e verificou-se o ajuste do modelo. Entretanto, a retirada da observação 18 não permitiu grandes alterações nas estimativas dos parâmetros, logo, opitou-se por conservar esse ponto no modelo.

O modelo final ajustado aos dados é o Gama com função de ligação identidade sem a variável Taxa de Fecundidade Total, em que o valor do AIC foi 136.

Análise do modelo

A partir das estimativas dos parâmetros do modelo Gama com ligação identidade, podemos agora, escrever o modelo ajustado

$$\mu = 84{,}09522 + 0{,}00043x_{t1} - 0{,}04610x_{t2} + 1{,}32883x_{t3} - 1{,}02196x_{t4}$$

A estimativa do parâmetro referente a variável PIB (0,00043) é positiva, porém, o seu valor é consideravelmente baixo. Essa estimativa significa que, para cada unidade positiva do PIB, a TMI aumenta em 0,043%. Mas, essa situação nos direciona ao seguinte questionamento: O PIB elevado sinaliza desenvolvimeto e, consequentemente melhores condições de vida da população?

Na verdade, o PIB per capita é um indicador que não leva em consideração as desigualdades sociais, já que divide toda produção de riquezas de um determinado espaço geográfico de forma igualitária entre seus habitantes. Segundo Viega (2006), a assimilação da riqueza ao PIB é equivocada, pois os métodos de mensuração sofrem alta margem de incerteza.

O PIB per capita provoca um viés de interpretação na relação com a mortalidade infantil, pois mesmo que a localidade apresente nível satisfatório, isso não implicará, necessáriamente, em distribuição de renda [2].

Contudo a teoria dos MLG mostrou-se bastante sensível na estimação desse parâmetro, fortalecendo o entendimento e exigindo cautela na interpretação da relação PIB - TMI.

Para a estimativa do parâmetro relacionada a variável Cobertura de esgotamento sanitário, observou-se uma relação inversa com a TMI, ou seja, o aumento da cobertura de esgotamento sanitário nos estados implica em diminuição da TMI naquela região. Teixeira & Guilhermino (2006) também comprovam essa relação.

Em estudos realizados por Benício & Monteiro (2000) e Teixeira & Guilhermino (2006), sinalizaram que condições precárias de saneamento torna-se um fator determinante para alta incidência de diarréia infantil no país, tida também como uma das causas da mortalidade infantil.

A terceira estimativa está relacionada a variável Taxa de analfabetismo. Esse parâmetro nos informa que para cada unidade positiva da Taxa de analfabetismo a TMI cresce na magnitude de 0,04610. Esta relação esta bem clara, pois quanto maior a Taxa de analfabetismo maior será a TMI, uma vez que, indivíduos com o baixo nível de instrução, sugerem que esses mesmos detêm um baixo nível de informação sobre os cuidados com a saúde e higiene pessoal.

Segundo Costa et al.(2003) a redução do analfabetismo contribui para a queda da mortalidade infantil.

Por ultimo, temos a Esperança de vida ao nascer, variável na qual teve sua estimativa do parâmetro de (-1,02196). Isto explica que para cada ano acrescentado na esperança de vida ao nascer significa em redução em aproximadamente uma unidade da TMI.

Em estudo realizado por Godinho (1999) a redução da mortalidade infantil contribui com o acrescimo da esperança de vida ao nascer, pois cada criança que sobrevive acrescenta anos de vida à população como um todo e aumenta a estimativa da sua vida média.

Nesse estudo o modelo apresentado para explicar os componentes da mortalidade infantil nos estados brasileiros foram expressos como combinações lineares entre indicadores socioeconômicos e demográficos. A relação de causa e consequência não esta perfeitamente caracterizada, sendo preferível o termo "possíveis determinantes" para as variáveis utilizadas no ajuste.

Conclusões

A teoria dos Modelos Lineares Generalizados mostrou-se uma ferramenta bastante eficaz para modelagem das Taxas de Mortalidade Infantil explicada por indicadores socioeconômicos e demográficos nos estados brasileiros referenciando o ano de 2004. Permitindo assim, ajustar os dados através do modelo Gama com função de ligação identidade.

As variáveis escolhidas e modeladas nesse estudo foram caracterizadas como possíveis determinantes da mortalidade infantil, e, esses determinantes ou indicadores sinalizaram as desigualdades existentes entre os estados.

