

# Modelo para Mortalidade Infantil no Brasil via Modelos Lineares Generalizados

Carlos Sérgio Araújo dos Santos\*    Alisson de Oliveira Silva†    Joseilme Fernandes Gouveia‡  
Gauss Moutinho Cordeiro§

## 1 Introdução

A Taxa de Mortalidade Infantil (TMI) evidencia, de maneira geral, o nível de vida e saúde e desenvolvimento socioeconômico de uma determinada região, sendo considerada um dos mais importantes indicadores epidemiológicos utilizados internacionalmente (MONTEIRO & NAZÁRIO, 2000).

Pode-se enfatizar que nas últimas décadas, a TMI vem sofrendo um significativo decréscimo na maior parte dos países. O Brasil também vem seguindo essa tendência mundial ocupando a 107<sup>a</sup> posição do ranking de mortalidade infantil - até cinco anos, segundo dados do fundo das Nações Unidas para a infância (UNICEF), com 22 óbitos por mil crianças nascidas vivas. Mesmo havendo uma queda no ranking, a UNICEF salienta que o país vem apresentando melhora nos seus índices desde 1990 que possuía uma taxa de mortalidade infantil de crianças até cinco anos de idade correspondente a 58 óbitos para mil crianças nascidas vivas. Já a taxa de crianças menores de um ano de idade que era de 49 óbitos para mil crianças nascidas vivas no mesmo período, hoje é de 20 para cada mil crianças nascidas vivas. Esse índice ainda apresenta-se distante do patamar dos países europeus a exemplo da Suécia e Islândia, e países da América do Norte, como Estados Unidos e Canadá, que possuem as menores taxas de mortalidade infantil do mundo.

Houve, portanto uma considerável redução de 62% na mortalidade de crianças até cinco anos e 59% na mortalidade de crianças de até um ano, sendo a 18<sup>a</sup> maior diminuição no índice entre todos os países. Os fatores determinantes dessa redução devem-se a melhoria das condições de vida, além de políticas voltadas a saúde infantil e desenvolvimento socioeconômico. Isso se deve ao fato da Taxa de Mortalidade Infantil ser o vértice das principais metas na área de saúde em diversos países (ROSANO et al., 2000). No Brasil, nota-se uma grande variação dessas taxas entre as regiões, devido ao alto índice de desigualdades dos níveis de saúde e desenvolvimento socioeconômico dos estados brasileiros. As regiões que possuem as maiores taxas são Norte-Nordeste enquanto que os Estados do eixo Centro-sul possuem

---

\*Departamento de Ciências da Natureza, IFPB, CEP: 58015-430, João pessoa, PB, Brasil,  
E-mail: *carlossergioaraujo@gmail.com*.

†Universidade Federal da Paraíba, UFPB, CEP: 58000-000, João Pessoa, PB, Brasil,  
E-mail: *alisson\_inocente@hotmail.com*.

‡Departamento de Informática e Estatística, UFPI, Teresina, PI, Brasil,  
E-mail: *juniorfg99@hotmail.com*.

§Departamento de Estatística e Informática, UFRPE, CEP: 52171-900, Recife, PE, Brasil,  
E-mail: *gauss@deinfo.ufrpe.com*.

as menores taxas. Então, é de suma importância buscar padrões de relacionamento que explique esse indicador, de modo que se encontre os determinantes dessas desigualdades nos Estados dessas regiões.

Portanto fazer uma modelagem a fim de Conhecer o perfil da Mortalidade Infantil é fundamental para a formulação de estratégias que permitam o seu controle, utilizando variáveis consideradas como fatores de risco e que sejam explicativas dessa taxa, para que possa ajudar, na tomada de decisões em projetos e planejamento das políticas de saúde infantil.

A modelagem das Taxas de Mortalidade Infantil dos Estados brasileiros em 2006 foi realizada por meio dos Modelos Lineares Generalizados (MLG). Considerado entre os Estatísticos uma ferramenta bastante eficiente, pois esses modelos não se restringem apenas à suposição de normalidade da variável resposta, sendo esta suposição ampliada para uma distribuição pertencente à família exponencial, tornando a modelagem mais flexível para diversos tipos de dados.

## 2 Objetivo

Objetiva-se assim, realizar uma modelagem das Taxas de Mortalidade Infantil para os Estados brasileiros aplicando a teoria dos MLG com referência ao ano de 2006, utilizando variáveis determinantes da mortalidade infantil.

## 3 Material e Métodos

Os indicadores básicos para realização da modelagem foram coletados das fontes oficiais de dados contínuos do DATASUS do Ministério da Saúde (Taxa de Mortalidade Infantil) e IBGE (Taxa de analfabetismo, Cobertura de esgotamento sanitário e Proporção de pobres). As observações foram os 26 estados da federação e Distrito Federal.

### 3.1 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  $x_1, \dots, x_p$ . O MLG é composto por três componentes (CORDEIRO & DEMÉTRIO, 2007).

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

$$E(y_i) = \mu_i, \quad i = 1, \dots, n.$$

em que  $\phi > 0$  um parâmetro de dispersão e o parâmetro  $\theta_i$  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  $b(\cdot)$  e  $c(\cdot)$  funções conhecidas.

