

USO DE UM MODELO BAYESIANO DE POISSON COM EXCESSO DE ZEROS NA ANÁLISE DE DADOS DE LESÕES MIOCÁRDICAS EM RECÉM-NASCIDOS COM CARDIOPATIAS CONGÊNITAS COMPLEXAS

Suleimy Cristina MAZIN¹
Marcela Silva de OLIVEIRA²
Edson Zangiacomi MARTINEZ³
Jorge Alberto ACHCAR³
Simone Gusmão RAMOS²

- **RESUMO:** Dados de contagens com muitos zeros são comuns em estudos da área médica. Neste artigo, apresentamos um modelo de Poisson com excesso de zeros voltado à análise dos dados de um estudo conduzido no Departamento de Patologia da Faculdade de Medicina de Ribeirão Preto -USP, onde há um grande número de zeros. Trata-se de um estudo sobre lesões miocárdicas em recém-nascidos com cardiopatias congênicas complexas submetidos a cirurgia. Com esta finalidade, apresentamos uma adaptação do modelo bayesiano baseado na distribuição de Poisson com excesso de zeros apresentado por Angers e Biswas (2003). Dado que o delineamento do estudo médico previa que um total de dez observações de contagem fossem obtidas a cada indivíduo, o que diferencia o modelo bayesiano aqui apresentado daquele proposto por Angers e Biswas (2003) é a inclusão de múltiplas observações por indivíduo e a presença de efeitos aleatórios que identificam uma dependência entre as observações de cada recém-nascido. O modelo proposto mostrou-se satisfatório, permitindo evidenciar, por exemplo, que recém-nascidos com cardiopatias congênicas não operados tendem a apresentar um número maior de eventos de edema intracelular. Para as estimativas dos parâmetros de interesse, utilizamos métodos Bayesianos baseados em algoritmo Monte Carlo em Cadeia de Markov (MCMC), com o auxílio do programa computacional Winbugs.

¹Centro de Métodos Quantitativos (CEMEQ), Departamento de Medicina Social, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, CEP: 14049-900, Ribeirão Preto, SP, Brasil. E-mail: *sumazin@pop.com.br*

²Departamento de Patologia, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, CEP: 14049-900, Ribeirão Preto, SP, Brasil. E-mail: *marcelaoliveira9@hotmail.com / sgramos@fmrp.usp.br*

³Departamento de Medicina Social, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, CEP: 14049-900, Ribeirão Preto, SP, Brasil. E-mail: *edson@fmrp.usp.br / achcar@fmrp.usp.br*

- PALAVRAS-CHAVE: Modelo de Poisson; excesso de zeros; métodos bayesianos; estatística médica.

1 Introdução

Dados de contagens que contêm muitos zeros são comuns em estudos da área médica. Por exemplo, Karazsia e van Dulmen (2008) apresentaram uma análise de dados longitudinais de contagens de atendimentos a crianças vítimas de lesões, onde muitos zeros estão presentes. Neste estudo, os autores enfatizaram a utilidade de um modelo baseado na distribuição de Poisson que considera o excesso de zeros na análise de seus dados. Comulada *et al.* (2007) utilizaram modelos com excesso de zeros para detectar a redução do uso abusivo de drogas por jovens portadores do vírus HIV e realçaram a importância desta técnica na análise de dados de estudos de comportamento. Ramis Prieto *et al.* (2007) utilizaram um modelo bayesiano com excesso de zeros para descrever o padrão geográfico da mortalidade por tumores hematológicos na Espanha.

Para a análise de dados de contagem com muitos zeros em sua generalidade, seja Y uma variável aleatória com função de probabilidades

$$\begin{aligned} P(Y = 0) &= p + (1 - p) \frac{b(0)}{c(\lambda)}, \\ P(Y = y) &= (1 - p) \frac{b(y)\lambda^y}{c(\lambda)}, \quad y = 1, 2, \dots, \end{aligned} \quad (1)$$

onde $b(y)$ é uma função real positiva de y , $c(\lambda) = \sum_{y=0}^{\infty} b(y)\lambda^y$, $0 \leq p < 1$ e $\lambda > 0$. Sendo Y_1, Y_2, \dots, Y_n variáveis aleatórias independentes e identicamente distribuídas com distribuição de probabilidade (1), a função de verossimilhança de $\theta = (p, \lambda)$ é dada por

$$L(\theta) \propto [pc(\lambda) + (1 - p)b(0)]^k (1 - p)^{n-k} \frac{\lambda^{s(y)}}{[c(\lambda)]^n},$$

onde $s(y) = \sum_{i=1}^n y_i$ e $k = \#\{i : y_i = 0, i = 1, \dots, n\}$.

