UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO

**Modelagem Bayesiana dos tempos entre
extrapolações do número de internações hospitalares:
associação entre queimadas de cana-de-açúcar e
doenças respiratórias**

MAYARA PIANI LUNA DA SILVA SICCHIERI

Ribeirão Preto

2012

MAYARA PIANI LUNA DA SILVA SICCHIERI

**Modelagem Bayesiana dos tempos entre
extrapolações do número de internações hospitalares:
associação entre queimadas de cana-de-açúcar e
doenças respiratórias**

Dissertação apresentada à Faculdade de
Medicina de Ribeirão Preto da Universidade
de São Paulo para a obtenção do título de
Mestre em Saúde na Comunidade.

VERSÃO CORRIGIDA

A versão original encontra-se na Biblioteca
do Departamento de Medicina Social-
FMRP-USP.

Área de Concentração: Saúde na Comu-
nidade

Orientador: Prof. Dr. Jorge Alberto Achcar

Ribeirão Preto

2012

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Ficha Catalográfica

Sicchieri, Mayara Piani Luna da Silva

Modelagem Bayesiana dos tempos entre extrapolações do número de internações hospitalares: associação entre queimadas de cana-de-açúcar e doenças respiratórias, 2012.

88 p. :il. ; 30cm

Dissertação de Mestrado, apresentada à Faculdade de Medicina de Ribeirão Preto/USP. Área de concentração: Saúde na Comunidade.

Orientador: Jorge Alberto Achcar

1. Tempo entre extrapolações. 2. Modelagem Bayesiana.
3. Distribuição exponencial. 4. Doenças respiratórias. 5. Queimadas.

Folha de Aprovação

Mayara Piani Luna da Silva Sicchieri

Modelagem Bayesiana dos tempos entre extrapolações do número de internações hospitalares: associação entre queimadas de cana-de-açúcar e doenças respiratórias

Dissertação apresentada ao Departamento de Medicina Social da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para a obtenção do título de Mestre em Saúde na Comunidade.

Área de Concentração: Saúde na Comunidade

Aprovado em: ____/____/____

Banca Examinadora

Prof.(a) Dr.(a): _____

Instituição: _____ Assinatura: _____

Prof.(a) Dr.(a): _____

Instituição: _____ Assinatura: _____

Prof.(a) Dr.(a): _____

Instituição: _____ Assinatura: _____

Dedicatória

Aos meus pais, Ailton e Inês, ao meu irmão Gabriel e ao meu marido Eder, pelo incentivo, apoio, compreensão e acima de tudo, pelo amor incondicional durante todos estes anos e por tudo que ainda virá.

Agradecimentos

Primeiramente a Deus, que sempre esteve ao meu lado dando paciência e perseverança.

Aos meus pais por sempre me apoiarem e não me deixarem desistir.

Ao meu marido que sempre esteve ao meu lado, dando força e carinho.

Aos amigos do CEMEQ, em especial à Estela, Juliana, Ana e Suleimy pelos bons momentos que passamos juntas e por toda palavra de apoio, carinho e companheirismo.

Ao Prof. Dr. Benedito Galvão Benze, que desde a graduação acreditou em mim e em meu potencial.

À Profa. Dra. Vera Lúcia Damasceno Tomazella, que me proporcionou o primeiro contato com a pesquisa acadêmica.

Ao Prof. Dr. Jorge Alberto Achcar, pela sua orientação, paciência, confiança e dedicação durante o desenvolvimento deste trabalho.

Ao Prof. Dr. Edson Zangiacomi Martinez, por todas as oportunidades, primeiramente oferecida no CEMEQ, pelo incentivo na carreira acadêmica, pelas discussões que contribuíram para a melhora desta dissertação e por todos os momentos de ajuda e transmissão de conhecimento.

Aos docentes que participaram da banca de qualificação, Prof. Dr. Elcio S. O. Vianna, Prof. Dr. Edson Z. Martinez e Prof. Dr. Ricardo Z. N. Vêncio, obrigada pelas valiosas sugestões.

Aos membros que participaram da banca de defesa, Prof. Dr. Elcio S. O. Vianna, Prof. Dr. Jorge A. Achcar e Dr. Fernando A. B. Colugnati, obrigada pela presença e contribuição.

A todos os funcionários que direta ou indiretamente contribuíram para realização deste trabalho, em especial à Rosane do CPDH, à Eliana da UPC, ao Gabriel e à Cristina da BCRP e aos secretários do programa, Sérgio e Paula.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro.

Resumo

SICCHIERI, M. P. L. S. **Modelagem Bayesiana dos tempos entre extrapolações do número de internações hospitalares: associação entre queimadas de cana-de-açúcar e doenças respiratórias.** 2012. 88f. Dissertação (Mestrado) - Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2012.

As doenças respiratórias e a poluição do ar são temas de muitos trabalhos científicos, porém a relação entre doenças respiratórias e queimadas de cana-de-açúcar ainda é pouco estudada. A queima da palha da cana-de-açúcar é uma prática comum em grande parte do Estado de São Paulo, com especial destaque para os dados da região de Ribeirão Preto. Os focos de queimadas são detectados por satélites do CPTEC/INPE (Centro de Previsão de Tempo e Estudos Climáticos do Instituto Nacional de Pesquisas Espaciais) e neste trabalho consideramos o tempo entre dias de extrapolação do número de internações diárias. Neste trabalho introduzimos diferentes modelos estatísticos para analisar dados de focos de queimadas e suas relações com as internações por doenças respiratórias. Propomos novos modelos para analisar estes dados, na presença ou não da covariável, que representa o número de queimadas. Sob o enfoque Bayesiano, usando os diferentes modelos propostos, encontramos os sumários a posteriori de interesse utilizando métodos de simulação de Monte Carlo em Cadeias de Markov. Também usamos técnicas Bayesianas para discriminar os diferentes modelos. Para os dados da região de Ribeirão Preto, encontramos modelos que levam à obtenção das inferências a posteriori com grande precisão e verificamos que a presença da covariável nos traz um grande ganho na qualidade dos dados ajustados. Os resultados a posteriori nos sugerem evidências de uma relação entre as queimadas e o tempo entre as extrapolações do número de internações, ou seja, de que quando observamos um maior número de queimadas anteriores à extrapolação, também observamos que o tempo entre as extrapolações é menor.

Palavras - chave: Tempo entre extrapolações, Modelagem Bayesiana, Distribuição exponencial, Doenças respiratórias.

Abstract

SICCHIERI, M. P. L. S. **Bayesian modelling of the times between peaks of hospital admissions: association between sugar cane plantation burning and respiratory diseases.** 2012. 88p. Dissertation (Master degree) - Faculty of Medicine of Ribeirão Preto, University of São Paulo, Ribeirão Preto, 2012.

Relations between respiratory diseases and air pollution has been the goals of many scientific works, but the relation between respiratory diseases and sugar cane burning still is not well studied in the literature. Pre-harvest burning of sugarcane fields used primarily to get rid of the dried leaves is common in most of São Paulo state, Southeast Brazil, especially in the Ribeirão Preto region. The locals of pre-harvest sugar cane burning are detected by surveillance satellites of the CPTEC/INPE (Center of Climate Prediction of the Space Research National Institute). In this work, we consider as our data of interest, the time in days, between peaks numbers of hospitalizations due to respiratory diseases. Different statistical models are assumed to analyze the data of pre-harvest burning of sugar cane fields and their relations with hospitalizations due to respiratory diseases. These new models are considered to analyze data sets in presence or not of covariates, representing the numbers of pre-harvest burning of sugar cane fields. Under a Bayesian approach, we get the posterior summaries of interest using MCMC (Markov Chain Monte Carlo) methods. We also use different existing Bayesian discrimination methods to choose the best model. In our case, considering the data of Ribeirão Preto region, we observed that the models in presence of covariates give accurate inferences and good fit for the data. We concluded that there is evidence of a relationship between respiratory diseases and sugar cane burning, that is, larger numbers of pre-harvest sugar cane burning, implies in larger numbers of hospitalizations due to respiratory diseases. In this case, we also observe small times (days) between extra numbers of hospitalizations.

Keywords: Times between peaks numbers of hospitalizations, Bayesian models, Exponential distribution, Respiratory diseases, Burning.

Lista de Figuras

1	Número de internações diárias no período de 01/jan./1998 a 30/nov./2007	17
2	Tempos entre dias de extrapolação de internações no período de 01/jan./1998 a 30/nov./2007	18
3	Tempos entre dias de extrapolação de internações e número de queimadas, no período de 01/mar./2005 a 30/nov./2007	19
4	Exemplos das funções de densidade, de sobrevivência e de risco para a distribuição exponencial	24
5	Diferença entre queimadas < 11 e ≥ 11 em relação ao tempo entre extrapolações	37
6	Tempos entre extrapolações do número de internações t_i versus i e médias ajustadas pelos modelos propostos	44
7	Valores ajustados versus valores observados	45
8	Gráfico para $diff_i$ versus i	49
9	Gráficos de Gelman-Rubin para os parâmetros do <i>Modelo 1</i>	66
10	Gráficos de Gelman-Rubin para os parâmetros do <i>Modelo 2</i>	67
11	Gráficos de Gelman-Rubin para os parâmetros do <i>Modelo 3</i>	68
12	Gráficos de Gelman-Rubin para os parâmetros do <i>Modelo 4</i>	69
13	Correlogramas dos resíduos de cada modelo proposto	70

Lista de Tabelas

1	Percentual das causas de internações hospitalares por doenças respiratórias no período de 01/jan./1998 a 30/nov./2007	16
2	Distribuição à priori dos parâmetros dos modelos propostos	42
3	Sumários a posteriori dos modelos 1 a 4	43
4	Valores de DIC e $D(j)$	43
5	Sumários a posteriori para dif_i , dos modelos 1 e 3, que não incluem o valor zero para os intervalos de credibilidade 95% (IC 95%)	48

Sumário

1	INTRODUÇÃO	12
1.1	Um pouco sobre os dados: doenças do aparelho respiratório	15
1.2	Objetivos	19
1.3	Comitê de Ética em Pesquisa	20
1.4	Alguns conceitos básicos em análise de sobrevivência	20
1.5	Função de verossimilhança	22
1.6	Um modelo especial em análise de sobrevivência: a distribuição exponencial	23
2	UMA BREVE INTRODUÇÃO À ANÁLISE BAYESIANA	26
2.1	Probabilidade condicional e a fórmula de Bayes	26
2.2	A fundamentação da análise Bayesiana	27
2.3	Distribuições a priori	29
2.4	Métodos de simulação MCMC	29
2.5	O algoritmo Metropolis-Hastings	30
2.6	O amostrador de Gibbs	31
2.7	Critério de Informação Deviance (DIC)	32
3	MODELAGEM ESTATÍSTICA	34
3.1	Modelos para tempos entre extrapolações	34
3.2	Modelos para tempos entre extrapolações e queimadas	36
3.3	Análise Bayesiana	38
4	APLICAÇÃO COM OS DADOS DE RIBEIRÃO PRETO	41
4.1	Modelagem das extrapolações do número de internações por doenças res- piratórias na região de Ribeirão Preto	41
4.2	Pontos múltiplos de mudanças	47
5	CONSIDERAÇÕES FINAIS	51
	REFERÊNCIAS	56
	APÊNDICE A - Programas	62
A.1	- Modelos para tempos entre extrapolações	62
A.1.1	- Modelo 1	62
A.1.2	- Modelo 2	63
A.2	- Modelos para tempos entre extrapolações e queimadas	64

A.2.1 - Modelo 3	64
A.2.2 - Modelo 4	65
APÊNDICE B - Gráficos de convergência	66
B.1 - Gráficos para os modelos dos tempos entre extrapolações	66
B.1.1 - Modelo 1	66
B.1.2 - Modelo 2	67
B.2 - Gráficos para os modelos dos tempos entre extrapolações e queimadas . . .	68
B.2.1 - Modelo 3	68
B.2.2 - Modelo 4	69
APÊNDICE C - Correlogramas dos resíduos	70
APÊNDICE D - Artigo Publicado: Revista Brasileira de Biometria	71
ANEXO A - Comitê de Ética em Pesquisa	88

1 INTRODUÇÃO

A contagem diária de internações hospitalares na rede de hospitais públicos da região de Ribeirão Preto sofre 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 está relacionado à modelagem dessas contagens diárias, especialmente para os casos onde o número de internações diárias excede um determinado limiar, o que pode implicar em muitos problemas nos hospitais como, por exemplo, falta de leitos e equipamentos.

Entre as várias doenças que levam às internações hospitalares, um conjunto se destaca: as doenças respiratórias. De acordo com dados obtidos no DATASUS (2009), as doenças do aparelho respiratório, classificadas no capítulo X da Décima Classificação Internacional de Doenças (CID-10) da Organização Mundial da Saúde (OMS, 1998), foram a sexta maior causa de internações hospitalares na cidade de Ribeirão Preto, em hospitais da rede do Sistema Único de Saúde (SUS), no ano de 2008.

Atualmente, as discussões relacionadas à queima da palha de cana-de-açúcar vêm aumentando no Estado de São Paulo. Por conta dos altos níveis de poluição que estas queimadas favorecem, as autoridades competentes vêm tentando mudar o rumo desta prática.

Em 2002, o Governo do Estado de São Paulo aprovou uma lei que estabeleceu prazos para a redução e eliminação da queima da palha de cana-de-açúcar. O prazo máximo para eliminação total das queimadas era o ano de 2031.

No ano de 2007, foi firmado um Protocolo Agroambiental entre o Governo do Estado de São Paulo, a Secretaria de Meio Ambiente, a Secretaria de Agricultura e Abastecimento e a União da Indústria de cana-de-açúcar (UNICA), onde foi reduzido o prazo para a erradicação total das queimadas em 2017.

Com base no fator de emissão utilizado por Zancul (1998) os valores de material particulado emitidos pela queima da cana foram atualizados para o ano de 2010. Segundo

a Secretaria de Meio Ambiente do Governo do Estado de São Paulo, foram cultivados 4728133 ha (hectares) de cana-de-açúcar na safra de 2010/2011 no Estado de São Paulo sendo que 2627023 ha foram colhidos mecanicamente, sem efetuar a queima. Totalizando aproximadamente 44% da área cultivada, foram queimados 2101110 ha, com isso, produzindo um 1680880 toneladas de palha e outros materiais queimados nas lavouras de cana-de-açúcar em 2010. Assumindo o fator de emissão de material particulado como sendo 3,5 kg/t, teremos ao final aproximadamente 58831 t de MP emitidos através da queima da cana.

Segundo o Relatório Anual de Qualidade do Ar 2010 - CETESB (2011), foram emitidos cerca de 5000 t de MP no ano de 2010 na Região Metropolitana de São Paulo, considerando fontes móveis e fixas de emissão.

Em comparação, o cultivo da cana-de-açúcar ocupa cerca de 19% da área do estado de São Paulo, enquanto que a Região Metropolitana de São Paulo, cerca de 3%, e cerca de 8,5% dos 19% plantados foram queimados na safra de 2010/2011 no estado de São Paulo.

Contudo, a região de Ribeirão Preto possui o maior pólo sucroalcooleiro do país, e é uma das mais avançadas no cumprimento do Protocolo Agroambiental, de acordo com Fredo et al. (2008).

Ainda existem poucos estudos onde se relacionam as queimadas da cana-de-açúcar com o número de internações por doenças respiratórias em geral. Podemos citar os trabalhos de Lopes e Ribeiro (2006), Roseiro e Takayanagui (2006), Arbex (2001) e Arbex et al. (2004), como pioneiros na área.

