

Análise da taxa de incidência da dengue no estado da Paraíba usando GAMLSS

Fernanda De Bastiani¹ e Alisson L. Brito²

Departamento de Estatística - UFPE, Recife, PE

A dengue é uma doença de etiologia viral e infecciosa, que representa um sério problema de saúde pública, causando sérios prejuízos relacionados à saúde e a economia de países. A enfermidade é causada por 4 sorotipos distintos (DENV1, DENV2, DENV3 e DENV4) e apresenta duas formas clínicas, dengue clássica e febre hemorrágica da dengue, sendo que, a segunda forma pode levar o indivíduo acometido a morte. No Brasil, a transmissão se dá, principalmente, por meio da picada do mosquito fêmea do gênero *Aedes Aegypti* contaminado com um dos sorotipos.

Os modelos aditivos generalizados de localização, escala e forma (GAMLSS) tem sido cada vez mais utilizados para a análise de dados. [1] discutem detalhadamente os GAMLSS, uma classe de modelos de regressão que estendem os modelos lineares generalizados e modelos aditivos generalizados. Nos GAMLSS todos os parâmetros da distribuição podem ser modelados em função das variáveis explicativas. Estas funções podem ser lineares e/ou funções de suavização.

Um aspecto importante no estudo de modelos de regressão é verificar se há desvio de pressuposições do modelo, e também verificar se há presença de observações influentes. [2] introduziu a análise de influência para modelos lineares, medindo o impacto das observações no ajuste do modelo. Diversos autores têm estudado e estendido essa ideia para identificar observações influentes nas classes dos modelos aditivos e semiparamétricos, como [3].

Este trabalho busca detectar observações influentes na classe dos modelos GAMLSS no estudo da taxa de incidência de dengue de 2008 a 2018 nos municípios do estado da Paraíba, Brasil. Foram consideradas como variáveis explicativas: o índice de desenvolvimento humano (IDH), a precipitação média anual em milímetros (PRECIP), o índice Gini (GINI), o número de domicílios ligados a rede geral de esgoto (RGE), a proporção de domicílios com saneamento básico (PSB), a temperatura média anual em graus Celsius (TEMP) e a taxa de desemprego (TD). Estas variáveis se mostraram ter relação com a taxa de incidência de dengue através do ajuste do modelo Gama Generalizado, $GG(\hat{\mu}, \hat{\sigma}, \hat{\nu})$. Na equação (1) é possível verificar para quais parâmetros da distribuição gamma generalizada as variáveis explicativas apresentaram algum tipo de relação com a variável resposta.

$$\begin{aligned} \log(\hat{\mu}) &= -2.2129 + 12,6966\text{IDH} - 0,0011\text{PRECIP} - 4.6035\text{GINI} \\ &\quad + h_{41}(\text{RGE}) + h_{51}(\text{PSB}) + h_{61}(\text{TEMP}) \\ \log(\hat{\sigma}) &= 2.5472 - 0,1049\text{TEMP} + h_{22}(\text{PRECIP}) - 0,03107\text{TD} \\ \hat{\nu} &= 1.6365 - 0.0016\text{PRECIP} \end{aligned} \tag{1}$$

em que $h(\cdot)$ representa uma função P-spline.

Na Figura 1 (bloco da esquerda), observa-se uma diminuição nos valores de $\log(\hat{\mu})$ a medida que há um aumento nos valores de precipitação e do índice de Gini. Por outro lado, os valores ajustados aumentam em municípios com maiores IDH. Já para o número de domicílios com rede geral de esgoto observa-se um decréscimo em $\log(\hat{\mu})$ para valores até 9.000 aproximadamente, a partir disto, os valores aumentam linearmente. Para maiores valores da proporção de domicílios

¹fernanda.bastiani@ufpe.br

²alisson.limab@ufpe.br

com saneamento básico ver-se um aumento constante nos valores de $\log(\hat{\mu})$. Na Figura 1 (bloco da direita) observa-se a relação ajustada entre as variáveis explicativas e os valores ajustados para $\log(\hat{\sigma})$. Especialmente interessante verificar como varia a relação da variável PRECIP em $\log(\hat{\sigma})$. O gráfico worm plot (omitido aqui), mostrou que o ajuste está adequado.

