

Prematuridade e fatores associados: Modelagem bayesiana por meio de diferentes funções de ligação

Rafaela Galo ¹, Robson Marcelo Robson ², Rosana Rosseto de Oliveira ³

Introdução

A necessidade de entender as relações entre duas ou mais variáveis em um conjunto de dados foi a base para a construção dos métodos de regressão. Com este método é possível verificar como a variação em determinadas variáveis pode influenciar a variável resposta da análise. Respostas de característica binária são muito utilizadas em análises estatísticas e em especial na área da saúde, visto que as relações com a resposta envolvem probabilidade. Diante disso, a seleção do modelo que descreverá de maneira adequada e simples esta relação é uma parte fundamental da modelagem estatística.

Nelder e Wedderburn (1972) mostraram que os modelos lineares generalizados (MLG) para análises em que a distribuição de probabilidade da variável resposta era diferente da normal apresenta-se de maneira mais eficaz para a estimação dos parâmetros do modelo, pois utiliza como base a probabilidade. Estes modelos possuem três características importantes: pertencer à família exponencial que forma a componente aleatória do modelo; ser formado por um conjunto de variáveis explicativas (independentes) que forma a componente sistemática do modelo e por fim, tem uma função de ligação que relaciona os componentes sistemáticos e aleatórios, respectivamente.

Este estudo, tem como objetivo realizar uma análise Bayesiana sobre fatores que podem ser considerados como risco nos nascimentos de crianças prematuras no ano de 2017, na cidade de Maringá, Paraná, utilizando modelos de regressão binária com as funções de ligação *logit*, *probit*, *cauchit*, *cloglog* e *loglog* (Tabela 1), sendo as três primeiras classificadas como simétricas e as demais como assimétricas, buscando a que melhor se adeque aos dados (ABANTO-VALLE; DEY; JIANG, 2015). Com a escolha da ligação, busca-se verificar se a diferença de aproximadamente oito vezes mais a quantidade de recém-nascidos não prematuros (Tabela 2) é melhor representada a partir de uma ligação assimétrica (ANYOSA, 2017).

Material e Métodos

Foram analisados 5.060 registros de nascimentos vivos ocorridos na cidade de Maringá, Paraná, no ano de 2017 e quatorze fatores possivelmente associados à prematuridade. Estas informações provêm das Declarações de Nascidos Vivos (DNV) e neste caso, cedidos pela 15^a Regional de Saúde do Estado do Paraná, por ainda não estarem disponíveis *online* no Datasus mediante aprovação pelo Comitê Permanente de Ética em Pesquisa, sob parecer 406.927/2013.

A Organização Mundial da Saúde (WHO, 2018) define como prematuros bebês nascidos vivos antes de completar 37 semanas de gestação. Com isso, a variável resposta

¹Programa de Pós-graduação em Bioestatística, Universidade Estadual de Maringá - UEM. e-mail: galo.rafaela@gmail.com

²Docente do Departamento de Estatística, Universidade Estadual de Maringá - UEM. e-mail: rm-rossi@uem.br

³Enfermeira. Doutora em Enfermagem. Docente do Programa de Pós-Graduação em Enfermagem da Universidade Estadual de Maringá - UEM. e-mail: rosanarosseto@gmail.com

prematuridade foi obtida a partir da classificação de cada recém-nascido em prematuro ($Y = 1$) ou não ($Y = 0$) a partir da semana gestacional em que mãe se encontrava no momento do parto. As variáveis explicativas utilizadas foram selecionadas a partir dos estudos de Oliveira (2015) e Ramos e Cuman (2009), os quais apresentam que os fatores associados a prematuridade podem ser os relacionados à mãe (idade, companheiro, escolaridade e paridade); gestão e parto (tipo de gravidez, tipo de parto, consultas pré-natal e local do parto); recém-nascido (sexo, raça/cor, apgar no primeiro e quinto minuto, peso ao nascer e malformação congênita), formando assim as variáveis iniciais das análises (Tabela 2). O banco de dados foi categorizado segundo os critérios apresentados por (OLIVEIRA, 2015) e (RAMOS; CUMAN, 2009), com intuito de determinar as categorias de referência e de risco.

