Risco Relativo e a Regressão Binomial

Diego Corrêa Alves [1], Juliana F.C.Gardelli [1] e Isolde Previdelli [1]

RESUMO

É comum na área da saúde e biológicas identificar fatores de risco, por meio de variáveis com desfechos binários. O objetivo é associar este desfecho binário à um fator de exposição. A estatística mais utilizada para verificar esta associação é a Razão de Chances (OR: *Odds ratio*). Contudo, a interpretação desta estatística não é muito intuitiva, sendo muitas vezes erroneamente interpretada como outra estatística, o Risco Relativo (RR: *Relative risk*). Ambas as estatísticas podem ser estimadas por um modelo de regressão Binomial, no entanto, com diferentes funções de ligação, logit (Modelo logístico: OR) e log (Modelo log-binomial: RR). Neste estudo, foi utilizada uma simulação para comparar o desempenho dos modelos logístico e log-binomial na estimação do risco relativo. Os conjuntos de dados simulados foram utilizados para ajustar um modelo log-binomial e logístico, sendo neste último, estimado o RR através da OR e do método delta. Com base nestes resultados, pode-se concluir que a primeira opção para obtenção de estimativas não viesadas para o RR, seria o modelo log-binomial, porém, este tende a apresentar falhas de convergência, em especial quando a probabilidade do desfecho está próxima de 1. Contudo, é possível aplicar o méto delta no modelo logístico para obter estimativas não viesadas para o RR, mesmo quando o modelo log-binomial falha em convergir.

Palavras chave: Regressão logística, Regressão log-binomial, Regressão Multinomial, Razão de prevalência, Razão de Incidência.

1 INTRODUÇÃO

De acordo com Wagner(2008) [2] , *Odds ratio* (OR) é a chance de desenvolver a doença nos indivíduos expostos, dividida pela chance de desenvolver a doença nos não-expostos. Já o Risco Relativo (RR) é a razão entre a incidência entre indivíduos expostos pela incidência entre os não-expostos.

Ambas as estatísticas podem ser estimadas por um modelo Binomial, no entanto, com diferentes funções de ligação, recebendo dois nomes diferentes: Modelo Logístico, quando a função de ligação é a logit e Modelo Log-Binomial, quando é usada a função logaritmo.

1.1 Modelo de Regressão Binomial

O modelo log-binomial, para k covariáveis é representado por $\pi = e^{\beta_0+\beta_1 X_1+\cdots+\beta_k X_k}$, aplicando a função de ligação *log*, a média deste modelo é linearizada.

$$log(\pi) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$$

Segundo Coutinho (2008)[1], o modelo log-binomial é um modelo linear generalizado onde a função de ligação é o logaritmo da proporção em estudo e a distribuição do erro é binomial. O RR estimado de uma dada covariável é e^{β}. Este modelo produz uma estimativa não viesada do RR, mas tem como desvantagem problemas de convergência.

Já o modelo Logístico é representado por $\pi = \frac{e^{\beta_0+\beta_1 x+\cdots+\beta_k X_k}}{1+e^{\beta_0+\beta_1 x+\cdots+\beta_k X_k}}$ e a função de ligação utilizada é a *logit*, definida por: $log\left[\frac{\pi}{1-\pi}\right]$. Então, a média da resposta é linearizada fazendo

$$log\left[\frac{\pi}{1-\pi}\right] = \beta_0 + \beta_1 x + \cdots + \beta_k X_k$$

Ainda, podemos estimar o RR através das estimativas de β_0 e β_1 de um modelo logístico, obtidas pelo método da máxima verossimilhança. Com base na propriedade de invariância dos estimadores de máxima verossimilhança, é possível obter o RR como uma função de β_0 e β_1, como segue:

$$RR = \frac{\pi_{x=1}}{\pi_{x=0}} = \frac{1/1+e^{-\beta_0-\beta_1}}{1/e^{-\beta_0}} = \frac{1+e^{-\beta_0}}{1+e^{-\beta_0-\beta_1}}$$

Dessa forma, podemos utilizar o método delta para estimar erro-padrão e o intervalo de confiança do RR. O método delta é capaz de aproximar a variância de uma função dos parâmetros, como a descrita acima. Este método consiste na expansão em série de Taylor, de segunda ordem, da função dos parâmetros. Assim, com base na normalidade aproximada dos estimadores de máxima verossimilhança e na variância aproximada, obtida pelo método delta, é possível estimar o seu intervalo de confiança.