O estudo de Toyoshima, Ito e Gouveia (2005) mostrou que existem padrões sazonais no número de internações por doenças respiratórias em geral. Ele verificou que os picos se dão nos meses do inverno e os vales no verão.

Godoy et al. (2001) fez um levantamento epidemiológico das doenças respiratórias que foram causas de internação hospitalar num serviço do SUS. Com isso, concluiu, que o tempo médio de internação dos pacientes com doenças respiratórias foi maior do que o tempo médio entre as internações por outras causas, no hospital estudado.

Martins et al. (2002) estudou a relação entre a poluição atmosférica e os atendimentos por pneumonia e gripe na cidade de São Paulo, mostrando que os poluentes provocam efeitos adversos na saúde dos idosos.

Os tempos entre extrapolações do número de internações são pouco estudados, porém sua modelagem é de grande valia para o sistema público de saúde e também para a saúde coletiva, pois esta modelagem pode trazer benefícios ao se detectar pontos de mudanças entre estes tempos, mostrando, por exemplo, que uma tomada de decisão por parte das autoridades competentes abaixou o número de internações causadas por uma determinada doença. O estudo de Achcar e Sicchieri (2010) possui estas características em relação à variável estudada e modelada.

Considere que num hospital o número de internações ultrapasse certo limiar, o objetivo deste trabalho está em modelar e analisar os tempos entre a extrapolação deste limiar previamente estabelecido, e que para este trabalho, será apresentado mais adiante.

Esta modelagem pode mostrar, por exemplo, que, se não houve mudanças em relação ao tempo entre as extrapolações de internações, implica que as políticas públicas existentes devem ser reavaliadas, pois estas não devem surtir o efeito necessário para redução dos níveis de poluentes, no caso deste trabalho, as reduções de queimadas.

Diferentes modelagens estatísticas podem ser consideradas para analisar os tempos entre as extrapolações do número de internações hospitalares devido às doenças respiratórias; este trabalho propõe a utilização sob o enfoque em análise de sobrevivência, modelar a taxa de risco do modelo exponencial.

As inferências de interesse, considerando diferentes modelagens para os tempos, serão obtidas usando métodos Bayesianos. Os sumários a posteriori de interesse serão obtidos via métodos de simulação MCMC (Monte Carlo em Cadeias de Markov) como o popular amostrador de Gibbs (GELFAND; SMITH, 1990) ou o algoritmo de Metropolis-Hastings (CHIB; GREENBERG, 1995).

1.1 Um pouco sobre os dados: doenças do aparelho respiratório

Os dados para este trabalho foram cedidos pelo Centro de Processamento de Dados Hospitalares (CPDH, 2010) do Departamento de Medicina Social da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo. Foi analisado o período entre Janeiro de 1998 e Novembro de 2007. Este banco de dados foi composto por variáveis de caracterização do paciente como sexo, idade, ocupação e cidade onde reside, e por variáveis de caracterização da internação como data da entrada e saída, condição de saída, principal afecção e cidade onde foi internado.

A queima de biomassa libera imediatamente para a atmosfera gases e material particulado (MP), o qual em sua maioria são de partículas finas e ultrafinas, sendo estas as mais nocivas à saúde humana, uma vez que as partículas finas e ultrafinas atingem os locais mais profundos do sistema respiratório, dando início a um processo inflamatório (ARBEX et al., 2004).

Outros elementos que são liberados pelas queimadas são o dióxido de carbono, monóxido de carbono, metano, hidrocarbonetos não metânicos, óxido nítrico e cloreto de metileno (MAGALHÃES, 2005).

No período do estudo, o Departamento Regional de Saúde de Ribeirão Preto (DRS XIII), abrangia um total de 25 municípios. Segundo Turn et al. (1997) as partículas finas emitidas pelas queimadas podem se deslocar para fora da região de origem, portanto foram selecionadas as internações cujos pacientes residissem no município de Ribeirão Preto e nas suas adjacências. Dessa forma, estão presentes na análise os residentes nas cidades de Brodowski, Cravinhos, Dumont, Guataporã, Jardinópolis, Ribeirão Preto, Serrana e Sertãozinho.

Foram registradas durante o período de estudo 80967 internações por doenças respiratórias, ou seja, todas as doenças referenciadas pelo capítulo X da CID-10. Sendo aproximadamente 54% das internações da rede pública e os outros 46%, da rede privada.

Assim como no estudo de Achcar e Sicchieri (2010), o valor do limiar foi escolhido

a partir de considerações empíricas. Uma vez que a média de internações diárias foi igual a 22,4, a mediana igual a 22 e o percentil 75 igual a 28, o valor escolhido foi 28. Consideramos que a partir deste valor (≥ 28) conseguimos representar o aumento das internações por doenças respiratórias agudas, que ocasionam os picos principalmente no período de estiagem, que consolidadamente é o período em que mais se agravam as doenças respiratórias.

Sendo o terceiro quartil considerado como o limiar para a extrapolação de internações, então se em um dia houver 28 ou mais internações, este dia será considerado como extrapolação, e o tempo analisado será o número de dias até a ocorrência da próxima extrapolação.

Um diferencial neste estudo é poder analisar os dados referentes ao sistema público e do sistema privado de saúde, sendo que os hospitais abrangidos foram os mesmos de onde os pacientes residissem, outra característica é que o paciente poderia residir em um município e ser internado em outro, desde que pertencesse à região de abrangência do estudo.

As principais causas de internações durante o período do estudo podem ser visualizadas na tabela 1.

Tabela 1 – Percentual das causas de internações hospitalares por doenças respiratórias no período de 01/jan./1998 a 30/nov./2007

Causa da internação (CID-10)	Percentual
Pneumonia não especificada (J189)	31,9
Doença pulmonar obstrutiva crônica não especificada (J449)	8,3
Hipertrofia das amígdalas com hipertrofia das adenóides (J353)	5,7
Asma não especificada (J459)	4,7
Desvio do septo nasal (J342)	4,1
Outras doenças do capítulo X	54,7
Total	100,0

Na figura 1 observa-se como estão distribuídas as internações no período de 01/Jan/1998 a 30/Nov/2007.

Figura 1 – Número de interações diárias no período de 01/jan./1998 a 30/nov./2007

Considerando que um número igual ou superior a 28 interações diárias caracteriza um dia cujo o número de interações foi extrapolado, observa-se na figura 2 como se comportam estes tempos no decorrer do período estudado. O tempo médio até a ocorrência de uma extrapolação é de, aproximadamente, 3,82, e o tempo mediano é de 2 dias.

Os dados sobre queimadas foram obtidos na página virtual do *Banco de Dados Queimadas* (BDQUEIMADAS, 2010) vinculado ao Centro de Previsão de Tempo e Estudos Climáticos/Instituto Nacional de Pesquisas Espaciais (CPTEC/INPE) no endereço eletrônico <<http://www.dpi.inpe.br/proarco/bdqueimadas/>>. Entretanto, não é possível identificar quais as queimadas que ocorreram exclusivamente em canaviais, acrescentando ao estudo um viés de informação. Porém, foram selecionadas somente as queimadas em áreas de vegetação denominada “não floresta”.

Segundo Baird (2002) o MP, tanto de partículas finas quanto de grossas, fica suspenso na atmosfera em média por 10 dias.

Por este motivo, para a modelagem considerando as queimadas, para cada dia extrapolado de interações relacionamos a quantidade de queimadas ocorridas nos últimos 10 dias, por exemplo, se o número de interações ficou acima ou igual a 28 no dia 27 de abril,

Figura 2 – Tempos entre dias de extrapolação de internações no período de 01/jan./1998 a 30/nov./2007

acumulamos a quantidade de queimadas entre os dias 18 e 27 de abril.

Para relacionar as queimadas ao tempo entre extrapolações será analisado o período entre 01/mar./2005 e 30/nov./2007, pois os dados coletados em relação às queimadas apresentavam muitas lacunas antes deste período.

Podemos observar na figura 3 a relação do tempo entre as extrapolações e a somatória das queimadas nos 10 dias que antecedem o evento. Neste caso, o tempo médio até a ocorrência de uma extrapolação é de 4,48, e o tempo mediano é 3 dias.

Figura 3 – Tempos entre dias de extrapolação de internações e número de queimadas, no período de 01/mar./2005 a 30/nov./2007

Pela figura 3 podemos observar que quanto menor o tempo entre extrapolações, maior é a quantidade de queimadas acumuladas nos últimos 10 dias.

1.2 Objetivos

Objetivo Geral:

Propor novos modelos estatísticos para o tempo entre os dias em que o número de internações hospitalares por doenças respiratórias foi maior ou igual a 28.

Objetivo Específico:

Estudar a relação entre as queimadas da palha de cana-de-açúcar e o tempo entre os dias de extrapolação do número de internações por doenças respiratórias na região de Ribeirão Preto.

1.3 Comitê de Ética em Pesquisa

Este trabalho foi submetido ao Comitê de Ética em Pesquisa do Hospital das Clínicas de Ribeirão Preto e Faculdade de Medicina de Ribeirão Preto, sob o título “MODELAGEM DOS EXCESSOS DE INTERNAÇÕES HOSPITALARES NA REGIÃO DE RIBEIRÃO PRETO: UMA APLICAÇÃO À DOENÇAS RESPIRATÓRIAS E EFEITO DAS QUEIMADAS DE CANA-DE-AÇÚCAR”. O trabalho foi aprovado de acordo com o processo HCRP nº 4290/2011 (ver anexo A).

1.4 Alguns conceitos básicos em análise de sobrevivência

Quando se fala em análise de sobrevivência logo relacionamos o tempo observado até o óbito de um paciente, porém esta área abrange muito mais, como análise industrial da duração de equipamentos e peças. Também é possível abordar o tempo até a ocorrência de uma extrapolação no número de internação, por exemplo, qual foi o tempo decorrido até a observação do número de internações por doenças respiratórias exceder o esperado.

Os tempos até a ocorrência de um evento pré determinado pelo pesquisador podem ser chamados de tempos de sobrevivência, de falha ou simplesmente de tempos de vida (COLOSIMO; GIOLO, 2006).

Uma característica frequente em dados de sobrevivência é a presença de censuras, que ocorrem quando a variável tempo possui informação incompleta, uma vez que o tempo parcial coletado pode não ser do tempo real até a ocorrência do evento. Num caso em que se observa a morte por pneumonia, a causa da morte do paciente pode ser de outra natureza, ou pode ser que o paciente tenha mudado de cidade; ou seja, todo o tempo até aquilo que for averso ao evento observado será um tempo censurado, ou simplesmente uma censura.

Existem vários tipos de censura entre as quais podem ser mencionadas: a censura do tipo I, censura do tipo II, censura aleatória, censura à esquerda e censura intervalar.

Os dados de sobrevivência reúnem não só os tempos de vida, mas também um conjunto de variáveis observáveis que podem ou não estar relacionadas com esses tempos. Essas variáveis são conhecidas como covariáveis ou variáveis explicativas.

A variável aleatória tempo de sobrevida, $T \geq 0$, que geralmente é uma variável contínua e sempre não negativa, pode ter seu comportamento definido pelas seguintes funções: função de densidade de probabilidade ($f(t)$), função de sobrevivência ($S(t)$), e função de risco ($h(t)$); estas funções podem ser utilizadas para descrever diferentes aspectos do tempos de sobrevida (LAWLESS, 1982).

A função densidade de probabilidade (fdp) é definida como o limite da probabilidade de uma observação falhar no intervalo de tempo $[t, t + \Delta t]$ por unidade de tempo, e é dada por (LEE, 1992)

$$f(t) = \lim_{\Delta t \rightarrow 0^+} \frac{P(t \leq T \leq t + \Delta t)}{\Delta t}, \quad (1)$$

onde, $f(t) \geq 0$, para todo t ; e a área contida entre a função de densidade e o eixo t é igual a 1.

A função de sobrevivência $S(t)$, é definida como a probabilidade de uma observação sobreviver mais que um determinado tempo t , ou seja, não falhar após um certo tempo, e é expressa por (LAWLESS, 1982)

$$S(t) = P(T \geq t) = 1 - F(t), \quad (2)$$

onde, $F(t) = \int_0^t f(u) du$, representa a função de distribuição acumulada.

Através da curva de sobrevivência pode-se comparar, por exemplo, dois ou mais tratamentos. Assim, uma curva de sobrevivência com declive acentuado representa baixa taxa ou curto tempo de vida, enquanto uma curva com leve declive representa alta taxa de sobrevida ou longo tempo de vida.

A função de risco $h(t)$, também conhecida como taxa de falha, é definida como o limite da probabilidade de uma observação falhar no intervalo de tempo $[t, t + \Delta t]$, dado que o mesmo tenha sobrevivido até o tempo t , e é dada por (COX; OAKES, 1984)

$$h(t) = \lim_{\Delta t \rightarrow 0^+} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}, \quad (3)$$

$$= \frac{f(t)}{S(t)}. \quad (4)$$

Esta função analisa a probabilidade de falha (risco) de uma determinada observação e descreve como a probabilidade instantânea de falha se modifica com o passar do tempo. Esta função pode apresentar diferentes comportamentos, pode ser crescente, decrescente, permanecer constante ou ser combinações destas.

1.5 Função de verossimilhança

Considere o caso em que as observações $t_1, \dots, t_n$ são uma amostra aleatória com fdp dada por $f(t; \theta)$, onde $\theta = (\theta_1, \dots, \theta_k)$ é um vetor de parâmetros desconhecidos assumindo valores em um conjunto Ω ; os t_i 's podem ser vetores, mas por simplicidade serão denotados como escalares. A função de verossimilhança para θ é definida por (LAWLESS, 1982)

$$L(\theta) = \prod_{i=1}^n f(t_i; \theta). \quad (5)$$

Seja $\hat{\theta}$ um ponto em Ω que maximiza $L(\theta)$; $\hat{\theta}$ é chamado de estimador de máxima verossimilhança (EMV) de θ . Em muitos modelos simples $\hat{\theta}$ existe e é único. É usual e conveniente trabalhar com $\log L(\theta)$, que também é maximizado em $\hat{\theta}$, e em muitos casos $\hat{\theta}$ pode ser prontamente encontrado pela resolução das então chamadas equações de máxima

verossimilhança $U_i(\theta) = 0$ ($i = 1, \dots, k$), onde

$$U_i(\theta) = \frac{\partial \log L(\theta)}{\partial \theta_i}. \quad (6)$$

1.6 Um modelo especial em análise de sobrevivência: a distribuição exponencial

Em análise de sobrevida muitas distribuições podem ser utilizadas na análise paramétrica de tempos relacionados a um certo evento de interesse. Dentre os principais modelos básicos de sobrevivência podemos mencionar algumas das distribuições mais populares: a exponencial, a Weibull, a log-normal, a log-logística, a gama e a distribuição valor extremo. Essas diferentes distribuições de probabilidade caracterizam diferentes formas de modelos podendo ter funções de risco constantes, crescentes, decrescentes e unimodais. A seguir serão explicitadas as funções de densidade de probabilidade, de sobrevivência e de risco para a distribuição exponencial.

A distribuição mais simples e importante do estudo da sobrevivência é a distribuição exponencial. Inicialmente, ela foi utilizada para modelar o tempo de vida de componentes eletrônicos. Também é tratada frequentemente como um modelo de falhas aleatórias. Ela é conhecida por ser a única que “não tem memória”, ou seja, o atual tempo de vida do animal ou do indivíduo não afeta a sobrevivência deles no futuro.

