

MODELO POISSON GLARMA APLICADO A ATENDIMENTOS HOSPITALARES POR CAUSAS RESPIRATÓRIAS NA REGIÃO DA GRANDE VITÓRIA - E.S.

Magno dos Santos Neto

Programa de Pós-Graduação em Ciências Contábeis - UFES/ES
Universidade Federal do Espírito Santo, Av. Fernando Ferrari, s/n, 29060-900, Vitória-ES
magno.sneto@yahoo.com.br

Mônica Kelley Bottoni de Souza

Secretaria de Estado da Saúde - SESA/ES
Rua Eng. Guilherme José Monjardim Varejão, 225, 29050-260, Ed. Enseada Plaza, Vitória - ES
monicabottoni@yahoo.com.br

RESUMO

Este trabalho conduz a uma apresentação de estudos de dados de contagem, onde o Modelo Autoregressivo Média Móvel Linear Generalizado é utilizado para investigar a associação entre o número de atendimentos diários por causas respiratórias em crianças de 0 a 6 anos de idade e as concentrações dos poluentes atmosféricos PTS, NO₂, O₃, SO₂ e PM₁₀, emitidos na Região da Grande Vitória (RGV), E.S., durante o período de 1º de janeiro de 2001 a 31 de julho de 2006. A concentração dos poluentes e as variáveis atmosféricas foram obtidas juntos ao Instituto Estadual de Meio Ambiente e Recursos Hídricos (IEMA), observados nas estações de monitoramento de ar, coordenados pela instituição. Os dados diários de atendimentos foram obtidos no hospital infantil de Vitória, principal pronto-socorro em atendimento infantil da RGV. Os resultados evidenciaram o aumento percentual do risco relativo do número de atendimentos diários por causas respiratórias em consequência dos poluentes.

PALAVRAS CHAVE. Séries Temporais, Poluição do ar, MLG - GLARMA(p, q).

ABSTRACT

This work leads to a presentation count data studies where the model Autoregressive Moving Average Generalized Linear is used to investigate the association between the number of daily visits for respiratory causes in children 0-6 years of age and concentrations of pollutants atmospheric PTS, NO₂, O₃, SO₂ and PM₁₀, issued in the Grande Vitória (RGV), ES, during the period 1 January 2001 to 31 July 2006. The concentration of pollutants and atmospheric variables were obtained together at the State Institute for the Environment and Water Resources, observed in air monitoring stations coordinated by the institution. The daily attendance data were obtained from the children's hospital in Victoria, emergency room principal child care of RGV. The results showed the percentage increase in the relative risk of the number of daily visits for respiratory causes as a result of pollutants.

KEYWORDS. Time series, air pollution, MLG - GLARMA(p, q).

1. Introdução

Os efeitos adversos dos poluentes atmosféricos na saúde humana têm sido uma preocupação dos órgãos regulamentadores e de saúde pública ao longo dos últimos 50 anos. A confirmação sobre os efeitos adversos provocados pelas concentrações dos poluentes na saúde pode ser obtida a partir de estudos populacionais utilizando métodos de investigação epidemiológica, orientando o desenvolvimento de normas visando o controle das emissões e da qualidade do ar.

[Nelder e Baker, 1972] propuseram o Modelo Linear Generalizado (MLG), que pode ser interpretado como uma generalização do modelo tradicional de regressão linear. A idéia básica deste modelo consiste em abrir um 'leque' de opções para a variável resposta, permitindo que a mesma pertença à família exponencial de distribuições. [McCullagh e Nelder, 1989] formalizaram a principal referência bibliográfica sobre os MLGs.

Vários artigos relacionados com os MLGs foram publicados desde 1972, pode-se citar [Wedderburn, 1974], [Zeger e Liang, 1986], [Jørgensen, 1983]. A partir de então, surge extensões do MLG, o Modelo Aditivo Generalizado (MAG), proposto por [Hastie e Tibshirani, 1990], o qual supõe que o preditor linear seja formado também por funções semi-paramétricas e os Modelos Lineares Mistos, proposto por [Breslow e Clayton, 1993], que admitem a inclusão de efeitos aleatórios Gaussianos no preditor linear. [Lee e Nelder, 1996] estenderam o trabalho de [Breslow e Clayton, 1993] propondo MLG Hierárquicos, em que o preditor linear pode ser formado por efeitos fixos e efeitos aleatórios não-Gaussianos.

Além desses estudos, surge uma derivação dos usuais modelos de séries temporais para dados de contagem, o Modelo Autoregressivo Média Móvel Linear Generalizado (GLARMA). Introduzido por [Shephard et al., 1995], o GLARMA realiza a junção do MLG com o Modelo Autoregressivo Média Móvel (ARMA). O autor apresenta os primeiros passos para a concepção do modelo, apresentando procedimentos de estimação e testes para o modelo, e conseqüentemente, aplicações em dados de saúde e financeiros.

Em seguida, [Baghishani, Tabatabaey & Nia, 2000] introduziram a função de verossimilhança para os modelos GLARMA, onde a estimação e inferência baseadas nesta verossimilhança são descritas, considerando o caso especial em que a variável resposta possui distribuição de Poisson.

Nos estudos epidemiológicos que envolvem a análise de séries temporais o que se pretende, usualmente, é verificar se os efeitos deletérios ou maléficos à saúde humana estão associados a um fator de risco, ou seja, no caso de estudos de poluição do ar, busca-se verificar se as concentrações dos poluentes atmosféricos estão diretamente associados aos atendimentos hospitalares, internações ou à mortalidade, por patologia respiratória ou até mesmo cardiovascular [Schwartz, 1997]. Apesar dos diversos estudos sobre a associação entre mortalidade e poluição do ar nos últimos anos, as atenções se voltaram principalmente para as internações e admissões hospitalares por causas respiratórias ou cardiovasculares. Visto que o aumento das concentrações de poluentes na atmosfera influenciam nos registros dos atendimentos e das internações hospitalares, principalmente em grupos de crianças e idosos, mais suscetíveis aos efeitos dos poluentes, [Pantazopoulou et al., 1995].