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

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

em que  $X = (x_1, \dots, x_n)^t$  a matriz do modelo,  $\beta = (\beta_1, \dots, \beta_n)^t$  o vetor de parâmetros e  $\eta = (\eta_1, \dots, \eta_n)^t$  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,

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

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

Para verificar se um modelo é adequado, temos que analisar se as pressuposições feitas para o desenvolvimento do modelo não estão sendo violadas. Para o melhor modelo iremos resaltar basicamente métodos gráficos, especificamente os gráficos das médias ajustadas versus valores observados, bem como os resíduos de Pearson versus os valores ajustados.

## 4 Resultados

Inicialmente verificou-se assimetria dos dados, a natureza contínua e o intervalo positivo de variação, características nas quais sugerem que a distribuição dos dados segue o modelo Gamma ou Normal Inverso. No entanto, o modelo Gamma não ajustou-se adequadamente aos dados.

Para ajustarmos o modelo Normal Inverso com função de ligação logarítmica a esses dados, assumiremos que  $y_1, \dots, y_{27}$  são independentes e seguem uma distribuição Normal Inversa com média  $\mu_t$ ,  $t = 1, \dots, 27$ , e parâmetro de dispersão  $\phi$  desconhecido.

Em seguida, as estimativas dos parâmetros mostraram que todas as variáveis foram significativas, produzindo um desvio residual na magnitude de 0,010 e AIC (Critério de Informação de Akaike) de 120,47. Relizou-se a análise de diagnóstico deste ajuste e verificou-se que as observações 3, 4, 6, 7, 9, 15 e 19 referentes aos Estados do Amazonas, Roraima, Amapá, Tocantins, Piauí, Sergipe e Rio de Janeiro, respectivamente, se configuraram como atípicas. Essas observações foram retiradas e o modelo foi novamente ajustado.

A retirada das observações melhorou significativamente o ajuste do modelo, reduzindo o Erro Padrão e o p-valor relacionado ao parâmetro de cada variável, assim como reduziu o desvio residual (0,003) e o AIC (74,75). As estimativas encontram-se na Tabela 1.

As medidas de diagnóstico para o novo modelo sem as observações citadas acima foram realizadas. Surgiram novos pontos atípicos, porém, a retirada dessas observações não causaram grandes mudanças na variação percentual das estimativas dos parâmetros. Logo, optou-se por conservar essas observações no modelo.

O gráfico de probabilidade meio-normal com envelope simulado (Figura 1), mostra que todos os pontos se encontram dentro da banda de confiança, indicando um bom ajuste dos dados.

Para verificar se o modelo é razoável para ajustar os dados das taxas de mortalidade infantil, apresentamos na Figura 2(a) o gráfico de dispersão dos valores observados versus médias ajustadas. Observa-se através da figura que os pontos apresentam-se de forma linear, indicando que os dados foram bem ajustados. Além disso, na Figura 2(b), os resíduos de Pearson apresentaram uma distribuição

aleatória quando feita a sua dispersão versus os valores ajustados, indicando assim, que os resíduos são não correlacionados, ou seja, hipótese de independência e variância constante para os resíduos são aceitas.

Tabela 1: Estimativas dos parâmetros do modelo Normal Inverso com função de ligação logarítmica ajustado para os dados de TMI sem as observações (3, 4, 6, 7, 9, 15 e 19).

Parâmetro	Estimativa	Erro Padrão	P-valor
$\beta_0$	2,48188	0,08773	0,00000
$\beta_1$	-0,00229	0,00083	0,01414
$\beta_2$	0,00944	0,00289	0,00488
$\beta_3$	0,03714	0,00996	0,00183
$\phi$	0,00020	0,37422	-

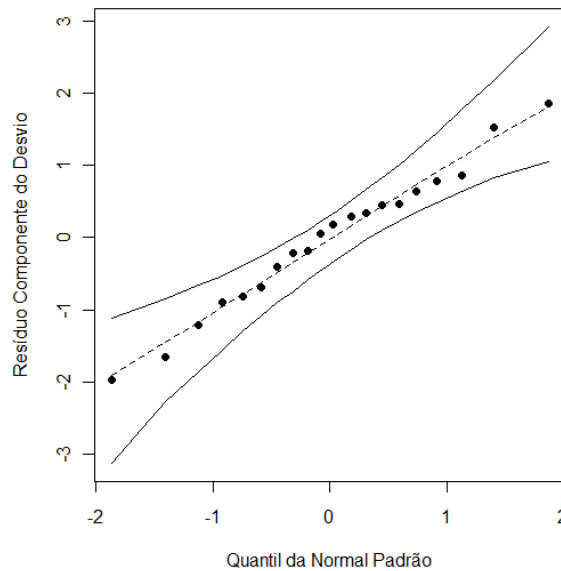


Figura 1: Percentil da Normal(0,1)

A partir das estimativas dos parâmetros do modelo Normal Inverso com ligação logarítmica, pode-se agora, escrever o modelo ajustado.

$$\hat{\mu} = \exp(2,48188 - 0,00229x_{t1} + 0,00944x_{t2} + 0,03714x_{t3})$$

em que  $\beta_0 = 2,48188$  corresponde ao intercepto,  $x_{t1}$  representa a Cobertura de esgotamento sanitário,  $x_{t2}$  a Proporção de pobres e  $x_{t3}$  a Taxa de analfabetismo, sendo  $t$  o respectivo Estado.