Em sua forma mais simples a distribuição exponencial possui somente um parâmetro, λ . A seguir, estão definidas a fdp, a função de sobrevivência e a função de risco, respectivamente:

$$f(t) = \frac{1}{\lambda} \exp \left\{ - \left(\frac{t}{\lambda} \right) \right\}, \quad (7)$$

$$S(t) = \exp \left\{ - \left(\frac{t}{\lambda} \right) \right\}, \quad (8)$$

$$h(t) = \frac{1}{\lambda}, \quad (9)$$

sendo que $t \geq 0$ e $\lambda > 0$.

Um valor grande de λ indica alto risco e um valor pequeno, baixo risco.

Na figura 4 observamos o comportamento da função de densidade, de sobrevivência e de risco, respectivamente, da distribuição exponencial para alguns valores de λ .

Fonte: Colosimo e Giolo (2005, p. 72)

Figura 4 – Exemplos das funções de densidade, de sobrevivência e de risco para a distribuição exponencial

Para analisar os dados relacionados às internações hospitalares em Ribeirão Preto, vamos considerar diferentes modelagens para o parâmetro λ da distribuição exponencial usada para analisar os tempos entre excessos de internações hospitalares.

2 UMA BREVE INTRODUÇÃO À ANÁLISE BAYESIANA

A análise Bayesiana tem por objetivo incorporar a informação a priori que o pesquisador acredita ter sobre um parâmetro de interesse θ . Observar que sob o enfoque clássico essa informação a priori não é considerada, pois o parâmetro é suposto como fixo e desconhecido, e utilizamos apenas a informação dos dados para obter as inferências de interesse.

Em geral, o valor verdadeiro de θ é desconhecido e o objetivo é fazer inferências sobre este parâmetro. Para representar os diferentes graus de incerteza sobre um parâmetro θ , diferentes modelos probabilísticos são elicitados; desta forma cada pesquisador pode formular um modelo estatístico baseado no seu grau de conhecimento sobre o parâmetro específico.

A informação acerca de um parâmetro θ é representada probabilisticamente por $\pi(\theta)$, também chamada de distribuição a priori, e incorporada ao estudo através do uso do teorema de Bayes, que combina a informação prévia do pesquisador com a informação contida nos dados, resultando na distribuição a posteriori.

Dentre os métodos para eliciações de distribuições a priori, podem ser citadas as prioris subjetivas, as conjugadas e as não informativas; neste trabalho serão utilizadas as prioris não informativas.

2.1 Probabilidade condicional e a fórmula de Bayes

Segundo Mood e Graybill (1974), sejam A e B dois eventos de um dado espaço de probabilidades denotado por Ω . A *probabilidade condicional* do evento B dado o evento A é denotada por $P(B|A)$, e é definida por

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \quad (10)$$

desde que $P(A) > 0$.

A fórmula de Bayes é baseada na probabilidade condicional apresentada em (10), mas sua maior característica é a de constituir uma regra de atualização das probabilidades, dada a informação de um evento A .

Suponha $B_1, \dots, B_n$ um conjunto de eventos em Ω , sendo $\Omega = \bigcup_{j=1}^n B_j$, $B_i \cap B_j = \phi$ (conjunto vazio) para todo $i \neq j$ e $P(B_j > 0)$ para todo $j = 1, \dots, n$. Seja A um evento qualquer tal que $P(A) > 0$. Então, para todo $j = 1, \dots, n$, define-se a *fórmula de Bayes* por

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{j=1}^n P(A|B_j)P(B_j)}. \quad (11)$$

2.2 A fundamentação da análise Bayesiana

Segundo interpretação de Ibrahim, Chen e Sinha (2001) a análise Bayesiana é baseada em especificar um modelo probabilístico para o vetor de tempos observados $\mathbf{T}$, dado um vetor de parâmetros $\boldsymbol{\theta}$, levando em consideração uma função de verossimilhança $L(\boldsymbol{\theta}|\mathbf{T})$. Assumindo, então, que $\boldsymbol{\theta}$ é aleatório, conseqüentemente tem-se uma distribuição a priori denotada por $\pi(\boldsymbol{\theta})$. A inferência sobre $\boldsymbol{\theta}$ é baseada numa distribuição a posteriori, a qual é obtida pelo teorema de Bayes, dado em (11). A distribuição a posteriori de $\boldsymbol{\theta}$ é dada por

$$\pi(\boldsymbol{\theta}|\mathbf{T}) = \frac{L(\boldsymbol{\theta}|\mathbf{T})\pi(\boldsymbol{\theta})}{\int_{\Theta} L(\boldsymbol{\theta}|\mathbf{T})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}} \quad (12)$$

onde Θ denota o espaço paramétrico de $\boldsymbol{\theta}$. Da equação (12) está evidente que $\pi(\boldsymbol{\theta}|\mathbf{T})$ é proporcional a verossimilhança multiplicada pela priori,

$$\pi(\boldsymbol{\theta}|\mathbf{T}) \propto L(\boldsymbol{\theta}|\mathbf{T})\pi(\boldsymbol{\theta})$$

e isso envolve uma contribuição dos dados observados através de $L(\boldsymbol{\theta}|\mathbf{T})$ e uma contribuição da informação a priori quantificada por $\pi(\boldsymbol{\theta})$.

A quantidade $m(\mathbf{T}) = \int_{\Theta} L(\boldsymbol{\theta}|\mathbf{T})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ é chamada de constante normalizadora de $\pi(\boldsymbol{\theta}|\mathbf{T})$.

Em muitos modelos e aplicações, $m(\mathbf{T})$ não apresenta uma forma analítica fechada, ou seja, uma distribuição conhecida, e portanto $\pi(\boldsymbol{\theta}|\mathbf{T})$ também não apresenta uma forma fechada. Além disso, na obtenção de outros sumários a posteriori precisamos resolver integrais múltiplas, muitas vezes, complicadas, o que exige o uso de métodos numéricos ou de aproximações de integrais, especialmente quando a dimensão do vetor de parâmetros é grande.

Daí surge a necessidade do uso de métodos computacionais poderosos, como o MCMC, e os algoritmos de Metropolis-Hastings e o amostrador de Gibbs (*Gibbs Sampler*).

Quando as distribuições condicionais a posteriori para cada parâmetro tem formas de distribuições conhecidas e são simples para gerar amostras, é mais usual utilizar-se do amostrador de Gibbs, que é baseado em um processo MCMC, o qual gera amostras das distribuições condicionais completas que convergem para a distribuição a posteriori de interesse; caso contrário o algoritmo de Metropolis-Hastings é utilizado no caso onde as distribuições condicionais a posteriori não possuem formas de distribuições conhecidas e simples para geração de amostras.

2.3 Distribuições a priori

As distribuições a priori carregam a informação que se sabe antes de um experimento ocorrer, ou seja, pode estar baseada na experiência pessoal de cada pesquisador. Porém, existem muitos métodos para a elicitacão de distribuições a priori adequados para cada experimento, dentre os quais se destacam:

- Distribuições a priori subjetivas

Consiste na quantificação de uma informação substancial, de um ou mais especialistas, sobre um parâmetro de interesse (PAULINO; TURKMAN; MURTEIRA, 2003).

- Distribuições a priori conjugadas

Atribuição de distribuições a priori que tenham a mesma classe de distribuições da posteriori (BERNARDO; SMITH, 1994).

- Distribuições a priori não informativas

Basicamente não incorpora informações relevantes sobre um parâmetro de interesse, seja por falta de informação sobre o mesmo, ou por descrença na informação obtida (BOX; TIAO, 1973).

2.4 Métodos de simulação MCMC

Os métodos MCMC são amplamente utilizados para a obtenção de resumos a posteriori em inferência Bayesiana. Como dito anteriormente, o amostrador de Gibbs é empregado quando as distribuições condicionais completas necessárias para o amostrador de Gibbs apresentam formas de distribuições conhecidas e fáceis para simulação de amostras.

2.5 O algoritmo Metropolis-Hastings

O algoritmo de Metropolis foi o primeiro a ser utilizado em métodos MCMC, foi introduzido por Metropolis et al.(1953) e generalizado por Hastings (1970). Tem por objetivo obter uma distribuição a posteriori através da construção de uma cadeia de Markov e de uma função de transição, das quais a distribuição estacionária gerada se iguala à distribuição a posteriori de interesse.

Supor que deseja-se simular uma densidade a posteriori $\pi(\theta|\mathbf{T})$. Um algoritmo de Metropolis-Hastings começa com um valor inicial θ^0 e especifica uma regra para a simulação do t -ésimo valor da sequência θ^t dado o $(t-1)$ -ésimo valor da sequência θ^{t-1} . Esta regra consiste em uma *densidade proposta* a qual simula um valor candidato θ^* e o cálculo de uma *probabilidade de aceitação* P , que indica a probabilidade de o valor candidato ser aceito para ser o próximo valor na sequência. Especificamente, esse algoritmo pode ser descrito da seguinte forma (ALBERT, 2007)

1. Simular um valor candidato θ^* de uma densidade proposta $p(\theta^*|\theta^{t-1})$.
2. Calcular a razão

$$R = \frac{\pi(\theta^*|\mathbf{T})p(\theta^{t-1}|\theta^*)}{\pi(\theta^{t-1}|\mathbf{T})p(\theta^*|\theta^{t-1})}$$

3. Calcular a probabilidade de aceitação $P = \min \{R, 1\}$.
4. Amostrar um valor θ^t tal que $\theta^t = \theta^*$ com probabilidade P , caso contrário $\theta^t = \theta^{t-1}$.

Sob certas condições de regularidade facilmente satisfeitas na densidade proposta $p(\theta^*|\theta^{t-1})$, a sequência simulada $\theta^1, \theta^2, \dots$ convergirá a uma variável aleatória que é distribuída de acordo com a distribuição a posteriori $\pi(\theta|\mathbf{T})$.

Para uma melhor compreensão do algoritmo, ver, por exemplo, Chib e Greenberg (1995).

2.6 O amostrador de Gibbs

O algoritmo de simulação Gibbs foi introduzido por Geman e Geman (1984) para a resolução de problemas ligados a reconstrução de imagens na área de física, mais tarde Gelfand e Smith (1990) mostraram como este algoritmo pode ser utilizado em estatística Bayesiana para simular distribuições a posteriori. O objetivo deste amostrador é gerar cadeias de Markov a partir das distribuições condicionais completas, com forma analítica conhecida, para finalmente obter resumos a posteriori dos parâmetros de interesse.

Suponha que $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ é um vetor de parâmetros aleatórios e $\mathbf{T}$ é o vetor dos dados observados; tem-se como objetivo, obter inferências sobre a distribuição a posteriori conjunta $\pi(\boldsymbol{\theta}|\mathbf{T}) = \pi(\theta_1, \dots, \theta_k|\mathbf{T})$ (BERNARDO; SMITH, 1994).

Dado um vetor arbitrário de valores iniciais

$$\theta_1^{(0)}, \dots, \theta_k^{(0)}$$

para as quantidades desconhecidas, pode-se implementar o seguinte procedimento iterativo:

obtem-se $\theta_1^{(1)}$ de $\pi(\theta_1|\mathbf{T}, \theta_2^{(0)}, \dots, \theta_k^{(0)})$,
 obtem-se $\theta_2^{(1)}$ de $\pi(\theta_2|\mathbf{T}, \theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_k^{(0)})$,
 obtem-se $\theta_3^{(1)}$ de $\pi(\theta_3|\mathbf{T}, \theta_1^{(1)}, \theta_2^{(1)}, \theta_4^{(0)}, \dots, \theta_k^{(0)})$,
 $\vdots$
 obtem-se $\theta_k^{(1)}$ de $\pi(\theta_k|\mathbf{T}, \theta_1^{(1)}, \dots, \theta_{k-1}^{(1)})$,
 obtem-se $\theta_1^{(2)}$ de $\pi(\theta_1|\mathbf{T}, \theta_2^{(1)}, \dots, \theta_k^{(1)})$,
 $\vdots$

e assim por diante.

Agora, suponha que este processo é continuado através de t iterações e é independentemente replicado m vezes para que ao final se tenha m replicações do vetor amostrado $\boldsymbol{\theta}^t = (\theta_1^{(t)}, \dots, \theta_k^{(t)})$, onde $\boldsymbol{\theta}^t$ é uma realização de uma cadeia de Markov com probabilidade

de transição dada por

$$p(\boldsymbol{\theta}^t, \boldsymbol{\theta}^{t+1}) = \prod_{l=1}^k \pi(\theta_l^{t+1} | \mathbf{T}, \theta_1^{t+1}, \dots, \theta_{l-1}^{t+1}, \theta_{l+1}^t, \dots, \theta_k^t).$$

Então, como $t \rightarrow \infty$, $(\theta_1^{(t)}, \dots, \theta_k^{(t)})$ tende em distribuição a um vetor aleatório cuja densidade conjunta é $\pi(\boldsymbol{\theta} | \mathbf{T})$, ou seja, a distribuição a posteriori de interesse. Em particular, $\theta_i^{(t)}$ tende em distribuição a uma quantidade aleatória cuja densidade é $\pi(\theta_i | \mathbf{T})$, também chamada de densidade marginal a posteriori de θ_i . Desta maneira, para um t grande, as replicações $(\theta_{i1}^{(t)}, \dots, \theta_{im}^{(t)})$ são aproximadamente uma amostra aleatória de $\pi(\theta_i | \mathbf{T})$.

Após a geração de amostras da distribuição a posteriori de interesse, utilizamos essas amostras para obter estimadores de Monte Carlo para sumários a posteriori de interesse como a média a posteriori, o desvio-padrão a posteriori e intervalos de credibilidade de interesse.

2.7 Critério de Informação Deviance (DIC)

Para a seleção de modelos, podemos usar diferentes critérios de seleção. Um critério Bayesiano muito popular e implementado no software WinBUGS (LUNN et al., 2000) é dado pelo critério de informação deviance (DIC).

O DIC é um critério útil para a seleção de modelos sob o enfoque Bayesiano, onde amostras da distribuição a posteriori para os parâmetros do modelo são obtidas usando métodos MCMC.

O deviance (desvio) é definido por

$$D(\boldsymbol{\theta}) = -2\log L(\boldsymbol{\theta}) + C \tag{13}$$

onde $\boldsymbol{\theta}$ é um vetor de parâmetros desconhecidos do modelo, $L(\boldsymbol{\theta})$ é a função de verossimilhança do modelo e C é uma constante que não é necessário que seja conhecida quando a comparação entre modelos é efetuada.

O critério DIC definido por Spiegelhalter et al. (2002) é dado por

$$DIC = D(\hat{\boldsymbol{\theta}}) + 2n_D \quad (14)$$

onde $D(\hat{\boldsymbol{\theta}})$ é o desvio avaliado na média a posteriori $\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}|\text{dados})$ e n_D é o número efetivo de parâmetros do modelo dado por $n_D = \bar{D} - D(\hat{\boldsymbol{\theta}})$, onde $\bar{D} = E(D(\boldsymbol{\theta})|\text{dados})$ é o desvio a posteriori que mede a qualidade dos dados ajustados pelo modelo. Menores valores de DIC indicam melhores modelos. Observar que esses valores podem ser negativos.

3 MODELAGEM ESTATÍSTICA

A modelagem estatística é fundamental quando se quer verificar a relação existente entre duas ou mais variáveis de interesse. No caso deste trabalho, tem-se por objetivo modelar a taxa de risco do modelo exponencial para tentar explicar como é o comportamento dos tempos entre os dias de extrapolação do número de internações hospitalares na região de Ribeirão Preto.

