UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO

**Modelos de séries temporais de dados de contagem
baseados na distribuição Poisson Dupla**

Davi Casale Aragon

Ribeirão Preto

2016

Davi Casale Aragon

Modelos de séries temporais de dados de contagem baseados na distribuição Poisson Dupla

Tese apresentada à Faculdade
de Medicina de Ribeirão
Preto da Universidade de São
Paulo para a obtenção do
título de Doutor em Ciências.
Programa: Saúde na Comuni-
dade.

Edson Zangiacomi Martinez

Orientador

Ribeirão Preto

2016

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

Aragon, Davi Casale

Modelos de séries temporais de dados de contagem baseados na distribuição Poisson Dupla

142 p.

Tese de Doutorado apresentada à Faculdade de Medicina de Ribeirão Preto - USP. Área de concentração: Saúde na Comunidade.

Orientador: Martinez, Edson Zangiacomi

1. Poisson dupla. 2. Séries temporais. 3. Métodos Bayesianos.

Folha de Aprovação

Davi Casale Aragon

Modelos de séries temporais de dados de contagem baseados na distribuição Poisson
Dupla

Tese apresentada à Faculdade de Medicina de Ribeirão
Preto da Universidade de São Paulo para a obtenção do
título de Doutor em Ciências.

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: _____

Dedico à Fernanda e ao pequeno Gabriel, com amor e gratidão, por compreenderem meus momentos de ausência e me apoiarem sempre, sendo os meus maiores incentivadores na elaboração deste trabalho e na vida. Obrigado por serem a minha família!

Agradecimentos

A Deus, por me fazer enxergar os caminhos corretos a serem percorridos nessa jornada.

Aos meus pais, pelos conselhos, apoio e amor incondicional.

Ao meu orientador, e grande amigo, Prof. Dr. Edson Zangiacomi Martinez, pelo incentivo, dedicação e paciência em me ensinar muito do que sei hoje. Qualquer agradecimento sempre será pouco para expressar minha gratidão.

Ao Prof. Dr. Jorge Alberto Achcar, pelas valiosas contribuições.

Aos meus irmãos, já doutores, pelo incentivo e eterna amizade.

Aos meus amigos, espalhados por tantas cidades, que me mostram que o tempo e a distância nunca nos fazem esquecer os bons momentos passados juntos.

“And in the end, the love you take is equal
to the love you make.”

Lennon/ McCartney

RESUMO

ARAGON, D. C. *Modelos de séries temporais de dados de contagem baseados na distribuição Poisson Dupla*. Ribeirão Preto, 2016, 142 p. Tese (Doutorado). Faculdade de Medicina de Ribeirão Preto. Universidade de São Paulo.

Dados de séries temporais são originados a partir de estudos em que se reportam, por exemplo, taxas de mortalidade, número de hospitalizações, de infecções por alguma doença ou outro evento de interesse, em períodos definidos (dia, semana, mês ou ano), objetivando-se observar tendências, sazonalidades ou fatores associados. Dados de contagem são aqueles representados pelas variáveis quantitativas discretas, ou seja, observações que assumem valores inteiros, no intervalo $\{0, 1, 2, 3, \dots\}$, por exemplo, o número de filhos de casais residentes em um bairro. Diante dessa particularidade, ferramentas estatísticas adequadas devem ser utilizadas, e modelos baseados na distribuição de Poisson apresentam-se como opções mais indicadas do que os baseados nos métodos propostos por Box e Jenkins (2008), usualmente utilizados para análise de dados contínuos, mas empregados para dados discretos, após transformações logarítmicas. Uma limitação da distribuição de Poisson é que ela assume média e variância iguais, sendo um obstáculo nos casos em que há superdispersão (variância maior que a média) ou subdispersão (variância menor que a média). Diante disso, a distribuição Poisson Dupla, proposta por Efron (1986), surge como alternativa, pois permite se estimarem os parâmetros de média e variância, nos casos em que a variância dos dados é menor, igual ou maior que a média, fornecendo grande flexibilidade aos modelos. Este trabalho teve como objetivo principal o desenvolvimento de modelos Bayesianos de séries temporais para dados de contagem, utilizando-se distribuições de probabilidade para variáveis discretas, tais como de Poisson e Poisson Dupla. Além disso, foi introduzido um modelo baseado na distribuição Poisson Dupla para dados de contagem com excesso de zeros. Os resultados obtidos pelo ajuste dos modelos de séries temporais baseados na distribuição Poisson Dupla foram comparados com aqueles obtidos por meio do uso da distribuição de Poisson. Como aplicações principais, foram apresentados resultados obtidos pelo ajuste de modelos para dados de registros de acidentes com picadas de cobras, no Estado de São Paulo, e picadas de escorpiões, na cidade de Ribeirão Preto, SP, entre os anos de 2007 e 2014. Com relação a esta última aplicação, foram consideradas covariáveis referentes a dados climáticos, como temperaturas máximas e mínimas médias mensais e precipitação. Nas situações em que a variância era diferente da média, modelos baseados na distribuição Poisson Dupla mostraram melhor ajuste aos dados, quando comparados aos modelos de Poisson.

Palavras-chave: Poisson Dupla; Séries Temporais; Métodos Bayesianos

ABSTRACT

ARAGON, D. C. *Count data time series models based on Double Poisson distribution*. Ribeirão Preto, 2016, 142 p. Thesis (Doctorate). Ribeirão Preto Medical School. University of São Paulo.

Time series data are derived from studies in which there are reported mortality, number of hospitalizations infections by disease or other event of interest per day, week, month or year, in order to observe trends, seasonality or associated factors. Count data are represented by discrete quantitative variables, i.e. observations that take integer values in the range $\{0, 1, 2, 3, \dots\}$. In view of this particular characteristic, such data must be analyzed by adequate statistical tools and the Poisson distribution is an option for modeling, being more suitable than models based on methods proposed by Box and Jenkins (2008), usually applied for continuous data, but used in the modeling of discrete data after logarithmic transformation. A limitation of the Poisson distribution is that it assumes equal mean and variance being an obstacle in cases which there are data overdispersion (variance higher than mean) or underdispersion (variance lower than mean). Therefore the Double Poisson distribution, proposed by Efron (1986), is an alternative because it allows to estimate the mean and variance parameters in cases with variance of the data is lower, equal, or higher than mean providing great flexibility to the models. This work aims to develop time series models for count data, under Bayesian approach using probability distributions for discrete variables such as Poisson and Double Poisson. Furthermore it will be introduced a zero-inflated Double Poisson model to excess zeros counting data. The results obtained by adjusting the time series models based on Double Poisson distribution are compared with those obtained by considering the Poisson distribution. As main applications modeling of snake bites reports in the State of São Paulo and scorpion stings in the city of Ribeirão Preto considering covariates as maximum and minimum average monthly temperatures and rainfall among the years 2007 and 2014 will be presented. Regression models based on double Poisson distribution showed a better fit to the data, when compared to Poisson models.

Keywords: Double Poisson; Time Series; Bayesian Methods

LISTA DE ABREVIATURAS E SIGLAS

AIC	Akaike Information Criterion
AR	Modelo Autorregressivo
ARMA	Modelo AR de Médias Móveis
ARIMA	Modelo AR Integrado de Média Móveis
CPO	Conditional Predictive Ordinate
DIC	Deviance Information Criterion
DP	Poisson Dupla
DVP	Desvio Padrão
EMV	Estimador de Máxima Verossimilhança
EP	Erro padrão
<i>IC</i> 95%	Intervalo de Confiança 95%
<i>ICr</i> 95%	Intervalo de Credibilidade 95%
ICPO	Inverse Conditional Predictive Ordinate
ICPV	Intervalo de Confiança do Perfil da Verossimilhança
LPML	Logarithm of the Pseudo Marginal Likelihood
MA	Moving Average - Modelo de Médias Móveis
MCMC	Markov Chain Monte Carlo
RV	Razão de Verossimilhanças
SARS	Síndrome Respiratória Aguda Grave
SD	Desvio Padrão
SARIMA	Modelo Sazonal AR Integrado de MA
SARIMAX	Modelo Sazonal AR Integrado de MA com Variável Exógena
SINAN	Sistema de Informação de Agravos de Notificação
ZIDP	Zero Inflated Double Poisson
ZIP	Zero Inflated Poisson

