

MODELAGEM BAYESIANA PARA OS EXCESSOS DE INTERNAÇÕES HOSPITALARES DIÁRIAS DEVIDO À PNEUMONIA NA CIDADE DE SÃO PAULO

Mayara Piani Luna da Silva SICCHIERI ^{1 2}
Jorge Alberto ACHCAR ¹

- RESUMO: Neste trabalho, introduzimos novos modelos estatísticos para analisar dados de contagem de internações diárias devido à pneumonia na rede pública de hospitais da cidade de São Paulo no período de 01/01/2002 à 31/12/2005. Para analisar esses dados, consideramos a modelagem dos tempos entre excessos de internações diárias causadas por pneumonia assumindo uma distribuição exponencial e diferentes estruturas para a função de risco. As inferências para esses modelos são obtidas sob o paradigma Bayesiano e usando métodos de simulação MCMC (Monte Carlo em Cadeias de Markov) com o auxílio do *software* WinBugs. Também obtemos inferências Bayesianas para pontos de mudança da contagem diária de internações hospitalares devido à pneumonia.
- PALAVRAS-CHAVE: Tempo entre excessos de internação; análise Bayesiana; distribuição exponencial; modelagem do risco

1 Introdução

A contagem diária de internações hospitalares na rede de hospitais públicos da cidade de São Paulo tem grandes variações devido à vários fatores: estações do ano, variação climática, variação nos índices de poluentes, entre vários outros fatores. Um grande interesse dos administradores da área de saúde pública é relacionado à modelagem dessas contagens diárias, especialmente para os casos de excessos, o que pode implicar em muitos problemas nos hospitais.

Entre as várias doenças que levam à internações hospitalares, uma se destaca entre várias outras: a pneumonia.

Como motivação para o presente estudo, consideramos as contagens diárias de internações hospitalares devido à pneumonia nos hospitais públicos da cidade de São Paulo, no período de 01/01/2002 à 31/12/2005 (dados obtidos no Sistema Único de Saúde (SUS) relativo à cidade de São Paulo). Neste período de tempo (1461 dias), observamos grande variabilidade nas contagens diárias de internações hospitalares devido à pneumonia (ver figura 1).

Neste trabalho, vamos considerar a modelagem dos excessos de internações hospitalares devido à pneumonia. Em lugar de considerarmos as contagens diárias das internações hospitalares, vamos considerar as contagens acima de um valor estipulado (aqui consideramos um número de contagens diárias ≥ 70), onde modelamos os tempos (dias) t_i entre as ocorrências de excessos de internações em razão da pneumonia (ver dados na tabela 1). O valor maior ou igual a 70 internações diárias foi escolhido a partir de considerações empíricas, pois os dados da tabela 1 apresentam uma média amostral igual a 61,1 e o valor do terceiro quartil é 74, isto é, onde 75 % das observações apresentam valores menores ou iguais à 74. Dessa forma é razoável assumir excessos de internações diárias para valores maiores ou iguais a 70, onde a capacidade hospitalar de atendimento atinge limites de saturação, especialmente na rede hospitalar pública do Brasil.

Diferentes modelagens estatísticas podem ser consideradas para analisar os tempos entre ocorrências de excessos de internações hospitalares devido à pneumonia.

¹Departamento de Medicina Social, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo - USP, CEP: 14048-900, Ribeirão Preto, SP, Brasil, E-mail: mayapiani@hotmail.com / achcar@fmrp.usp.br

²Agradecimento à FAEPA/HCFMRP/USP pelo apoio financeiro.

