

Modelo para Mortalidade Infantil no Brasil via Modelos Lineares Generalizados

Carlos S. A. dos Santos, Alisson de O. Silva, Joseilme F. Gouveia, Gauss M. Cordeiro

Douglas Vinícius

Universidade Federal de Rondônia

03 de Dezembro de 2019

- Introdução
- Objetivo
- Material e Métodos
 - Modelos Lineares Generalizados
- Resultados
- Conclusões
- Referências

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).

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 18ª 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).

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.

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.

Os dados foram coletados das fontes:

- 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.

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 μ_1, \dots, μ_n , ou seja,

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

em que $\phi > 0$ 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:

$$f(y_i; \theta_i, \phi) = \exp\{\phi^{-1}[y_i\theta_i - b(\theta_i)] + c(y_i, \phi)\} \quad (2)$$

$b(\cdot)$ e $c(\cdot)$ funções conhecidas.

Modelos Lineares Generalizados

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

$$\eta_i = \sum_{r=1}^p x_{ir} \beta_r = x_i^t \beta \quad \text{ou} \quad \eta = X\beta \quad (3)$$

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,

$$\eta_i = g(\mu_i) \quad (4)$$

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.

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 μ_t , $t = 1, \dots, 27$, e parâmetro de dispersão ϕ 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. Realizou-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.

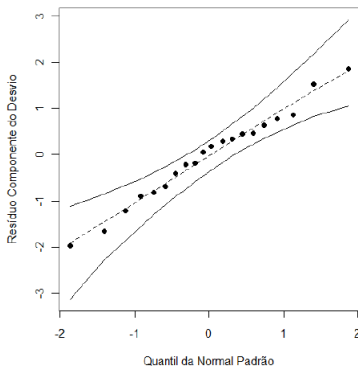
A retirada das observações melhorou significativamente o ajuste do modelo, reduzindo o Erro Padrão e o p-valor relacionado ao parâmetro de cada variável, assim como reduziu o desvio residual (0,003) e o AIC (74,75). As estimativas encontram-se na Figura 1.

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

Parâmetro	Estimativa	Erro Padrão	P-valor
β_0	2,48188	0,08773	0,00000
β_1	-0,00229	0,00083	0,01414
β_2	0,00944	0,00289	0,00488
β_3	0,03714	0,00996	0,00183
ϕ	0,00020	0,37422	-

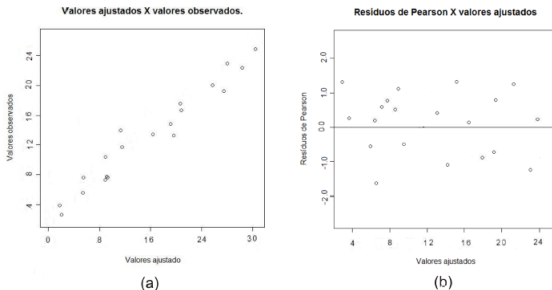
O gráfico de probabilidade meio-normal com envelope simulado (Figura 2), mostra que todos os pontos se encontram dentro da banda de confiança, indicando um bom ajuste dos dados.

Figure 2: Percentil da Normal(0,1)



A 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.

Figure 3: Gráficos de verificação da qualidade do ajuste.



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

$$\hat{\mu} = \exp\{2,48188 - 0,00229x_{t1} + 0,00944x_{t2} + 0,03714x_{t3}\} \quad (5)$$

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 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).

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.

- Instituto Brasileiro de Geografia e Estatística, Síntese dos Indicadores Sociais. *Instituto Brasileiro de Geografia e Estatística, Síntese dos Indicadores Sociais*. Acesso em 12 Mai. 2008.. IBGE. <http://www.ibge.gov.br>.
- Monteiro, C. and C. Nazário (2000). "Evolução da agenda tradicional da saúde pública: mortalidade infantil e materna. Declínio da mortalidade infantil e equidade social: o caso da cidade de São Paulo entre 1973 e 1993". In: *Monteiro CA, organizador. Velhos e novos males da saúde no Brasil: a evolução do país e suas doenças* 2, pp. 173–85.
- Nelder, J. and R. Wedderburn (1972). "Generalized linear models, journal of the Royal Statistical Society, A, 135".
- Rosano, A, L. D. Botto, B. Botting, et al. (2000). "Infant mortality and congenital anomalies from 1950 to 1994: an international perspective". In: *Journal of Epidemiology & Community Health* 54.9, pp. 660–666.
- Teixeira, J. C. and R. L. Guilhermino (2006). "Análise da associação entre saneamento e saúde nos estados brasileiros, empregando dados secundários do banco de dados indicadores e dados básicos para a saúde 2003–IDB 2003". In: *Engenharia Sanitária e Ambiental*.