Inicialmente, considerando as 14 variáveis explicativas, utilizou-se uma regressão binária frequentista padrão (ligação *logit*) para pré-seleção por meio do AIC (Critério de Informação de Akaike) as variáveis significativas a 10%. Posteriormente, foram ajustados os modelos apresentados na Tabela 1, sob o enfoque Bayesiano a fim de se analisar e escolher o melhor ajuste.

Tabela 1: Funções de ligação

Nome	F	F^{-1}
<i>logit</i>	$p_i = \frac{\exp(\eta_i)}{1 + \exp \eta_i}$	$\log \left(\frac{p_i}{1 - p_i} \right)$
<i>probit</i>	$p_i = \Phi(\eta_i)$	$\Phi^{-1}(p_i) = \eta_i$
<i>cauchit</i>	$p_i = 0,5 + \frac{\arctan(\eta_i)}{\pi}$	$\tan(\pi(p_i - 0,5)) = \eta_i$
<i>cloglog</i>	$p_i = 1 - \exp(-\exp(\eta_i))$	$\log(-\log(1 - p_i)) = \eta_i$
<i>loglog</i>	$p_i = \exp(-\exp(-\eta_i))$	$-\log(-\log(p_i)) = \eta_i$

Φ é a função de distribuição acumulada normal padrão

Fonte: Anyosa (2017)

em que $F(\cdot)$ denota a função distribuição acumulada (fda), F^{-1} a função de ligação e η_i o i -ésimo preditor linear, ou seja, $\eta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik}$, sendo $i = 1, \dots, 5.060$ e k possíveis variáveis explicativas.

Foram considerados no processo de estimação, distribuições *a priori* normais não informativas para os parâmetros, isto é, $\beta_0 \sim N(0, 10^{-6})$ e $\beta_k \sim N(0, 10^{-6})$, ($\sigma^2 = \tau^{-1}$, parametrização OpenBugs). A seleção do modelo foi feita segundo o critério de informação DIC (Deviance Information Criterion) proposto por Spiegelhalter et al. (2002).

A obtenção das distribuições marginais condicionais *a posteriori* para os parâmetros foram obtidas por meio do programa R (R CORE TEAM, 2018) através do pacote (não oficial) *BioBayes* disponível no link GitHub “<https://github.com/rmrossidesuem/BioBayes>”. Para os modelos com ligação *logit*, *probit* e *loglog* foram gerados 10.000 valores considerando um período de descarte amostral de 1.000 e saltos de tamanho 10 e para os modelos com ligação *cauchit* e *cloglog* foram gerados 50.000 valores considerando um período de descarte amostral de 1.000 e saltos de tamanho 50 para reduzir a correlação. A análise de convergência das cadeias se dá pelo critério de Heidelberger e Welch e técnicas gráficas.

Como estimativa para os parâmetros foi utilizada a mediana *a posteriori* da distribuição marginal condicional e suas significâncias foram verificadas caso o valor zero não pertence ao seus respectivos intervalos HPD (*Highest Posterior Density*).

Resultados e Discussões

A Tabela 2 apresenta uma análise descritiva das variáveis explicativas que compõem o modelo obtido na etapa de seleção, a qual observa-se que as porcentagens associadas a prematuros são menores do que as relacionadas a não prematuros. Com estas variáveis explicativas, os modelo de regressão binária com as ligações apresentadas da Tabela 1 foram ajustados via funções do pacote *BioBayes*.

Tabela 2: Análise descritiva dos dados.