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Método de Box e Jenkins	14
1.2	Dados de Contagem	17
2	OBJETIVOS	19
3	JUSTIFICATIVA	20
4	MÉTODOS	21
4.1	Distribuição de Poisson	21
4.2	Distribuição Poisson Dupla	24
4.2.1	Estimadores de Máxima Verossimilhança	26
4.2.2	Estudo de Simulação - $c(\theta, \mu)$, $E(Y)$ e $Var(Y)$	30
4.2.3	Estimadores Bayesianos	33
4.2.4	Modelo de Regressão Baseado na Distribuição Poisson Dupla . .	36
4.2.5	Modelos Baseados na Distribuição Poisson Dupla para Dados com Excesso de Zeros	36
4.2.6	Comparações entre os Modelos	38
5	RESULTADOS	41
5.1	Simulações	41
5.2	Exemplos com Dados Reais	43
5.2.1	Registros de Picadas de Escorpiões em Ribeirão Preto, SP, Brasil	43
5.2.2	Registros de Picadas de Cobras em Benjamin Constant, AM, Brasil	47
5.2.3	Número de Visitas ao Médico durante o Primeiro Trimestre da Gravidez	47
5.2.4	Número de Bebês Nascidos de Mulheres Sobreviventes ao Câncer de Mama	49
5.3	Modelos de Séries Temporais de Dados de Contagem Baseados na Dis- tribuição Poisson Dupla	51

5.3.1	Registros de Acidentes com Cobras no Estado de São Paulo, entre 2007 e 2014	51
5.3.2	Registros de Acidentes com Escorpiões em Ribeirão Preto, São Paulo, entre 2007 e 2014	77
6	CONCLUSÃO	108
	REFERÊNCIAS BIBLIOGRÁFICAS	110
	APÊNDICES	116

1 INTRODUÇÃO

Dados de séries temporais, em saúde, são originados a partir de estudos em que se reportam taxas de mortalidade, número de hospitalizações ou de infecções por alguma doença, quantidades de acidentes e fraturas por causas específicas, número de nascidos vivos ou crianças vacinadas e taxas de dispensação de medicamentos, em períodos definidos (dia, semana, mês ou ano) (SHUMWAY e STOFFER, 2006).

A modelagem de dados em série faz-se importante na área da saúde pelo fato de, por meio dela, ser possível se observarem tendências, sazonalidades ou fatores associados à variável de interesse, permitindo aos gestores de saúde tomar medidas cautelares antecipadas ao saberem, por exemplo, que, se o volume de chuva, em uma cidade, ultrapassou um ponto crítico, haverá uma epidemia de dengue algum tempo após o evento. Ou, ainda, se a umidade do ar está abaixo de um certo valor, levará a uma superlotação dos postos de saúde, por pessoas com problemas respiratórios, nos dias seguintes.

Uma série temporal, observada como um conjunto de n variáveis aleatórias em tempos $t_1, t_2, \dots, t_n$, para qualquer n inteiro, é dada pela função de distribuição conjunta, avaliada como a probabilidade de os valores desta série serem menores do que as constantes $c_1, c_2, \dots, c_n$, ou seja,

$$F(c_1, c_2, \dots, c_n) = P(x_{t_1} < c_1, x_{t_2} < c_2, \dots, x_{t_n} < c_n).$$

Essa função de distribuição possui n argumentos, e qualquer gráfico das funções de distribuições multivariadas seria praticamente impossível. Assim, considerem-se funções de distribuições unidimensionais $F_t(x) = P(x_t < x)$, e a função densidade correspondente, $f_t(x) = \frac{\partial F_t(x)}{\partial x}$. A função da média é dada por:

$$\mu_{x_t} = E(x_t) = \int_{-\infty}^{\infty} x f_t(x) dx. \quad (1)$$

Uma série estritamente estacionária é aquela cujo comportamento probabilístico de um conjunto de valores $\{x_{t_1}, x_{t_2}, \dots, x_{t_k}\}$ é o mesmo observado em outro conjunto, com defasagem em h unidades no tempo, $\{x_{t_1+h}, x_{t_2+h}, \dots, x_{t_k+h}\}$. Assim, μ_{x_t} e $\sigma_{x_t}^2$ são constantes para todo t . Na prática, este conceito é considerado rigoroso. Portanto, define-se uma série fracamente estacionária, x_t , como um processo com variância finita, tal que:

- A função da média é definida como $\mu_{x_t} = E(x_t) = \mu$, para todo t .
- A função de autocovariância, dada por $\gamma_x(t, s) = E[(x_t - \mu_t)(x_s - \mu_s)]$, para quaisquer tempos t e s , depende de t e s apenas por meio de sua diferença $|t - s|$. A autocovariância mede a dependência linear entre dois pontos, na mesma série observada em diferentes tempos.

A função de autocorrelação é dada por

$$\rho(t, s) = \frac{\gamma_x(t, s)}{\sqrt{\gamma_x(t)\gamma_x(s)}}, \text{ em que } -1 \leq \rho(t, s) \leq 1,$$

em que $\gamma_x(t, s)$ é a função de autocovariância para quaisquer tempos t e s da série x ; $\gamma_x(t)$ e $\gamma_x(s)$ são as variâncias dos dados, nos tempos t e s , respectivamente. Essa função mede a correlação de uma predição linear de um valor x_t da série, a partir de um valor x_s . Assim, quando dois valores, x_t e x_s , da série, são correlacionados, afirma-se que existe uma autocorrelação de defasagem ("lag") $s - t$.

Podem-se, também, fazer previsões de uma série y_t a partir de outra série x_t . A função de covariância cruzada entre duas séries y_t e x_t é dada por

$$\gamma_{xy}(t, s) = E[(x_t - \mu_{x_s})(y_t - \mu_{y_s})],$$

e a função de autocorrelação cruzada é dada por

$$\rho_{xy}(t, s) = \frac{\gamma_{xy}(t, s)}{\sqrt{\gamma_x(t)\gamma_y(s)}}.$$

Uma série temporal é chamada de ruído branco quando é formada por uma coleção de variáveis aleatórias (w_t) não correlacionadas, com média zero e uma variância σ_w^2 , ou seja, as autocorrelações $\rho(t, s)$ são iguais a zero para quaisquer tempos t e s . Um caso particular é o ruído branco Gaussiano, em que se assume que a série é composta por variáveis aleatórias independentes e identicamente distribuídas por uma Normal com média zero e variância σ^2 .

1.1 Método de Box e Jenkins

O método mais usual para a análise de dados em série foi proposto por Box e Jenkins (2008). Ele é baseado em modelos que levam em conta variáveis dependentes contínuas, ruídos brancos Gaussianos, séries estacionárias (flutuam em torno de uma média constante), identificação de sazonalidade e componentes autorregressivos, por

meio de gráficos de funções de autocorrelações, antes da definição de qual o melhor modelo a ser ajustado.

Um modelo autorregressivo de ordem p , denotado por $AR(p)$, tem a seguinte forma:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + w_t,$$

em que x_t é uma série estacionária; $\phi_1, \phi_2, \dots, \phi_p$ são coeficientes autorregressivos; w_t é um ruído branco (ou erro aleatório) e $w_t \sim N(0, \sigma_w^2)$. Assim, neste modelo, assume-se que o valor atual da série x_t pode ser explicado pelos p valores passados ($x_{t-1}, x_{t-2}, \dots, x_{t-p}$).

Um modelo de médias móveis de ordem q , denotado por $MA(q)$, tem a seguinte forma:

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \dots + \theta_q w_{t-q},$$

em que x_t é uma série estacionária, na qual se têm q lags nas médias móveis; $\theta_1, \theta_2, \dots, \theta_q$ são parâmetros de médias móveis, e w_t é um ruído branco, com $w_t \sim N(0, \sigma_w^2)$. Assim, em um modelo de médias móveis de ordem q , o valor atual da série depende dos q erros aleatórios passados.

Considerando-se os modelos $AR(p)$ e $MA(q)$, pode-se definir um modelo autorregressivo de ordem p e médias móveis de ordem q , denotado por $ARMA(p, q)$:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \dots + \theta_q w_{t-q},$$

em que x_t é uma série estacionária; $\phi_1, \phi_2, \dots, \phi_p$ são coeficientes autorregressivos; $\theta_1, \theta_2, \dots, \theta_q$ são parâmetros de médias móveis; w_t é um ruído branco, com $w_t \sim N(0, \sigma_w^2)$.

Muitos dados podem compor séries temporais que não são totalmente estacionárias, e o uso de modelos autorregressivos de médias móveis não é adequado para ajustá-los. Assim, é preciso diferenciarem-se d vezes a série, de tal maneira que esta se torne estacionária. A série diferenciada é denotada por:

$$y_t = \nabla^d x_t = (1 - B)^d x_t,$$

em que x_t é a série não estacionária; y_t é a série diferenciada a partir de x_t ; $B^k x_t = x_{t-k}$ é o operador de defasagem para $k > 1$; $\nabla^d = (1 - B)^d$ é a diferença integrada de ordem

d . Nota-se que, se $d = 0$, tem-se um modelo $ARMA(p, q)$. Assim, um modelo autorregressivo integrado de médias móveis de ordem (p, d, q) , denotado por $ARIMA(p, d, q)$, é escrito como:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + w_t + \beta_1 w_{t-1} + \beta_2 w_{t-2} + \dots + \beta_q w_{t-q}.$$

Como exemplo, Jiang et al. (2014), utilizando dados coletados na Austrália, entre 1935 e 2006, encontraram uma associação entre o consumo per capita de álcool com a mortalidade por doenças hepáticas. Ramirez et al. (2014) utilizaram um modelo ARIMA, sob um enfoque proposto por Box e Jenkins (2008), para apresentar o comportamento dos casos de morbidade e mortalidade por malária, na Colômbia, entre 1990 e 2011. Earnest et al. (2005) utilizaram o modelo ARIMA para predizerem e monitorarem a quantidade de leitos ocupados em um hospital de Singapura, durante uma epidemia de SARS, entre março e abril de 2003. Os autores concluíram que o modelo é capaz de auxiliar os gestores de saúde em outras situações de epidemias.