O modelo de Poisson com excesso de zeros é caracterizado por $c(\lambda) = e^\lambda$, $\lambda > 0$ e $b(y) = 1/y!$, o que resulta em

$$\begin{aligned} P(Y = 0) &= p + (1 - p)e^{-\lambda}, \\ P(Y = y) &= (1 - p) \frac{e^{-\lambda}\lambda^y}{y!}, \quad y = 1, 2, \dots \end{aligned}$$

Neste caso, a média e a variância de Y são dadas respectivamente por

$$\begin{aligned} E(Y) &= (1 - p)\lambda, \text{ e} \\ Var(Y) &= (1 - p)\lambda(1 + p\lambda) = E(Y)(1 + p\lambda) \end{aligned}$$

(ver, por exemplo, Lambert, 1992). Notar que o modelo contempla a superdispersão gerada pelo excesso de zeros, dado que $Var(Y) > E(Y)$.

Estimadores de máxima verossimilhança e testes de razão de verossimilhanças para o modelo de regressão com excesso de zeros foram propostos por Hall (2000). Angers e Biswas (2003) introduziram um método bayesiano para a estimação dos parâmetros do modelo de Poisson com excesso de zeros, enquanto Ghosh *et al.* (2006) apresentaram um modelo bayesiano mais flexível, que permite a modelagem de dados com muitos zeros com base em outras distribuições discretas. Em um contexto bayesiano, outras importantes contribuições incluem os artigos publicados por Rodrigues (2003, 2006).

No presente artigo apresentamos uma adaptação do modelo bayesiano apresentado por Angers e Biswas (2003), objetivando a análise de dados de lesões miocárdicas em recém-nascidos com cardiopatias congênicas complexas, operados no Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo (USP). O delineamento deste estudo previa que um total de dez observações de contagens seriam obtidas a cada indivíduo. Assim, o que torna diferente o modelo bayesiano aqui apresentado daquele proposto por Angers e Biswas (2003) é a inclusão de múltiplas observações por indivíduo e a presença de efeitos aleatórios que identificam uma dependência entre as observações de cada recém-nascido.

2 Apresentação dos dados

O banco de dados utilizado no presente estudo foi obtido a partir da contagem de cortes de miocárdio humano, emblocados em parafina, processados para análise histológica e corados com Azan Heidenhain, provenientes de necropsias de recém-nascidos (RN). Esse material foi dividido em quatro grupos:

- *G1*: RN submetidos à cirurgia cardíaca para correção de cardiopatias congênicas com circulação extracorpórea (CEC),
- *G2*: RN submetidos a cirurgia cardíaca sem CEC,
- *G3*: RN com cardiopatias congênicas não operados e
- *G4*: grupo de controles, RN sem malformação cardíaca.

Notar que esta alocação dos indivíduos em grupos não é aleatória, ou seja, é dependente das condições clínicas do RN. A quantificação de células lesadas foi realizada usando análise estereológica em microscopia de luz (Mandarim de Lacerda, 2003). Para aplicação desse método os cortes em estudo devem ser isotrópicos, ou seja, os diversos elementos que o constituem devem estar distribuídos ao acaso e homoganeamente no espaço. Atualmente são utilizadas “grades” estereológicas, ou sistemas teste, que facilitam o emprego dessa metodologia. As grades são figuras, geralmente retangulares, de área conhecida, contendo em seu interior um número predeterminado de pontos e segmentos de retas que podem ser sobrepostas às estruturas analisadas para a contagem de pontos e interseções das linhas que caem sobre o elemento em estudo.