As características de dados de séries temporais ambientais exigem métodos estatísticos complexos e suficientemente sensíveis, para detectar os efeitos combinados de outras covariáveis variantes no tempo. Isto é, no caso de estudos sobre a poluição do ar e saúde existem estruturas de tendências e flutuações sazonais, que interferem na dinâmica das variáveis observadas tais como fluxo de veículos, dias da semana, feriados, número de indústrias, entre outras. Além disso, influências de variáveis meteorológicas, como temperatura e umidade relativa do ar, também estão presentes na dispersão dos poluentes na atmosfera e na natureza de diversas doenças respiratórias [Schwartz, 2000] e [Gouveia et al., 2003].

Estudo realizado por [Almeida, 2006] para verificar, através do MAG, a relação existente entre os poluentes atmosféricos PM_{10} , O_3 , SO_2 e NO_2 e o número de atendimentos diários por

causas respiratórias da população infantil da Região da Grande Vitória (RGV), E.S., tem evidenciado para o grupo investigado no período de 2001 a 2004 que os níveis da poluição atmosférica são significativos, com destaque para os danos causados por PM_{10} e NO_2 . Dessa forma, pode-se definir como objetivo deste estudo a apresentação da modelagem, para dados referentes ao número de atendimentos hospitalares por causas respiratórias, através do MAG, que é uma das diversas técnicas de modelagem para dados discretos.

O presente trabalho investiga a associação entre as concentrações dos poluentes atmosféricos (PTS , NO_2 , O_3 , SO_2 , PM_{10}) e internações por causas respiratórias de crianças de 0 a 6 anos de idade na Região da Grande Vitória (RGV), tomando como base a utilização do Modelo Autoregressivo Média Móvel Linear Generalizado (GLARMA) para dados de contagem. Além disso, foi calculado os efeitos da poluição sobre a saúde, quantificado pelo Risco Relativo (RR), a probabilidade de um indivíduo desenvolver uma doença relativo à exposição ao fator de risco.

2. Metodologia

2.1. Dados Utilizados

O estudo considerou o número de atendimentos hospitalares por causas respiratórias em crianças de 0 a 6 anos, no período de janeiro de 2001 a julho de 2006, no Hospital Infantil Nossa Senhora da Glória (HINSG), onde se encontra o principal pronto-socorro infantil da cidade de Vitória, referência secundária e terciária para o atendimento pediátrico da RGV.

Os dados referentes aos atendimentos hospitalares foram obtidos junto ao sistema de registro de atendimentos do hospital.

As variáveis atmosféricas consideradas no estudo foram a temperatura e a umidade relativa do ar. Ambas variáveis foram disponibilizadas pelo IEMA.

A concentração diária da temperatura é expressa em graus Celsius ($^{\circ}C$). Para a realização do estudo utilizou-se a temperatura média mínima, temperatura média e temperatura média máxima referentes às estações de Carapina e Cariacica, as únicas estações na região que medem essas variáveis.

A umidade relativa do ar foi medida pelas estações de Carapina e Cariacica, esta expressa em porcentagens (%). Com a umidade relativa do ar de cada estação foi possível calcular a umidade relativa média diária.

Os dados diários de concentração dos poluentes atmosféricos PTS , NO_2 , O_3 , SO_2 , PM_{10} e os dados meteorológicos, temperatura e umidade relativa do ar foram obtidos na Rede de Monitoramento Automática da Qualidade do Ar da Grande Vitória (RAMQar) para o período de 1^o de janeiro de 2001 a 31 de julho de 2006, atualmente operada pelo IEMA.

2.2. Modelos Lineares Generalizados

Nelder e Wedderburn (1972) apresentaram uma extensão dos modelos lineares de regressão múltipla, os modelos lineares generalizados. Em comparação dos modelos lineares generalizados com os modelos de regressão múltipla, permitiu alargar as hipóteses admitidas. A variável resposta do modelo passa a poder provir de uma população que siga uma lei de distribuição da família exponencial, deixando de ter obrigatoriamente uma distribuição Normal. Além disso, se nos modelos lineares de regressão clássica, a relação entre o valor médio da variável resposta e a combinação linear das variáveis explicativas é a função identidade, nos modelos lineares generalizados a relação entre o valor médio da variável resposta e a combinação linear das variáveis explicativas pode ser estabelecida por qualquer função monótona e diferenciável.

Segundo [Nelder e Baker, 1972], um modelo linear generalizado é caracterizado pelas seguintes estruturas: família de distribuição exponencial, preditor linear e função de ligação dada por

$$f(y_i|\theta_i, \phi) = \exp \left\{ \frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi) \right\} \quad (1)$$

onde $a_i(\cdot)$, $b(\cdot)$ e $c(\cdot)$ são funções específicas para cada distribuição, e $E(Y_i) = \mu_i = b'(\theta_i)$, $Var(Y_i) = \phi^{-1}V_i$, $V_i = d\mu_i/d\theta$ é a função de variância. Se ϕ for conhecido, tem-se uma distribuição da família exponencial com parâmetro canônico. Caso ϕ seja desconhecido, a distribuição pode ou não pertencer à família exponencial, então têm-se que considerar apenas os casos em que a distribuição pertence aquela família. A função $a_i(\phi)$ tem a seguinte forma, na maioria dos casos, $a_i(\phi) = \frac{\phi}{\omega_i}$, onde assume-se que ω_i é conhecido, quanto ϕ , denominado *parâmetro de dispersão ou escala*, pode ser ou não conhecido.