Para os casos onde se quer verificar o efeito das queimadas de cana-de-açúcar, é imprescindível a modelagem estatística, pois somente através desta é possível chegar aos resultados esperados.

3.1 Modelos para tempos entre extrapolações

Sejam $T_1, T_2, \dots, T_n$ os tempos entre os dias de extrapolação de internações hospitalares causadas por doenças respiratórias supostos como variáveis aleatórias independentes com densidade exponencial, dada por

$$f(t_i|\lambda_i) = \lambda_i e^{-\lambda_i t_i} \quad (15)$$

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 em que internações estivessem acima de 28 (≥ 28) em razão das doenças respiratórias no período de 3621 dias (01/01/1998 à 30/11/2007) na região de Ribeirão Preto. Note que em (15) foi utilizada uma forma reparametrizada da distribuição exponencial.

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

$$\begin{aligned}\lambda_1 &= w_1 \\ \lambda_2 &= \{(\phi_1 t_1) \cdot [\alpha(N-1)]\} + w_2\end{aligned}$$

$$\lambda_i = \{(\phi_1 t_{i-1} + \phi_2 t_{i-2}) \cdot [\alpha(N-i+1)]\} + w_i \quad (16)$$

onde $0 < \phi_1, \phi_2 < 1$ e $0 < \alpha, N < \infty$, para $i = 3, 4, \dots, n$ onde w_i é um fator aleatório não observado com distribuição uniforme dada por

$$w_i \sim U(0, b) \quad (17)$$

onde $U(a, b)$ denota uma distribuição uniforme com média $\frac{a+b}{2}$ e variância $\frac{(b-a)^2}{12}$; assumir b conhecido. Os parâmetros α , ϕ_1 e ϕ_2 definidos em (16) são supostos desconhecidos. Denotar o modelo definido por (15), (16) e (17) como “modelo 1”.

Observar que apesar da independência assumida entre os tempos avaliados (uma suposição discutível), os termos ϕ_1 e ϕ_2 dados em (16), fornecem uma contribuição autoregressiva de ordem 1 e 2, respectivamente, do valor da variável tempo entre extrapolações na taxa de risco da distribuição exponencial com densidade (15).

É importante salientar que o “modelo 1” e as generalizações desse modelo que serão introduzidos a seguir são novos modelos propostos nesta dissertação para avaliar o comportamento dos tempos entre as internações hospitalares acima do esperado, sejam as internações por doenças respiratórias, como é o caso desta aplicação, seja por qualquer outra doença de interesse da saúde coletiva.

Observar que no modelo 1 consideramos o termo $(N-i+1)$ na modelagem da função de risco λ_i , $i = 3, 4, \dots, n$, o qual faz a taxa de risco decrescer a cada extrapolação.

Uma alternativa seria assumir o termo $(N+i)$ em lugar do termo $(N-i+1)$ o que implica em taxas de riscos λ_i crescentes, isto é, tempos médios θ_i decrescentes; dessa

forma, consideramos o “modelo 2” dado por

$$\begin{aligned}\lambda_1 &= w_1 \\ \lambda_2 &= \{(\phi_1 t_1) \cdot [\alpha (N + 2)]\} + w_2 \quad \lambda_3 = \{(\phi_1 t_2 + \phi_2 t_1) \cdot [\alpha (N + 3)]\} + w_3 \\ \lambda_i &= \{(\phi_1 t_{i-1} + \phi_2 t_{i-2} + \phi_3 t_{i-3}) \cdot [\alpha (N + i)]\} + w_i\end{aligned}\tag{18}$$

onde $0 < \phi_1, \phi_2, \phi_3 < 1$ e $0 < \alpha, N < \infty$, para $i = 4, 5, \dots, n$ onde w_i é um fator aleatório não observado com distribuição uniforme dada em (17). Denotar este modelo como “modelo 2”.

Observar que considerando cada um dos modelos 1 e 2, a função de verossimilhança é dada por

$$L = \prod_{i=1}^n \lambda_i e^{-\lambda_i t_i}.\tag{19}$$

3.2 Modelos para tempos entre extrapolações e queimadas

A fim de verificar a associação entre as queimadas de cana-de-açúcar e o tempo entre extrapolações do número de internações por doenças respiratórias foi acrescentada ao modelo uma variável X , do tipo “*dummy*”, dada por,

$$x_i = \begin{cases} 0, & \text{se o número de queimadas nos últimos 10 dias à extrapolação for } < 11 \\ 1, & \text{se o número de queimadas nos últimos 10 dias à extrapolação for } \geq 11. \end{cases}$$

Este ponto de corte foi determinado com base na maior diferença média encontrada entre os tempos de extrapolações. Portanto, quando assumimos o valor 11 como ponto de corte encontramos uma diferença média de 3,84 dias entre os grupos < 11 e ≥ 11 , ou seja, é o ponto que melhor consegue diferenciar os tempos de extrapolações. Podemos observar este comportamento através da figura 5.

Figura 5 – Diferença entre queimadas < 11 e ≥ 11 em relação ao tempo entre extrapolações

Desta forma, os modelos propostos na seção 3.1 foram acrescidos da variável X e de seu respectivo coeficiente, β , e se apresentam a seguir.

Inicialmente considerar o modelo

$$\begin{aligned}\lambda_1 &= w_1 \\ \lambda_2 &= \{(\phi_1 t_1) \cdot [\alpha (N - 1) + \beta x_2]\} + w_2\end{aligned}$$

$$\lambda_i = \{(\phi_1 t_{i-1} + \phi_2 t_{i-2}) \cdot [\alpha (N - i + 1) + \beta x_i]\} + w_i \quad (20)$$

onde $0 < \phi_1, \phi_2 < 1$ e $0 < \alpha, N, \beta < \infty$, para $i = 3, 4, \dots, n$ onde w_i é um fator aleatório

não observado com distribuição uniforme dada por (17)

Denotar o modelo definido por (15), (20) e (17) como “modelo 3”.

O próximo modelo proposto para a função de risco, é dado por

$$\begin{aligned}\lambda_1 &= w_1 \\ \lambda_2 &= \{(\phi_1 t_1) \cdot [\alpha (N + 2) + \beta x_2]\} + w_2 \\ \lambda_3 &= \{(\phi_1 t_2 + \phi_2 t_1) \cdot [\alpha (N + 3) + \beta x_3]\} + w_3 \\ \\ \lambda_i &= \{(\phi_1 t_{i-1} + \phi_2 t_{i-2} + \phi_3 t_{i-3}) \cdot [\alpha (N + i) + \beta x_i]\} + w_i\end{aligned}\tag{21}$$

onde $0 < \phi_1, \phi_2, \phi_3 < 1$ e $0 < \alpha, N, \beta < \infty$, para $i = 4, 5, \dots, n$ onde w_i é um fator aleatório não observado com distribuição uniforme dada em (17). Denotar este modelo como “modelo 4”.

3.3 Análise Bayesiana

Para as modelagens introduzidas nas seções 3.1 e 3.2, considera-se independência a priori para os parâmetros dos modelos.

Observa-se que a abordagem Bayesiana é particularmente útil ao problema descrito para os diferentes modelos propostos, todos considerando a presença de efeitos aleatórios ou variáveis não-observadas (BERNARDO; SMITH, 1994).

Amostras das distribuições a posteriori para cada modelo introduzido, são simuladas usando métodos MCMC. Uma grande simplificação é dada pelo uso do *software* WinBUGS que só requer a especificação da distribuição para os dados e as distribuições a priori para os parâmetros dos modelos considerados. Dessa forma, não serão especificadas 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 (15), (16) e (17), foram propostas as seguintes distribuições a priori para os parâmetros α , ϕ_1 e ϕ_2 :

$$\begin{aligned}\alpha &\sim U(b_1, c_1) \\ \phi_1 &\sim Beta(d_1, e_1) \\ \phi_2 &\sim Beta(d_1, e_1) \\ N &\sim U(f_1, g_1)\end{aligned}\tag{22}$$

onde $Beta(d, e)$ denota uma distribuição $Beta$ com média $\frac{d}{(d+e)}$ e variância $\frac{de}{(d+e+1)(d+e)^2}$; os hiperparâmetros b_1, c_1, d_1, e_1, f_1 e g_1 são hiperparâmetros conhecidos.

As distribuições a priori dadas em (22) foram consideradas levando em conta a variação dos parâmetros α , ϕ_1 e ϕ_2 (ver (16)).

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

É importante salientar que a escolha dos hiperparâmetros das distribuições a priori (22) podem ser baseadas em opinião de especialistas ou usando métodos Bayesianos empíricos (CARLIN; LOUIS, 2000). Em alguns casos, os valores dos hiperparâmetros foram escolhidos de forma a que levassem a distribuições a priori aproximadamente não-informativas.

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

$$\begin{aligned}\alpha &\sim U(b_2, c_2) \\ \phi_1 &\sim Beta(d_2, e_2) \\ \phi_2 &\sim Beta(d_2, e_2) \\ \phi_3 &\sim Beta(d_2, e_2) \\ N &\sim U(f_2, g_2)\end{aligned}\tag{23}$$

Considerando os modelos 3 e 4, assumimos as mesmas distribuições a priori consideradas nos modelos 1 e 2, respectivamente.

Para a comparação dos diferentes modelos, usou-se técnicas Bayesianas existentes;

uma dessas técnicas é dada pelo critério DIC introduzido na seção 2.7.

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 . Dessa forma, pode-se comparar as somas das diferenças $|\hat{\theta}_i - t_i|$, $i = 1, \dots, n$ para cada modelo considerado.

4 APLICAÇÃO COM OS DADOS DE RIBEIRÃO PRETO

Para a simulação de amostras de Gibbs, gerou-se inicialmente 3000 amostras, que foram descartadas para eliminar o efeito dos valores iniciais (“*burn-in-sample*”), e foram consideradas 5000 amostras finais tomadas de 25 em 25. Este salto de 25 em 25 visa retirar uma possível autocorrelação entre as amostras geradas. Também foram simuladas três cadeias independentes e simultâneas para cada modelo apresentado, e sua convergência foi verificada pelo método gráfico de Gelman-Rubin (ver apêndice B). Todo o procedimento foi realizado com o auxílio do *software* WinBUGS.

Além de testes de convergências dos modelos, também foi realizada uma análise de resíduos, no qual estes foram avaliados por um correlograma (ver apêndice C), que é um método gráfico para verificar se os resíduos estão correlacionados.

4.1 Modelagem das extrapolações do número de internações por doenças respiratórias na região de Ribeirão Preto

Foi realizada uma análise Bayesiana dos tempos entre os dias em que ocorreram extrapolações no número de internações por doenças respiratórias no período entre 01/jan./1998 e 30/nov./2007, totalizando 3621 dias, sendo que, o número total de extrapolações no período foi igual a 948.

No entanto, para analisar uma possível relação entre queimadas e o tempo entre extrapolações, foi analisado o período entre 01/mar./2005 e 30/nov./2007, pois os dados coletados em relação às queimadas apresentavam muitas lacunas antes deste período, embora que, pelas informações obtidas no CPTEC/INPE, estas lacunas representam que não houve foco de queimada registrado pelo satélite. Portanto, a fim de diminuir um possível viés na modelagem, foi utilizado um tempo menor de observações, totalizando

223 extrapolações.

Amostras das distribuições a posteriori para cada modelo introduzido nas seções 3.1 e 3.2 foram obtidas pelo uso do *software* WinBUGS. Dessa forma, não serão especificadas as distribuições condicionais necessárias para o amostrador de Gibbs para cada modelo.

Na tabela 2 temos as distribuições a priori assumidas nos modelos de 1 a 4. Vale ressaltar que todas as distribuições e hiperparâmetros foram escolhidos de forma com que não se colocasse informação sobre os dados, ou seja, foram admitidas prioris não-informativas, respeitando-se os limites de variação de cada parâmetro dos modelos propostos na seção 3.

Tabela 2 – Distribuição à priori dos parâmetros dos modelos propostos

Modelo	Prioris
<i>Modelo 1</i>	$\alpha \sim U(0, 100)$
	$\phi_1 \sim Beta(1, 1)$
	$\phi_2 \sim Beta(1, 1)$
	$N \sim U(948, 100000)$
	$w_i \sim U(0, 100)$
<i>Modelo 2</i>	$\alpha \sim U(0, 100)$
	$\phi_1 \sim Beta(1, 1)$
	$\phi_2 \sim Beta(1, 1)$
	$\phi_3 \sim Beta(1, 1)$
	$N \sim U(0, 100000)$
<i>Modelo 3</i>	$w_i \sim U(0, 100)$
	$\alpha \sim U(0, 100)$
	$\phi_1 \sim Beta(1, 1)$
	$\phi_2 \sim Beta(1, 1)$
	$N \sim U(223, 100000)$
<i>Modelo 4</i>	$w_i \sim U(0, 100)$
	$\beta \sim U(0, 10)$
	$\alpha \sim U(0, 100)$
	$\phi_1 \sim Beta(1, 1)$
	$\phi_2 \sim Beta(1, 1)$
	$\phi_3 \sim Beta(1, 1)$
<i>Modelo 4</i>	$N \sim U(0, 100000)$
	$w_i \sim U(0, 100)$
	$\beta \sim U(0, 10)$

Após estimar os parâmetros de interesse, é necessário avaliar qual o modelo que melhor se ajustou aos dados, para isso, além do valor de DIC apresentado na seção 2.7, foram utilizados os valores das somas das diferenças obtidas através da equação (24)

$$D(j) = \sum_{i=1}^n \left| \hat{\theta}_i - t_i \right|, \quad (24)$$

onde j indexa o modelo.

Na tabela 3 são apresentadas as estimativas dos parâmetros dos modelos 1 a 4. Enquanto que na tabela 4 são apresentados os estimadores de Monte Carlo para o DIC, obtidos diretamente do *software* WinBUGS, e os valores das somas das diferenças obtidas através equação 24.

Tabela 3 – Sumários a posteriori dos modelos 1 a 4

Modelo	Parâmetro	Média	Desvio padrão	Intervalo de Credibilidade 95%
<i>Modelo 1</i>	α	1,08E-6	1,771E-5	(8,53E-10 ; 4,13E-6)
	ϕ_1	0,3545	0,286	(0,005656 ; 0,9423)
	ϕ_2	0,3668	0,289	(0,006323 ; 0,9542)
	N	17710	23510	(10000 ; 85920)
<i>Modelo 2</i>	α	1,15E-7	3,378E-7	(4,96E-10 ; 7,441E-7)
	ϕ_1	0,4154	0,2857	(0,01392 ; 0,9607)
	ϕ_2	0,4187	0,2855	(0,01492 ; 0,9619)
	ϕ_3	0,4422	0,2879	(0,01734 ; 0,9673)
	N	19230	24730	(96,72 ; 88000)
<i>Modelo 3</i>	α	0,00111	0,01928	(9,704E-9 ; 5,992E-3)
	ϕ_1	0,07514	0,1631	(1,743E-5 ; 0,6372)
	ϕ_2	0,08658	0,1772	(2,453E-5 ; 0,6874)
	N	15410	23180	(250 ; 85300)
	β	1,104	2,08	(6,25E-4 ; 8,027)
<i>Modelo 4</i>	α	1,252E-5	0,001012	(1,712E-9 ; 2,295E-5)
	ϕ_1	0,2973	0,2681	(0,003602 ; 0,9241)
	ϕ_2	0,3155	0,2744	(0,004548 ; 0,9331)
	ϕ_3	0,3032	0,2726	(0,003811 ; 0,9323)
	N	14000	22450	(17,46 ; 83100)
	β	0,01374	0,1316	(1,248E-4 ; 0,06949)

Tabela 4 – Valores de DIC e $D(j)$

Modelo	DIC	$D(j)$
<i>Modelo 1</i>	4696,310	193,243
<i>Modelo 2</i>	5051,050	206,181
<i>Modelo 3</i>	-35926,900	425,698
<i>Modelo 4</i>	-649,087	750,159

Figura 6 – Tempos entre extrapolações do número de interações t_i *versus* i e médias ajustadas pelos modelos propostos

Dos resultados da tabela 4, observa-se que o melhor modelo ajustado, no geral, utilizando o critério DIC (menor DIC estimado) é o “modelo 3”; mas pelo critério de comparação dos valores das somas das diferenças $D(j)$, observamos que o “modelo 1” é o melhor modelo ajustado (menor valor de $D(j)$ para $j = 1$).