Liu et al. (2011) utilizaram o modelo ARIMA para fazerem previsões de casos de febre hemorrágica com síndrome renal, a partir de casos obtidos entre 1975 e 2008. Os autores concluíram que os resultados foram muito importantes para a vigilância epidemiológica, e esperava-se um aumento dos casos da doença nos anos posteriores. Além disso, Razvodovsky (2015) associou as mudanças no consumo de vários tipos de bebidas alcoólicas com a taxa de incidência de psicoses causadas pelo álcool, no período de 1970 e 2012. Para tal, ajustou modelo ARIMA e concluiu que o alto consumo de vodka está associado à incidência de psicoses alcoólicas.

Existem modelos que contemplam o comportamento sazonal (cíclico) da série temporal não estacionária. São os modelos multiplicativos sazonais autorregressivos integrados de médias móveis, denotados por $SARIMA(p, d, q) \times (P, D, Q)_s$, em que s é o componente sazonal. Por exemplo, se $s = 12$, então o comportamento da série tem uma tendência semelhante a cada 12 pontos observados no tempo.

Vários autores investigaram o padrão temporal da disseminação da dengue, em diferentes populações, utilizando modelos SARIMA (WONGKOON et al., 2007; SILAWAN et al., 2008; GHARBI et al., 2011; MARTINEZ e SILVA, 2011 e MARTINEZ et al., 2011). Em todas essas aplicações, os autores obtiveram predições satisfatórias para a incidência de dengue e sugeriram que esses modelos podem ser utilizados para tomada de decisões na vigilância da doença e gerenciamento de risco. Zhang et al. (2015), entre 2000 e 2012, estudaram o comportamento do número mensal de mortes por acidentes de carro na China e, utilizando um modelo SARIMA, concluíram que as taxas vêm caindo, mas apresentam um padrão sazonal, com mais mortes nos últimos

3 meses do ano. Bras et al. (2014) encontraram um padrão sazonal na incidência de tuberculose em Portugal, com um pico no número de casos no mês de março e uma queda em dezembro. Esses autores utilizaram um modelo SARIMA, estratificando as análises por sexo, idade e áreas com baixas e altas incidências.

Quando se faz necessária a incorporação de covariáveis para se explicar uma série de dados, o modelo ARIMAX surge como uma opção. Ele permite que se utilize um componente linear que contempla as informações de variáveis exógenas. Este modelo também permite a inclusão do componente sazonal e passa a ser chamado de SARIMAX. Lee et al. (2013) utilizaram um modelo SARIMAX para avaliar a incidência de brucelose em pessoas residentes na Coreia do Sul, entre os anos de 2005 e 2010. Como covariável, os autores consideraram a taxa de incidência da doença nos rebanhos e concluíram que ela está diretamente associada com o aumento do número de infecções nos humanos. E, ainda, Gao et al. (2014) publicaram um estudo no qual avaliaram a influência de variáveis climáticas, como temperaturas mínima e máxima, umidade relativa do ar, pressão do ar, pluviosidade e velocidade do vento, na incidência de casos de disenteria bacilar na cidade de Changsha, China, entre os anos de 2004 e 2010, ajustando um modelo SARIMAX. Chadsuthi et al. (2012), utilizando um modelo SARIMAX, associaram a média mensal de temperatura e pluviosidade com a incidência de leptospirose na Tailândia, entre 2003 e 2010. Modarres et al. (2012), publicaram um estudo, utilizando um modelo SARIMAX, relacionando a sazonalidade das incidências de fraturas de quadril, em idosos canadenses, entre 1985 e 2005, com variáveis climáticas. Os autores puderam verificar uma relação entre períodos frios e maiores taxas de incidências de fraturas.

1.2 Dados de Contagem

Dados de contagem são aqueles representados pelas variáveis quantitativas discretas, ou seja, observações que assumem valores inteiros, no conjunto $\{0, 1, 2, 3, \dots\}$. São exemplos desse tipo de dado o número de filhos de casais residentes em um bairro, número de acidentes de trânsito em uma região, em dado período, número de bactérias em amostras de alimentos para análise, número de internações hospitalares por infarto do miocárdio, cigarros consumidos por indivíduos, em um dia. Diante dessa característica particular, esse tipo de dado deve ser analisado por ferramentas estatísticas adequadas.

Alguns autores introduziram modelos de séries temporais baseados em distribuições discretas, visando analisar a relação entre a incidência semanal ou mensal de dengue e as variáveis climáticas (LU et al., 2009; PINTO et al., 2011 e HII et al., 2012). Arbex et al. (2014) publicaram um estudo que associou o número de internações à variação diária do total de partículas suspensas no ar, durante o plantio de cana-de-

açúcar, com períodos com e sem queimadas, e a incidência de visitas a um importante hospital da cidade de Araraquara, SP, cujo motivo principal era o diagnóstico de pneumonia. Com base em resultados originados pelo ajuste de um modelo de regressão de Poisson, os autores concluíram que, nos períodos de queimadas, o hospital recebe um maior número de pessoas com o diagnóstico da doença. Usualmente, estes modelos consideram que o número de casos reportados em cada intervalo de tempo segue uma distribuição de Poisson e permitem a especificação de estruturas autorregressivas que levam em conta a autocorrelação entre as sucessivas unidades de tempo.

Uma limitação da abordagem proposta por Box e Jenkins é o fato de os modelos, originalmente, poderem ser ajustados apenas para variáveis contínuas e, por esse motivo, muitas vezes, variáveis discretas são transformadas em logaritmos. Essa prática, além de ser utilizada para se tentar tornar a série estacionária, faz com que uma variável discreta possa ser utilizada como contínua, caso a amplitude dos dados seja grande.

2 OBJETIVOS

Este trabalho tem como objetivo principal desenvolver modelos de séries temporais para dados de contagem, sob enfoque Bayesiano, utilizando-se a distribuição de probabilidade Poisson Dupla, introduzida por Efron (1986), para variáveis discretas.

Os objetivos específicos são:

- Introduzir um modelo baseado na distribuição Poisson Dupla, para dados de contagem com excesso de zeros.
- Comparar os resultados obtidos por meio do ajuste dos modelos de séries temporais, baseados na distribuição Poisson Dupla, com aqueles obtidos por meio do uso da distribuição de Poisson.
- Modelar dados de registros de acidentes com picadas de cobras no Estado de São Paulo, entre os anos de 2007 e 2014, utilizando-se as distribuições Poisson Dupla e Poisson e comparando-se os resultados dos ajustes.
- Modelar dados de registros de picadas de escorpiões na cidade de Ribeirão Preto, SP, entre os anos de 2007 e 2014, considerando-se covariáveis, como temperaturas máximas e mínimas médias mensais e precipitação, utilizando-se as distribuições Poisson Dupla e Poisson e comparando-se os resultados por meio dos respectivos ajustes.

3 JUSTIFICATIVA

Dentre as vantagens do uso de modelos de Poisson, em séries temporais de dados de contagens, estão a maior flexibilidade para se introduzirem covariáveis e a simplicidade na interpretação do seu único parâmetro (que descreve tanto a média quanto a variância). Uma desvantagem é que, como a média e a variância são iguais, dados com superdispersão (*overdispersion*) ou subdispersão (*underdispersion*) não são ajustados satisfatoriamente por tais modelos.

Diante dessa limitação, o modelo baseado na distribuição Poisson Dupla apresenta-se como opção. Essa distribuição possui um parâmetro adicional, que permite ajustarem-se dados com superdispersão, subdispersão e dados que poderiam ser ajustados por meio de modelos de Poisson, com média igual à variância, mostrando grande flexibilidade para análise de dados de contagem com diferentes variabilidades.

Além disso, poucos trabalhos são encontrados, na literatura, sobre a distribuição Poisson Dupla. O trabalho original de Efron (1986) traz, somente, uma breve descrição da distribuição, sendo necessários desenvolvimentos mais específicos, obtendo-se estimadores de máxima verossimilhança, bem como seus erros padrão para obtenção de intervalos de confiança, estimadores bayesianos e mesmo uma extensão dessa distribuição, para acomodar dados com excesso de zeros.

Esta tese demonstra, a partir de diferentes conjuntos de dados, a utilidade da distribuição Poisson Dupla na análise estatística de dados epidemiológicos.