2 Simulação

O estudo de simulação considerou um desfecho binário (Y), com um fator de exposição binário (X). O risco do grupo não exposto foi fixado em $exp(\beta_0)$, onde β_0 foi fixado em -1, enquanto que o risco relativo à exposição foi $exp(\beta_1)$, onde β_1

foi gerado aleatoriamente de uma distribuição normal com média 1 e desvio-padrão $0,2$. Para os casos com β_1 maior do que 1, o risco do grupo exposto foi limitado à 1. Metade da amostra foi exposta ($X = 1$) e a outra metade não exposta ($X = 0$), considerado uma amostra de tamanho 500. O conjunto de dados foi utilizado para estimar a OR através do ajuste de um modelo logístico e o RR através de um modelo log-binomial e de um modelo logístico, utilizando o método delta. Este procedimento foi repetido 10000 vezes. Com base nestas estimativas, foram calculados o viés médio e mediano e a probabilidade de cobertura dos intervalos de confiança de 95% do risco relativo. Toda a programação foi realizada no ambiente R (R Core Team, 2016)

3 Resultados

Para os conjuntos de dados simulados, o modelo log-binomial convergiu apenas em 57,92% das vezes, enquanto que o modelo logístico convergiu em todos os casos (Tabela 1). Considerando apenas os conjuntos de dados para o qual o modelo log-binomial convergiu, podemos observar que os modelos log-binomial e logístico (RR) apresentaram desempenho equivalente, com viés irrelevante e percentual de cobertura do intervalo de confiança próximos de 95% (Tabela 1). Em contrapartida, o modelo logístico (OR) obteve estimativas positivamente viesadas, com baixo percentual de cobertura dos IC (Tabela 1). Já para os conjuntos de dados para os quais o modelo log-binomial não convergiu, destaque-se os desempenhos desastroso do modelo logístico (OR) e razoável do modelo logístico (RR) (Tabela 1). O modelo logístico (RR) foi capaz de gerar estimativas pouco viesadas e com percentual de cobertura dos IC de 85,58%, para os conjuntos de dados onde o modelo log-binomial falhou em convergir (Tabela 1).

Tabela 1: Estimativas obtidas pelos três modelos.

Conjunto de dados	Modelos	Viés Médio	Viés Mediano	IC Cobertura	Convergência(%)
Completo	Log-binomial	0.0093	-0.0052	54.92 %	57.92 %
	Logístico (OR)	88233040.0000	5.7607	21.19 %	100.00 %
	Logístico (RR)	-0.0320	-0.0250	91.05 %	100.00 %
Com Convergência	Log-binomial	0.0093	-0.0052	94.82 %	-
	Logístico (OR)	3.1908	2.7403	8.86 %	-
	Logístico (RR)	0.0093	-0.0052	95.03 %	-
Sem Convergência	Log-binomial	-	-	-	-
	Logístico (OR)	209679300.0000	40.0558	38.17 %	-
	Logístico (RR)	-0.0887	-0.0619	85.58 %	-

Na Figura 1 podemos observar que os três modelos foram capazes de obter estimativas acuradas quando o RR verdadeiro foi próximo de 1. Entretanto, conforme o RR verdadeiro aumenta, a regressão logística (OR) começa a superestimar o RR. Para RR verdadeiro maiores do que $2,5$, o risco do grupo exposto se aproxima de 1, e nestes casos, o modelo log-binomial falha em convergir (risco dos expostos maiores do que 0,9).

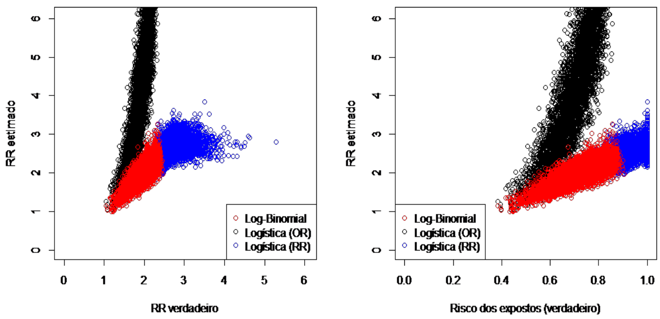

Figura 1: Estimativas obtidas pelos três modelos

4 Conclusão

Com base nos resultados apresentados neste estudo, podemos concluir que o modelo logístico estima a estatística OR que superestima o RR. A primeira opção para obtenção de estimativas não viesadas para o RR, seria a regressão log-binomial, porém, esta tende a apresentar falhas na convergência, em especial quando a probabilidade do desfecho está próxima de 1. Contudo, é possível utilizar a regressão logística e o método delta para obter estimativas não viesadas para o RR, mesmo quando a regressão log-binomial falha em convergir.

Ressaltamos que este é um estudo preliminar, pois outros cenários avaliando outros tamanhos de amostra e diferentes tamanhos de efeito serão avaliados.

Referências

[1] L. M. Coutinho, M. Scazufca, P. R. Menezes, et al. Métodos para estimar razão de prevalência em estudos de corte transversal. *Rev Saude Publica*, 42(6):992–8, 2008.

[2] M. B. Wagner and S. M. Callegari-Jacques. Medidas de associação em estudos epidemiológicos: risco relativo e odds ratio. *Jornal de pediatria. Rio de Janeiro. Vol. 74, no. 3 (1998), p. 247-251.*, 1998.