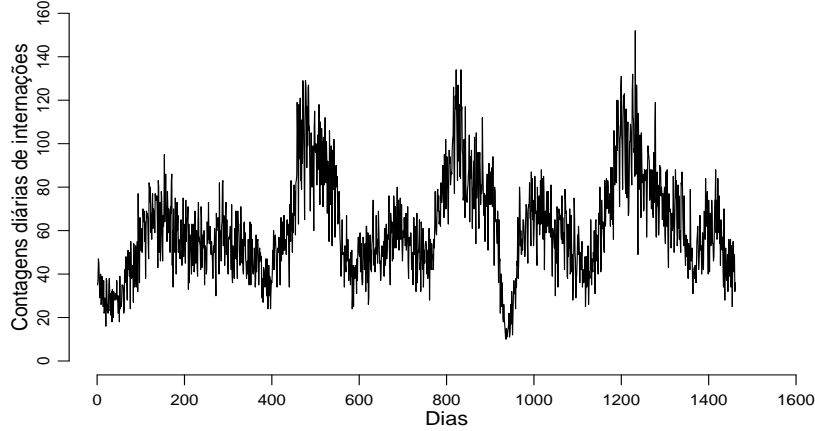


Figura - 1: Contagens diárias de internações hospitalares devido à pneumonia na cidade de São Paulo no período de 01/01/2002 a 31/12/2005.

As inferências de interesse considerando diferentes modelagens para os tempos entre ocorrências de excessos t_i são obtidos usando métodos Bayesianos. Os sumários a posteriori de interesse são obtidos via métodos de simulação MCMC (Monte Carlo em Cadeias de Markov) como o popular amostrador de Gibbs (ver por exemplo, Gelfand e Smith, 1990) ou o algoritmo de Metropolis-Hastings (ver por exemplo, Chib e Greenberg, 1995).

Este trabalho é organizado da seguinte forma: na seção 2 introduzimos diferentes modelagens estatísticas para t_i , $i=1, \dots, n$; na seção 3, introduzimos uma análise Bayesiana para os modelos propostos; na seção 4, introduzimos uma análise para os tempos entre excessos de internações hospitalares devido à pneumonia na cidade de São Paulo, e finalmente na seção 5, apresentamos algumas conclusões e discussão dos resultados obtidos.

2 Modelagem Estatística

Sejam T_1, T_2, \dots, T_n os tempos entre excessos de internações hospitalares devido à pneumonia supostos como variáveis aleatórias independentes com densidade exponencial, dada por,

$$f(t_i|\lambda_i) = \lambda_i e^{-\lambda_i t_i} \quad (1)$$

onde $t_i > 0$; a média de T_i é dada por $\theta_i = 1/\lambda_i$ e a variância de T_i é dada por $1/\lambda_i^2$, $i = 1, \dots, n$, e n é o número de dias com excesso de internações (≥ 70) em razão da pneumonia no período de 1461 dias (01/01/2002 à 31/12/2005) na cidade de São Paulo.

Diferentes modelos podem ser considerados para a função de risco λ_i ; inicialmente vamos considerar o modelo,

$$\lambda_1 = \exp\{\alpha + \eta_1\}$$

$$\lambda_i = \exp\{\alpha + \phi(\eta_{i-1} - \alpha)\} \quad (2)$$

onde $-\infty < \alpha < \infty$; $0 < \phi < 1$, para $i = 2, 3, \dots, n$ onde η_i é um fator aleatório não observado com distribuição normal dada por,

$$\eta_i \sim N(0, a^2) \quad (3)$$

onde $N(\mu, \sigma^2)$ denota uma distribuição normal com média μ e variância σ^2 ; vamos assumir a conhecido. Os parâmetros α e ϕ definidos em (2) são supostos desconhecidos. Vamos denotar o modelo definido por (1), (2) e (3) como “modelo 1”.

Observar que apesar de estarmos assumindo independência entre os tempos entre excessos de internações hospitalares devido à pneumonia (uma suposição discutível), o termo $\phi(\eta_{i-1} - \alpha)$ dado em (2), fornece uma contribuição auto-regressiva do valor prévio da variável latente não-observada na taxa de risco da distribuição exponencial com densidade (1).

O “modelo 2” é denotado por

$$\lambda_1 = N\alpha + w_1$$

$$\lambda_i = \alpha(N - i + 1) + w_i \quad (4)$$

onde $0 < \alpha < \infty$, pois a taxa de risco λ_i deve ser positiva; observar que neste caso, o termo $(N-i+1)$ leva a um decaimento da taxa ou função de risco do início ao término do período de tempo observado, $i = 2, 3, \dots, n$ e w_i tem distribuição Gama dada por

$$w_i \sim Gama(b, c) \quad (5)$$

onde b e c são hiperparâmetros conhecidos; $Gama(b, c)$ denota uma distribuição Gama com média b/c e variância b/c^2 .