Distribuição de Poisson: No caso da distribuição de Poisson a função de probabilidade é

$$f(y|\lambda) = \frac{e^{-\lambda}\lambda^y}{y!} \quad (2)$$

onde $y=0,1,2, \dots$ e $\lambda > 0$. Então, $f(y|\lambda) = \exp\{y \ln(\lambda) - \lambda - \ln(y!)\}$. Fazendo $\theta = \ln(\lambda)$, temos $f(y|\theta, \phi) = \exp\{y\theta - e^\theta - \ln(y!)\}$, obtendo-se assim, para o caso da distribuição de Poisson, $a(\phi) = 1$, $b(\theta) = e^\theta$, $c(y, \phi) = -\ln(y!)$, $E(Y) = e^\theta$ e $Var(Y) = e^\theta$. A função de ligação para a distribuição Normal é $\eta = \theta \Leftrightarrow g(\mu) = \mu$, a função de ligação canônica é a identidade, assim, tem-se o modelo linear de regressão clássica. Já para o caso de Y ter distribuição de Poisson $\eta = \theta \Leftrightarrow g(\mu) = \ln(\mu)$, visto que no exemplo acima, se concluiu que $\mu = e^\theta$.

Para estimar o modelo há que estimar os parâmetros β_j ($j = 1, 2, \dots, p$) e, no caso da distribuição ter parâmetro de escala desconhecido, também é necessário estimar ϕ . Para obter os estimadores dos β_j , $j = 1, 2, \dots, p$, vai seguir o método da máxima verossimilhança. Este método consiste em maximizar a função de verossimilhança de Y em ordem a β ou, de forma equivalente, determinar o máximo absoluto do logaritmo da função verossimilhança, uma vez que a função logaritmo da função de verossimilhança para o vetor aleatório \vec{Y} . Então

$$L(\theta, \phi) = \sum_{i=1}^n \left[\frac{Y_i \theta_i - b(\theta_i)}{\alpha_i(\phi)} + c(Y_i, \phi) \right]. \quad (3)$$

2.3. Modelo Autoregressivo Média Móvel Linear Generalizado, Poisson-GLARMA(p, q)

Para introduzir o modelo, assume-se que a observação Y_t dado o passado $\mathcal{F}_{t-1} = \sigma(Y_s, s \leq t-1)$ tem distribuição de Poisson com média μ_t , e é denotado por

$$Y_t | \mathcal{F}_{t-1} \sim P(\mu_t). \quad (4)$$

Assim

$$f(y_i; \mu_t | \mathcal{F}_{t-1}) = \frac{\exp(-\mu_t)\mu_t^{y_t}}{y_t!}, t = 1, \dots, n. \quad (5)$$

Colocando a equação 5 na forma da equação 1 tem-se

$$f(y_i; \mu_t | \mathcal{F}_{t-1}) = \exp\{y_t \log(\mu_t) - \log(y_t!)\}, t = 1, \dots, n, \quad (6)$$

onde $E[Y_t | \mathcal{F}_{t-1}] = \mu_t$, $b(\theta_t) = \mu_t = \exp(\theta_t)$, $V(\mu_t) = \mu_t$, $\phi = 1$ e $\gamma_t = 1$. A relação canônica é

$$g(\mu_t) = \theta_t(\mu_t) = \log(\mu_t) = \eta_t = \mathbf{Z}'_{t-1}\beta. \quad (7)$$

Estendendo essa nomenclatura ao trabalho de [Davis et al., 1999], pode considerar a equação 7 como o vetor de covariáveis $\mathbf{Z}_{t-1} = (\mathbf{X}_t, e_{t-1}, \dots, e_{t-q})$. Assume-se ainda que o processo $\log(\mu_t)$ segue um modelo linear nas variáveis explicativas com os resíduos que tem uma estrutura

média móvel. A média móvel será uma sequência da diferença marginal gerada a partir dos dados. Formalmente, o processo é dado por

$$W_t = \log(\mu_t) = \mathbf{x}_t^T \beta + \sum_{i=1}^q \theta_i e_{t-i} \quad (8)$$

onde

$$e_t = (Y_t - \mu_t) / \mu_t^\lambda, \lambda \geq 0, \lambda \in [0, 1] \quad (9)$$

A média móvel infinita neste modelo pode ser especificada em termos de um número finito de parâmetros de várias maneiras. Neste trabalho serão consideradas as estruturas distribuídas com *lag* que são gerados pelo preditor linear para processos autoregressivos com média móvel. Seja $\{U_t\}$ um processo ARMA com erros dados pela diferença marginal da sequência $\{e_t\}$, isto é,

$$U_t = \phi_1 U_{t-1} + \dots + \phi_p U_{t-p} + e_t + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q}, \quad (10)$$

onde o polinômios AR e MA, $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ e $\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$, respectivamente, possuem todos os zeros fora do círculo unitários. Então, o melhor preditor de U_t baseado em passado infinitos $\{U_{t-1}, U_{t-2}, \dots\}$ é

$$\hat{U}_t = \sum_{i=1}^{\infty} \tau_i e_{t-i} \quad (11)$$

onde

$$\begin{aligned} \sum_{i=1}^{\infty} \tau_i z^j &= \left((1 - \sum_{i=1}^p \phi_i z^i)^{-1} (1 + \sum_{i=1}^q \theta_i z^i) \right) - 1 \\ &= \phi(z)^{-1} \theta(z) - 1 \end{aligned} \quad (12)$$

O modelo para $\log(\mu_t)$ então é

$$W_t = \mathbf{x}_t^T \beta + Z_t = \mathbf{x}_t^T \beta + \sum_{i=1}^{\infty} \tau_i e_{t-i} \quad (13)$$

onde $Z_t = \hat{U}_t$. Referimos ao modelo acima para os dados $\{Y_t\}$ como ARMA Linear Generalizado Poisson de ordem (p, q) , $GLARMA(p, q)$. O ajuste do modelo é feito por estágio, Z_t é computado usando recursões ARMA. Especificamente, para $t \leq 0$, $e_t = 0$ e $Z_t = 0$, e para $t > 0$, as recursões são aplicadas da seguinte forma

$$\begin{aligned} \hat{Z}_t &= \phi(\hat{Z}_{t-1} + e_{t-1} + \dots + \phi_p(\hat{Z}_{t-p} + e_{t-p})) + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} \\ W_t &= \mathbf{x}_t^T \beta + \hat{Z}_t \\ e_t &= (Y_t - \mu_t) / \mu_t^\lambda \end{aligned} \quad (14)$$

A estrutura deste modelo é similar à formulada em [Shephard et al., 1995].