4 MÉTODOS

4.1 Distribuição de Poisson

A distribuição de Poisson foi desenvolvida por Siméon-Denis Poisson, quando estudava a distribuição binomial, e foi publicada no trabalho intitulado *Recherches sur la probabilité des jugements en matières criminelles et matière civile* ("Inquérito sobre a probabilidade em julgamentos sobre matérias criminais e civis"), em 1837.

Uma variável aleatória Y segue uma distribuição de Poisson com parâmetro λ (denota-se $Y \sim \text{Poisson}(\lambda)$), se ela assume valores inteiros, no intervalo $\{0, 1, 2, 3, \dots\}$ e possui função de probabilidade dada por:

$$P(Y = y) = \frac{\lambda^y e^{-\lambda}}{y!}, \text{ para } \lambda > 0.$$

A esperança de Y , denotada por $E[Y]$, é dada por:

$$E[Y] = \sum_{y=0}^{\infty} y P(Y = y) = \sum_{y=0}^{\infty} y \frac{\lambda^y e^{-\lambda}}{y!} = 0 \frac{\lambda^0 e^{-\lambda}}{0!} + 1 \frac{\lambda^1 e^{-\lambda}}{1!} + 2 \frac{\lambda^2 e^{-\lambda}}{2!} + \dots$$

Observar que

$$E[Y] = \sum_{y=1}^{\infty} \lambda \frac{\lambda^{y-1} e^{-\lambda}}{(y-1)!} = \lambda e^{-\lambda} \sum_{y=1}^{\infty} \frac{\lambda^{y-1}}{(y-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda,$$

dado que

$$\sum_{y=1}^{\infty} \frac{\lambda^{y-1}}{(y-1)!} = e^{\lambda}.$$

A respectiva variância é dada por:

$$\begin{aligned}
Var[Y] &= E[Y^2] - [E(Y)]^2 = E[Y^2 - Y] + E[Y] - [E(Y)]^2 \\
&= \left[\sum_{y=0}^{\infty} y(y-1)P(Y=y) \right] + \lambda - \lambda^2 \\
&= \lambda - \lambda^2 + \lambda^2 e^{-\lambda} \sum_{y=1}^{\infty} \frac{\lambda^{y-2}}{(y-2)!} = \lambda - \lambda^2 + \lambda^2 = \lambda.
\end{aligned}$$

A função de verossimilhança para n observações independentes e identicamente distribuídas $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ é dada por:

$$L(\lambda) = \prod_{i=1}^n \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} = \frac{\lambda^{\sum_{i=1}^n y_i} e^{-n\lambda}}{\prod_{i=1}^n y_i!}$$

e, aplicando-se o logaritmo natural, tem-se

$$l(\lambda) = \ln[L(\lambda)] = -n\lambda + \sum_{i=1}^n y_i \ln(\lambda) - \sum_{i=1}^n \ln(y_i!), \quad \text{para } \lambda > 0.$$

O estimador de máxima verossimilhança (EMV) para λ é dado pelo valor do parâmetro que maximiza a expressão de $L(\lambda)$. Assim, de

$$\frac{\partial l(\lambda)}{\partial \lambda} = -n + \frac{1}{\lambda} \sum_{i=1}^n y_i = 0,$$

temos que o EMV é dado por

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n Y_i = \bar{Y}.$$

Para a estimação bayesiana de λ , será considerada uma distribuição a priori Gama com hiperparâmetros $a > 0$ e $b > 0$, denotada por $\lambda \sim Gama(a, b)$, sendo a e b conhecidos. Assim,

$$\pi(\lambda) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda} \propto \lambda^{a-1} e^{-b\lambda}.$$

Do teorema de Bayes, tem-se que a distribuição a posteriori para λ é dada

por

$$\pi(\lambda \mid \mathbf{y}) \propto \pi(\lambda)L(\lambda) \propto \lambda^{a+\sum_{i=1}^n y_i-1}e^{-(b+n)\lambda}.$$

Notar que $\pi(\lambda)$ é uma distribuição conjugada para a função de verossimilhança $L(\lambda)$, dado que a distribuição a posteriori $\pi(\lambda \mid \mathbf{y})$ também é escrita na forma da função densidade de probabilidade de uma distribuição Gama, ou seja, $\lambda \mid \mathbf{y} \sim \text{Gama}(a + \sum_{i=1}^n y_i, b + n)$. Neste caso, o estimador bayesiano para λ , dado pela média da distribuição a posteriori, é dado por

$$\hat{\lambda}_{Bayes} = \frac{a + \sum_{i=1}^n Y_i}{b + n}.$$

Se a e b são próximos a zero, o EMV para λ e o respectivo estimador bayesiano são próximos.

A priori de Jeffreys (JEFFREYS, 1946) é bastante útil quando não se tem conhecimento sobre os parâmetros a serem estimados, sendo não informativa e, portanto, permite a comparação de resultados provenientes da modelagem frequentista, que só utiliza informações da amostra. Ela é dada pela raiz quadrada do determinante da matriz de informação de Fisher. Assim, considerando-se o parâmetro λ , em toda reta, tem-se que

$$I(\theta) = E \left\{ \left[\frac{d \ln L(\lambda)}{d\lambda} \right]^2 \right\}.$$

Para qualquer transformação ψ , um a um, de λ , tem-se

$$I(\psi) = I[\lambda(\psi)] \left(\frac{d\lambda}{d\psi} \right)^2.$$

Assim, a variância é constante na aproximação assintótica da distribuição a posteriori de ψ , ou seja, $I(\psi)$ é constante. Portanto,

$$\left(\frac{d\lambda}{d\psi} \right)^2 = I^{-1}[\lambda(\psi)] \quad \text{e} \quad \frac{d\lambda}{d\psi} = I^{-\frac{1}{2}}[\lambda(\psi)].$$

Nessa parametrização, a função de verossimilhança só se altera, em locação, para amostras diferentes e de mesmo tamanho, portanto, uma priori não-informativa para ψ é dada por

$$\pi_{\psi}(\psi) \propto \text{constante}$$

e, na parametrização em λ , tem-se que uma distribuição a priori não informativa correspondente é dada por

$$\begin{aligned}\pi_\lambda(\lambda) &= I[\lambda(\psi)] \left[\frac{d\psi}{d\lambda} \right] \\ \pi_\lambda(\lambda) &\propto \text{constante} \left[\frac{d\psi}{d\lambda} \right].\end{aligned}$$

Portanto, considerando que $\frac{\partial\psi}{\partial\lambda} = I^{\frac{1}{2}}(\lambda)$, uma distribuição a priori não informativa de Jeffreys para λ , é dada por

$$\pi(\lambda) = I^{\frac{1}{2}}(\lambda).$$

Assim,

$$E \left[-\frac{\partial l(\lambda)}{\partial \lambda^2} \right] = E \left[\frac{1}{\lambda^2} \sum_{i=1}^n Y_i \right] = \frac{n\lambda}{\lambda^2} = \frac{n}{\lambda},$$

e a distribuição a priori de Jeffreys para λ é

$$\pi(\lambda) \propto \sqrt{\frac{1}{\lambda}}.$$

Notar que essa distribuição é imprópria, dado que $\lambda^{-\frac{1}{2}}$ não é integrável no intervalo $[0, \infty)$. Esta distribuição de Jeffreys pode ser representada por $\lambda \sim Gama\left(\frac{1}{2}, 0\right)$ o que não é, a rigor, uma distribuição de probabilidade. Entretanto, ainda que imprópria, leva a uma distribuição a posteriori para λ que é própria, tal que $\lambda \mid \mathbf{y} \sim Gama\left(\frac{1}{2} + \sum_{i=1}^n y_i, n\right)$, sendo que o respectivo estimador bayesiano é dado por

$$\hat{\lambda}_{Bayes} = \frac{\frac{1}{2} + \sum_{i=1}^n y_i}{n}.$$

4.2 Distribuição Poisson Dupla

Efron (1986) introduziu a família dupla exponencial de distribuições, permitindo operações com parâmetros de média e de dispersão com as propriedades da família exponencial. Essa família inclui, como um caso especial, a distribuição Poisson Dupla, cuja vantagem é permitir a modelagem de dados com superdispersão e subdispersão. Essa distribuição envolve uma aproximação na sua função massa de probabilidade, tal que a soma das probabilidades não é exatamente um, embora essa diferença não

seja tão grande para quaisquer valores dos parâmetros. (CAMERON e JOHANSSON, 1997).

