



Fatores associados à Sífilis em gestantes residentes em cidades de porte médio no Estado do PR nos períodos de 2012 a 2016

Stéfane Lele¹, Robson M. Rossi², Larissa Falavina³, Thaís A. F. Mathias⁴

¹Discente do programa de mestrado PBE - Departamento de Estatística - UEM

²Docente do Departamento de Estatística - UEM

³Discente do programa de mestrado PSE - Departamento de Enfermagem - UEM

⁴Docente do Departamento de Enfermagem - UEM

RESUMO

Quando os dados relativos a uma análise possuem informações sobre a contagem do desfecho da variável resposta e sobre o total de ensaios, podemos utilizar uma regressão Binomial com função de ligação “logit”. O enfoque bayesiano foi utilizado para analisar os dados devido à capacidade de contornar pressupostos frequentistas quando as suposições não são satisfeitas, dados estes que continham informações sobre a quantidade de gestantes com Sífilis nos municípios contendo entre 20 e 80 mil habitantes do Estado do Paraná entre os períodos de 2012 a 2016. Na regressão, as variáveis explicativas referentes à porcentagem de área do município coberta pelo programa Saúde da Família, gestantes de raça não-branca, gestantes adolescentes com idade menor ou igual a 19 anos, gestantes chefes de família sem ensino fundamental e indivíduos extremamente pobres no município mostraram-se significativas, além de indicarem que os municípios com área coberta pelo programa de saúde e com indivíduos extremamente pobres possuem um fator de proteção (*Odds ratio* = 0,997 e 0,899, respectivamente) em relação ao contágio da doença. Já as demais variáveis, verificou-se que a partir de seu aumento, o risco das gestantes apresentarem Sífilis também aumenta com uma *Odds ratio* de 1,011, 1,042 e 1,008 para raça não-branca, adolescentes e mães chefes, respectivamente.

Palavras chave: logit, Binomial, odds ratio, inferência bayesiana.

1 INTRODUÇÃO

A Sífilis é uma doença causada pela bactéria *Treponema pallidum*, transmitida na maior parte das vezes por contato íntimo sem proteção. Os primeiros sintomas são feridas indolores das quais se não forem tratadas, desaparecem e reaparecem após algum tempo, em fases mais avançadas. Quando a infecção surge durante a gravidez, ela pode infectar o feto, podendo causar má-formação, aborto, ou até morte do bebê (SEDICIAS, 2017).

Com o objetivo de compreender o comportamento do fenômeno, ou seja, como as gestantes no Estado do Paraná são afetadas pela Sífilis, uma análise de regressão com enfoque bayesiano foi utilizada para identificar variáveis que contribuem com o contágio da doença.

2 METODOLOGIA

O banco de dados referente a este trabalho foi obtido por meio do programa de mestrado em Enfermagem - UEM em que a variável resposta Y diz respeito à quantidade de gestantes diagnosticadas com Sífilis no período de 2012 a 2016 residentes nos municípios com 20.000 a 80.000 habitantes no Estado do Paraná, totalizando 68 cidades a serem analisadas.

As variáveis explicativas disponíveis para as análises foram: Renda média municipal, IDHM, Cobertura de saúde, População baixa renda, Raça não-branca, Consulta pré-natal, Adolescente, Escolaridade < 8 anos, Sem companheiro, Porcentagem de mães chefes, Taxa de analfabetismo, Porcentagem de pobreza, Porcentagem de extrema pobreza, Porcentagem de vulneráveis à pobreza e Taxa de desemprego.

Para o caso das gestantes com Sífilis nos municípios de porte médio do Paraná (entre 20.000 a 80.000 habitantes), foi realizada uma análise assumindo uma distribuição Binomial que comporta o formato dos dados, que

contêm além da contagem do número de gestantes com Sífilis por município, o número total de gestantes por município, com o objetivo de compreender o comportamento do fenômeno identificando os fatores que influenciam no diagnóstico destas mulheres.

O enfoque bayesiano foi adotado devido à capacidade de contornar os pressupostos frequentistas comumente assumidos como normalidade dos dados, amostras de tamanho suficientemente grandes, entre outros, para se obter, sem errôneas suposições, um modelo adequado capaz de explicar o comportamento da Sífilis no Estado do Paraná (NTZOUFRAS, 2011).

Foi considerado que $Y_i \sim Bin(n, p)$ já que o banco de dados fornece informações sobre a quantidade de gestantes com o desfecho (sucesso) e o número de gestantes por município (n).

A função de ligação utilizada para associar os valores da resposta ao preditor linear do modelo de forma que esta relação possa ser modelada por uma regressão linear foi a *logit*, dada por $logit(p) = \log\left(\frac{p}{1-p}\right) = X\beta$, sendo X a matriz de covariáveis explicativas $[n(k+1)]$ e β o vetor de parâmetros $(\beta_1, \beta_2, \dots, \beta_k)$.