Na prática, podemos ter dificuldades na escolha de valores dos hiperparâmetros b e c em (5); uma alternativa seria considerar b e c como quantidades aleatórias e usar métodos Bayesianos hierárquicos.

Considerando os modelos 1 a 2, a função de verossimilhança é dada por,

$$L = \prod_{i=1}^n \lambda_i e^{-\lambda_i t_i}. \quad (6)$$

3 Uma Análise Bayesiana

Para os modelos introduzidos na seção 2, consideramos independência a priori para os parâmetros dos modelos.

Observar que a abordagem Bayesiana é particularmente útil ao problema abordado considerando os modelos introduzidos na seção 2, com a presença de efeitos aleatórios ou variáveis não-observadas (ver por exemplo Bernardo e Smith, 1995).

Amostras das distribuições a posteriori para cada modelo proposto, são simuladas usando métodos MCMC. Uma grande simplificação é dada pelo uso do *software* Winbugs (Spiegelhalter *et al.*, 2001) que só requer a especificação da distribuição para os dados e as distribuições a priori para os parâmetros dos modelos considerados. Dessa forma, não vamos especificar as distribuições condicionais necessárias para o amostrador de Gibbs para cada modelo.

Diferentes distribuições a priori são consideradas para cada caso. Considerando o “modelo 1” definido por (1), (2) e (3), assumimos as seguintes distribuições a priori para os parâmetros α e ϕ :

$$\alpha \sim N(0, b^2) \quad \phi \sim Beta(c_1, d_1) \quad (7)$$

onde $Beta(c_1, d_1)$ denota uma distribuição Beta com média $c_1/(c_1 + d_1)$ e variância $c_1 d_1 / [(c_1 + d_1 + 1)(c_1 + d_1)^2]$; os hiperparâmetros b , c_1 e d_1 são conhecidos.

As distribuições a priori dadas em (7) foram consideradas levando em conta a variação dos parâmetros α e ϕ (ver (2)).

De forma similar, consideramos as distribuições a priori para o “modelo 2”.

É importante salientar que a escolha dos hiperparâmetros das distribuições a priori (7) podem ser baseadas em opinião de especialistas ou usando métodos Bayesianos empíricos (ver por exemplo, Carlin e Louis, 2000). Em alguns casos, escolhemos os valores dos hiperparâmetros que levam a distribuições a priori aproximadamente não-informativas.

Considerando o “modelo 2”, assumimos as seguintes distribuições a priori para os parâmetros do modelo:

$$\begin{aligned}\phi &\sim Beta(m_4, n_4) \\ \alpha &\sim U(o_4, p_4) \\ N &\sim U(g_2, h_2)\end{aligned}\tag{8}$$

Para a comparação dos modelos, usamos técnicas Bayesianas existentes; uma dessas técnicas é dada pelo critério DIC (Deviance Information Criterion) introduzido por Spiegelhalter *et al.*(2002).

Outra possibilidade para verificar o melhor ajuste, é comparar as médias ajustadas $\hat{\theta}_i = 1/\hat{\lambda}_i$ para cada modelo, onde $\hat{\lambda}_i$ é a estimativa de Monte Carlo da média a posteriori para λ_i baseada nas amostras simuladas de Gibbs, com os tempos observados t_i entre as ocorrências de excessos de internações devido à pneumonia. Dessa forma, podemos comparar as somas das diferenças $|\hat{\theta}_i - t_i|$, $i = 1, \dots, n$ para cada modelo considerado.

4 Análise dos tempos entre excessos de internações devido à pneumonia na cidade de São Paulo

Para uma análise Bayesiana dos tempos entre ocorrências de excessos de internações em razão da pneumonia na cidade de São Paulo, no período entre 01/01/2002 a 31/12/2005 (1461 dias), consideramos inicialmente o “modelo 1” com distribuições a priori dadas em (7) com $b = 1, c_1 = d_1 = 1$. Observar que $Beta(1, 1)$ corresponde a uma distribuição uniforme no intervalo (0, 1). Dessa forma estamos considerando distribuições aproximadamente não-informativas. Na tabela 1, temos os sumários a posteriori de interesse, obtidos usando o *software* WinBugs.