Tabela 1 - Dados do estudo

i	grupo	sexo	Número de eventos de edema intracelular									
1	G1	1	0	0	0	0	0	0	0	0	0	0
2	G1	2	0	0	0	0	0	0	0	0	0	0
3	G1	1	0	0	0	0	0	0	0	0	0	0
4	G1	2	0	0	0	0	0	0	0	0	0	0
5	G1	1	0	0	0	0	0	0	0	0	0	0
6	G1	2	13	0	0	0	0	0	0	0	0	0
7	G1	2	0	0	0	9	4	21	51	9	2	2
8	G1	1	14	2	3	26	15	1	0	0	0	0
9	G1	1	0	0	0	0	0	0	0	0	0	0
10	G1	1	0	0	0	0	0	0	0	0	0	13
11	G1	1	8	0	0	0	0	36	0	0	0	0
12	G1	1	0	0	0	0	0	0	0	0	0	0
13	G1	1	0	0	0	0	0	35	31	1	0	0
14	G1	1	0	0	0	0	0	0	0	31	21	19
15	G1	1	0	0	0	0	0	0	0	0	0	0
16	G1	2	0	0	0	0	0	0	0	0	0	0
17	G2	2	0	0	0	0	0	0	6	7	15	6
18	G2	1	0	0	0	0	0	0	0	0	0	0
19	G2	2	0	0	0	0	0	0	0	0	0	0
20	G2	1	2	2	11	3	89	27	0	0	0	9
21	G3	1	0	1	97	47	0	26	9	0	0	62
22	G3	1	39	45	14	20	13	0	0	0	0	0
23	G3	2	0	0	0	0	0	0	0	0	0	0
24	G3	2	0	0	0	0	0	0	0	0	0	0
25	G3	1	0	0	0	0	0	0	0	0	0	0
26	G3	2	26	30	25	3	0	18	5	3	11	2
27	G3	1	32	42	40	17	7	5	33	5	1	3
28	G3	2	0	0	0	0	0	0	0	0	0	0
29	G3	1	78	81	70	42	51	0	0	0	0	0
30	G4	2	0	0	0	0	0	0	0	0	0	0
31	G4	1	0	0	0	0	0	0	0	0	0	0
32	G4	1	0	0	0	0	0	0	0	0	0	0
33	G4	1	0	0	0	0	0	0	0	0	0	0
34	G4	2	0	0	0	0	0	0	0	0	0	0

Sexo: 1 = feminino, 2 = masculino.

Foi utilizado um microscópio de luz acoplado a uma lente ocular com graticulo indexado de 100 pontos. Cada campo coberto pelo graticulo delimita uma área equivalente a 0,0625 mm² no aumento final de 400x. Foram contadas as áreas correspondentes a edema intracelular, um dos primeiros eventos da lesão isquêmica provocado pela depleção de ATP celular (Contran *et al*, 1999).

A Tabela 1 mostra o número de eventos de edema intracelular em 10 campos aleatórios não coincidentes, para $n = 34$ recém-nascidos. Observa-se um grande número de contagens iguais a zero, sendo que todas as contagens para os RN sem malformação cardíaca (grupo G4) são nulas. O objetivo deste estudo clínico é comparar os quatro grupos, com base na hipótese que RN operados para correção de cardiopatias congênicas com CEC e sem CEC apresentariam mais lesões miocárdicas do que o grupo não operado. Diversos tipos de lesões foram estudados (por exemplo, necrose com bandas de contração e necrose de coagulação), mas, para os propósitos do presente artigo, usamos os dados de edema intracelular, que é uma lesão teoricamente reversível, presente em pacientes com insuficiência cardíaca.

3 Modelo Bayesiano

Genericamente, sejam \mathbf{x}_i e \mathbf{w}_i respectivamente observações de \mathbf{X}_i e \mathbf{W}_i , vetores (conhecidos) de covariáveis associadas ao i -ésimo indivíduo, tal que $\mathbf{X}_i = (1, X_{1i}, X_{2i}, \dots, X_{mi})$ e $\mathbf{W}_i = (1, W_{1i}, W_{2i}, \dots, W_{gi})$ (ver, por exemplo, Lambert, 1992 ou Lam *et al.*, 2006). Sendo $\boldsymbol{\lambda}(\mathbf{w}_i) = (\lambda_1(\mathbf{w}_i), \lambda_2(\mathbf{w}_i), \dots, \lambda_n(\mathbf{w}_i))'$ e $\mathbf{p}(\mathbf{x}_i) = (p_1(\mathbf{x}_i), p_2(\mathbf{x}_i), \dots, p_n(\mathbf{x}_i))'$, o modelo de Poisson para os dados da Tabela 1 é descrito como

$$\begin{aligned} P(Y_{ij} = 0 | \mathbf{x}_i, \mathbf{w}_i) &= p_i(\mathbf{x}_i) + [1 - p_i(\mathbf{x}_i)] e^{-\lambda_i(\mathbf{w}_i)}, \\ P(Y_{ij} = y_{ij} | \mathbf{x}_i, \mathbf{w}_i) &= [1 - p_i(\mathbf{x}_i)] \frac{e^{-\lambda_i(\mathbf{w}_i)} [\lambda_i(\mathbf{w}_i)]^{y_{ij}}}{y_{ij}!}, \quad y_{ij} = 1, 2, \dots, \end{aligned}$$