Foi possível verificar que a diminuição da mortalidade infantil está diretamente relacionada com o aumento da cobertura de esgotamento sanitário, com a redução da taxa de analfabetismo e com a elevação da esperança de vida ao nascer.

Observamos também que não basta aumentar o PIB, tem que verificar como esse PIB está sendo distribuído, quem está se apropriando da riqueza.

Espera-se que este estudo possa contribuir para definição de estratégias que reduzam a mortalidade infantil nos estados brasileiros, subsidiando o planejamento e delimitando políticas publicas nas áreas de saúde materna e infantil. Fazendo com que a TMI no Brasil, atinja um patamar ideal, isto é, proxima das taxas dos países desenvolvidos.

Agradecimentos

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e ao Conselho de Apoio à Pesquisa do Ensino Superior (CAPES) pelo apoio financeiro.

Referências

[1] M. H. D'A. Benício, C. A. Monteiro , Tendência secular da diarréia na infância na cidade de São Paulo (1984-1996).*Rev. Saúde Pública*,34(6),sulp.,(2000) 83-90.

[2] J. G. Bezerra-Filho et. al., Mortalidade infantil e condições sociodemográficas no Ceará, em 1991 e 2000. *Rev. Saúde Pública*, 41(6), (2007) 1023-31.

[3] G. M. Cordeiro, Modelos lineares generalizados. VII SINAPE, UNICAMP, 1986.

[4] G. M. Cordeiro, C. G. B. Demétrio, Modelos lineares generalizados. Minicurso para o 12° SEAGRO e a 52° Reunião Anual da RBRAS, UFSM, Santa Maria, RS, 2007.

[5] M. C. N. Costa et. al., Mortalidade infantil no Brasil em períodos recentes de crise econômica.*Rev. Saúde Pública*, 37(6), (2003) 699-706.

[6] R. E. Godinho, A trajetória das Estatísticas Vitais nos anos 80 e seus reflexos no fim do milênio no Estado de São Paulo. Tese de Doutorado apresentada à Faculdade de Filosofia, Letras e Ciências Humanas - USP, São Paulo, 1999. 165p.

[7] IBGE - Instituto Brasileiro de Geografia e Estatística, Síntese dos Indicadores Sociais. Disponível em: *http://www.ibge.gov.br* , Acesso em 12 Mai. 2008.

[8] IPEA - Instituto Pesquisas Econômicas Aplicadas. Disponível em: *http://www.ipeadata.gov.br* , Acesso em 12 Mai. 2008.

[9] P. McCullagh, J. A. Nelder, "Generalized Linear Models", Chapman and Hall, London, 2.ed 1989.

[10] Ministério da Saúde - DATASUS, Indicadores de Saúde. Disponível em: *http://tabnet.datasus.gov.br* , Acesso em 12 Mai. 2008.

[11] C. A. Monteiro, C. L. Nazário, Evolução da agenda tradicional da saúde pública: mortalidade infantil e materna. Declínio da mortalidade infantil e equidade social: o caso de São Paulo 1973 e 1993 In: Monteiro, organizador, Velhos e novos males da saúde no Brasil: a evolução do país e suas doênças. 2.ed São Paulo, *Hucitec*,(2000) 173-85.

[12] J. A. Nelder, R. W. M. Wedderburn, Generalized linear models.*Journal of the Royal Statistical Society*, A, 135 (1972) 370-384.

[13] G. A. Paula, Modelos de Regressão com Apoio Computacional. IME/USP, São Paulo, 2004.

[14] R Development Core Team, "R: A language and environment for statistical computing" *http://www.R-project.org*, ISBN 3-900051-07-0, 2007.

[15] A. Rosano, L. D. Botto, P. Mastroiacovo, Infant mortality and congenital anomalies from 1950 to 1994: an perspective.*J. Epidemiol. Community Health*, 54 (2000) 660-6.

[16] J. E. Viega, Nada justifica o apego ao PIB. *Revista Valor Econômico*, 20 Out. 2006.

[17] J. C. Teixeira, R. L. Guilhermino, Análise de associação entre saneamento e saúde nos estados brasileiros, empregando dados secundários do banco de dados de indicadores e dados básicos para a saúde 2003 - IDB 2003.*Eng. Sanit. ambient.*, Vol. II, n. 3, Jul./Set. (2006) 277-82.