2.3.1. Estimação

Segundo [Davis et al., 2003] a função verossimilhança e a primeira e segunda derivadas são facilmente calculadas recursivamente utilizando o procedimento de Newton-Raphson. Os erros padrão para as estimativas de parâmetros que devidamente para a dependência de série também estão disponíveis. Os detalhes a seguir.

Seja $\delta = (\beta^T, \gamma^T)^T$ e define-se $L_t(\delta) = \log f(y_t | \mathcal{F}_{t-1})$, onde f é a distribuição condicional Poisson de Y_t dado \mathcal{F}_{t-1} . A função log-verossimilhança pode ser escrita como $\sum_{t=1}^n L_t(\delta)$, que ignorando os termos que não envolvem os parâmetros, torna-se

$$L(\delta) = \sum \left(Y_t W_t(\delta) - e^{W_t(\delta)} \right), \quad (15)$$

onde $\log(\mu_t) = W_t(\delta) = \mathbf{x}_t^T \beta + \sum_{i=1}^{\infty} \tau_i(\gamma) e_{t-i}(\delta)$, e $e_t(\delta) = \frac{Y_t - \mu_t}{\mu_t^\lambda}$.

Para inicializar as recursões de Newton-Raphson descobrimos usando as estimativas MLG se os termos autoregressivos médias móveis, juntamente com valor inicial zero para e_t , $t \leq 0$, dá valores iniciais razoáveis. A matriz de covariância dos estimadores é estimado por

$$\hat{\Omega} = - \left(\frac{\partial^2 L(\hat{\theta})}{\partial \delta \partial \delta^T} \right)^{-1}. \quad (16)$$

2.4. Risco Relativo

O risco relativo (*relative risk*) tornou-se uma das medidas mais utilizadas na investigação biomédica e recentemente adotada nos estudos dos efeitos da poluição do ar na saúde. O risco relativo é utilizado com frequência na análise estatística dos desfechos binários. Alternativamente, é utilizado para comparar o risco de desenvolvimento de um efeito colateral em pessoas que recebem uma droga, em comparação às pessoas que não estão recebendo o tratamento (ou a receber um placebo). É particularmente interessante devido a sua utilização em análises de regressão de contagens, tipicamente quando a distribuição dos dados segue uma *Poisson*. As medidas de associação buscam responder se existe uma associação entre uma exposição e um desfecho (algum evento de saúde como: atendimentos, internações, mortes, etc). São medidas do tipo razão que comparam duas medidas de frequência e medem a força da relação estatística entre uma variável e a frequência da doença. Portanto, o risco relativo (*RR*) é o coeficiente de incidência - a incidência de uma doença é, estritamente, representada pelo número de novos casos surgidos a cada ano na população - de casos expostos e o coeficiente de incidência de casos não expostos, definida como:

$$RR = \frac{E(Y|X = 1)}{E(Y|X = 0)}. \quad (17)$$

Como o risco relativo é uma razão de taxas de incidência, que conseqüentemente são contagens, a distribuição de *Poisson* é um modelo onde é possível utilizar o risco relativo, pois a média da *Poisson* é uma taxa (λ) com função de ligação dado pelo logaritmo. Para reduzir o risco relativo nos modelos de regressão com distribuição de *Poisson*, considere o modelo:

$$\begin{aligned} \log(E[Y|X = x]) &= \alpha + \beta X \\ \log(E[Y|X = 0]) &= \alpha \end{aligned} \quad (18)$$

A interpretação do α é clara quando $\beta=0$ (não significativo), pois nenhuma variação seria registrada na equação 18, ou seja, o efeito do poluente na saúde é nulo, portanto ($RR = 1$). Agora fazendo ($X = 1$) para um acréscimo de uma unidade na variável explicativa, temos que: $\log(E[Y|X = 1]) = \alpha + \beta$. Assim, usando as propriedades do logaritmo, temos:

$$RR = \frac{E[Y|X = 1]}{E[Y|X = 0]} = e^{(\beta)}. \quad (19)$$

Logo, para variações acima de uma unidade a expressão do risco relativo fica definida por:

$$RR = e^{(k\hat{\beta}_i)} \quad \text{com } i=1,2,\dots,p \quad (20)$$

onde k e $\hat{\beta}$ são respectivamente a variação da concentração do poluente que pode ser ($10\mu\text{g}/\text{m}^3$, variação interquartilica, etc.) e o coeficiente estimado associado ao poluente em estudo.

3. Resultados

Esta seção apresenta as análises descritivas dos dados utilizados e os resultados do ajuste do modelo GLARMA(p, q).

3.1. Análise Descritiva

A seguir são apresentadas tabelas e gráficos dos dados meteorológicos, poluentes atmosféricos e atendimentos por doenças por causas respiratórias em crianças de 0 a 6 anos de idade utilizados na presente trabalho. Durante o período de estudo ocorreram 52036 atendimentos doenças por causas respiratórias em crianças de 0 a 6 anos de idade na Região da Grande Vitória, ES, Brasil. A Tabela 1 mostra que o número médio de atendimentos diários foi igual a 25.55 atendimentos com um desvio padrão de 16.57.