$i = 1, 2, \dots, n$, $j = 1, 2, \dots, 10$. Sendo $0 \leq p_i(\mathbf{x}_i) < 1$, $i = 1, \dots, n$, podemos considerar este modelo como uma mistura entre uma distribuição de Bernoulli, sendo $\mathbf{p}(\mathbf{x}_i)$ as probabilidades de sucesso, e uma distribuição de Poisson com parâmetro $\lambda_i(\mathbf{w}_i) > 0$ para cada indivíduo. Assim, para um dado $\lambda_i(\mathbf{w}_i)$, interpretamos $p_i(\mathbf{x}_i)$ como um parâmetro associado à proporção do componente desta mistura de duas distribuições relativas ao excesso de zeros. Temos assim

$$\begin{aligned} \ln \left[\frac{p_i(\mathbf{x}_i)}{1 - p_i(\mathbf{x}_i)} \right] &= \mathbf{x}_i \boldsymbol{\alpha} + \omega_i = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \dots + \alpha_m x_{mi} + \omega_i \quad \text{e} \\ \ln [\lambda_i(\mathbf{w}_i)] &= \mathbf{w}_i \boldsymbol{\beta} + \omega_i = \beta_0 + \beta_1 w_{1i} + \beta_2 w_{2i} + \dots + \beta_g w_{gi} + \omega_i, \end{aligned}$$

onde $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_m)'$ e $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_g)'$ são vetores de efeitos fixos e ω_i são efeitos aleatórios independentes com distribuição normal, tal que $\omega_i \sim N(0, \sigma_\omega^2)$. A função de verossimilhança L no ponto $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, dadas as observações y_{ij} ,

as variáveis latentes ω_i não observadas e as covariáveis \mathbf{X}_i e \mathbf{W}_i é dada por

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{\{i,j:y_{ij}=0\}} \left\{ \frac{\exp(\mathbf{x}_i \boldsymbol{\alpha} + \omega_i) + \exp[-(\mathbf{w}_i \boldsymbol{\beta} + \omega_i)]}{1 + \exp(\mathbf{x}_i \boldsymbol{\alpha} + \omega_i)} \right\} \times \\ \times \prod_{\{i,j:y_{ij}>0\}} \left[\frac{\exp[-(\mathbf{w}_i \boldsymbol{\beta} + \omega_i)] [\mathbf{w}_i \boldsymbol{\beta} + \omega_i]^{y_{ij}}}{1 + \exp(\mathbf{x}_i \boldsymbol{\alpha} + \omega_i) y_{ij}!} \right]. \quad (2)$$

Por simplicidade, consideraremos $\mathbf{W}_i = \mathbf{X}_i$, o que implica $\ln[\lambda_i(\mathbf{x}_i)] = \mathbf{X}_i \boldsymbol{\beta} + \omega_i$. No banco de dados apresentado estão presente duas covariáveis, sexo e grupo (Tabela 1), que introduziremos no modelo, por simplicidade de notação, como

$$\ln \left[\frac{p_i(\mathbf{x}_i)}{1 - p_i(\mathbf{x}_i)} \right] = \alpha_0 + \alpha_{1,j_i} + \alpha_{2,l_i} + \omega_i \quad e \\ \ln[\lambda_i(\mathbf{x}_i)] = \beta_0 + \beta_{1,j_i} + \beta_{2,l_i} + \omega_i,$$

$j_i = 1$ se o i -ésimo indivíduo for do sexo feminino e $j_i = 2$ for do sexo masculino, e l_i é o grupo ao qual o i -ésimo indivíduo pertence, $l_i = 1, 2, 3, 4$ (ou seja, respectivamente, grupos $G1, G2, G3$ e $G4$). Definimos $\alpha_{1,1} = \alpha_{2,1} = \beta_{1,1} = \beta_{2,1} = 0$.

Considerando que $\alpha_0, \alpha_{1,2}, \alpha_{2,2}, \alpha_{2,3}, \alpha_{2,4}, \beta_0, \beta_{1,2}, \beta_{2,2}, \beta_{2,3}$ e $\beta_{2,4}$ assumem valores reais e $\sigma_\omega^2 > 0$, para a análise bayesiana dos dados, especificamos as seguintes distribuições a priori:

$$\alpha_0 \sim N(a_{\alpha_0}, b_{\alpha_0}^2), \quad a_{\alpha_0} \text{ e } b_{\alpha_0} \text{ conhecidos,} \\ \alpha_{1,2} \sim N(a_{\alpha_{1,2}}, b_{\alpha_{1,2}}^2), \quad a_{\alpha_{1,2}} \text{ e } b_{\alpha_{1,2}} \text{ conhecidos,} \\ \alpha_{2,l_i} \sim N(a_{\alpha_{2,l_i}}, b_{\alpha_{2,l_i}}^2), \quad a_{\alpha_{2,l_i}} \text{ e } b_{\alpha_{2,l_i}} \text{ conhecidos, } l_i = 2, 3, 4, \\ \beta_0 \sim N(a_{\beta_0}, b_{\beta_0}^2), \quad a_{\beta_0} \text{ e } b_{\beta_0} \text{ conhecidos,} \\ \beta_{1,2} \sim N(a_{\beta_{1,2}}, b_{\beta_{1,2}}^2), \quad a_{\beta_{1,2}} \text{ e } b_{\beta_{1,2}} \text{ conhecidos,} \\ \beta_{2,l_i} \sim N(a_{\beta_{2,l_i}}, b_{\beta_{2,l_i}}^2), \quad a_{\beta_{2,l_i}} \text{ e } b_{\beta_{2,l_i}} \text{ conhecidos, } l_i = 2, 3, 4, \\ \sigma_\omega^2 \sim GI(c, d), \quad c \text{ e } d \text{ conhecidos,}$$

sendo que $N(a, b^2)$ genericamente denota uma distribuição normal com média 0 e desvio padrão b e $GI(c, d)$ denota uma distribuição gama inversa com distribuição de probabilidade

$$\pi(\sigma_\omega^2) \propto (\sigma_\omega^2)^{-(c+1)} \exp\left(-\frac{d}{\sigma_\omega^2}\right), \quad c, d, \sigma_\omega^2 > 0.$$

Assim, sob o enfoque bayesiano os efeitos aleatórios w_i são considerados como parâmetros do modelo. Considerando independência a priori entre os parâmetros,

a distribuição conjunta a priori é

$$\begin{aligned} \pi(\boldsymbol{\theta}_1) \propto & \pi(\sigma_\omega^2) \times \prod_{i=1}^n (\sigma_\omega^2)^{-\frac{1}{2}} \exp\left(-\frac{\omega_i^2}{2\sigma_\omega^2}\right) \\ & \times \exp\left[-\frac{1}{2}\left(\frac{(\alpha_0 - a_{\alpha_0})^2}{b_{\alpha_0}^2} + \frac{(\alpha_{1,2} - a_{\alpha_{1,2}})^2}{b_{\alpha_{1,2}}^2} + \right.\right. \\ & \left. + \sum_{l=2}^4 \frac{(\alpha_{2,l_i} - a_{\alpha_{l_i}})^2}{b_{\alpha_{l_i}}^2} + \frac{(\beta_0 - a_{\beta_0})^2}{b_{\beta_0}^2} + \right. \\ & \left. \left. + \frac{(\beta_{1,2} - a_{\beta_{1,2}})^2}{b_{\beta_{1,2}}^2} + \sum_{l=2}^4 \frac{(\beta_{2,l_i} - a_{\beta_{l_i}})^2}{b_{\beta_{l_i}}^2}\right)\right], \end{aligned} \quad (3)$$

onde $\boldsymbol{\theta}_1 = (\alpha_0, \alpha_{1,2}, \alpha_{2,2}, \alpha_{2,3}, \alpha_{2,4}, \beta_0, \beta_{1,2}, \beta_{2,2}, \beta_{2,3}, \beta_{2,4}, \sigma_\omega^2)$. A distribuição a posteriori conjunta para os parâmetros é obtida combinando a distribuição a priori (3) com a função de verossimilhança (2) a partir da fórmula de Bayes.

4 Resultados

Para a análise bayesiana dos dados da Tabela 1 foi considerado $a_{\alpha_0} = a_{\alpha_{1,2}} = a_{\alpha_{2,2}} = a_{\alpha_{2,3}} = a_{\alpha_{2,4}} = a_{\beta_0} = a_{\beta_{1,2}} = a_{\beta_{2,2}} = a_{\beta_{2,3}} = a_{\beta_{2,4}} = 0$ e $b_{\alpha_0}^2 = b_{\alpha_{1,2}}^2 = b_{\alpha_{2,2}}^2 = b_{\alpha_{2,3}}^2 = b_{\alpha_{2,4}}^2 = b_{\beta_0}^2 = b_{\beta_{1,2}}^2 = b_{\beta_{2,2}}^2 = b_{\beta_{2,3}}^2 = b_{\beta_{2,4}}^2 = 10$. E ainda, consideramos $c = d = 1$. Esta escolha dos hiperparâmetros foi motivada pela obtenção de distribuições a priori aproximadamente não informativas.