Uma família dupla exponencial com parâmetros θ , μ e n foi definida por Efron (1986), de uma maneira geral, como:

$$f_{\theta,\mu,n}(y) = c(\theta, \mu, n) \sqrt{\theta} [g_{\mu,n}(y)]^\theta [g_{y,n}(y)]^{1-\theta} [dG_n(y)], \quad (2)$$

em que $c(\theta, \mu, n)$ é uma constante normalizadora escolhida, para que a densidade integrada resulte no valor 1. Assim,

$$g_{\mu,n}(y) = \exp \{n[\eta y - \psi(\mu)]\} [dG_n(y)]$$

é uma família exponencial de funções densidades de um parâmetro, $\psi(\mu)$ é uma função normalizadora, η é uma função monótona de μ ; n é o tamanho amostral, e $G_n(y)$ é um operador para a família exponencial, que satisfaz a expressão

$$P_\mu\{A\} = \int_A g_{\mu,n}(y) dG_n(y)$$

para conjuntos mensuráveis A . Considere-se que $g_{\mu,n}(y)$ segue uma distribuição de Poisson, ou seja,

$$g_{\mu,n}(y) = g_\mu(y) = \frac{e^{-\mu} \mu^y}{y!}, \quad (3)$$

para $y = 0, 1, 2, \dots$. Efron (1986) argumenta que n pode ser descartado de (3) desde que a família Poisson seja fechada em convoluções e $g_{\mu,n}(y)$ seja a mesma família para todos os valores de n . Considerando que

$$g_{y,n}(y) = \frac{e^{-y} y^y}{y!},$$

a expressão (2) resulta em

$$f_{\theta,\mu}(y) = c(\theta, \mu) \sqrt{\theta} \left[\frac{e^{-\mu} \mu^y}{y!} \right]^\theta \left[\frac{e^{-y} y^y}{y!} \right]^{1-\theta},$$

para $y = 0, 1, 2, \dots$. Assim, a função massa de probabilidade de uma variável aleatória Y que segue uma distribuição Poisson Dupla (DP) pode ser escrita da forma:

$$P(Y = y) = c(\theta, \mu) e^{-\theta\mu} \sqrt{\theta} \frac{e^{y(\theta-1)} y^y}{y!} \left(\frac{\mu}{y} \right)^{\theta y}, \quad (4)$$

$y = 0, 1, 2, \dots$, em que $\theta > 0$, $\mu > 0$ e $c(\theta, \mu)$ é uma constante normalizadora que

assegura que $P(Y = y)$ some 1. Efron (1986) obteve uma aproximação para esta constante, utilizando uma expansão de Edgeworth, dada por:

$$\frac{1}{c(\theta, \mu)} \cong 1 + \frac{1 - \theta}{12\theta\mu} \left(1 + \frac{1}{\theta\mu}\right). \quad (5)$$

Segundo Efron (1986), essa constante pode ser aproximada, satisfatoriamente, para o valor 1, para quaisquer valores de θ e $\mu > 1$.

Assumindo-se essa aproximação, tem-se que

$$E(Y) \cong \mu \quad \text{e} \quad Var(Y) \cong \frac{\mu}{\theta}.$$

Assim, a distribuição descreve dados com superdispersão se $\theta < 1$ e, com subdispersão, se $\theta > 1$. Se $\theta = 1$, a expressão (4) corresponde à função massa de probabilidade de uma variável aleatória que segue uma distribuição Poisson com média e variância iguais a μ .

A distribuição Poisson Dupla pertence à família exponencial de distribuições com dois parâmetros, expressa, de forma geral, como:

$$f_{\theta, \mu}(y) = h(y)d(\theta, \mu) \exp[\omega_1(\theta, \mu)t_1(y) + \omega_2(\theta, \mu)t_2(y)],$$

em que $h(y) = 1/y!$ é uma função de y , $d(\theta, \mu) = c(\theta, \mu)e^{-\theta\mu}\sqrt{\theta} \geq 0$, $\omega_1(\theta, \mu) = \theta - 1 + \theta \ln \mu$ e $\omega_2(\theta, \mu) = 1 - \theta$ são funções dos parâmetros θ e μ , $t_1(y) = y$ e $t_2(y) = y \ln y$ são funções de y e não dependem de θ e μ . Consequentemente, $T_1 = \sum_{i=1}^n Y_i$ e $T_2 = \sum_{i=1}^n Y_i \ln Y_i$ são estatísticas suficientes para θ e μ .

4.2.1 Estimadores de Máxima Verossimilhança

Considerando-se uma amostra aleatória de tamanho n , a função de verossimilhança para θ e μ dada é por:

$$L(\mu, \theta) = [c(\theta, \mu)]^n e^{-n\theta\mu} \theta^{n/2} e^{(\theta-1) \sum_{i=1}^n y_i} \mu^{\sum_{i=1}^n y_i} \prod_{i=1}^n \frac{y_i^{y_i(1-\theta)}}{y_i!},$$

e, aplicando-se o logaritmo, tem-se

$$\begin{aligned}\ln L(\mu, \theta) &= n \ln [c(\theta, \mu)] - n\theta\mu + \frac{n}{2} \ln \theta + (\theta - 1) \sum_{i=1}^n y_i \\ &\quad + (1 - \theta) \sum_{i=1}^n y_i \ln y_i - \sum_{i=1}^n \ln y_i! + \theta (\ln \mu) \sum_{i=1}^n y_i.\end{aligned}\quad (6)$$

Estimadores de máxima verossimilhança (EMVs) para θ e μ são obtidos a partir da maximização de $\ln L(\mu, \theta)$. Normalmente, $c(\theta, \mu)$ é descartada, dado que pode ser uma fonte significativa de não-linearidade (CAMERON e TRIVEDI, 2013). Assim, derivando-se a função log-verossimilhança (6) em relação a θ e μ , têm-se as seguintes equações:

$$\frac{\partial}{\partial \theta} \ln L(\mu, \theta) = \frac{n}{2\theta} - \sum_{i=1}^n (\ln y_i) y_i + \sum_{i=1}^n y_i - n\mu + (\ln \mu) \sum_{i=1}^n y_i$$

e

$$\frac{\partial}{\partial \mu} \ln L(\mu, \theta) = \frac{\theta}{\mu} \sum_{i=1}^n y_i - n\theta.$$

A partir dessas equações, obtêm-se os seguintes estimadores de máxima verossimilhança:

$$\hat{\mu}_{ML} = \frac{1}{n} \sum_{i=1}^n y_i = \bar{y}$$

e

$$\hat{\theta}_{ML} = \frac{1}{2 \left[\sum_{i=1}^n \frac{y_i (\ln y_i)}{n} - \bar{y} (\ln \bar{y}) \right]}.$$

Como esperado, esses estimadores são funções das estatísticas suficientes $T_1 = \sum_{i=1}^n Y_i$ e $T_2 = \sum_{i=1}^n Y_i \ln Y_i$ para θ e μ . Como uma observação y , da variável aleatória Y , pode ser igual a zero, a expressão para $\hat{\theta}_{ML}$ traz um inconveniente devido à presença do termo $\ln y_i$. Entretanto, considerando-se que

$$\lim_{y \rightarrow 0} y (\ln y) = 0,$$

pode-se assegurar que a aproximação $y_i (\ln y_i)$ é igual a 0 quando $y_i = 0$ para todo $i = 1, \dots, n$. As variâncias assintóticas dos EMVs são dadas pelos elementos da matriz de informação de Fisher inversa $I(\mu, \theta)$. A segunda derivada parcial da função log-

verossimilhança é dada por:

$$\frac{\partial^2}{\partial \theta^2} \ln L(\mu, \theta) = -\frac{n}{2\theta^2},$$

$$\frac{\partial^2}{\partial \mu^2} \ln L(\mu, \theta) = -\frac{\theta}{\mu^2} \sum_{i=1}^n y_i$$

e

$$\frac{\partial^2}{\partial \mu \partial \theta} \ln L(\mu, \theta) = -n + \frac{1}{\mu} \sum_{i=1}^n y_i.$$

Substituindo-se os parâmetros pelos seus EMVs correspondentes, tem-se a matriz observada 2×2 de informação de Fisher, $I(\hat{\mu}, \hat{\theta})$, dada por:

$$I(\hat{\mu}, \hat{\theta}) = \begin{bmatrix} \frac{n}{2\bar{y}} \left[\sum_{i=1}^n \frac{y_i (\ln y_i)}{n} - \bar{y} (\ln \bar{y}) \right]^{-1} & 0 \\ 0 & n \left[\sum_{i=1}^n \frac{y_i (\ln y_i)}{n} - \bar{y} (\ln \bar{y}) \right] \end{bmatrix}. \quad (7)$$

Dado que a diagonal secundária da matriz $I(\hat{\mu}, \hat{\theta})$ possui elementos iguais a zero, pode-se afirmar que os EMVs, $\hat{\mu}_{ML}$ e $\hat{\theta}_{ML}$ são assintoticamente independentes. Considerando-se que a matriz de variância-covariância pode ser aproximada pela inversa de (7), os erros padrão estimados para $\hat{\mu}_{ML}$ e $\hat{\theta}_{ML}$, são, respectivamente,

$$\hat{e}p(\hat{\mu}_{ML}) = \sqrt{\frac{2\bar{y}}{n} \left[\sum_{i=1}^n \frac{y_i (\ln y_i)}{n} - \bar{y} (\ln \bar{y}) \right]} = \sqrt{\frac{\bar{y}}{n\hat{\theta}_{ML}}}$$

e

$$\hat{e}p(\hat{\theta}_{ML}) = \sqrt{\frac{1}{n} \left[\sum_{i=1}^n \frac{y_i (\ln y_i)}{n} - \bar{y} (\ln \bar{y}) \right]^{-1}} = \sqrt{\frac{2\hat{\theta}_{ML}}{n}}.$$

Intervalos de confiança $100(1 - \alpha)\%$, aproximados, para μ e θ são dados, respectivamente, por:

$$\hat{\mu}_{ML} \pm z_{\alpha/2} \sqrt{\frac{\bar{y}}{n\hat{\theta}_{ML}}} \quad \text{e} \quad \hat{\theta}_{ML} \pm z_{\alpha/2} \sqrt{\frac{2\hat{\theta}_{ML}}{n}},$$

em que $z_{\alpha/2}$ é o $100(\alpha)$ -ésimo percentil de uma distribuição Normal padronizada. Esses intervalos são comumente chamados de intervalos de confiança de Wald.