Tabela 1: Estatísticas descritivas dos poluentes, variáveis meteorológicas e atendimentos por doenças respiratória na RGV, jan/2001 a jul/2006.

	n	média	dp	mínimo	Percentis			máximo
					25	50	75	
PM ₁₀ ($\mu\text{g}/\text{m}^3$)	2037	28.07	7.22	6.78	23.31	27.48	31.99	64.15
SO ₂ ($\mu\text{g}/\text{m}^3$)	2037	23.59	10.13	2.38	16.27	22.21	29.59	64.05
O ₃ ($\mu\text{g}/\text{m}^3$)	2037	29.89	9.79	10.17	22.51	28.09	35.55	78.08
NO ₂ ($\mu\text{g}/\text{m}^3$)	2037	22.62	7.51	6.59	17.09	21.99	27.27	61.12
PTS ($\mu\text{g}/\text{m}^3$)	2037	46.45	14.44	9.63	6.65	45.76	54.72	119.06
Temperatura mínima ($^{\circ}\text{C}$)	2037	20.74	2.35	14.35	18.85	20.85	22.75	25.75
Temperatura média ($^{\circ}\text{C}$)	2037	23.99	2.30	16.60	22.09	24.10	25.94	28.84
Temperatura máxima ($^{\circ}\text{C}$)	2037	28.25	3.09	19.00	25.90	28.35	30.50	35.60
Umidade relativa do ar (%)	2037	77.55	7.73	47.30	72.50	77.70	82.48	95.63
Atendimentos por DAR	2037	25.55	16.57	0.00	3.00	23.00	34.00	119.00

Nota-se na Figura 1 um comportamento sazonal e verifica-se que no período da primavera para o inverno ocorreu o aumento do número de atendimentos por causas respiratórias. Esse aumento se deve a dois fatores principais, as baixas temperaturas e os aumentos nas concentrações dos poluentes primários.

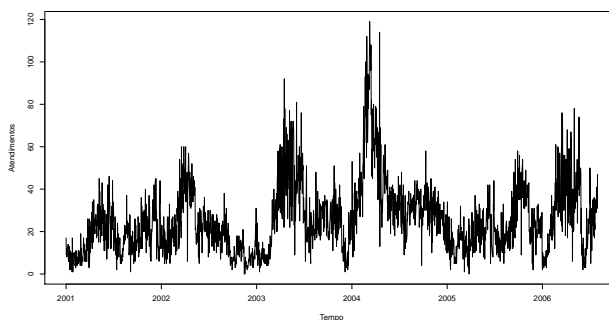


Figura 1: Série do número de atendimentos hospitalares por causas respiratórias diários de crianças de 0 a 6 anos da RGV.

As próximas figuras apresentam as séries das médias diárias dos poluentes PTS, NO₂, O₃, SO₂, PM₁₀. Pelos gráficos percebe-se a presença de componentes sazonais em todos os poluentes, exceto no SO₂.

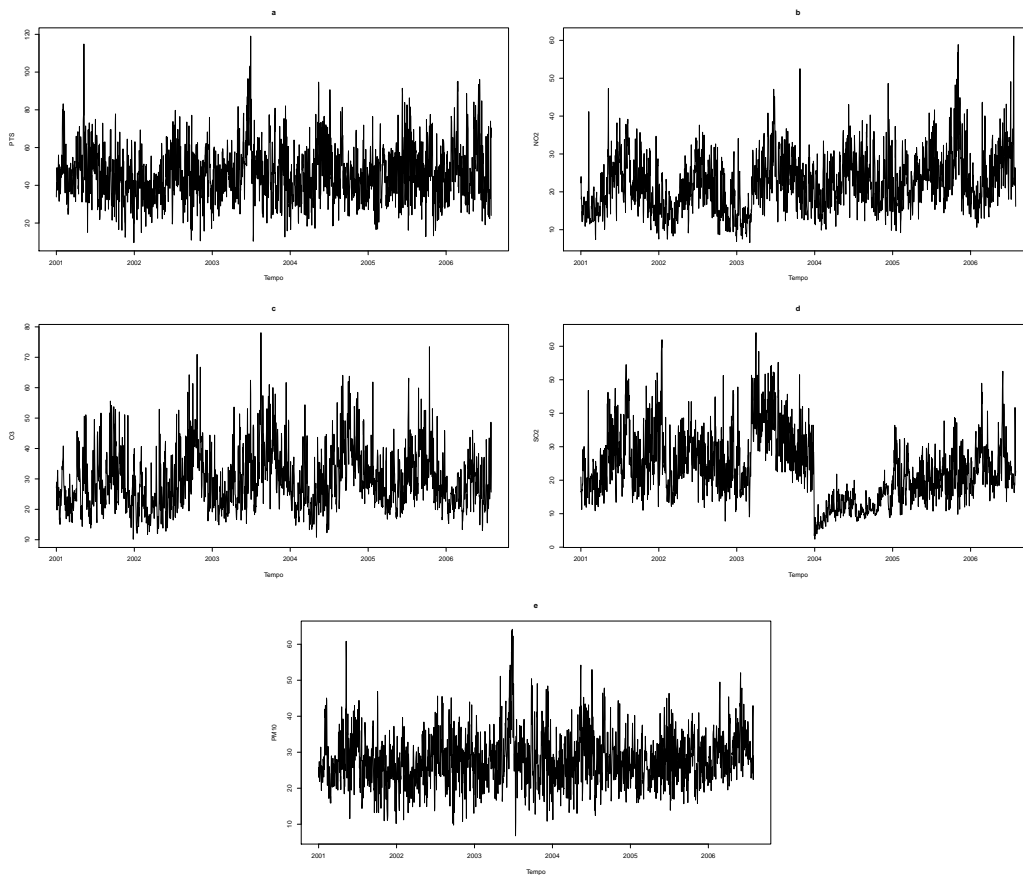


Figura 2: (a)-Média diária de Partículas Totais em Suspensão(PTS); (b)-Média diária de Dióxido de Nitrogênio(NO₂);(c)-Média diária de Ozônio(O₃);(d)-Média diária de Dióxido de Enxofre(SO₂);(e)-Média diária de Material Particulado Inalável(PM₁₀).