Para obtenção das estimativas dos parâmetros do modelo baseado no método bayesiano via amostrador de Gibbs, foi utilizado o programa *Winbugs* (ver código no apêndice). Os parâmetros foram estimados via algoritmos Gibbs-Sampling e Metropolis-Hastings. Foram geradas 505.000 amostras, das quais 5.000 foram descartadas ("burn in samples") com a finalidade de eliminar o efeito dos valores iniciais usados no algoritmo de simulação. Além disso, foram consideradas as iterações 50^a; 100^a; 150^a;..., resultando em uma amostra final de 10.000 cadeias simuladas para cada parâmetro. A convergência do algoritmo foi verificada através de gráficos temporais das amostras geradas e utilizando técnicas usuais existentes na literatura (Gelman e Rubin, 1992). A Tabela 2 mostra os resumos a posteriori para os efeitos fixos do modelo, obtidos do programa *Winbugs* (medianas das distribuições a posteriori como estimativas dos parâmetros e percentis 2,5% e 97,5% representando intervalos de credibilidade 95%). Dentre os parâmetros $\alpha_{2,2}$, $\alpha_{2,3}$ e $\alpha_{2,4}$, relativos ao efeito dos grupos de RN sobre as proporções p_i , nota-se que $\alpha_{2,4}$ é estimado em $-2,9$, a estimativa de maior magnitude, o que sugere um efeito maior do grupo *G4* sobre o excesso de zeros quando comparado ao grupo *G1* (embora o intervalo de credibilidade 95% seja bastante largo, e cubra o valor zero).

Dentre os parâmetros $\beta_{2,2}$, $\beta_{2,3}$ e $\beta_{2,4}$, destaca-se a magnitude da estimativa de $\beta_{2,3}$, 1,507, com intervalo de credibilidade 95% dado por (0,422; 2,886), o que sugere que o número médio de eventos de edema intracelular para os RN do grupo

Tabela 2 - Resumos a posteriori para os efeitos fixos

Parâmetro	estimativa	IC 95%	
α_0	-0,1043	(-1,382	1,177)
$\alpha_{1,2}$	-0,9585	(-2,877	0,838)
$\alpha_{2,2}$	0,4990	(-2,095	3,131)
$\alpha_{2,3}$	0,6889	(-1,247	2,647)
$\alpha_{2,4}$	-2,9000	(-7,603	3,094)
β_0	1,2020	(0,118	1,878)
$\beta_{1,2}$	-0,5902	(-1,698	0,470)
$\beta_{2,2}$	0,6508	(-0,998	2,070)
$\beta_{2,3}$	1,5070	(0,422	2,886)
$\beta_{2,4}$	-0,6388	(-7,664	5,979)

IC: intervalo de credibilidade

Tabela 3 - Resumos a posteriori para as médias do número de eventos de edema intracelular em cada grupo

Grupo	estimativa	IC 95%	
$G1$	2,292	(2,062	2,531)
$G2$	4,417	(3,791	5,103)
$G3$	11,15	(10,47	11,86)
$G4$	0	(0	0,02)

IC: intervalo de credibilidade.

$G3$ é expressivamente maior, quando comparado ao grupo $G1$. Inferências sobre as médias do número de eventos de edema intracelular com base nos parâmetros em α e β consideram a presença dos efeitos aleatórios ω_i , ou seja,

$$E(Y_i|\mathbf{x}_i) = E[E(Y_i|\mathbf{x}_i, \omega_i)].$$

A partir das amostras geradas pelo algoritmo computacional para os parâmetros em α e β e para os efeitos aleatórios ω_i , simulamos as distribuições a posteriori para as médias do número de eventos de edema intracelular em cada um dos quatro grupos do estudo, cujos resumos encontram-se na Tabela 3 (medianas e percentis 2,5% e 97,5%). Evidentemente, o grupo que apresenta o maior número médio de eventos de edema é o $G3$ (RN com cardiopatia e sem cirurgia), seguido do $G2$ e do $G1$, sendo que no grupo $G4$ não foram observados eventos de edema. Isso pode refletir que os RN do grupo $G3$ tinham patologias mais graves, repercutindo seriamente no miocárdio e levando à insuficiência cardíaca precoce, não havendo tempo hábil de se submeterem à cirurgia (é um grupo com idade mais precoce quanto ao óbito, cerca de 10 dias).