Uma desvantagem desses estimadores é que eles não são resultados exatos, pois a constante normalizadora $c(\theta, \mu)$ é considerada igual a 1. Ao mesmo tempo, se $c(\theta, \mu)$ for incluída na expressão (4), as primeiras derivadas de $\log L(\mu, \theta)$ com relação a θ e μ são, respectivamente:

$$\frac{\partial}{\partial \theta} \ln L(\mu, \theta) = \frac{n}{2\theta} - \sum_{i=1}^n (\ln y_i) y_i + \sum_{i=1}^n y_i - n\mu + (\ln \mu) \sum_{i=1}^n y_i - \frac{ne^{-\theta\mu} (2\theta\mu - 1)}{2\sqrt{\theta} - 2\theta e^{-\theta\mu}}$$

e

$$\frac{\partial}{\partial \mu} \ln L(\mu, \theta) = \frac{\theta}{\mu} \sum_{i=1}^n y_i - n\theta + \frac{n\theta^{\frac{3}{2}} e^{-\theta\mu}}{\sqrt{\theta} e^{-\theta\mu} - 1}.$$

Igualando-se essas expressões a zero, suas soluções resultam nos EMVs. Nota-se que, neste caso, não é possível se obterem expressões explícitas para $\hat{\mu}_{ML}$ e $\hat{\theta}_{ML}$, e métodos numéricos iterativos, como o de Newton-Raphson, são necessários para se encontrarem os EMVs. Assim, utilizando-se rotinas disponíveis em softwares, como SAS ou R, por exemplo, podem-se encontrar os EMVs com relativa facilidade.

Intervalos de confiança para μ e θ também podem ser obtidos pelo perfil da verossimilhança (ICPV), como descrito por Millar (2011). Esse método consiste em se inverter o teste de razão de verossimilhanças para obtenção o intervalo de confiança para o parâmetro de interesse. Assim, um intervalo aproximado $100(1 - \alpha)\%$, para μ , por exemplo, é definido como um conjunto de valores tais, que, em um teste bicaudal de hipóteses, com hipótese nula $H_0 : \mu = \mu_0$ (μ_0 conhecido), esta não seria rejeitada a um nível de significância fixado em α . Os ICPV não assumem normalidade do estimador, mas são baseados em aproximações assintóticas, dado que consideram que a estatística de razão de verossimilhanças (RV), dada por

$$RV_{\mu|\theta} = -2 \ln \frac{L(\hat{\mu}_{ML}, \hat{\theta}_{ML})}{L(\mu_0, \hat{\theta}_{ML})}$$

converge em distribuição para uma variável aleatória seguindo uma qui-quadrado central com um grau de liberdade. Assim, rejeita-se H_0 a um nível de significância α se $RV_{\mu|\theta} > \chi_1^2(1 - \alpha)$, em que $\chi_1^2(1 - \alpha)$ é o $100(1 - \alpha)$ -ésimo percentil superior da

distribuição qui-quadrado com um grau de liberdade. Portanto, o $100(1 - \alpha)\%$ ICPV para μ é dado por todos os valores μ que satisfazem

$$\ln L(\mu, \hat{\theta}_{ML}) \geq \ln L(\hat{\mu}_{ML}, \hat{\theta}_{ML}) - \frac{1}{2}\chi_1^2(1 - \alpha)$$

e, analogamente, o $100(1 - \alpha)\%$ ICPV para θ é dado por todos os valores de θ que satisfazem

$$\ln L(\hat{\mu}_{ML}, \theta) \geq \ln L(\hat{\mu}_{ML}, \hat{\theta}_{ML}) - \frac{1}{2}\chi_1^2(1 - \alpha).$$

Em algumas situações, o ICPV pode ser uma melhor opção aos intervalos de confiança de Wald, pois eles não incluem valores que extrapolam o espaço paramétrico.

4.2.2 Estudo de Simulação - $c(\theta, \mu)$, $E(Y)$ e $Var(Y)$

Por se tratarem de aproximações, justifica-se um estudo sobre a constante $c(\theta, \mu)$ e as quantidades $E(Y)$ e $Var(Y)$.

Assim, sendo Y uma variável aleatória que assume valores discretos, no intervalo $0, 1, 2, 3, \dots$, considerando-se $c(\theta, \mu)$ e a função de probabilidade dada por (4), para $y = 0, 1, 2, 3, \dots$, pode-se escrever a seguinte expressão:

$$c(\theta, \mu)^{-1} = e^{-\theta\mu} \sqrt{\theta} \sum_{y=0}^{\infty} \frac{e^{y(\theta-1)} y^y}{y!} \left(\frac{\mu}{y}\right)^{\theta y},$$

isto é,

$$c(\theta, \mu)^{-1} = \frac{\sqrt{\theta}}{e^{\theta\mu}} \left[1 + \sum_{y=1}^{\infty} \left(\frac{y\mu^{\theta} e^{(\theta-1)}}{y^{\theta}} \right)^y \frac{1}{y!} \right]. \quad (8)$$

Supondo-se diferentes sequências de valores discretos e, ainda, valores arbitrários para μ e θ , a Tabela 1 apresenta um estudo de simulação para $c(\theta, \mu)$, trazendo resultados que a consideram da mesma forma proposta por Efron, dada por (5) e, também, desenvolvida, como mostrada em (8).

Nota-se que a aproximação proposta por Efron, utilizando a expansão de Edgeworth, é bastante satisfatória. Também, devem-se observar que os valores atípicos (distantes de 1), nos painéis (2), (3) e (4), podem ser resultados incoerentes com os valores propostos para os parâmetros, nas simulações. Por exemplo, em uma sequência

Tabela 1: Resultados do estudo de simulação - Constante da distribuição Poisson Dupla.

		Painel (1): $\mu = 1$ e $\theta = 1$		Painel (2): $\mu = 3$ e $\theta = 1$	
y		Efron	Desenvolvida	Efron	Desenvolvida
$n = 4$	$0, \dots, 3$	1,00	1,019	1,00	1,53
$n = 6$	$0, \dots, 5$	1,00	1,0005	1,00	1,09
$n = 8$	$0, \dots, 7$	1,00	1,00001	1,00	1,01
$n = 10$	$0, \dots, 9$	1,00	1,00	1,00	1,001
$n = 12$	$0, \dots, 11$	1,00	1,00	1,00	1,0001
$n = 14$	$0, \dots, 13$	1,00	1,00	1,00	1,00
$n = 16$	$0, \dots, 15$	1,00	1,00	1,00	1,00
$n = 18$	$0, \dots, 17$	1,00	1,00	1,00	1,00
$n = 20$	$0, \dots, 19$	1,00	1,00	1,00	1,00
$n = 22$	$0, \dots, 21$	1,00	1,00	1,00	1,00
		Painel (3): $\mu = 10$ e $\theta = 2$		Painel (4): $\mu = 5$ e $\theta = 10$	
y		Efron	Desenvolvida	Efron	Desenvolvida
$n = 4$	$0, \dots, 3$	1,004	2569,881	1,015	151,22
$n = 6$	$0, \dots, 5$	1,004	68,891	1,015	1,29
$n = 8$	$0, \dots, 7$	1,004	7,903	1,015	1,015
$n = 10$	$0, \dots, 9$	1,004	2,36	1,015	1,015
$n = 12$	$0, \dots, 11$	1,004	1,32	1,015	1,015
$n = 14$	$0, \dots, 13$	1,004	1,07	1,015	1,015
$n = 16$	$0, \dots, 15$	1,004	1,01	1,015	1,015
$n = 18$	$0, \dots, 17$	1,004	1,005	1,015	1,015
$n = 20$	$0, \dots, 19$	1,004	1,004	1,015	1,015
$n = 22$	$0, \dots, 21$	1,004	1,004	1,015	1,015

$0, \dots, 3$ ($n = 4$), não seria possível se ter uma variância igual a 5, dado que $Var(Y) \cong \frac{\mu}{\theta}$.