A estimativa do parâmetro referente a variável Cobertura de esgotamento sanitário (-0,00229), apresenta uma relação inversa com a TMI, ou seja, o aumento da cobertura de esgotamento sanitário

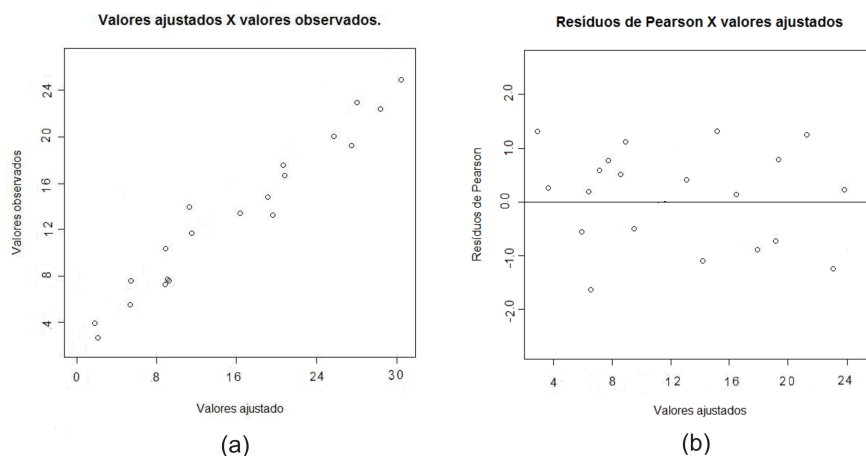


Figura 2: Gráficos de verificação da qualidade do ajuste.

nos estados, implica em diminuição da TMI naquela região. Teixeira & Guilhermino (2006) também comprovam essa relação.

A segunda estimativa está relacionada com a variável Proporção de pobres, variável na qual teve sua estimativa do parâmetro de 0,00944. Essa estimativa significa que, para cada unidade positiva da Proporção de pobres, a TMI aumenta em 0,944%.

Por último, tem-se a Taxa de analfabetismo. Esse parâmetro nos informa que para cada unidade positiva da Taxa de analfabetismo, a TMI cresce na magnitude de 0,03714. A redução do analfabetismo contribui para a queda da mortalidade infantil, 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.

Nesse estudo o modelo apresentado para explicar os componentes da mortalidade infantil nos estados brasileiros foram expressos como combinações lineares entre indicadores socioeconômicos. A relação de causa e consequência não está perfeitamente caracterizada, sendo preferível o termo “possíveis determinantes” para as variáveis utilizadas no ajuste. Porém, em MLG, um modelo parcimonioso é, também, uma exigência, no sentido de que o número de parâmetros seja tão pequeno quanto possível (CORDEIRO & DEMÉTRIO, 2007).

## 5 Conclusões

A teoria dos MLG mostrou-se uma ferramenta bastante eficaz para modelagem das Taxas de Mortalidade Infantil explicada por indicadores socioeconômicos nos Estados brasileiros referenciando o ano de 2006. Permitindo assim, ajustar os dados por meio do modelo Normal Inverso com ligação logarítmica.

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 diminuição da Proporção de pobres na população.

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 públicas nas áreas de saúde materna e infantil. Fazendo com que a TMI no Brasil, atinja um patamar ideal, isto é,

próxima das taxas dos países desenvolvidos.

## Referências

- [1] CORDEIRO, G. M., DEMÉTRIO, C. G. B. Modelos lineares generalizados. Minicurso para o 12º SEAGRO e a 52º Reunião Anual da RBRAS, UFSM, Santa Maria, RS, 2007.
- [2] IBGE - Instituto Brasileiro de Geografia e Estatística, Síntese dos Indicadores Sociais. Disponível em: <http://www.ibge.gov.br> , Acesso em 12 Mai. 2008.
- [3] MINISTÉRIO DA SAÚDE - DATASUS, Indicadores de Saúde. Disponível em: <http://tabnet.datasus.gov.br> , Acesso em 12 Mai. 2008.
- [4] MONTEIRO, C. A., NAZÁRIO, C. L. , 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 doenças. 2.ed São Paulo, *Hucitec*,(2000) 173-85.
- [5] NELDER, J. A., WEDDERBURN, R. W. M. Generalized linear models. *Journal of the Royal Statistical Society*, A, 135 (1972) 370-384.
- [6] ROSANO, A., BOTTO L. D., MASTROIACOVO, P. Infant mortality and congenital anomalies from 1950 to 1994: an perspective. *J. Epidemiol. Community Health*, 54 (2000) 660-6.
- [7] TEIXEIRA, J. C., GUILHERMINO, R. L. 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.