A Tabela 4 exibe resumos a posteriori para as distribuições simuladas das diferenças entre as médias do número de eventos de edema intracelular segundo grupos. Todos os intervalos de credibilidade 95% excluem o valor zero, evidenciando que os diferentes grupos apresentam médias para os números de eventos de edema

Tabela 4 - Resumos a posteriori para as comparações entre as médias do número de eventos de edema intracelular segundo grupos

Comparação	estimativa	IC 95%	
$G1 - G2$	-2,125	(-2,856	-1,462)
$G1 - G3$	-8,848	(-9,593	-8,134)
$G1 - G4$	2,29	(2,059	2,530)
$G2 - G3$	-6,717	(-7,685	-5,766)
$G2 - G4$	4,416	(3,79	5,101)
$G3 - G4$	11,14	(10,47	11,85)

IC: intervalo de credibilidade.

intracelular diferentes, sendo que os RN com cardiopatias congênitas não operados são os que possuem a maior média em relação aos demais grupos.

Um outro resultado do modelo é que RN de sexo feminino têm, em média, maior número de eventos de edema intracelular, 6,1 (IC95%: 5,7 a 6,4), em relação aos RN de sexo masculino, que têm, em média, 2,1 eventos de edema intracelular (IC95%: 1,8 a 2,3).

5 Comentários finais

Dados de contagem são comuns em estudos da área médica, e não são raras as situações onde encontramos um grande número de zeros. O uso de métodos bayesianos é uma alternativa promissora para análise de dados desse tipo, conforme se observa nos resultados obtidos. O uso de métodos de simulação de amostras para a distribuição a posteriori de interesse via amostrador de Gibbs não requer grande custo computacional e pode facilmente ser implementado usando programas computacionais disponíveis. Uma limitação que encontramos no uso do programa *Winbugs* é que ele não permite a determinação do valor do DIC (*Deviance Information Criterion*, Spiegelhalter *et al.*, 2002) em análises baseadas em modelos com misturas de distribuições, como o aqui apresentado (ver <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml>). Isto pode ser um obstáculo em situações onde se pretende comparar diferentes modelos, o que pode ser contornado com alternativas como o fator de Bayes.

Agradecimentos

A pesquisa de S.C.M. recebe apoio da Fundação de Amparo ao Ensino, Pesquisa e Assistência (FAEPA) do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto (HC-FMRP), Universidade de São Paulo (USP). Somos muito gratos aos dois pareceristas que, de forma anônima e profissional, trouxeram grandes contribuições a este artigo.

MAZIN, S. C.; OLIVEIRA, M. S.; MARTINEZ, E. Z.; ACHCAR, J. A.; RAMOS, S. G. Data analysis of a study of myocardial injuries in newborns with congenital cardiopathies using a Bayesian Poisson model with an excess of zero-valued observations *Rev. Bras. Biom.*, São Paulo, v.26, n.4, p.113-123, 2008.

- **ABSTRACT:** *Count data with many zeros are common in studies in the medical field. In this article, we present a Poisson model with excess of zeros for the analysis of data from a study conducted in the Department of Pathology, Faculty of Medicine of Ribeirão Preto (FMRP-USP), where there is a huge excess of counts equal to zero. This is a study of myocardial injury in newborns with congenital heart diseases undergoing complex surgery. To this end, we present an adaptation of the Bayesian model based on Poisson distribution with excess of zeros by Angers and Biswas (2003). Given that the medical study design predicted that ten data counts were obtained for each individual, the difference between the Bayesian model presented here and that proposed by Angers and Biswas (2003) is the inclusion of multiple observations by individual and the presence of random effects that identify the dependency relationships between the observations of each newborn. The model proposed here showed to be satisfactory for the analysis of these data, allowing evidence, for example, that non-operated newborns with congenital heart diseases tend to have a greater number of events of intracellular edema. We used Bayesian methods based on Monte Carlo Markov Chain (MCMC) algorithm with the aid of the program Winbugs in order to obtain estimates of the parameters of interest.*

- **KEYWORDS:** *Poisson distribution; excess zeros; Bayesian methods; medical statistics.*

Referências

- ANGERS, J.F.; BISWAS, A. A Bayesian analysis of zero-inflated generalized Poisson model. *Comput. Stat. Data Anal.*, Amsterdam, v.42, n.1-2, p.37-46, 2003.
- COMULADA, W. S.; WEISS, R. E.; CUMBERLAND, W.; ROTHERAM-BORUS, M. J. Reductions in drug use among young people living with HIV. *Am. J. Drug. Alcohol. Abuse*, New York, v.33, n.3, p.493-501, 2007.
- CONTRAN, R. S.; KUMAR, V.; ROBBINS, C.T. *Pathologic basis of the disease*. 6^a. ed. Edition. Rio de Janeiro: Guanabara Koogan, 1999.
- GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. *Stat. Sci.*, Hyward, v.7, n.4, p.457-472.
- GHOSH, S.K.; MUKHOPADHYAY, P.; LU, J.C. Bayesian analysis of zero-inflated regression models. *J. Stat. Plann. Inference*, Amsterdam, v.136, n.4, p.1360-1375, 2006.
- HALL, D.B. Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics*, Washington, v.56, n.4, p.1030-1039, 2000.