Variável	Categoria	Prematuridade			
		Não		Sim	
		Fi	%	Fi	%
Companheiro	0 (sim)	3.310	65,4	459	9
	1 (não)	1.152	22,7	139	2,7
Tipo de gravidez	0 (única)	4.423	87,4	481	9,5
	1 (múltipla)	39	0,7	117	2,3
Tipo de parto	0 (vaginal)	1.051	20,7	97	1,9
	1 (cesárea)	3.411	67,4	501	9,9
Peso	0 ($\geq 2500\text{g}$)	4.357	86,1	278	5,4
	1 ($< 2500\text{g}$)	105	2	320	6,3
Consultas pré-natal	0 (≥ 7)	3.850	76	421	8,3
	1 (< 7)	612	12	177	3,4
Índice apgar no 1º min	0 (≥ 7)	4.234	83,6	496	9,8
	1 (< 7)	228	4,5	102	2
Índice apgar no 5º min	0 (≥ 7)	4.436	87,6	572	11,3
	1 (< 7)	26	0,5	26	0,5
Raça	0 (branco)	3.142	62	454	8,9
	1 (não branco)	1.320	26	144	2,8
Geral	-	4.462	88,2	598	11,8

Fonte: Os autores

Os valores dos DIC's dos modelos ajustados estão apresentados na Tabela 3, em que se verifica alguns resultados próximos. Modelos com menores valores de DIC podem ser considerados mais adequados, já que apresentam ajuste ponderado pelo grau de complexidade do modelo (SPIEGELHALTER et al., 2002). Assim, pode-se concluir que os

modelos com ligações *logit*, *probit* e *loglog* são ditos equivalentes, isto é, qualquer um destes modelos ajustam-se bem aos dados.

Tabela 3: Valores dos DIC's e tempo de processamento (TP) para os modelos analisados.

Função de ligação	DIC	TP(min)
<i>logit</i>	2.444	83
<i>probit</i>	2.445	7
<i>cauchit</i>	2.465	9
<i>cloglog</i>	2.470	35
<i>loglog</i>	2.447	2

Fonte: Os autores

Levando em consideração o amplo conhecimento e uso da função de ligação em diferentes áreas do conhecimento, a escolha do modelo final seria o que tem a ligação *logit*. Entretanto, a simulação das cadeias para esta modelo demanda um tempo de processamento computacional maior que as demais ligações, o que reduz as opções de escolha do modelo com as funções de ligação *probit* (7 minutos) ou *loglog* (2 minutos). Com isso, o modelo de regressão binária escolhido será o que apresenta menor tempo de processamento, ou seja, o que tem função de ligação *loglog* que é dado na Equação 1.

$$\begin{aligned} \loglog(p_i) = & \beta_0 + \beta_1 \text{ companheiro} + \beta_2 \text{ tip. de gravidez} + \beta_3 \text{ tip. de parto} \\ & + \beta_4 \text{ peso} + \beta_5 \text{ cons. pré-natal} + \beta_6 \text{ apgar 1} + \beta_7 \text{ apgar 5} + \beta_8 \text{ raça} \end{aligned} \quad (1)$$

O resumo das estimativas obtidas *a posteriori* para o modelo está representando na Tabela 4. É possível verificar que apenas o parâmetro β_3 , associado ao tipo de parto, apresenta-se não significativo, visto que o valor zero está incluído no intervalo HPD.

Tabela 4: Estimativas *a posteriori* dos parâmetros com seus respectivos intervalos HPD e das *Odds Ratio* (OR) do modelo com ligação *loglog*.

Parâmetros	Estimativas	Intervalo HPD	OR
β_0	-1,149	(-1,265 ; -1,045)	0,317
β_1	-0,111	(-0,191 ; -0,009)	0,895
β_2	1,291	(0,936 ; 1,576)	3,636
β_3	0,079	(-0,028; 0,183)	1,083
β_4	1,987	(1,788 ; 2,215)	7,295
β_5	0,318	(0,200 ; 0,420)	1,374
β_6	0,365	(0,207 ; 0,514)	1,440
β_7	0,515	(0,034 ; 0,906)	1,673
β_8	-0,116	(-0,219 ; -0,032)	0,890