Agora, considerando que $E(Y) = \sum_{y=0}^{\infty} yP(Y = y)$, tem-se que

$$\begin{aligned}
 E(Y) &= \frac{c(\theta, \mu) \sqrt{\theta}}{e^{\mu\theta}} \sum_{y=0}^{\infty} y \frac{e^{y(\theta-1)}}{y!} y^y \left(\frac{\mu}{y}\right)^{\theta y} \text{ e, se} \\
 c(\theta, \mu)^{-1} &= e^{-\theta\mu} \sqrt{\theta} \sum_{y=0}^{\infty} \frac{e^{y(\theta-1)} y^y}{y!} \left(\frac{\mu}{y}\right)^{\theta y}, \text{ tem-se que} \\
 E(Y) &= \frac{\sum_{y=0}^{\infty} y \frac{e^{y(\theta-1)}}{y!} y^y \left(\frac{\mu}{y}\right)^{\theta y}}{\sum_{y=0}^{\infty} \frac{e^{y(\theta-1)}}{y!} y^y \left(\frac{\mu}{y}\right)^{\theta y}}.
 \end{aligned} \tag{9}$$

A Tabela 2 traz um estudo de simulação para $E(Y)$, considerando-se algumas sequências de valores discretos para diferentes propostas para μ e θ .

Assim, também pode-se dizer que a aproximação proposta por Efron (1986) é satisfatória, e os valores mais discrepantes, observados nos painéis (1) e (2) podem ser resultantes de proposições pouco prováveis na prática, como já visto anteriormente.

Considerando-se a expressão de $E(Y)$, dada por (9), pode-se obter o desen-

Tabela 2: Resultados do estudo de simulação - $E(Y)$.

θ	Painel (1): $\mu = 3$ e $y = 0, \dots, 6$		Painel (2): $\mu = 10$ e $y = 0, \dots, 15$	
	Efron	Desenvolvida	Efron	Desenvolvida
1	3,00	2,84	10,00	9,63
3	3,00	3,00	10,00	9,98
5	3,00	3,00	10,00	10,00
7	3,00	3,00	10,00	10,00
9	3,00	3,00	10,00	10,00
11	3,00	3,00	10,00	10,00
13	3,00	3,00	10,00	10,00
15	3,00	3,00	10,00	10,00
17	3,00	3,00	10,00	10,00
19	3,00	3,00	10,00	10,00
θ	Painel (3): $\mu = 5$ e $y = 0, \dots, 20$		Painel (4): $\mu = 10$ e $y = 0, \dots, 20$	
	Efron	Desenvolvida	Efron	Desenvolvida
1	5,00	4,99	10,00	9,98
3	5,00	5,00	10,00	10,00
5	5,00	5,00	10,00	10,00
7	5,00	5,00	10,00	10,00
9	5,00	5,00	10,00	10,00
11	5,00	5,00	10,00	10,00
13	5,00	5,00	10,00	10,00
15	5,00	5,00	10,00	10,00
17	5,00	5,00	10,00	10,00
19	5,00	5,00	10,00	10,00

volvimento da variância de Y , denotada por $Var(Y)$. Assim, sendo o segundo momento de Y , dado por

$$E(Y^2) = E[Y(Y-1)] + E(Y),$$

e, ainda,

$$Var(Y) = E(Y^2) - [E(Y)]^2,$$

faz-se necessário se obter a quantidade $E[Y(Y-1)]$, que é dada por

$$E[Y(Y-1)] = \frac{c(\theta, \mu)\sqrt{\theta}}{e^{\mu\theta}} \sum_{y=0}^{\infty} y(y-1) \frac{e^{y(\theta-1)}}{y!} y^y \left(\frac{\mu}{y}\right)^{\theta y}$$

$$E[Y(Y-1)] = \frac{c(\theta, \mu)\sqrt{\theta}}{e^{\mu\theta}} \sum_{y=2}^{\infty} \left(\frac{y\mu^{\theta}e^{\theta-1}}{y^{\theta}}\right)^y \frac{1}{(y-2)!}$$

Substituindo $c(\theta, \mu)$ pela expressão dada por (8), tem-se que

$$E[Y(Y-1)] = \frac{\sum_{y=2}^{\infty} \left(\frac{y\mu^{\theta}e^{\theta-1}}{y^{\theta}} \right)^y \frac{1}{(y-2)!}}{\left[1 + \sum_{y=1}^{\infty} \left(\frac{y\mu^{\theta}e^{\theta-1}}{y^{\theta}} \right)^y \frac{1}{y!} \right]}. \text{ Assim,}$$

$$E[Y^2] = \frac{\sum_{y=2}^{\infty} \left(\frac{y\mu^{\theta}e^{\theta-1}}{y^{\theta}} \right)^y \frac{1}{(y-2)!} + \sum_{y=1}^{\infty} \left(\frac{y\mu^{\theta}e^{\theta-1}}{y^{\theta}} \right)^y \frac{1}{(y-1)!}}{\left[1 + \sum_{y=1}^{\infty} \left(\frac{y\mu^{\theta}e^{\theta-1}}{y^{\theta}} \right)^y \frac{1}{y!} \right]}.$$

Com essas expressões, pode-se obter $Var(Y)$. A Tabela 3 traz um estudo de simulação para $Var(Y)$, considerando-se algumas sequências de valores discretos para diferentes propostas para μ e θ .

Tabela 3: Resultados do estudo de simulação - $Var(Y)$.

θ	Painel (1): $\mu = 3$ e $y = 0, \dots, 6$		Painel (2): $\mu = 10$ e $y = 0, \dots, 15$	
	Efron	Desenvolvida	Efron	Desenvolvida
1	3,00	2,34	10,00	7,68
3	1,00	0,98	3,33	3,25
5	0,60	0,60	2,00	1,99
7	0,42	0,42	1,42	1,42
9	0,33	0,32	1,11	1,11
11	0,27	0,25	0,91	0,91
13	0,23	0,18	0,77	0,77
15	0,20	0,14	0,66	0,66
17	0,17	0,10	0,59	0,59
19	0,15	0,07	0,52	0,52
θ	Painel (3): $\mu = 5$ e $y = 0, \dots, 20$		Painel (4): $\mu = 10$ e $y = 0, \dots, 20$	
	Efron	Desenvolvida	Efron	Desenvolvida
1	5,00	5,00	10,00	10,00
3	1,66	1,66	3,33	3,33
5	1,00	1,00	2,00	2,00
7	0,71	0,71	1,43	1,43
9	0,55	0,55	1,11	1,11
11	0,45	0,45	0,91	0,91
13	0,38	0,38	0,77	0,77
15	0,33	0,32	0,66	0,66
17	0,29	0,27	0,59	0,59
19	0,26	0,23	0,52	0,52

A partir desses resultados, pode-se observar que a variância de Y pode ser aproximada por μ/θ de forma satisfatória.

4.2.3 Estimadores Bayesianos

Para a estimação Bayesiana dos parâmetros μ e θ , é necessário se estabelecerem distribuições a priori e, dado que não se tem conhecimento sobre os mesmos,

a priori não informativa conjunta de Jeffreys apresenta-se como uma opção. Assim, considerando-se $c(\theta, \mu) = 1$, tem-se que

$$E \left[-\frac{\partial^2 \ln L(\mu, \theta)}{\partial \theta^2} \right] = \frac{n}{2\theta^2} \quad \text{e}$$

$$E \left[-\frac{\partial^2 \ln L(\mu, \theta)}{\partial \mu^2} \right] = \frac{\theta}{\mu^2} E \left(\sum_{i=1}^n y_i \right) = \frac{n\theta}{\mu}$$

e

$$E \left[-\frac{\partial^2 \ln L(\mu, \theta)}{\partial \mu \partial \theta} \right] = 0.$$

Portanto, a matriz de informação de Fisher é dada por

$$I(\mu, \theta) = \begin{pmatrix} \frac{n\theta}{\mu} & 0 \\ 0 & \frac{n}{2\theta^2} \end{pmatrix}$$

e a distribuição a priori conjunta de Jeffreys, para μ e θ , é

$$\pi(\mu, \theta) \propto \sqrt{\frac{1}{\mu\theta}}. \quad (10)$$

Combinando-se (10) com a função de verossimilhança $L(\mu, \theta)$ dada por (6), tem-se a densidade conjunta a posteriori

$$\begin{aligned} \pi(\mu, \theta | \mathbf{y}) &\propto \pi(\mu, \theta) L(\mu, \theta) \\ &= k [c(\theta, \mu)]^n e^{-n\theta\mu} \theta^{(n-1)/2} e^{(\theta-1) \sum_{i=1}^n y_i} \mu^{\theta \sum_{i=1}^n y_i - \frac{1}{2}} \prod_{i=1}^n y_i^{y_i(1-\theta)}, \end{aligned}$$

em que k é a constante normalizadora que torna $\pi(\mu, \theta | \mathbf{y})$ uma função densidade de probabilidade própria. As distribuições condicionais a posteriori para μ e θ , utilizadas no algoritmo de amostradores de Gibbs, são dadas, respectivamente, por:

$$\pi(\mu | \theta, \mathbf{y}) \propto [c(\theta, \mu)]^n e^{-n\theta\mu} \mu^{\theta \sum_{i=1}^n y_i - \frac{1}{2}}$$

e

$$\pi(\theta | \mu, \mathbf{y}) \propto [c(\theta, \mu)]^n e^{-n\theta\mu} \theta^{(n-1)/2} e^{(\theta-1) \sum_{i=1}^n y_i} \mu^{\theta \sum_{i=1}^n y_i - \frac{1}{2}} \prod_{i=1}^n y_i^{y_i(1-\theta)}.$$

Dado que essas duas distribuições condicionais a posteriori possuem formas desconhecidas, o algoritmo Metropolis-Hastings (CHIB e GREENBERG, 1995) é utilizado nas simulações das amostras a partir das distribuições conjuntas a posteriori para μ e θ , para obtenção das medidas a posteriori de interesse.