3.2. Ajuste do Modelo

Nesta seção será apresentado o ajuste do modelo GLARMA(p, q) aos dados de atendimentos por doenças respiratórias em criança de 0 a 6 anos de idade na RGV.

A Tabela 2 mostra o ajuste do modelo utilizado para explicar os atendimentos hospitalares. Os coeficientes são todos significativos ao nível de 5%, e como critério de escolha do modelo observou-se o modelo ajustado apresentado na tabela 2, possui o menor *AIC*. O modelo em questão é constituído por: Intercept: Intercepto; Terca, Sexta: dias da semana Novo: Feriado ano Novo Carn: Feriado Carnaval Cinzas: Feriado dia de Cinzas; Paix: Feriado dia da Paixão; Pascoa: Feriado Páscoa; Tirad: Feriado Tiradentes; Trab: Feriado dia do Trabalho; Indep: Independência; Crianca: Feriado dia das Crianças; Penha: Feriado Penha; tmpmin: Temperatura mínima; tmpmedl2: Temperatura média defasada dois dias; tmpmaxl1: Temperatura máxima defasada um dia; tmpmaxl2: Temperatura máxima defasada dois dias; Umr11: Umidade relativa defasada um dia; PTS: Partículas totais suspensas; PTS12: Partículas totais suspensas defasada dois dias; NO2: Dióxido de Nitrogênio; NO211: Dióxido de Nitrogênio defasado um dia; NO213: Dióxido de Nitrogênio defasado três dias; O313: Ozônio defasado três dias; SO13: Dióxido de Enxofre defasado três dias; PM: Material particulado (PM₁₀) PML1: Material particulado defasado um dia; PML2: Material particulado defasado dois dias; PML3: Material particulado defasado três dias; PML4: Material particulado defasado quatro dias PML5: Material particulado defasado cinco dias; AR(1): Autorregressivo de ordem 1; MA(1): Média móvel de ordem 1.

Tabela 2: Resultado da Estimação do Modelo GLARMA

Variável	Coef.	E.P.	P-valor	Variável	Coef.	E.P.	P-valor
Intercept	2.6132	0.0891	0.0000	Umr1	0.0029	0.0006	0.0000
Terça	0.0370	0.0146	0.0116	PTS	-0.0044	0.0005	0.0000
Sexta	-0.0714	0.0145	0.0000	PTS12	-0.0017	0.0005	0.0014
Novo	-0.5008	0.1053	0.0000	NO2	0.0042	0.0009	0.0000
Carn	0.5410	0.0559	0.0000	NO211	0.0048	0.0010	0.0000
Cinzas	0.4992	0.0689	0.0000	NO213	0.0046	0.0008	0.0000
Paix	0.4573	0.0684	0.0000	O313	-0.0053	0.0005	0.0000
Pascoa	0.3478	0.0731	0.0000	SO13	-0.0050	0.0005	0.0000
Tirad	0.3793	0.0757	0.0000	PM	-2.6935	1.1684	0.0212
Trab	0.6333	0.0756	0.0000	PML1	-2.6970	1.1684	0.0210
Indep	0.2584	0.0821	0.0017	PML2	-4.4946	0.7745	0.0000
Crianca	-0.2248	0.1026	0.0285	PML3	-3.0739	0.5604	0.0000
Penha	0.4111	0.0677	0.0000	PML4	-1.9838	0.4091	0.0000
tmpmin	0.0152	0.0035	0.0000	PML5	-1.3913	0.2563	0.0000
tmpmed12	0.0275	0.0067	0.0000	AR(1)	0.88802	0.01711	$< 2e - 16$
tmpmax11	-0.0154	0.0025	0.0000	MA(1)	-0.68053	0.02584	$< 2e - 16$
tmpmax12	-0.0227	0.0042	0.0000				

(a)

(b) continuação

A Figura 3 mostra o ajuste do modelo $GLARMA(1, 1)$ para a série de atendimentos hospitalares por causas respiratórias de crianças de 0 a 6 anos de idade na R.G.V. A figura mostra que o ajuste do modelo é capaz explicar a média dos dados de atendimentos hospitalares.

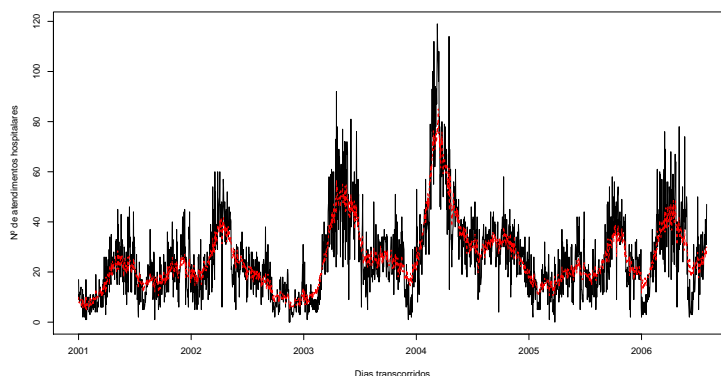


Figura 3: — Série de atendimentos hospitalares - - - Série ajustada

3.3. Risco Relativo

A seguir são apresentados os cálculos dos Riscos Relativos (que foram estatisticamente significante ao nível de 5%) para os aumentos nos atendimentos hospitalares por doenças respiratórias, em crianças de 0 a 6 anos de idade, devido aos efeitos das concentrações dos poluentes.