Fonte: Os autores

Das características socioeconômicas e demográficas, Oliveira (2015) apresenta que diferentes estudos em distintas localidades, a raça/cor materna preta tem sido considerada como fator de risco para o nascimento prematuro, assim como a ausência de companheiro durante a gestação. Entretanto, neste estudo as estimativas das OR para os parâmetros β_1 e β_8 são menores que 1, indicando que estas variáveis explicativas reduzem os riscos de ocorrer a prematuridade, ou seja, para mulheres que tem companheiros e para recém-nascidos brancos têm-se aproximadamente 12% mais de chance da prematuridade. Acredita-se que esta inversão possa ter ocorrido devido às características do município

em estudo, onde a maior parte dos nascimentos prematuros ocorrem no período limítrofe (34 à < 37 semanas), oriundos de partos cesáreas, que são mais frequentes entre mulheres brancas, com melhores condições socioeconômicas e com parto realizado na rede privada de saúde.

Os nascimentos prematuros podem ser evitáveis quando se têm uma assistência pré-natal de qualidade à gestante, que vai além de um número mínimo de consultas, mas com profissionais de conhecimentos clínicos e científicos substanciais para fornecer auxílio necessário para intervenções em situações adversas, como a gravidez múltipla (OLIVEIRA, 2015). Os resultados desse estudo, mostraram que o número insuficiente de consultas pré-natal foi associado significativamente ao nascimento prematuro, indicando que uma chance 37% maior, além do índice apgar menor que 7 no 1º e 5º minutos, que apresentaram uma maior chance de 44% e 67%, respectivamente. Ainda, fatores que são ditos como agravos a prematuridade (RAMOS; CUMAN, 2009), a gravidez múltipla apresentou aproximadamente 4 vezes mais chance para o desfecho e o baixo peso ao nascer aproximadamente 7 vezes mais quando comparados ao nível baseline (não risco).

Conclusão

O modelo com função de ligação assimétrica *loglog* apresentou-se um dos com menor DIC e também com menor tempo de processamento para a estimação dos parâmetros via métodos Bayesianos.

Como principais conclusões relacionadas ao desfecho do estudo, a prematuridade esteve associada a mãe ter companheiro, a gravidez ser múltipla, o número de consultas pré-natal inadequado, o recém-nascido ser não branco, um baixo peso, malformação congênita além dos índices apgar no primeiro e quinto minutos apresentarem abaixo do nível preconizado.

Agradecimentos

Agradecemos a CAPES pelo suporte financeiro.

Referências

- ABANTO-VALLE, C. A.; DEY, D. K.; JIANG, X. Binary state space mixed models with flexible link functions: a case study on deep brain stimulation on attention reaction time. *Statistics and Its Interface*, International Press of Boston, v. 8, n. 2, p. 187–194, 2015.
- ANYOSA, S. A. C. *Regressão binária usando ligações potência e reversa de potência*. Tese (Doutorado) — Universidade de São Paulo, 2017.
- NELDER, J. A.; WEDDERBURN, R. W. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, Wiley Online Library, v. 135, n. 3, p. 370–384, 1972.
- OLIVEIRA, R. R. d. Nascimento prematuro no estado do paran  e no munic pio de maring . Universidade Estadual de Maring , 2015.

R CORE TEAM. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2018. Disponível em: <https://www.R-project.org/>.

RAMOS, H. d. C.; CUMAN, R. K. N. Fatores de risco para prematuridade: pesquisa documental. *Esc Anna Nery Rev Enferm*, SciELO Brasil, v. 13, n. 2, p. 297–304, 2009.

SPIEGELHALTER, D. J. et al. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Wiley Online Library, v. 64, n. 4, p. 583–639, 2002.

WHO. *World Health Organization. Preterm birth*. 2018. Disponível em: <http://www.who.int/news-room/fact-sheets/detail/preterm-birth>. Acesso em: 24 agosto de 2018.