Por meio do programa R (R CORE TEAM, 2017), todas as análises estatísticas foram realizadas por meio do pacote não-oficial *BioBayes*, ainda em desenvolvimento, capaz de fazer automaticamente uma ligação direta do R ao programa *OpenBUGS*, gerando relatórios com informações e gráficos a respeito das cadeias e estimativas bayesianas.

A verificação de variáveis colineares entre si faz-se necessária pois se $X'X$ é singular (não tem inversa), ou seja, algumas covariáveis estão correlacionadas com outras, não há estimadores de Mínimos Quadrados único para os parâmetros (SEBER, 2012). Nesse caso, se tratando de covariáveis contínuas (taxas e proporções), o teste de correlação de Pearson disponível no programa R foi utilizado.

Seguindo a sugestão de Hosmer et al. (2013), foi realizada uma análise bivariada entre a resposta e cada variável preditora pois as variáveis que apresentarem valor- p menor que 0,25 são candidatas a entrarem no modelo inicial.

Para o ajuste do modelo com resultados consistentes via método bayesiano foram fixadas 10.000 iterações em um processo MCMC (Monte Carlo Cadeias de Markov), 1.000 valores para descarte amostral (*burn*), e saltos de tamanho 10 (*jump*), além de distribuições *a priori*s não-informativas para os coeficientes de regressão $\beta \sim N(0, 0, 000001)$ com convergência das cadeias verificadas a partir dos critérios de Heidelberg e Welch (1983).

A significância das variáveis foi verificada a partir dos intervalos de credibilidade (95%) para a média dos parâmetros, de forma que se o mesmo contempla o valor zero, indica que β_j não é significativo.

Identificando as variáveis significativas, o cálculo da *Odds Ratio* (e^{β_j}), $j = 0, 1, \dots, k$ nos dá interpretações diretas a respeito do risco/proteção que cada fator representa em relação ao desfecho, utilizando-se a média estimada dos parâmetros quando *a posteriori* é simétrica e a mediana estimada quando *a posteriori* for assimétrica.

3 RESULTADOS E DISCUSSÕES

A Tabela 1 apresenta um estatística descritiva na qual pôde-se verificar que apenas 2 municípios de porte médio possuem gestantes livres de Sífilis.

Tabela 1: Quantidade de municípios para cada número de casos de Sífilis.

| | | | | | | | | | | | | | | | | | | |
|--------------------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|
| Nº de casos | 0 | 1 | 3 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 20 |
| Quantidade de municípios | 2 | 1 | 1 | 3 | 1 | 9 | 2 | 2 | 3 | 1 | 1 | 4 | 1 | 4 | 2 | 2 | 1 | 2 |
| Nº de casos | 21 | 25 | 26 | 27 | 28 | 29 | 34 | 36 | 37 | 40 | 48 | 49 | 51 | 58 | 59 | 61 | 65 | 112 |
| Quantidade de municípios | 2 | 3 | 1 | 1 | 2 | 3 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

Por meio do teste de correlação de Pearson foram identificadas algumas covariáveis com multicolinearidade, apresentadas na Tabela 2:

Tabela 2: Tabela dos resultados dos testes de correlação entre as variáveis multicolineares

| Variáveis | Correlação (%) |
|---|----------------|
| Renda média x Taxa de pobreza | ≈ -55 |
| Taxa pobreza x Taxa extrema pobreza | ≈ 90 |
| Taxa pobreza x Taxa vulneráveis à pobreza | ≈ 90 |

As variáveis que obtiveram aproximadamente 90% de correlação foram retiradas, permanecendo apenas “Taxa de extrema pobreza” para representar as demais.

Tabela 3: Variáveis candidatas a entrarem no modelo inicial

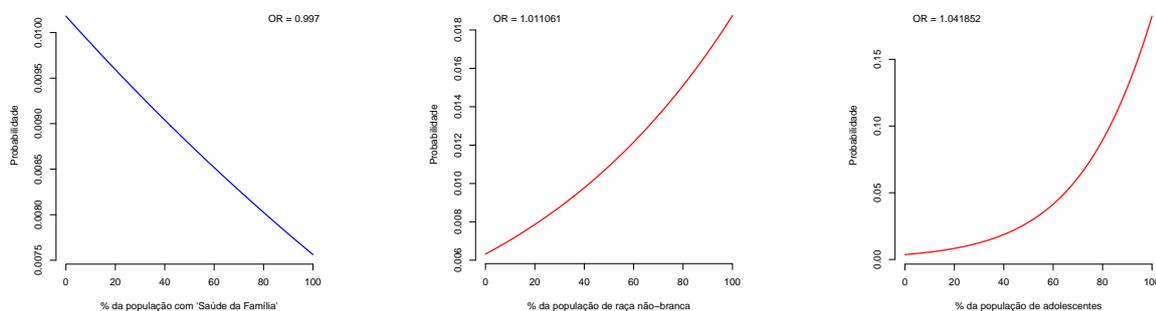
| Variável explicativa | valor-p |
|-----------------------|----------|
| Cobertura de Saúde | 0,0822 |
| Raça não-branca | 7,86e-07 |
| Consulta pré-natal | 0,0525 |
| Adolescente | 1,35e-05 |
| Escolaridade < 8 anos | 1,66e-05 |
| Mães chefes | 0,000113 |
| Extrema pobreza | 0,0768 |