Na figura 6, podem ser observados os gráficos dos tempos entre extrapolações do número de interações causadas por doenças respiratórias *versus* i , e sobrepostas, as médias estimadas pelos modelos. Podemos observar graficamente que todos os modelos apresentam estimativas muito próximas das observações. Também visualizamos este ajuste pela figura 7, onde podem ser observados os gráficos entre os valores ajustados

contra os observados para cada modelo proposto, traçando uma reta, a qual representa a igualdade entre os valores dos dois eixos, verifica-se que os valores estimados pelos modelos 1 e 2 são os que mais se aproximam dos valores observados, embora quando o modelo é ajustado pelas queimadas, o “modelo 3” é mais próximo que o “modelo 4”.

Figura 7 – Valores ajustados *versus* valores observados

A partir dos modelos 1 e 3 interpretamos os resultados obtidos na tabela 3. Lembrando que as interpretações são facilmente extensíveis aos modelos 2 e 4, respectivamente.

Primeiramente tomemos o “modelo 1”. Observamos que todos os parâmetros são significativos pelos intervalos de credibilidade 95%, ou seja, o valor zero não está contido em toda sua amplitude. O mesmo observou-se para os demais modelos avaliados.

Quanto ao parâmetro α , ele está diretamente relacionado com a taxa de risco ($\hat{\lambda}$) do modelo exponencial, e portanto inversamente relacionado com a média ($\hat{\theta}$) dos tempos de extrapolações estimados, consequentemente, quanto maior é α , menor é $\hat{\theta}$.

O parâmetro N também exerce o mesmo efeito de α . Os parâmetros ϕ_1 e ϕ_2 representam uma contribuição dos tempos anteriores ao tempo atual estimado. ϕ_1 representa o efeito que o tempo imediatamente anterior possui sobre a taxa de risco, enquanto que ϕ_2 , representa um efeito auto-regressivo de ordem 2, ou seja, é dependente do penúltimo tempo imediatamente anterior. Pelos modelos 2 e 4 temos ϕ_3 , que contribui com efeito auto-regressivo de ordem 3. Como o tempo é estimado através de $\hat{\theta} = \frac{1}{\lambda}$, quanto maiores as contribuições auto-regressivas, menores serão os tempos estimados.

Por fim, interpretamos o parâmetro β , que assim como os demais, quanto maior é sua estimativa, menor é o tempo entre extrapolações. Como suas estimativas são maiores que zero e seus respectivos intervalos de credibilidade 95% não possuem o valor zero, podemos dizer que quando o número de queimadas acumuladas é > 10 , o efeito que β exerce sobre o tempo estimado faz com que esse valor diminua; o que é totalmente plausível com a literatura, uma vez que as queimadas prejudicam o sistema respiratório.

Em relação à convergência dos modelos e autocorrelação dos parâmetros, estes foram verificados e suas suposições foram satisfatórias. Os correlogramas dos resíduos também mostraram resultados favoráveis.

4.2 Pontos múltiplos de mudanças

É interessante salientar que é possível detectar pontos múltiplos de mudanças para os tempos entre ocorrências de extrapolação nas internações. A partir desses dados podemos verificar quando ocorreu uma diferença significativa entre os dias de extrapolações e verificar se na época ocorreu algum tipo de intervenção, seja ela política, econômica ou climática.

Para isso, são construídos intervalos de credibilidade 95% para as diferenças médias calculadas através de (25)

$$dif_i = \theta_i - \theta_{i-1} \quad (25)$$

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

Com base nos valores de DIC e $D(j)$, conclui-se que os modelos 1 e 3 foram os que melhor se ajustaram aos dados dos tempos entre as extrapolações do número de internações por doenças respiratórias na região de Ribeirão Preto. Sendo assim, são apresentados na tabela 5 os intervalos de credibilidade 95% para as diferenças dif_i dadas em (25) que não incluem o valor zero, pois isto é um indicativo de múltiplos pontos de mudanças.

Tabela 5 – Sumários a posteriori para dif_i , dos modelos 1 e 3, que não incluem o valor zero para os intervalos de credibilidade 95% (IC 95%)

Modelo	Parâmetro	Média	IC 95%	Mudança entre os dias
<i>Modelo 1</i>	dif_7	-9,394	(-41,09 ; -0,4766)	02/03/1998 e 03/03/1998
	dif_{10}	-10,82	(-47,86 ; -0,6443)	17/03/1998 e 18/03/1998
	dif_{229}	24,98	(1,867 ; 115,7)	16/11/1998 e 17/11/1998
	dif_{242}	-12,25	(-54,38 ; -1,047)	24/01/2000 e 25/01/2000
	dif_{328}	44,14	(4,93 ; 192,00)	18/12/2000 e 19/12/2000
	dif_{342}	-8,748	(-38,6 ; -0,09691)	05/04/2001 e 06/04/2001
	dif_{374}	-9,219	(-39,13 ; -0,2057)	25/04/2001 e 26/04/2001
	dif_{401}	-12,07	(-52,41 ; -0,9752)	18/09/2001 e 19/09/2001
	dif_{506}	16,55	(1,897 ; 72,02)	29/09/2002 e 30/09/2002
	dif_{516}	-21,39	(-86,97 ; -2,022)	16/01/2003 e 17/01/2003
	dif_{520}	-10,45	(-46,08 ; -0,4866)	05/02/2003 e 06/02/2003
	dif_{527}	-8,798	(-39,26 ; -0,1515)	06/03/2003 e 07/03/2003
	dif_{549}	11,45	(0,6437 ; 50,65)	11/05/2003 e 12/05/2003
	dif_{632}	-52,71	(-192,7 ; -7,596)	29/02/2004 e 01/03/2004
	dif_{635}	12,04	(0,8973 ; 51,55)	15/03/2004 e 16/03/2004
	dif_{699}	-9,924	(-44,76 ; -0,3281)	02/08/2004 e 03/08/2004
	dif_{720}	13,71	(1,255 ; 60,75)	21/10/2004 e 22/10/2004
	dif_{725}	-54,71	(-196,8 ; -4,333)	27/02/2005 e 28/02/2005
	dif_{727}	-24,86	(-87,24 ; -4,906)	04/04/2005 e 05/04/2005
	dif_{777}	31,87	(4,631 ; 136,5)	02/10/2005 e 03/10/2005
	dif_{786}	34,52	(1,094 ; 146,9)	12/03/2006 e 13/03/2006
	dif_{860}	19,89	(1,079 ; 89,09)	19/10/2006 e 20/10/2006
<i>Modelo 3</i>	dif_2	-374,5	(-1366 ; -8,46)	04/04/2005 e 05/04/2005
	dif_{52}	29,55	(4,564 ; 123,7)	02/10/2005 e 03/10/2005
	dif_{61}	32,38	(1,049 ; 129,5)	12/03/2006 e 13/03/2006
	dif_{135}	18,60	(1,307 ; 77,11)	19/10/2006 e 20/10/2006

Na figura 8, temos o gráfico dos estimadores de Monte Carlo para as médias a posteriori das diferenças dif_i , que foram obtidas através de amostras simuladas de Gibbs, usando o *software* WinBUGS, $i = 1, 2, \dots, n$.

Figura 8 – Gráfico para dif_i versus i

Embora sejam apresentados, na tabela 5, apenas os resultados dos modelos 1 e 3, os modelos 2 e 4 também apresentaram os mesmos pontos de mudanças dos modelos 1 e 3, porém com algumas variações nas estimativas.

Com base na tabela 5 podemos identificar um aumento no tempo entre extrapolações no dif_{506} , este pode ser uma consequência da promulgação da Lei n.º 11.241, de 19 de setembro de 2002 de São Paulo, que a saber dispõe sobre a eliminação da queima da palha de cana-de-açúcar. Pode ser um indício de a lei surtiu o efeito proposto nos dias consecutivos à publicação, pois a estimativa de dif_{506} é positiva, e sabemos que o tempo anterior à extrapolação é menor e em seguida aumenta, mostrando que após um pequeno período sem extrapolações do número de internações, a próxima extrapolação ocorreu após um longo período.

5 CONSIDERAÇÕES FINAIS

O Brasil cultiva cana-de-açúcar desde os tempos da colonização. Os portugueses trouxeram a planta a fim de levar açúcar para a Europa e, desde então, o cultivo foi expandido para o nordeste do país devido a proximidade com a Europa, porém não deixou de ser cultivado no sudeste, onde teve origem o primeiro engenho de açúcar, mais precisamente em São Vicente, estado de São Paulo (SACHS; MARTINS, 2007).

O plantio da cana-de-açúcar teve sua expansão em 1975 com a crise do petróleo e com o lançamento do Programa Nacional do Álcool (PROÁLCOOL), que teve como objetivo incentivar a utilização do álcool procedente da cana-de-açúcar como combustível automotivo. Um novo impulso ao PROÁLCOOL surgiu após a crise do açúcar e do petróleo em 1980, que coincidiu com a popularização dos carros totalmente movidos a álcool, os quais até então eram movidos à gasolina, e o álcool era utilizado em apenas 22% na mistura final do combustível (NITSCH, 1991).

Entretanto, com o aumento das plantações de cana-de-açúcar chegam os problemas relacionados às queimadas, que têm por objetivo, eliminar a palha para facilitar a colheita manual e espantar animais peçonhentos. Porém, não só os trabalhadores são afetados pela fuligem e pelos componentes nocivos à saúde, mas toda a comunidade que reside em torno das áreas cultivadas (MAGALHÃES; BRUNS; VASCONCELLOS, 2007).

Uma das relações mais evidentes e encontradas na literatura é a associação das queimadas com o aumento do número de internações por doenças respiratórias, principalmente as ocorrências em crianças e idosos (CANÇADO, 2003; NASCIMENTO; MEDEIROS, 2012; RIGUERA; ANDRE; ZANETTA, 2011).

No entanto, os tempos entre extrapolações do número de internações são pouco estudados, e neste trabalho propomos novos modelos para estes tempos em relação às doenças respiratórias na região de Ribeirão Preto, além de fazermos uma associação com a presença de queimadas. Sob o enfoque Bayesiano, obtemos as estimativas a posteriori dos parâmetros de interesse por métodos de simulações via MCMC com o auxílio do *software*

WinBUGS. Os resultados dos modelos que não incorporaram a covariável mostraram estimativas mais precisas, porém, com a adição da informação das queimadas, os modelos obtiveram um grande ganho em relação à sua qualidade, visto que os valores de DIC para estes modelos abaixaram significativamente.

Os modelos que foram ajustados pela covariável X , sugerem evidências de que as queimadas de cana-de-açúcar exercem uma influência no tempo entre extrapolações do número de internações, ou seja, a presença de 11 ou mais queimadas nos últimos 10 dias que antecedem a extrapolação, leva a um menor tempo entre as extrapolações do número de internações hospitalares por doenças respiratórias na região de Ribeirão Preto, isso mostra que medidas contra as queimadas devem ser mantidas e fiscalizadas pelos órgãos competentes para uma diminuição no número de internações, uma vez que as queimadas podem contribuir para o aumento deste número.

É importante salientar que devemos interpretar o efeito das queimadas com certa cautela, pois outros fatores não controlados neste estudo poderiam interferir ou mesmo confundir estes efeitos, como citado por Carneseca, Achcar e Martinez (2012), que também mostraram um possível efeito de confundimento quando associaram o número de nebulizações/inalações com o MP e os efeitos climáticos.

Embora os fatores de efeito autorregresivo não sugiram um grande efeito sobre os tempos estimados, estes fatores são indispensáveis nos ajustes dos modelos, pois são de grande importância no controle da autocorrelação dos resíduos.

Esses novos modelos introduzidos na seção 3 apresentam a presença de efeitos aleatórios, o que leva à grande flexibilidade de ajuste e permite estimar de forma satisfatória, múltiplos pontos de mudança, pois grande parte da variabilidade é explicada pelos efeitos aleatórios.

A detecção desses pontos de mudança é de grande valia num sistema que deseja verificar se um método de intervenção está funcionando da forma correta e esperada, pois podemos detectar, por exemplo, se uma determinada lei que foi promulgada está surtindo o efeito a que se propõe, por exemplo, diminuir o número de queimadas, e consequentemente

diminuir o número de internações.

A metodologia proposta neste trabalho poderia ser comparada com outras metodologias existentes na literatura, em geral envolvendo processos de Poisson não-homogêneos para a contagem dos dias em que ocorreram extrapolações do número de internações por doenças respiratórias (ou qualquer outra doença), mas a simplicidade da metodologia proposta neste trabalho pode ser um fato de grande interesse na análise de dados da saúde.

Observar que a presença de efeitos aleatórios nos modelos propostos pode trazer dificuldade utilizando-se de métodos estatísticos clássicos usuais baseados na teoria assintótica de estimadores de máxima verossimilhança.

Dessa forma, o uso de métodos Bayesianos via simulação MCMC pode ser uma alternativa eficaz para ajustar esses modelos aos dados de internações hospitalares, uma vez que não é necessário especificar as distribuições condicionais para o amostrador de Gibbs, o método se torna de fácil implementação por parte dos profissionais da saúde.

O uso de programas computacionais disponíveis e de domínio público, como o *software* WinBUGS simplifica a obtenção dos sumários a posteriori de interesse.

Um grande diferencial deste estudo foi avaliar a relação entre internações e queimadas através das informações dos hospitais da rede pública e privada, enquanto que vários estudos que relacionam internações com queima de biomassa ou poluição apresentam apenas dados do SUS (AMANCIO; NASCIMENTO, 2012; JASINSKI; PEREIRA; BRAGA, 2011; SILVA et al., 2010).

Como não há estudos no Brasil relacionando os tempos entre extrapolações do número de internações, faremos uma comparação com a literatura, cujos estudos relacionam o número de internações e as queimadas, que é análogo às interpretações que fazemos acerca dos resultados obtidos neste estudo.

Os achados neste trabalho corroboram a literatura, ou seja, podemos sugerir um efeito das queimadas da palha da cana-de-açúcar com o aumento no número de internações por doenças respiratórias, assim como em Lopes e Ribeiro (2006).

Os resultados aqui descritos também poderiam ser comparados aos resultados da literatura em relação à queima de biomassa na Amazônia, porém é necessário ter certo cuidado, pois as exposições em relação ao indivíduo podem ser diferentes, uma vez que o número de indústrias e a poluição proveniente dos veículos automotores é uma característica que destoa da região Sudeste (CARMO et al., 2010), principalmente da região nordeste do estado de São Paulo.