A Tabela 3 exhibe as estimativas de efeitos simples e cumulativos do poluente PTS. Por essa tabela, verifica-se que para um aumento de $18.07 \mu g/m^3$ (intervalo interquartilico), a defasagem de 3 dias do poluente, está associado a um aumento de 4.0% no número médio de atendimentos diários por causas respiratórias em crianças de 0 e 6 anos. As exposições acumuladas de 4 e 7 dias também

estão associados a aumentos de 2.8% e 0.9% no número médio de atendimentos diários por causas respiratórias, respectivamente. Os efeitos estimados de *PTS*, com defasagem de 3 dias, e com exposição acumulada de 4 e 7 dias, são estatisticamente significativos ao nível de 10%. A Tabela 4 exibe as estimativas de efeitos simples e cumulativos de NO_2 . Verifica-se que, para um aumento de $10.18\mu\text{g}/\text{m}^3$, com exposição de 7 dias, está associado a um aumento de 2.10% no número médio de atendimentos diários por causas respiratórias em crianças de 0 a 6 anos. O efeito estimado de NO_2 , com exposição acumulada de 7 dias, é estatisticamente significante ao nível de 10%.

Tabela 3: Poluente PTS .

	RR	%RR	P-valor
Dia Corrente	1.003	0.3	0.859
Defasagem de 1 dia	0.995	-0.5	0.858
Defasagem de 2 dia	1.0432	4.3	0.638
Defasagem de 3 dia	1.04	4.0	0.007
Média Móvel 2 dias	0.009	-99.1	0.894
Média Móvel 3 dias	1.0001	0.0	0.164
Média Móvel 4 dias	1.028	2.8	0.007
Média Móvel 5 dias	1.037	3.7	0.515
Média Móvel 6 dias	1.008	0.8	0.992
Média Móvel 7 dias	1.009	0.9	0.080

Tabela 4: poluente NO_2 .

	RR	%RR	P-valor
Dia Corrente	0.998	-0.20	0.557
Defasagem de 1 dia	0.999	-0.10	0.557
Defasagem de 2 dia	1.009	0.90	0.636
Defasagem de 3 dia	1.032	3.20	0.819
Média Móvel 2 dias	1.000	0.01	0.710
Média Móvel 3 dias	0.997	-0.30	0.449
Média Móvel 4 dias	1.005	0.50	0.815
Média Móvel 5 dias	1.010	1.00	0.281
Média Móvel 6 dias	1.043	4.30	0.290
Média Móvel 7 dias	1.021	2.10	0.015

A Tabela 5 exibe as estimativas de efeitos simples e cumulativos do poluente O_3 . Por essa tabela, verifica-se que para um aumento de $13.04\mu\text{g}/\text{m}^3$ (intervalo interquartilico), a defasagem de 3 dias do poluente, está associado a um aumento de 1.13% no número médio de atendimentos diários por causas respiratórias em crianças de 0 e 6 anos. As exposições acumuladas de 3, 4, 5 e 7 dias também estão associados a aumentos de 3.2% e 4.9%, 4.6% e 3.7% no número médio de atendimentos diários por causas respiratórias em crianças de 0 e 6 anos, respectivamente. Os efeitos estimados de O_3 , com defasagem de 3 dias, e com exposição acumulada de 3, 4, 5 e 7 dias, são estatisticamente significativos ao nível de 10%. A Tabela 6 exibe as estimativas de efeitos simples e cumulativos do poluente SO_2 . Por essa tabela, verifica-se que para um aumento de $13.32\mu\text{g}/\text{m}^3$ (intervalo interquartilico), a defasagem de 2 dias do poluente, está associado a um aumento de 1.28% no número médio de atendimentos diários por causas respiratórias em crianças de 0 e 6 anos. As exposições acumuladas de 3, 5, 6 e 7 dias também estão associados a aumentos de 0.89% e 3.4%, 4.1% e 3.8% no número médio de atendimentos diários por causas respiratórias em crianças de 0 e 6 anos, respectivamente. Os efeitos estimados de NO_2 , com defasagem de 2 dias, e com exposição

Tabela 5: Poluente O_3 .

	RR	%RR	P-valor
Dia Corrente	0.981	-1.90	0.21
Defasagem de 1 dia	0.986	-1.37	0.21
Defasagem de 2 dia	1.011	1.13	0.95
Defasagem de 3 dia	1.013	1.30	0.01
Média Móvel 2 dias	1.001	0.10	0.13
Média Móvel 3 dias	1.032	3.20	0.05
Média Móvel 4 dias	1.049	4.90	0.01
Média Móvel 5 dias	1.046	4.60	0.00
Média Móvel 6 dias	1.003	0.25	0.14
Média Móvel 7 dias	1.037	3.70	0.02

Tabela 6: Poluente SO_2 .

	RR	%RR	P-valor
Dia Corrente	1.023	2.3	0.176
Defasagem de 1 dia	1.103	10.32	0.176
Defasagem de 2 dias	1.013	1.28	0.061
Defasagem de 3 dia	0.997	-0.3	0.823
Média Móvel 2 dias	1.013	1.25	0.801
Média Móvel 3 dias	1.009	0.89	0.014
Média Móvel 5 dias	1.034	3.4	0.000
Média Móvel 6 dias	1.041	4.1	0.050
Média Móvel 7 dias	1.038	3.8	0.010

acumulada de 3, 5, 6 e 7 dias, são estatisticamente significativos ao nível de 10%. A Tabela 7 exibe as estimativas de efeitos simples e cumulativos do poluente PM_{10} . Por essa tabela, verifica-se que para um aumento de $8.68\mu g/m^3$ (intervalo interquartilico), o dia corrente, e as defasagens de 1 e 3 dias do poluente, estão associados a um aumento de 3.9%, 5.1% e 3.99% no número médio de atendimentos diários por causas respiratórias em crianças de 0 e 6 anos, respectivamente. As exposições acumuladas de 2, 3, 4, 5 e 6 dias também estão associados a aumentos de 2.1%, 5.46%, 4.86%, 2.3% e 4.91% no número médio de atendimentos diários por causas respiratórias em crianças de 0 e 6 anos, respectivamente. Os efeitos estimados de NO_2 , com defasagens de 1 e 3 dias, e com exposição acumulada de 2, 3, 4, 5 e 6 dias, são estatisticamente significativos ao nível de 10%.