Tabela 4: Estimativas *a posteriori* dos parâmetros.

| | Média | Erro padrão | Mediana | ICr(95%) |
|-----------|--------|-------------|---------|------------------|
| β_0 | -5,739 | 0,174 | -5,734 | [-6,094; -5,413] |
| β_1 | -0,003 | 0,001 | -0,003 | [-0,006; -0,001] |
| β_2 | 0,011 | 0,002 | 0,011 | [0,007; 0,016] |
| β_3 | -0,007 | 0,003 | -0,007 | [-0,013; 0] |
| β_4 | 0,041 | 0,014 | 0,04 | [0,014; 0,069] |
| β_5 | 0,006 | 0,006 | 0,006 | [-0,007; 0,018] |
| β_6 | 0,008 | 0,004 | 0,008 | [0,001; 0,016] |
| β_7 | -0,106 | 0,018 | -0,107 | [-0,138; -0,072] |

Nota-se que as variáveis Cobertura de Saúde, Raça não-branca, Adolescente, Mães chefes e Extrema pobreza referentes a β_1 , β_2 , β_4 , β_6 e β_7 , respectivamente, foram as variáveis significativas pois estas não contemplam o zero em seus intervalos de credibilidade (ICr(95%)).

Fixando as demais variáveis significativas no intercepto em seus valores médios e variando cada uma delas individualmente, podemos observar as seguintes probabilidades apresentadas na Figura 1:



(a) Cobertura de saúde

(b) Raça não-branca

(c) Adolescente

Figura 1: Probabilidades de ocorrência de Sífilis considerando a evolução de Cobertura de Saúde (a), Raça não-branca (b), Adolescente (c), Mães chefe (d) e Extrema pobreza (e).



(d) Mães chefes

(e) Extrema pobreza

Como pode ser visto na Figura 1(a), $OR_{cobertura} = 0,997$, ou seja, a cada aumento de 1% na porcentagem de população coberta pela Estratégia Saúde da Família, há uma diminuição de 0,003% no risco da gestante vir a desenvolver Sífilis, sendo o Programa de Saúde um fator de proteção.

Já na Figura 1(b), $OR_{raça} = 1,011$ indicando que a cada aumento de 1% na porcentagem de gestantes com raça não-branca, há um aumento no risco destas virem a ser diagnosticadas com Sífilis de 0,11%.

Pela Figura 1(c), $OR_{adolescente} = 1,041$. Portanto, vemos que a cada aumento de 1% na porcentagem de gestantes adolescentes (com idade menor ou igual a 19 anos), há um aumento de 0,41% no risco destas obterem diagnóstico positivo para Sífilis.

Já pela Figura 1(d), $OR_{mãeschefe} = 1,008$. Dessa forma, a cada aumento de 1% na porcentagem de mães chefes de família que não têm ensino fundamental, há um aumento no risco de diagnóstico de Sífilis de 0,08%.

Por conseguinte, como visto na Figura 1(e), $OR_{extremapobreza} = 0,899$, indicando que a cada aumento de 1% na porcentagem indivíduos em extrema pobreza no município (renda *per capita*: 70 reais), há uma diminuição no risco de diagnóstico positivo para Sífilis nas gestantes de 10%, ou seja, existe um fator de proteção relacionado a esta variável.

4 CONCLUSÃO

As variáveis Cobertura de Saúde, Raça não-branca, Adolescente, Mães chefes e Extrema pobreza são significativamente associadas à prevalência de Sífilis em gestantes nos municípios de porte médio (20 a 80 mil habitantes) no Estado do Paraná.

5 AGRADECIMENTOS

CNPq e Fundação Araucária.

Referências

- [1] ROSSI, R. M.; LELE, S. Biobayes. Pacote de análises Bayesianas (Material didático). v.1, 2017.
- [2] HEIDELBERGER, P.; WELCH P. D. Simulation run length control in the presence of an initial transient. *Operations Research*, v.31, p.1109-1144, 1983.
- [3] HOSMER J.; DAVID W.; LEMESHOW, S.; STURDIVANT, R.X. *Applied logistic regression*. John Wiley & Sons, 2013.
- [4] NTZOUFRAS, I. *Bayesian modeling using WinBUGS*. John Wiley & Sons, 2011.
- [5] R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL <<https://www.R-project.org/>>.
- [6] SEBER, G. A.; Lee, A. J. *Linear regression analysis* (Vol. 936). John Wiley & Sons, 2012.
- [7] SEDICIAS, S. Sífilis - o que é, sintomas e tratamento. Tua Saúde, 2017. Disponível em <<https://www.tuasaude.com/sifilis/>>