Um grande problema que pode interferir na relação entre os tempos e as queimadas de cana-de-açúcar são as queimadas urbanas, que são de difícil identificação por satélite, pois muitas vezes estas não são registradas pelo Banco de Dados Queimadas do CPTEC/INPE porque o satélite pode não detectar áreas pouco extensas. Neste caso, sugerimos que campanhas contra as queimadas urbanas sejam feitas para conscientizar a população sobre os malefícios que as queimadas causam à sua própria saúde, ao meio ambiente e até à economia, quando o governo tem que investir em atendimento de saúde, quando este poderia ser evitado. Este investimento, por sua vez, está deixando de ser aplicado no atendimento em outras áreas e especialidades que demandariam maior atenção no momento.

Não poderíamos deixar de citar o lado dos agricultores, pois muito se fala da extinção das queimadas, mas há poucos estudos que avaliam o impacto negativo que a extinção das queimadas pode trazer para os trabalhadores. A colheita da cana-de-açúcar deve ser totalmente mecanizada até 2017, mas e os trabalhadores, quais são suas perspectivas? O estudo de Ribeiro e Ficarelli (2010) na cidade de Macatuba, estado de São Paulo, mostra que apenas metade dos entrevistados apresenta planos alternativos de trabalho e a outra metade mostrou disposição para retornar aos estudos. Porém, a oferta de cursos profissionalizantes gratuitos é escassa na cidade de Macatuba, e isto dificulta que estes trabalhadores tenham uma recolocação no mercado de trabalho. Mesmo assim, alguns acharam bom o fato da erradicação das queimadas por melhorar a sua saúde e qualidade de vida, mas principalmente a de seus filhos, pois são as crianças que sofrem mais.

Dessa forma, conclui-se que os modelos apresentados podem ser bastante úteis na elaboração de políticas de saúde pública e planos de intervenção, como combate às queimadas agrícolas e urbanas; remanejamento no número de leitos hospitalares para melhor atendi-

mento da demanda, uma vez que, sabendo-se das condições climáticas e surgimento de queimadas, é possível estabelecer que, de certo em certo tempo, o número de internações pode ser elevado.

Outro aspecto que certamente pode ser evidenciado é a que as queimadas prejudicam a qualidade de vida das pessoas afetadas, pode não estar direta e unicamente associada às internações, mas causa muitos transtornos no dia a dia, como a sujeira da “fuligem” que entra nas casas, a fumaça que atrapalha o trânsito nas rodovias e a sua contribuição para a mudança climática, que prejudica não só a região onde a queimada é observada, mas que pode se estender a um contexto global.

REFERÊNCIAS

ALBERT, J. **Bayesian computation with R**. New York: Springer, 2007.

AMANCIO, C. T.; NASCIMENTO, L. F. C. Asma e poluentes ambientais: um estudo de séries temporais. **Revista da Associação Médica Brasileira**, São Paulo, v. 58, n. 3, jun. 2012. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-42302012000300009&lng=en&nrm=iso>. Acesso em: 11 out 2012.

ARBEX, M. A. **Avaliação dos efeitos do material particulado proveniente da queima da plantação de cana-de-açúcar sobre a morbidade respiratória na população de Araraquara-SP**. 2001. 188f. Tese (Doutorado) - Faculdade de Medicina, Universidade de São Paulo, São Paulo, 2001.

ARBEX, M. A.; CANÇADO, J. R. D.; PEREIRA, L. A. A.; BRAGA, A. L. F.; SALDIVA, P. H. N. Queima de biomassas e efeito sobre a saúde. **Jornal Brasileiro de Pneumologia**, São Paulo, v. 30, n. 2, p. 158-175, mar./abr. 2004.

ACHCAR, J. A.; SICCHIERI, M. P. L. S. Comparação de modelos para os excessos de internações hospitalares diárias por pneumonia na cidade de São Paulo. **Revista Brasileira de Biometria**, São Paulo, v. 28, n. 1, p. 87-103, 2010.

BAIRD, C. **Química ambiental**. 2. ed. Porto Alegre: Bookman, 2002.

Banco de Dados Queimadas (BDQUEIMADAS). Disponível em:<<http://www.dpi.inpe.br/proarco/bdqueimadas/>>. Acesso em: 03 set. 2010.

BERNARDO, J. M.; SMITH, A. F. M. **Bayesian theory**. Chichester: Wiley, 1994.

BOX, G. E. P.; TIAO, G. C. **Bayesian inference in statistical analysis**. New York: Addison-Wesley, 1973.

CANÇADO, J. E. D. **A poluição atmosférica e sua relação com a saúde humana na região canavieira de Piracicaba SP**. 2003. 201 f. Tese (Doutorado em Ciências) - Faculdade de Medicina, Universidade de São Paulo, São Paulo, 2003.

CARLIN, B. P.; LOUIS, T. A. **Bayes and empirical Bayes methods for data analysis**. 2nd ed. Boca Raton: Chapman and Hall, 2000.

CARMO, C. N.; HACON, S.; LONGO, K. M.; FREITAS, S.; IGNOTTI, E.; PONCE DE LEON, A.; ARTAXO P. Associação entre material particulado de queimadas e doenças respiratórias na região sul da Amazônia brasileira. **Revista Panamericana de Salud Pública**, Washington, v. 27, n. 1, p. 106, 2010.

CARNESECA, E. C.; ACHCAR, J. A.; MARTINEZ, E. Z. Association between particulate matter air pollution and monthly inhalation and nebulization procedures in Ribeirão Preto, São Paulo State, Brazil. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 28, n. 8, ago. 2012. Disponível em: <http://www.scielo.org/scielo.php?script=sci_arttext&pid=S0102-311X2012000800017&lng=en&nrm=iso>. Acesso em: 11 out. 2012.

CENTRO DE PROCESSAMENTO DE DADOS HOSPITALARES (CPDH). Faculdade de Medicina de Ribeirão Preto. Universidade de São Paulo. 2010.

COMPANHIA DE TECNOLOGIA DE SANEAMENTO AMBIENTAL (CETESB). **Qualidade do ar no estado de São Paulo 2010**. Disponível em: <<http://www.cetesb.sp.gov.br/ar/qualidade-do-ar/31-publicacoes-e-relatorios>>. Acesso em: 05 set. 2011.

CHIB, S.; GREENBERG, E. Understanding the Metropolis-Hastings algorithm. **The American Statistician**, v. 49, n. 4, p. 327-335, 1995.

COLOSIMO, E. A.; GIOLO, S. R. **Análise de sobrevivência aplicada**. São Paulo: ABE - Projeto Fisher, 2006.

COX, D. R.; OAKES, R. **Analysis of survival data**. London: Chapman & Hall, 1984.

FREDO, C. E., VICENTE, M. C. M., BAPTISTELLA, C. S.L., VEIGA, J. E. R. Índice de mecanização na colheita da cana-de-açúcar no estado de São Paulo e nas regiões produtoras paulistas, junho de 2007. **Análises e Indicadores do Agronegócio**, v. 3, n. 3., p. - , mar. 2008.

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.

GEMAN, S.; GEMAN, D. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 6, p. 721-741. 1984.

GELMAN, A.; RUBIN, D.R. Inference from iterative simulation using multiple sequences. **Statistical Science**, v. 7, n. 4, p. 457-472, 1992.

GODOY, D. V.; DAL ZOTTO, C.; BELLICANTA, J.; WESCHENFELDER, R. F.; NACIF, S. B. Doenças respiratórias como causa de internações hospitalares de pacientes do Sistema Único de Saúde num serviço terciário de clínica médica na região nordeste do Rio Grande do Sul. **Jornal de Pneumologia**, São Paulo, v. 27, n. 4, p. 193-198, jul./ago. 2001.

HASTINGS, W. K. Monte Carlo sampling methods using Markov chains and their applications. **Biometrika**, v. 57, p. 97-109, 1970.

IBRAHIM, J. G.; CHEN, M.-H.; SINHA, D. **Bayesian survival analysis**. New York: Springer Verlag, 2001.

JASINSKI, R.; PEREIRA, L. A. A.; BRAGA, A. L. F. Poluição atmosférica e internações hospitalares por doenças respiratórias em crianças e adolescentes em Cubatão, São Paulo, Brasil, entre 1997 e 2004. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 27, n. 11, nov. 2011. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2011001100017&lng=en&nrm=iso>. Acesso em: 11 out. 2012.

LAWLESS, J. F. **Statistical models and methods for lifetime data**. New York: Wiley, 1982.

LEE, E. T. **Statistical methods for survival data analysis**. 2nd ed. New York: John Wiley and Sons, 1992.

LOPES, F. S.; RIBEIRO, H. Mapeamento de internações hospitalares por problemas respiratórios e possíveis associações à exposição humana aos produtos da queima da palha de cana-de-açúcar no estado de São Paulo. **Revista Brasileira de Epidemiologia**, São Paulo, v. 9, n. 2, p. 215-225, 2006.

LUNN, D. J., THOMAS, A., BEST, N., and SPIEGELHALTER, D. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. **Statistics and Computing**, v. 10, p. 325-337, 2000.

MAGALHÃES, D. **Determinação de compostos orgânicos em material particulado emitido em queimadas de cana de açúcar**. 2005. 65 f. Dissertação (Mestrado em Tecnologia Nuclear e Materiais) - Instituto de Pesquisas Energéticas e Nucleares, São Paulo, 2005.

MAGALHÃES, D.; BRUNS, R. E.; VASCONCELLOS, P. C. Hidrocarbonetos policíclicos aromáticos como traçadores da queima de cana-de-açúcar: uma abordagem estatística. **Química Nova**, São Paulo, v. 30, n. 3, jun. 2007. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-40422007000300014&lng=en&nrm=iso>. Acesso em: 13 out. 2012.

MARTINS, L. C.; LATORRE, M. R. D. O.; CARDOSO, M. R. A.; GONÇALVES, F. L. T.; SALDIVA, P. H. N.; BRAGA, A. L. F. Poluição atmosférica e atendimentos por pneumonia e gripe em São Paulo, Brasil. **Revista de Saúde Pública**, v. 36, n. 1, p. 88-94, 2002.

METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H., TELLER, E. Equation of state calculations by fast computing machines. **Journal of Chemical Physics**, vol. 21, p. 1087-1092, 1953

Ministério da Saúde. Secretaria Executiva. DATASUS. Informações de Saúde. Morbidade Hospitalar do SUS por local de internação. Disponível em: <<http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sih/cnv/nisp.def>> Acesso em: 14 out. 2009.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of statistics**. 3rd ed. New York: McGraw-Hill, 1974.

NASCIMENTO, L. F. C.; MEDEIROS, A. P. P. Admissions due to pneumonia and

biomass burning: a spatial approach. **Jornal de Pediatria (Rio J.)**, Porto Alegre, v. 88, n. 2, abr. 2012. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0021-75572012000200014&lng=en&nrm=iso>. Acesso em: 13 out. 2012.

NITSCH, M. O programa de biocombustíveis Proálcool no contexto da estratégia energética brasileira. **Revista de Economia Política**, São Paulo, v. 11, n. 2, p. 123-138, abr./jun. 1991.

ORGANIZAÇÃO MUNDIAL DA SAÚDE (OMS). **Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde (CID-10)**. 10. Rev. São Paulo: OPS/OMS/ EdUSP, 1998. v. 1.

PAULINO, C. D.; TURKMAN, M. A. A. ; MURTEIRA, B. **Estatística Bayesiana**. Lisboa: Fundação Calouste Gulbenkian, 2003.

RIBEIRO, H.; FICARELLI, T. R. A. Queimadas nos canaviais e perspectivas dos cortadores de cana-de-açúcar em Macatuba, São Paulo. **Saude e Sociedade.**, São Paulo, v. 19, n. 1, mar. 2010. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-12902010000100005&lng=en&nrm=iso>. Acesso em: 11 out. 2012.

RIGUERA, D.; ANDRE, P. A.; ZANETTA, D. M. T. Poluição da queima de cana e sintomas respiratórios em escolares de Monte Aprazível, SP. **Revista de Saúde Pública**, São Paulo, v. 45, n. 5, out. 2011. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-89102011000500009&lng=en&nrm=iso>. Acesso em: 13 out. 2012.

ROSEIRO, M. N. V.; TAKAYANAGUI, A. M. M. Morbidade por problemas respiratórios em Ribeirão Preto (SP) de 1995 a 2001, segundo indicadores ambientais, sociais e econômicos. **Revista Paulista de Pediatria**, v. 24, n. 2, p. 163-170, jun. 2006.

SACHS, R. C. C.; MARTINS, V. A. Análise da cultura da cana-de-açúcar, por escritório de desenvolvimento rural, estado de São Paulo, 1995-2006. **Informações Econômicas**, São Paulo, v. 37, n. 9, p. 41-52, 2007. Disponível em <<http://www.iea.sp.gov.br/out/LerTexto.php?codTexto=9071>>. Acesso em: 13 out. 2012.

SILVA, A. M. C.; MATTOS, I. E.; FREITAS, S. R.; LONGO, K. M.; HACON, S. S. Material particulado (PM_{2.5}) de queima de biomassa e doenças respiratórias no sul da Amazônia

nia brasileira. **Revista Brasileira de Epidemiologia**, São Paulo, v. 13, n. 2, jun. 2010. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-790X2010000200015&lng=en&nrm=iso>. Acesso em: 11 out. 2012.

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.

SPIEGELHALTER, D. J.; THOMAS, A.; BEST, N. G.; GILKS, W. R. **BUGS**: Bayesian inference Using Gibbs Sampling, Version 0.50. Cambridge: MRC Biostatistics Unit. 1995.

TOYOSHIMA, M. T. K., ITO, G. M., GOUVEIA, N. Morbidade por doenças respiratórias em pacientes hospitalizados em São Paulo/SP. **Revista da Associação Médica Brasileira**, São Paulo, v. 51, n. 4, p. 209-213, jul./ago 2005.

TURN, S. Q.; JENKINS, B. M.; CHOW, J. C., PRITCHETT, L. C.; CAMPBELL, D., CAHILL, T.; WHALEN, S. A. Elemental characterization of particulate matter emitted from biomass burning: Wind tunnel derived source profiles for herbaceous and wood fuels. **Journal of Geophysical Research**, v. 102, n. D3, p. 3683-3699, 1997.

ZANCUL, A. **O efeito da queimada de cana-de-açúcar na qualidade do ar na região de Araraquara**. 1998. 99f. Dissertação (Mestrado em Hidráulica e Saneamento) - Escola de Engenharia de São Carlos, São Carlos, 1998.

APÊNDICE A - Programas

Nesse apêndice são apresentados os programas computacionais utilizados nesta dissertação de mestrado.