Para a simulação de amostras de Gibbs, geramos inicialmente 5000 amostras, que foram descartadas para eliminar o efeito dos valores iniciais (“*burn-in-sample*”), e consideramos 5000 amostras finais tomadas de 100 em 100. Esse procedimento de simulação de amostras utilizando o *software* WinBugs, também foi considerado para os outros modelos que serão dados a seguir. A convergência do algoritmo foi verificada através dos gráficos das séries simuladas.

Tabela - 1: Sumários a posteriori do “modelo 1”.

Parâmetro	Média	Desvio padrão	Intervalo de Credibilidade 95%
α	-1,96	0,2643	(-2,542 ; -1,521)
ϕ	0,6332	0,0425	(0,5491 ; 0,7182)

Para o “modelo 2”, consideramos $w_i \sim U(0, 100)$, $i = 1, \dots, n$; $\alpha \sim U(0, 1)$ e $N \sim U(445, 100000)$. Na tabela 2, temos os sumários a posteriori de interesse considerando o “modelo 2”.

Tabela - 2: Sumários a posteriori do “modelo 2”.

Parâmetro	Média	Desvio padrão	Intervalo de Credibilidade 95%
N	2876	2533	(469,2 ; 9180)
α	0,000004	0,0000056	(0,00000006 ; 0,00002)

Tabela - 3: Valores de DIC e $D(j)$.

Modelo j	DIC	$D(j)$
Modelo 1	1647,53	1071
Modelo 2	2009,4	946,9

Na tabela 3, temos os estimadores de Monte Carlo para o DIC, obtidos diretamente do *software* WinBugs, e os valores das somas das diferenças $D(j) = \sum_{i=1}^n |\hat{\theta}_i - t_i|$, onde j indexa o modelo.

Dos resultados dados na tabela 3, observamos que o melhor modelo ajustado usando o critério DIC é dado pelo “modelo 1” (menor DIC estimado); usando como critério de comparação os valores das somas das diferenças $D(j)$, observamos que o “modelo 2” é o melhor modelo ajustado pelos dados (menor valor de $D(j)$ para $j = 2$).

Na figura 2, temos os gráficos dos tempos entre excessos de internações devido à pneumonia *versus* i , $i = 1, 2, \dots, 445$, e sobrepostas, as médias ajustadas pelos modelos.

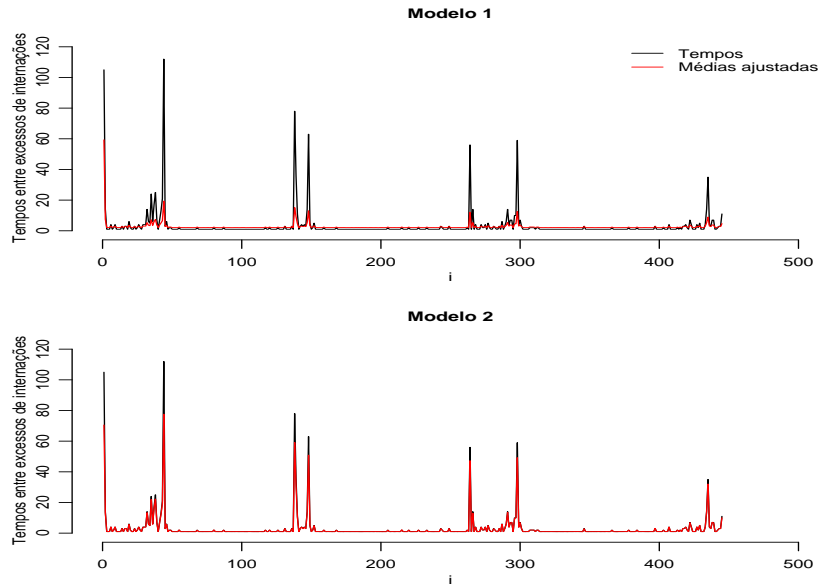


Figura - 2: Tempos entre ocorrências de excessos de internações t_i *versus* i e médias ajustadas pelos modelos propostos, $i = 1, \dots, 445$.

É interessante salientar que também podemos detectar pontos múltiplos de mudanças para os tempos entre ocorrências de excessos de internações, um problema de grande interesse para a rede pública de saúde.

Para isso, podemos construir intervalos de credibilidade para as diferenças entre as médias,

$$diff_i = \theta_i - \theta_{i-1} \quad (9)$$

para $i = 2, 3, \dots, n$.