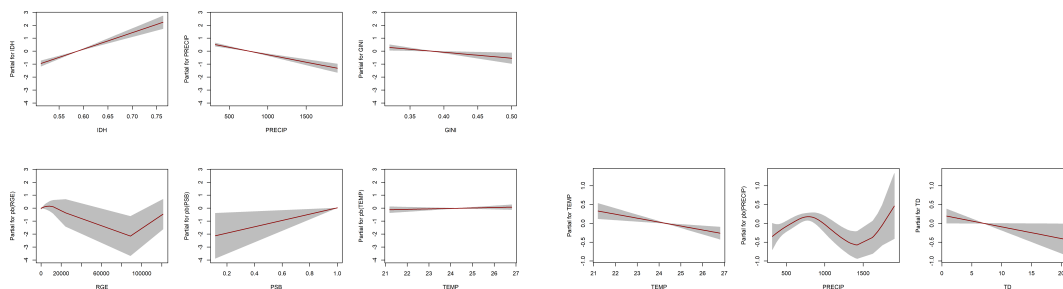


Figura 1: Relação entre as variáveis explicativas e os valores ajustados para $\hat{\mu}$ (bloco da esquerda). Relação entre as variáveis explanatórias e os valores ajustados para $\hat{\sigma}$ (bloco da direita).

A Figura 2 (esquerda) mostra uma novidade, o gráfico do incremento do desvio, e nota-se que os pontos potencialmente influentes são 77, 123 e 163 que se referem aos municípios de Dona Inês, Monte Horebe e Riacho dos Cavalos, respectivamente. Na Figura 2 (direita) nota-se que os municípios que apresentaram maiores taxas para a incidência da dengue foram Várzea seguido por Santa Luzia. Municípios com baixa precipitação no período de estudo.

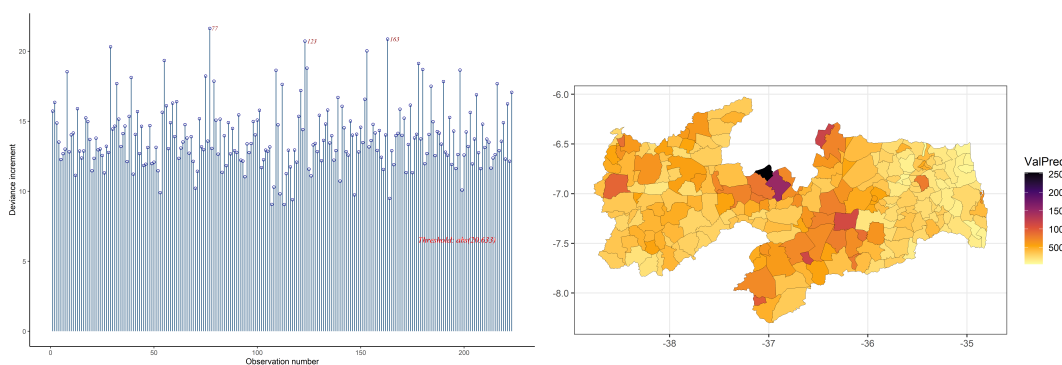


Figura 2: Gráfico do incremento de desvio para detectar potenciais observações influentes. Valores ajustados para o parâmetro μ por meio do modelo ajustado com a distribuição gama generalizada.

Referências

- [1] M D Stasinopoulos, R A Rigby, G Z Heller, V Voudouris e F De Bastiani. **Flexible Regression and Smoothing: Using GAMLSS in R**. CRC Press, 2017.
- [2] R D Cook. “Detection of Influential Observation in Linear Regression”. Em: **Technometrics** 19.1 (1977), pp. 15–18.
- [3] S Türkan e O Toktamis. “Detection of influential observations in semiparametric regression model”. Em: **Revista Colombiana de Estadística** 36.2 (2013), pp. 271–284.