A.1 - Modelos para tempos entre extrapolações

A.1.1 - Modelo 1

Listagem 1 – Programa Desenvolvido no software *WinBUGS* (Modelo 1)

```

1  model
2  {
3    t[1] ~ dexp(lambda[1])
4    lambda[1] <- w[1]
5    w[1] ~ dunif(0,100)
6    theta[1] <- 1/lambda[1]
7
8    t[2] ~ dexp(lambda[2])
9    lambda[2] <- (phi1*t[1])*(alpha*(N-1))+w[2]
10   w[2] ~ dunif(0,100)
11   theta[2] <- 1/lambda[2]
12
13   for (i in 3:n)
14   {
15     t[i] ~ dexp(lambda[i])
16     lambda[i] <- (phi1*t[i-1]+phi2*t[i-2])*(alpha*(N-i+1))+w[i]
17     theta[i] <- 1/lambda[i]
18     w[i] ~ dunif(0,100)
19   }
20
21   alpha ~ dunif(0,1)
22   N ~ dunif(948,100000)
23   phi1 ~ dbeta(1,1)
24   phi2 ~ dbeta(1,1)
25
26   for (i in 2:n)
27   {
28     dif[i] <- theta[i]-theta[i-1]
29   }
30 }
```

A.1.2 - Modelo 2

Listagem 2 – Programa Desenvolvido no software *WinBUGS* (Modelo 2)

```

1 model
2 {
3   t[1] ~ dexp(lambda[1])
4   lambda[1] <- w[1]
5   w[1] ~ dunif(0,100)
6   theta[1] <- 1/lambda[1]
7
8   t[2] ~ dexp(lambda[2])
9   lambda[2] <- (phi1*t[1])*(alpha*(N+2))+w[2]
10  w[2] ~ dunif(0,100)
11  theta[2] <- 1/lambda[2]
12
13  t[3] ~ dexp(lambda[3])
14  lambda[3] <- (phi1*t[2]+phi2*t[1])*(alpha*(N+3))+w[3]
15  w[3] ~ dunif(0,100)
16  theta[3] <- 1/lambda[3]
17
18  for (i in 4:n)
19  {
20    t[i] ~ dexp(lambda[i])
21    lambda[i] <- (phi1*t[i-1]+phi2*t[i-2]+phi3*t[i-3])*(alpha*(N+i))+w[i]
22    theta[i] <- 1/lambda[i]
23    w[i] ~ dunif(0,100)
24  }
25
26  alpha ~ dunif(0,10)
27  N ~ dunif(0,100000)
28  phi1 ~ dbeta(1,1)
29  phi2 ~ dbeta(1,1)
30  phi3 ~ dbeta(1,1)
31
32  for (i in 2:n)
33  {
34    diff[i] <- theta[i]-theta[i-1]
35  }
36 }

```

A.2 - Modelos para tempos entre extrapolações e queimadas

A.2.1 - Modelo 3

Listagem 3 – Programa Desenvolvido no software *WinBUGS* (Modelo 3)

```

1 model
2 {
3   t[1] ~ dexp(lambda[1])
4   lambda[1] <- w[1]
5   w[1] ~ dunif(0,100)
6   theta[1] <- 1/lambda[1]
7
8   t[2] ~ dexp(lambda[2])
9   lambda[2] <- (phi1*t[1])*(alpha*(N-1)+beta*x[2])+w[2]
10  w[2] ~ dunif(0,100)
11  theta[2] <- 1/lambda[2]
12
13  for (i in 3:n)
14  {
15    t[i] ~ dexp(lambda[i])
16    lambda[i] <- (phi1*t[i-1]+phi2*t[i-2])*(alpha*(N-i+1)+beta*x[i])+w[i]
17    theta[i] <- 1/lambda[i]
18    w[i] ~ dunif(0,100)
19  }
20
21  alpha ~ dunif(0,1)
22  beta ~ dunif(0,10)
23  N ~ dunif(948,100000)
24  phi1 ~ dbeta(1,1)
25  phi2 ~ dbeta(1,1)
26
27  for (i in 2:n)
28  {
29    diff[i] <- theta[i]-theta[i-1]
30  }
31 }

```

A.2.2 - Modelo 4

Listagem 4 – Programa Desenvolvido no software *WinBUGS* (Modelo 4)

```

1 model
2 {
3   t[1] ~ dexp(lambda[1])
4   lambda[1] <- w[1]
5   w[1] ~ dunif(0,100)
6   theta[1] <- 1/lambda[1]
7
8   t[2] ~ dexp(lambda[2])
9   lambda[2] <- (phi1*t[1])*(alpha*(N+2)+beta*x[2])+w[2]
10  w[2] ~ dunif(0,100)
11  theta[2] <- 1/lambda[2]
12
13  t[3] ~ dexp(lambda[3])
14  lambda[3] <- (phi1*t[2]+phi2*t[1])*(alpha*(N+3)+beta*x[3])+w[3]
15  w[3] ~ dunif(0,100)
16  theta[3] <- 1/lambda[3]
17
18  for (i in 4:n)
19  {
20    t[i] ~ dexp(lambda[i])
21    lambda[i] <- (phi1*t[i-1]+phi2*t[i-2]+phi3*t[i-3])*(alpha*(N+i)+beta*x[i])+w[i]
22    theta[i] <- 1/lambda[i]
23    w[i] ~ dunif(0,100)
24  }
25
26  alpha ~ dunif(0,10)
27  N ~ dunif(0,100000)
28  phi1 ~ dbeta(1,1)
29  phi2 ~ dbeta(1,1)
30  phi3 ~ dbeta(1,1)
31  beta ~ dunif(0,10)
32
33  for (i in 2:n)
34  {
35    diff[i] <- theta[i]-theta[i-1]
36  }
37 }

```

APÊNDICE B - Gráficos de convergência

Nesse apêndice são apresentados os os gráficos de convergência das três cadeias independentes geradas para a simulação de amostras.

B.1 - Gráficos para os modelos dos tempos entre extrapolações

B.1.1 - Modelo 1

Figura 9 – Gráficos de Gelman-Rubin para os parâmetros do *Modelo 1*

B.1.2 - Modelo 2

(a) α (b) N (c) ϕ_1 (d) ϕ_2 (e) ϕ_3 Figura 10 – Gráficos de Gelman-Rubin para os parâmetros do *Modelo 2*

B.2 - Gráficos para os modelos dos tempos entre extrapolações e queimadas

B.2.1 - Modelo 3

Figura 11 – Gráficos de Gelman-Rubin para os parâmetros do *Modelo 3*

B.2.2 - Modelo 4

(a) α (b) N (c) ϕ_1 (d) ϕ_2 (e) ϕ_3 (f) β Figura 12 – Gráficos de Gelman-Rubin para os parâmetros do *Modelo 4*

APÊNDICE C - Correlogramas dos resíduos

Figura 13 – Correlogramas dos resíduos de cada modelo proposto

APÊNDICE D - Artigo Publicado

Revista Brasileira de Biometria

COMPARAÇÃO DE MODELOS PARA OS EXCESSOS DE INTERNAÇÕES HOSPITALARES DIÁRIAS POR PNEUMONIA NA CIDADE DE SÃO PAULO

Jorge Alberto ACHCAR¹

Mayara Piani Luna da Silva SICCHIERI¹

- **RESUMO:** Neste artigo, 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.

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

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).

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.

Para a modelagem dos dados de contagens diárias das internações hospitalares devido à pneumonia, podemos considerar diferentes modelagens estatísticas: o uso de processos de Poisson homogêneos ou não homogêneos (ver por exemplo, Javits, 1980; Raftery, 1989; Achcar *et al.*, 2008); o uso de modelos de séries temporais (ver por exemplo, Flaum *et al.*, 1996); o uso de modelos de Cadeias de Markov (ver por exemplo Piegorsch *et al.*, 1998) entre várias outras modelagens.

Neste artigo, 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 (dados na Tabela 1).

Tabela 1 - Tempos entre ocorrências de excessos de internações hospitalares (≥ 70)

105	14	1	1	1	4	1	2	4	1	1	1	1	3	1	3
3	1	6	2	1	1	3	1	2	4	2	1	4	4	4	14
6	5	24	6	17	25	7	1	7	13	21	112	3	6	1	2
2	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1
1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	2
1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	2	1	1	2	1	1	1	1	1	2	1	1
1	1	3	1	1	1	1	3	1	78	36	11	1	3	4	3
4	3	14	63	5	1	2	5	1	1	1	1	1	1	2	1
1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1
1	1	1	1	1	1	2	1	1	1	1	2	1	1	1	1
1	1	2	1	1	1	1	1	2	1	1	1	1	1	1	1
1	1	3	2	1	1	1	1	3	1	1	1	1	1	1	1
1	1	1	1	1	2	1	56	1	14	1	4	1	1	1	4
2	2	4	1	5	2	1	1	3	2	1	1	3	1	6	1
3	6	14	4	7	7	1	10	10	59	4	7	3	1	1	1
1	1	2	2	2	2	1	2	2	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	3	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1
1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	2
1	1	1	1	1	1	1	1	1	1	1	1	3	1	1	1
1	1	2	1	1	1	4	1	1	1	1	1	2	1	2	1
3	3	4	2	1	7	4	1	1	1	4	2	5	1	1	1
4	14	35	4	3	7	7	1	1	2	3	3	11			

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).

O artigo 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 $i = 2, 3, \dots, n$ e 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).

É importante salientar que o “modelo 1” e as generalizações desse modelo que serão introduzidas a seguir são novos modelos propostos neste artigo para modelar excessos de internações hospitalares e não são apresentados na literatura.

Um segundo modelo proposto para a função de risco, é dado por,

$$\lambda_1 = N\alpha e^{\eta_1}$$

$$\lambda_i = (N - i + 1) \alpha \exp \{ \phi (\eta_{i-1} - \alpha) \} \quad (4)$$

onde $0 < \alpha < \infty$, pois a taxa de risco λ_i deve ser positiva e $0 < \phi < 1$, para $i = 2, 3, \dots, n$ e onde η_i é um fator aleatório com distribuição normal dada em (3); observar que neste caso, introduzimos o termo $(N-i+1)$ que leva a um decaimento da taxa ou função de risco do início ao término do período de tempo observado. Vamos denotar o modelo definido por (1), (3) e (4) como “modelo 2”.

Observar que o parâmetro N em (4) pode ser interpretado como um número (desconhecido) total de excessos de internações hospitalares num período de tempo muito grande.

Um terceiro modelo proposto é dado por,

$$\lambda_1 = N\alpha e^{\eta_1}$$

$$\lambda_i = [(N - i + 1) \alpha + \phi_2 t_{i-1}] \exp \{ \phi_1 (\eta_{i-1} - \alpha) \} \quad (5)$$

onde $0 < \alpha < \infty$; $0 < \phi_1, \phi_2 < 1$; o termo $\phi_2 t_{i-1}$ fornece uma contribuição auto-regressiva do tempo prévio na taxa de excessos de internações hospitalares devido à pneumonia, para $i = 2, 3, \dots, n$ e onde η_i é um efeito aleatório com distribuição normal dada em (3). Vamos denotar este modelo como “modelo 3”.

Outro modelo, denotado por “modelo 4”, é dado por,

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

$$\lambda_i = \alpha (N - i + 1) + w_i + \phi t_{i-1} \quad (6)$$

onde $0 < \alpha < \infty$; $0 < \phi < 1$; $i = 2, 3, \dots, n$; w_i é um efeito aleatório, com uma distribuição dada por,

$$w_i \sim Gama(b, c) \quad (7)$$

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 (7); uma alternativa seria considerar b e c como quantidades aleatórias e usar métodos Bayesianos hierárquicos.

Um outro modelo denotado por “modelo 5” é considerado retirando-se o termo ϕt_{i-1} em (6), isto é,

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

$$\lambda_i = \alpha (N - i + 1) + w_i \quad (8)$$

onde $0 < \alpha < \infty$, $i = 2, 3, \dots, n$ e w_i tem distribuição Gama dada em (7).

Observar que nos modelos 2, 3, 4 e 5, consideramos o termo $(N - i + 1)$ na modelagem da função de risco λ_i , $i = 1, \dots, n$.

Uma alternativa seria assumir o termo $(N + i)$ em lugar do termo $(N - i + 1)$ o que implica em taxas de riscos λ_i crescentes, isto é, tempos médios θ_i decrescentes; dessa forma, consideramos o “modelo 6” dado por,

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

$$\lambda_i = \alpha (N + i) + w_i \quad (9)$$

onde $0 < \alpha < \infty$, $i = 2, 3, \dots, n$ e w_i é um efeito aleatório com distribuição Gama dada em (7).

Observar que considerando cada um dos modelos 1 a 6, a função de verossimilhança é dada por,

$$L = \prod_{i=1}^n \lambda_i e^{-\lambda_i t_i}. \quad (10)$$

3 Uma análise Bayesiana

Para todos 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 diferentes modelos introduzidos na seção 2, todos considerando 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 :

$$\alpha \sim N(0, b^2) \quad \phi \sim \text{Beta}(c_1, d_1) \quad (11)$$

onde $\text{Beta}(c_1, d_1)$ 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 (11) foram consideradas levando em conta a variação dos parâmetros α e ϕ (ver (2)).

De forma similar, consideramos as distribuições a priori para os outros modelos.

É importante salientar que a escolha dos hiperparâmetros das distribuições a priori (11) 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.

A distribuição a posteriori conjunta para α, ϕ e $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)$ é obtida combinando-se (10) com (3) e (11), isto é,

$$\begin{aligned} \pi(\alpha, \phi, \boldsymbol{\eta} | \mathbf{t}) &\propto \exp\{\alpha + \eta_1 - t_1 e^{\alpha + \eta_1}\} \times \\ &\times \exp\left\{(n-1)\alpha + \phi \sum_{i=2}^n (\eta_{i-1} - \alpha) - \sum_{i=1}^n t_i \exp[\alpha + \phi(\eta_{i-1} - \alpha)]\right\} \times \\ &\times \exp\left\{-\frac{1}{2a^2} \sum_{i=1}^n \eta_i^2 - \frac{\alpha^2}{2b^2}\right\} \phi^{c_1-1} (1-\phi)^{d_1-1} \end{aligned} \quad (12)$$

onde $\mathbf{t}=(t_1, \dots, t_n)$; $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)$, $-\infty < \alpha < \infty$, $-\infty < \eta_i < \infty$ e $0 < \phi < 1$.

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

$$\begin{aligned} \phi &\sim U(c_2, d_2) \\ \alpha &\sim \text{Gama}(e_2, f_2) \\ N &\sim U(g_2, h_2) \end{aligned} \quad (13)$$

onde $U(c_2, d_2)$ denota uma distribuição uniforme no intervalo (c_2, d_2) ; c_2, d_2, e_2, f_2, g_2 e h_2 são hiperparâmetros conhecidos.

Para o “modelo 3”, consideramos as mesmas distribuições a priori para α e N consideradas em (13) e distribuições a priori uniforme para os parâmetros ϕ_1 e ϕ_2 com hiperparâmetros conhecidos.

Considerando o “modelo 4”, assumimos as seguintes distribuições a priori para os parâmetros do modelo:

$$\begin{aligned}
\phi &\sim \text{Beta}(m_4, n_4) \\
\alpha &\sim U(o_4, p_4) \\
N &\sim U(g_2, h_2)
\end{aligned} \tag{14}$$

Considerando os modelos 5 e 6, assumimos as mesmas distribuições a priori consideradas no “modelo 4”.

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

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 (11) com $b = 1, c_1 = d_1 = 1$. Observar que $\text{Beta}(1, 1)$ corresponde a uma distribuição uniforme no intervalo $(0, 1)$. Dessa forma estamos considerando distribuições aproximadamente não-informativas. Na Tabela 2, 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 2 - 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)

Na Figura 2, temos os traços das cadeias das amostras simuladas da posteriori dos parâmetros obtidos do *software* WinBugs e considerando três valores iniciais

para os parâmetros. Dos gráficos da Figura 2, observamos a convergência do algoritmo. Também temos na Figura 2, os gráficos de autocorrelações para as cadeias simuladas. Observamos que as amostras de Gibbs simuladas são aproximadamente não-correlacionadas (ver progama WinBugs no apêndice 2).

Figura 2 - Traços das cadeias das amostras de Gibbs simuladas e gráficos de autocorrelações.