Quando consideramos o “modelo 2”, na tabela 4 temos os intervalos de credibilidade 95% para as diferenças $diff_i$ dadas em (9) que não incluem o valor zero, pois isto é um indicativo de múltiplos pontos de mudanças.

De forma geral, podemos considerar o “modelo 2” mais apropriado para estimar os tempos entre excessos de internações devido à pneumonia, pois é o que melhor se ajusta aos dados tanto pelo critério $D(j)$, quanto graficamente.

Tabela - 4: Sumários a posteriori para $diff_i$, do “modelo 2”, que não incluem o valor zero para os intervalos de credibilidade 95%.

Parâmetro	Média	Desvio padrão	Intervalo de credibilidade 95%
$diff_3$	-12,07	17,17	(-52,3 ; -0,9447)
$diff_{45}$	-74,75	78,63	(-249 ; -15,94)
$diff_{138}$	58,17	57,87	(12,67 ; 202)
$diff_{141}$	-9,521	14,94	(-42,83 ; -0,5261)
$diff_{149}$	-45,99	53,18	(-171,4 ; -4,531)
$diff_{264}$	46,23	57,28	(8,702 ; 175,5)
$diff_{265}$	-46,26	57,27	(-175,5 ; -8,724)
$diff_{266}$	12,07	16,83	(1,187 ; 52,46)
$diff_{267}$	-12,07	16,78	(-52,32 ; -1,069)
$diff_{296}$	8,584	12,44	(0,0503 ; 39,65)
$diff_{299}$	-45,31	52,36	(-175,4 ; -5,278)

Notas Conclusivas

Neste trabalho, introduzimos novos modelos para analisar dados de internações hospitalares. Esses novos modelos introduzidos na seção 2 apresentam a presença de efeitos aleatórios, o que leva à grande flexibilidade de ajuste e permite estimar de forma precisa, múltiplos pontos de mudança, pois grande parte da variabilidade é explicada pelos efeitos aleatórios.

A detecção desses pontos de mudança é de grande interesse para os administradores e médicos da rede de saúde.

A metodologia proposta neste artigo poderia ser comparada com outras metodologias existentes na literatura, em geral envolvendo processos de Poisson não-homogêneos para a contagem dos excessos de internações devido à pneumonia (ou qualquer outra doença), mas a simplicidade da metodologia proposta neste artigo pode ser um fator de grande interesse na análise de dados da saúde.

Observar que a presença de efeitos aleatórios nos modelos propostos pode levar à dificuldades usando métodos de estatística clássica usuais baseado na teoria assintótica de estimadores de máxima verossimilhança (ver por exemplo, Cox e Hinkley, 1974).

Dessa forma, o uso de métodos Bayesianos via simulação MCMC pode ser uma alternativa eficaz para ajustar esses modelos à dados de internações hospitalares.

O uso de programas computacionais disponíveis, como o *software* WinBugs simplifica a obtenção dos sumários a posteriori de interesse.

Referências

- BERNARDO, J.M.; SMITH, A.F.M. *Bayesian theory*. New York: John Wiley and Sons, 1995.
- CARLIN, B.P.; LOUIS, T.A. *Bayes and empirical Bayes methods for data analysis*. 2.ed. Boca Raton: Chapman and Hall, 2000.
- CHIB, S.; GREENBERG, E. *Analysis of multivariate probit models*. *Biometrika*, v.85, n.2, p.347-361, 1998.
- COX, D.R.; HINKLEY, D.V. *Theoretical statistics*. Boca Raton: Chapman and Hall, 2000.
- GELFAND, A. E.; SMITH, A. F. M. *Sampling-based approaches to calculating marginal densities*. *Journal of the American Statistical Association*, v.85, p.398-409, 1990.
- SPIEGELHALTER, D. J.; THOMAS, A.; BEST, N. G.; GILKS, W. R. *WinBugs: Bayesian inference using Gibbs sampling*. Cambridge: MRC Biostatistics Unit., 2001.
- SPIEGELHALTER, D. J.; BEST, N. G.; CARLIN, B. P.; VAN DER LINDE, A. *Bayesian measures of model complexity and fit*. *Journal of the Royal Statistical Society, series B*, v.64, p.583-639, 2002.