- KARAZSIA, B.T.; VAN DULMEN, M.H. Regression models for count data: illustrations using longitudinal predictors of childhood injury. *J. Pediatr. Psychol.*, Washington, v.33, n.10, p.1076-1084, 2008.
- LAM, K. F.; XUE, H.; CHEUNG, Y.B. Semiparametric analysis of zero-inflated count data. *Biometrics*, Washington, v.62, n.4, p.996-1003, 2006.
- LAMBERT, D. Zero-inflated Poisson regression with application to defects in manufacturing. *Technometrics*, Washington, v.34, p.1-4, 1992.
- LI, C.S.; LU, J.C.; PARK, J.; KIM, K.M.; BRINKLEY, P.A.; PETERSON, J. A multivariate zero-inflated Poisson distribution and its inference. *Technometrics*, Washington, v.41, n.1, p.29-38, 1999.
- MANDARIM DE LACERDA, C.A. Fisiologia e fisiopatologia. Estereologia do rim: determinação do Vv, Nv e volume médio do glomérulo. *J. Bras. Nefrol.*, v.22, n.2, p.103-109, 2000.
- RAMIS PRIETO, R.; GARCÍA-PÉREZ, J.; POLLÁN, M.; ARAGONÉS, N.; PÉREZ-GÓMEZ, B.; LÓPEZ-ABENTE, G. Modelling of municipal mortality due to haematological neoplasias in Spain. *J. Epidemiol. Commun. Health.*, London, v.61, n.2, p.165-171, 2007.
- RODRIGUES, J. Bayesian analysis of zero-inflated distributions. *Commun. Stat. Part A, Theor. Methods*, New York, v.32, n.2, p.281-289, 2003.
- RODRIGUES, J. Full Bayesian significance test for zero-inflated distributions. *Commun. Stat. Part A, Theor. Methods*, New York, v.35, n.2, 299-307, 2006.
- SPIEGELHALTER, D. J.; BEST, N. G.; CARLIN, B. P.; VAN DER LINDE A. Bayesian measures of model complexity and fit (with discussion). *R. Soc. Stat. J. Ser. B, Stat. Methodol.*, Oxford, v.64, p.583-640, 2002.

Recebido em 15.09.2008.

Aprovado após revisão em 04.12.2008.

Apêndice: programa Winbugs

```
model {
# Verossimilhança:
for(i in 1:n) {
mu[i] <- u[i]*lambda[i]
u[i] ~ dbern(p[i])
logit(p[i]) <- alpha0+alpha1[sexo[i]]+alpha2[grupo[i]]+w[i]
log(lambda[i]) <- beta0+beta1[sexo[i]]+beta2[grupo[i]]+w[i]
w[i] ~ dnorm(0,sigmaw)
for(j in 1:R) {
y[i,j] ~ dpois(mu[i])
}}
# Priors:
sigmaw ~ dgamma(1,1)
alpha0 ~ dnorm(0,.1)
beta0 ~ dnorm(0,.1)
alpha1[1] <- 0
beta1[1] <- 0
alpha2[1] <- 0
beta2[1] <- 0
alpha1[2] ~ dnorm(0,.1)
beta1[2] ~ dnorm(0, .1)
for(j in 2:4) {
alpha2[j] ~ dnorm(0,.1)
beta2[j] ~ dnorm(0, .1)
}}
# Dados
list(n=34, R=10,
sexo=c(1,2,1,2,1,2,2,1,1,1,1,1,1,1,1,2,2,1,2,1,1,1,2,2,1,2,1,2,1,2,1,1,1,2),
grupo=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,2,2,2,3,3,3,3,3,3,3,3,4,4,4,4,4),
y=structure(.Data =
c(0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,
13,0,0,0,0,0,0,0,0,0,0,
0,0,0,9,4,21,51,9,2,2,
14,2,3,26,15,1,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,13,
8,0,0,0,0,36,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,
```

0,0,0,0,0,35,31,1,0,0,
0,0,0,0,0,0,0,31,21,19,
0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,6,7,15,6,
0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,
2,2,11,3,89,27,0,0,0,9,
0,1,97,47,0,26,9,0,0,62,
39,45,14,20,13,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,
26,30,25,3,0,18,5,3,11,2,
32,42,40,17,7,5,33,5,1,3,
0,0,0,0,0,0,0,0,0,0,
78,81,70,42,51,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0),
.Dim = c(34,10))