Assumindo o “modelo 2” com as distribuições a priori dadas em (13) com $c_2 = 0, d_2 = 10; e_2 = 1, f_2 = 1000; g_2 = 445$ e $h_2 = 100000$, temos na Tabela 3, os sumários a posteriori obtidos. Observar que a distribuição $Gama(1, 100)$ equivale à distribuição exponencial.

Também observar que a priori especificada para ϕ , uma distribuição $Gama(1, 1000)$, implica em uma esperança igual à $1/1000$ e uma variância igual à $1/1000^2$. Essa escolha dos hiperparâmetros da priori para ϕ levou à convergência do algoritmo usando o *software* WinBugs. Possivelmente usando outra linguagem de programação isso não seja necessário.

Tabela 3 - Sumários a posteriori do “modelo 2”

Parâmetro	Média	Desvio padrão	Intervalo de Credibilidade 95%
N	32680	26690	(3067; 93240)
α	0,0000374	0,000045	(0,0000052 ; 0,000172)
ϕ	0,6355	0,0454	(0,5483 ; 0,7263)

Para o “modelo 3”, consideramos as distribuições a priori $\phi_1 \sim U(0, 10); \phi_2 \sim U(0; 0, 2)$ e $\alpha \sim Gama(1, 100)$ e $N \sim U(445, 100000)$. Na Tabela 4, temos os sumários a posteriori obtidos.

Tabela 4 - Sumários a posteriori do “modelo 3”

Parâmetro	Média	Desvio padrão	Intervalo de Credibilidade 95%
N	32330	26740	(2787 ; 93370)
α	0,0000402	0,0000498	(0,00000515 ; 0,000193)
ϕ_1	0,6392	0,04547	(0,5519 ; 0,7303)
ϕ_2	0,0005	0,000513	(0,0000126 ; 0,001846)

Considerando o “modelo 4”, assumimos duas distribuições diferentes para os efeitos aleatórios w_i : $w_i \sim Gama(1, 1)$ (denotar como modelo 4*) e $w_i \sim U(0, 100)$ (denotar como modelo 4**) e as distribuições a priori, $\phi \sim Beta(1, 1)$; $\alpha \sim U(0, 1)$ e $N \sim U(445, 100000)$. Nas Tabelas 5 e 6 temos os sumários a posteriori obtidos para cada modelo.

Tabela 5 - Sumários a posteriori do “modelo 4*”

Parâmetro	Média	Desvio padrão	Intervalo de Credibilidade 95%
N	22480	25480	(642,6 ; 89000)
α	0,00000495	0,0000085	(0,0000001 ; 0,00003)
ϕ	0,000212	0,000196	(0,00000596 ; 0,000733)

Tabela 6 - Sumários a posteriori do “modelo 4**”

Parâmetro	Média	Desvio padrão	Intervalo de Credibilidade 95%
N	2862	2525	(473,6 ; 1834)
α	0,00000355	0,00000513	(0,00000005 ; 0,0000182)
ϕ	0,000233	0,0002	(0,0000077 ; 0,000758)

Assumindo o “modelo 5”, consideramos $w_i \sim U(0, 100)$, $i = 1, \dots, n$; $\alpha \sim U(0, 1)$ e $N \sim U(445, 100000)$. Na Tabela 7, temos os sumários a posteriori de interesse considerando o “modelo 5”.

Tabela 7 - Sumários a posteriori do “modelo 5”

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

Finalmente, considerando o “modelo 6”, assumimos $w_i \sim U(0, 100)$; $\alpha \sim U(0, 1)$ e $N \sim U(0, 100000)$. Na Tabela 8, temos os sumários a posteriori obtidos usando o *software* WinBugs.

Na Tabela 9, 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.

Tabela 8 - Sumários a posteriori do “modelo 6”

Parâmetro	Média	Desvio padrão	Intervalo de Credibilidade 95%
N	15250	23260	(29,6 ; 84760)
α	0,0000049	0,0000095	(0,000000012 ; 0,000033)

Tabela 9 - Valores de DIC e $D(j)$

Modelo j	DIC	$D(j)$
Modelo 1	1647,53	1071
Modelo 2	1296,67	1090
Modelo 3	1230,21	1090
Modelo 4*	2011,18	1088
Modelo 4**	2011,18	928,7
Modelo 5	2009,4	946,9
Modelo 6	1898,83	943,4

Dos resultados dados na Tabela 9, observamos que o melhor modelo ajustado usando o critério DIC é dado pelo “modelo 3” (menor DIC estimado); usando como critério de comparação os valores das somas das diferenças $D(j)$, observamos que o “modelo 4**” é o melhor modelo ajustado pelos dados (menor valor de $D(j)$ para $j = 4$).

Na Figura 3, 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.

Dos gráficos da Figura 3, também observamos que o “modelo 6” é bem ajustado aos dados quando comparamos as médias ajustadas $\hat{\theta}_i$ com os dados observados t_i , $i = 1, 2, \dots, 445$.

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,

$$diff_i = \theta_i - \theta_{i-1} \quad (15)$$

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

Baseado em $D(j)$ e na complexidade do modelo, assumimos o “modelo 5” para efeito de conclusões. Sendo assim, temos na Tabela 10, os intervalos de credibilidade 95% para as diferenças $diff_i$ dadas em (15) que não incluem o valor zero, pois isto é um indicativo de múltiplos pontos de mudanças.

Figura 3 - Tempos entre ocorrências de excessos de interações t_i versus i e médias ajustadas pelos modelos propostos, $i = 1, \dots, 445$.

Na Figura 4, temos o gráfico dos estimadores de Monte Carlo para as médias a posteriori das diferenças $diff_i$, que foram obtidas através de amostras simuladas de Gibbs, usando o *software* WinBugs, $i = 1, 2, \dots, 445$.

De forma geral, podemos considerar o “modelo 5” mais apropriado para estimar os tempos entre excessos de interações devido à pneumonia, pois tem uma forma mais simples e seu $D(j)$ se mostrou próximo aos dos modelos 4** e 6, que apresentam maior complexidade.

Tabela 10 - Sumários a posteriori para $diff_i$, do “modelo 5”, 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)

Figura 4 - Gráfico para $diff_i$ versus i , $i = 2, \dots, 445$.

Notas Conclusivas

Neste artigo, 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 baseados 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.

Agradecimentos

Os autores agradecem aos revisores desse artigo pelas sugestões que levaram à grande melhoria do texto apresentado.

ACHCAR, J. A.; SICCHIERI, M. P. L. S. Model comparison for the excesses of daily hospitalizations due to pneumonia in the city of São Paulo. *Rev. Bras. Biom.*, São Paulo, v.28, n.1, p.87-103, 2010.

■ **ABSTRACT:** *In this paper we introduce new models to analyze counting data related to daily hospitalizations due to pneumonia in the public hospitals of the city of São Paulo in the period ranging from 01/01/2002 to 12/31/2005. To analyze this data set, we consider modeling the times between the excesses of daily hospitalizations due to pneumonia assuming an exponential distribution and different structures for the hazard function. Inferences for these models are obtained under the Bayesian paradigm and using simulation MCMC (Markov Chains Monte Carlo) methods with the help of the WinBugs software. We also obtain Bayesian inferences for the change-points of the average daily countings of hospitalizations due to pneumonia.*

■ **KEYWORDS:** *Time between excess hospitalization; Bayesian analysis; exponential distribution; hazard function modeling.*

Referências

- ACHCAR, J. A.; FERNÁNDEZ-BREMAUNTZ, A. A.; RODRIGUES, E. R.; TZINTZUN, G. Estimating the number of ozone peaks in Mexico City using a nonhomogeneous Poisson model. *Environmetrics*, Chicester, v.19, p.469-485, 2008.
- BERNARDO, J. M.; SMITH, A. F. M. *Bayesian theory*. New York: John Wiley and Sons, 1995. 608p.
- CARLIN, B. P.; LOUIS, T. A. *Bayes and empirical Bayes methods for data analysis*. 2.ed. Boca Raton: Chapman and Hall, 2000. 397p.
- CHIB, S.; GREENBERG, E. Analysis of multivariate probit models. *Biometrika*, London, v.85, n.2, p.347-361, 1998.
- COX, D.R.; HINKLEY, D. V. *Theoretical statistics*. Boca Raton: Chapman and Hall, 2000. 511p.
- FLAUM, J. B.; RAO, S. T.; ZURBENKO, I. G. Moderating influence of meteorological conditions on ambient ozone concentrations. *J. Air Waste Manage. Assoc.*, Pittsburg, v.46, p.33-46, 1996.
- GELFAND, A. E.; SMITH, A. F. M. Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.*, Alexandria, v.85, p.398-409, 1990.
- JAVITS, J. S. Statistical interdependencies in the ozone national ambient air quality standard. *J. Air Pollus Control Assoc.*, Pittsburg, v.30, p.58-59, 1980.
- PIEGORSCH, W. W.; SMITH, E. P., EDWARDS, D.; SMITH, R. L. Statistical advances in environmental sciences. *Stat. Sci.*, Hayward, v.13, p.186-208, 1998.
- RAFTERY, A. E. Air the ozone exceedance rate decreasing? Comment of the paper "Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone" by R. L. Smith. *Stat. Sci.*, Hayward, v.4, p.378-381, 1989.
- 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. *J. R. Stat. Soc., Ser. B Methodol.*, Londres, v.64, p.583-639, 2002.

Recebido em 21.09.2009.

Aprovado após revisão em 05.02.2010.

Apêndice

(1) Critério de informação deviance (DIC)

O critério de informação deviance (DIC) é um critério útil para a seleção de modelos sob o enfoque Bayesiano onde amostras da distribuição a posteriori para os parâmetros do modelo são obtidas usando métodos MCMC.

O deviance (desvio) é definido por,

$$D(\boldsymbol{\theta}) = -2\log L(\boldsymbol{\theta}) + C \quad (16)$$

onde $\boldsymbol{\theta}$ é um vetor de parâmetros desconhecidos do modelo, $L(\boldsymbol{\theta})$ é a função de verossimilhança do modelo e C é uma constante que não é necessário que seja conhecida quando a comparação entre modelos é efetuada.

O critério DIC definido por Spiegelhalter *et al.*(2002) é dado por,

$$DIC = D(\hat{\boldsymbol{\theta}}) + 2n_D \quad (17)$$

onde $D(\hat{\boldsymbol{\theta}})$ é o desvio (deviance) avaliado na média a posteriori $\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}|\text{dados})$ e n_D é o número efetivo de parâmetros do modelo dado por $n_D = \bar{D} - D(\hat{\boldsymbol{\theta}})$, onde $\bar{D} = E(D(\boldsymbol{\theta})|\text{dados})$ é o desvio (deviance) a posteriori que mede a qualidade dos dados ajustados pelo modelo. Menores valores de DIC indicam melhores modelos. Observar que esses valores podem ser negativos.

(2) Programa WinBugs - “Modelo 1”

```
model {  
  t[1] ~ dexp(lambda[1])  
  lambda[1] <- exp(alpha+eta[1])  
  theta[1]<- 1/lambda[1]  
  eta[1] ~ dnorm(0,1)  
  for (i in 2:n) {  
    t[i] ~ dexp(lambda[i])  
    lambda[i] <- exp(alpha+phi*(eta[i-1]-alpha))  
    eta[i] ~ dnorm(0,1)  
    theta[i]<- 1/lambda[i]  
  }  
  phi ~ dbeta(1,1)  
  alpha ~ dnorm(0,1)  
  for (i in 1:n) {  
    theta.t[i]<-abs(theta[i]-t[i])  
  }  
  soma<-sum(theta.t[])  
}
```

```

        for (i in 2:n){
          diff[i]<=theta[i]-theta[i-1]
        }
      }

```

```

list(t= c(105,14,1,1,1,4,1,2,4,1,1,1,1
,3,1,3,3,1,6,2,1,1,3,1,2,4
,2,1,4,4,4,14,6,5,24,6,17,25,7
,1,7,13,21,112,3,6,1,2,2,1,1,1
,1,1,2,1,1,1,1,1,1,1,1,1,1,1
,1,1,2,1,1,1,1,1,1,1,1,1,1,1
,1,2,1,1,1,1,1,1,2,1,1,1,1,1
,1,1,1,1,1,1,1,1,1,1,1,1,1,1
,1,1,1,1,1,1,1,1,1,1,1,1,2
,1,1,2,1,1,1,1,1,2,1,1,1,1,1
,3,1,1,1,1,3,1,78,36,11,1,3,4
,3,4,3,14,63,5,1,2,5,1,1,1,1
,1,1,2,1,1,1,1,1,1,1,1,2,1
,1,1,1,1,1,1,1,1,1,1,1,1,1
,1,1,1,1,1,1,1,1,1,1,1,1,1
,1,1,1,1,1,1,1,1,1,2,1,1,1
,1,1,1,1,1,1,2,1,1,1,1,2,1
,1,1,1,1,1,2,1,1,1,1,1,2,1
,1,1,1,1,1,1,1,1,3,2,1,1,1
,1,3,1,1,1,1,1,1,1,1,1,1,1
,1,2,1,56,1,14,1,4,1,1,1,4,2
,2,4,1,5,2,1,1,3,2,1,1,3,1
,6,1,3,6,14,4,7,7,1,10,10,59,4
,7,3,1,1,1,1,1,2,2,2,2,1,2
,2,1,1,1,1,1,1,1,1,1,1,1,1
,1,1,1,1,1,1,1,1,1,1,1,1,1
,1,1,1,1,1,1,1,3,1,1,1,1,1
,1,1,1,1,1,1,1,1,1,1,1,1,1
,1,2,1,1,1,1,1,1,1,1,1,1,1
,2,1,1,1,1,1,2,1,1,1,1,1,1
,1,1,1,1,1,1,3,1,1,1,1,2
,1,1,1,4,1,1,1,1,1,2,1,2,1
,3,3,4,2,1,7,4,1,1,1,4,2,5
,1,1,1,4,14,35,4,3,7,7,1,1,2
,3,3,11), n = 445)

```


ANEXO A - Comitê de Ética em Pesquisa

HOSPITAL DAS CLÍNICAS DA FACULDADE DE MEDICINA
DE RIBEIRÃO PRETO DA UNIVERSIDADE DE SÃO PAULO

www.hcrp.usp.br

Ribeirão Preto, 11 de maio de 2011

Ofício nº 1605/2011
CEP/MGV

Prezados Senhores,

O trabalho intitulado **“MODELAGEM DOS EXCESSOS DE INTERNAÇÕES HOSPITALARES NA REGIÃO DE RIBEIRÃO PRETO: UMA APLICAÇÃO À DOENÇAS RESPIRATÓRIAS E EFEITO DAS QUEIMADAS DE CANA DE AÇÚCAR”** foi analisado pelo Comitê de Ética em Pesquisa, em sua 323ª Reunião Ordinária realizada em 09/05/2011 e enquadrado na categoria: **APROVADO**, de acordo com o Processo HCRP nº 4290/2011.

Este Comitê segue integralmente a Conferência Internacional de Harmonização de Boas Práticas Clínicas (IGH-GCP), bem como a Resolução nº 196/96 CNS/MS.

Lembramos que devem ser apresentados a este CEP, o Relatório Parcial e o Relatório Final da pesquisa.

Atenciosamente.

DRª MARCIA GUIMARÃES VILLANOVA
Coordenadora do Comitê de Ética em
Pesquisa do HCRP e da FMRP-USP

Ilustríssimos Senhores

MAYARA PIANI LUNA DA SILVA SICCHIERI

JORGE ALBERTO ACHCAR (Orientador)

Depto. de Medicina Social