Tabela 7: Poluente PM_{10} .

	RR	%RR	P-valor
Dia Corrente	1.039	3.90	0.077
Defasagem de 1 dia	1.051	5.10	0.078
Defasagem de 3 dias	1.040	3.99	0.085
Média Móvel 2 dias	1.021	2.10	0.009
Média Móvel 3 dias	1.055	5.46	0.054
Média Móvel 4 dias	1.049	4.86	0.086
Média Móvel 5 dias	1.023	2.30	0.022
Média Móvel 6 dias	1.049	4.91	0.060

4. Conclusão

Este trabalho tem como objetivo apresentar uma metodologia de modelagem para dados de contagem, Modelo Autoregressivo Média Móvel Linear Generalizado ($GLARMA(p, q)$). Através do $GLARMA(p, q)$ e o Risco Relativo, verificou-se a relação existente entre os poluentes atmosféricos PTS , NO_2 , O_3 , SO_2 , PM_{10} e o número de atendimentos diários por causas respiratórias, entre 0 e 6 anos de idade, na RGV.

Os efeitos estimados dos poluentes foram significativamente associados com atendimentos hospitalares por causas respiratórias em crianças de 0 a 6 anos de idade com a seguinte relação: O PTS com defasagem simples de 3 dias (4.0%), média móvel de 4 dias (2.8%) e média móvel de 7 dias (0.9%). O NO_2 com a média móvel de 7 dias (2.10%). O poluente O_3 com defasagem simples de 3 dias (1.30%), com média móvel de 3 dias (3.20%), média móvel de 4 dias (4.90%), média móvel de 5 dias (4.60%) e com média móvel de 7 dias (3.70%). O SO_2 com defasagem simples de 2 dias (1.28%), com média móvel de 3 dias (0.89%), média móvel de 5 dias (3.4%), média móvel de 6 dias (4.1%) e com média móvel de 7 dias (3.8%). E o poluente PM_{10} com dia corrente (3.90%), defasagem simples de 1 dia (5.1%), defasagem simples de 3 dias (3.99%), média móvel de 2 dias (2.1%), média móvel de 3 dias (5.46%), média móvel de 4 dias (4.86%), média móvel de 5 dias (2.30%) e com média móvel de 6 dias (4.91%).

Estudos futuros podem ser desenvolvidos também visando empregar outras técnicas de modelagem como os Modelos Aditivos Generalizados (MAG) e o Modelo Valores Inteiros Autoregressivo (INAR).

Referências

Almeida, M. (2006). *Modelo Aditivo Generalizado (MAG) no estudo da relação entre o número de atendimentos hospitalares por causas respiratórias e a qualidade do ar*. PhD thesis, Dissertação de Mestrado, Vitória: Programa de Pós-Graduação em Engenharia Ambiental: Universidade Federal do Espírito Santo.

- Breslow, N. E. e Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American statistical Association*, 88(421):9–25.
- Davis, R. A., Dunsmuir, W. T., e Streett, S. B. (2003). Observation-driven models for poisson counts. *Biometrika*, 90(4):777–790.
- Davis, R. A., Dunsmuir, W., e Wang, Y. (1999). Modeling time series of count data. *STATISTICS TEXTBOOKS AND MONOGRAPHS*, 158:63–114.
- Gouveia, N., Mendonça, G. A., Leon, A. P. d., Correia, J. E. d. M., Junger, W. L., Freitas, C. U. d., Daumas, R. P., Martins, L. C., Giussepe, L., Conceição, G., et al. (2003). Poluição do ar e efeitos na saúde nas populações de duas grandes metrópoles brasileiras. *Epidemiologia e Serviços de Saúde*, 12(1):29–40.
- Hastie, T. J. e Tibshirani, R. J. (1990). Generalized additive models, volume 43 of monographs on statistics and applied probability.
- Jørgensen, B. (1983). Maximum likelihood estimation and large-sample inference for generalized linear and nonlinear regression models. *Biometrika*, 70(1):19–28.
- Lee, Y. e Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, p. 619–678.
- McCullagh, P. e Nelder, J. (1989). Generalised linear models. Chapman and Hall. London, UK.
- Nelder, J. A. e Baker, R. J. (1972). Generalized linear models. *Encyclopedia of Statistical Sciences*.
- Pantazopoulou, A., Katsouyanni, K., Kourekremastinou, J., e Trichopoulos, D. (1995). Short-term effects of air pollution on hospital emergency outpatient visits and admissions in the greater athens, greece area. *Environmental research*, 69(1):31–36.
- Schwartz, J. (1997). Air pollution and hospital admissions for cardiovascular disease in tucson. *Epidemiology*, p. 371–377.
- Schwartz, J. (2000). Harvesting and long term exposure effects in the relation between air pollution and mortality. *American journal of epidemiology*, 151(5):440–448.
- Shephard, N. et al. (1995). Generalized linear autoregressions. Technical report, Technical report, Nuffield College, Oxford University.
- Wedderburn, R. W. (1974). Quasi-likelihood functions, generalized linear models, and the gauss-newton method. *Biometrika*, 61(3):439–447.
- Zeger, S. L. e Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, p. 121–130.