Assim, para se obter uma aproximação de (10), assumem-se as seguintes distribuições a priori:

$$\mu \sim Gama(0.5, a)$$

e

$$\theta \sim Gama(0.5, b),$$

em que a e b são hiperparâmetros conhecidos e razoavelmente próximos de zero. Nesse caso, têm-se

$$\pi(\mu) \propto \sqrt{\frac{1}{\mu}} e^{-a\mu} \quad \text{e} \quad \pi(\theta) \propto \sqrt{\frac{1}{\theta}} e^{-b\theta},$$

e, assumindo-se independência entre μ e θ , tem-se que $\pi(\mu, \theta) = \pi(\mu)\pi(\theta)$ é satisfatoriamente próxima a (10). Nessa aproximação, a distribuição a priori conjunta é o produto das prioris independentes e não uma distribuição bivariada a priori formal. Ela é válida desde que os parâmetros μ e θ sejam ortogonais.

Alternativamente, pode-se considerar uma reparametrização de μ e θ , dada por $\log \mu = \lambda_1$ e $\log \theta = \lambda_2$. Portanto, λ_1 e λ_2 seguem uma distribuição a priori normal, ou seja, $\lambda_1 \sim N(a_1, b_1)$ e $\lambda_2 \sim N(a_2, b_2)$, em que $a_1, a_2, b_1 > 0$ e $b_2 > 0$ são hiperparâmetros conhecidos, e considera-se independência a priori entre λ_1 e λ_2 . Também, podem-se considerar as distribuições Gama e Uniforme, sendo que $\mu \sim Gama(c_1, d_1)$, $\theta \sim Gama(c_2, d_2)$ ou $\mu \sim U(e_1, f_1)$, $\theta \sim U(e_2, f_2)$, em que $c_1, c_2, d_1, d_2, e_1, e_2, f_1 > 0$ e $f_2 > 0$. Combinando-se $L(\mu, \theta)$ com as distribuições a priori e aplicando-se o teorema de Bayes, obtêm-se as distribuições conjuntas a posteriori para os parâmetros de interesse.

Para a simulação amostras para a distribuição conjunta a posteriori, considera-se o uso do amostrador de Gibbs via algoritmo MCMC (Markov Chain Monte Carlo). O software OpenBugs (LUNN et al., 2009) apresenta-se como uma prática ferramenta, pois necessita apenas que sejam especificadas a distribuição conjunta dos dados e as distribuições a priori para os parâmetros. Os intervalos de credibilidade 95% são definidos pelos 2,5-ésimo e pelo 97,5-ésimo percentis das respectivas distribuições a posteriori. A convergência do algoritmo é verificada por meio de gráficos de séries temporais (Gelman

e Rubin, 1992).

4.2.4 Modelo de Regressão Baseado na Distribuição Poisson Dupla

Se $x_1, x_2, \dots, x_k$ são observações de um vetor de covariáveis de k dimensões, uma regressão sobre μ pode ser escrita como:

$$\log \mu_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki},$$

com $i = 1, \dots, n$, em que $\beta_0, \beta_1, \dots, \beta_k$ são parâmetros desconhecidos. E, ainda, se $w_1, w_2, \dots, w_m$ são observações de um vetor de covariáveis de m -dimensões, uma regressão sobre θ assume a seguinte forma:

$$\log \theta_i = \gamma_0 + \gamma_1 w_{1i} + \dots + \gamma_m w_{mi},$$

com $i = 1, \dots, n$, em que $\gamma_0, \gamma_1, \dots, \gamma_m$ são parâmetros desconhecidos. Deve-se observar que os vetores de covariáveis $\mathbf{x}$ e $\mathbf{w}$ podem ser iguais.

Na análise Bayesiana, pode-se assumir distribuição normal a priori para os parâmetros $\beta_0, \beta_1, \dots, \beta_k, \gamma_0, \gamma_1, \dots, \gamma_m$, com média e variância conhecidas.

4.2.5 Modelos Baseados na Distribuição Poisson Dupla para Dados com Excesso de Zeros

Modelos para dados com excesso de zeros baseados na distribuição de Poisson são bastante utilizados e podem ser entendidos como modelos de mistura, que combinam um componente de contagem e um ponto de massa em zero. Nesta Seção, introduziu-se o modelo para dados com excesso de zeros, com base na distribuição Poisson Dupla (ZIDP - *zero inflated double Poisson*) como uma alternativa viável para modelagem de dados de contagem com superdispersão ou subdispersão, com excesso de zeros. A função massa de probabilidade (*pmf*) desse modelo é dada por:

$$P(Y = 0) = p + (1 - p)c(\theta, \mu) e^{-\theta\mu} \sqrt{\theta}$$

e

$$P(Y = y) = (1 - p)c(\theta, \mu) e^{-\theta\mu} \sqrt{\theta} \frac{e^{y(\theta-1)} y^y}{y!} \left(\frac{\mu}{y}\right)^{\theta y}, \quad y = 1, 2, 3, \dots,$$

em que o parâmetro p é responsável pela modificação do modelo devido ao excesso de zeros. Nota-se que, quando $p = 0$, a *pmf* é equivalente à da distribuição Poisson Dupla, e, quando $p = 1$, a *pmf* degenera-se em uma distribuição pontual em zero. Ainda, quando $p = 0$ e $\theta = 1$, a distribuição ZIDP se reduz à uma distribuição Poisson padrão

com média μ . Quando $\theta = 1$, a distribuição ZIDP é equivalente à uma distribuição Poisson com excessos de zeros (ZIP), como descrita por Lambert (1992). A média e a variância de uma variável aleatória que segue uma distribuição ZIDP são dadas por:

$$E(Y) = \mu(1 - p) \quad \text{e} \quad \text{Var}(Y) = \mu(1 - p) \left(p\mu + \frac{1}{\theta} \right).$$

Assim, a variável Y é distribuída com superdispersão se $\theta < (1 - p\mu)^{-1}$ e com subdispersão se $\theta > (1 - p\mu)^{-1}$.

Para se assegurar que $0 < p < 1$, é conveniente se utilizar a transformação logito, tal que:

$$p = \frac{e^\beta}{1 + e^\beta},$$

em que β é um número real. Portanto,

$$P(Y = 0) = \frac{e^\beta + c(\theta, \mu) e^{-\theta\mu} \sqrt{\theta}}{1 + e^\beta},$$

e a função de verossimilhança para θ, μ e β é

$$\begin{aligned} L(\theta, \mu, \beta) &= \prod_{i:y_i=0} \left[\frac{e^\beta + c(\theta, \mu) e^{-\theta\mu} \sqrt{\theta}}{1 + e^\beta} \right] \times \prod_{i:y_i>0} \left[\frac{c(\theta, \mu) e^{-\theta\mu}}{1 + e^\beta} \sqrt{\theta} \frac{e^{y_i(\theta-1)} y_i^{y_i}}{y_i!} \left(\frac{\mu}{y_i} \right)^{\theta y_i} \right] \\ &= \frac{1}{(1 + e^\beta)^n} \prod_{i=1}^n \left[e^\beta + c(\theta, \mu) e^{-\theta\mu} \sqrt{\theta} \right]^{1-h_i} \times \\ &\quad \times \left[c(\theta, \mu) e^{-\theta\mu} \sqrt{\theta} \frac{e^{y_i(\theta-1)} y_i^{y_i}}{y_i!} \left(\frac{\mu}{y_i} \right)^{\theta y_i} \right]^{h_i}, \end{aligned}$$

em que

$$h_i = \begin{cases} 0 & \text{se } y_i = 0 \\ 1 & \text{caso contrário} \end{cases}, \quad i = 1, \dots, n.$$

A função log-verossimilhança é dada por:

$$\begin{aligned} \ln L(\theta, \mu, \beta) &= -n \ln(1 + e^\beta) + n_0 \ln \left[e^\beta + c(\theta, \mu) e^{-\theta\mu} \sqrt{\theta} \right] \\ &\quad + n_1 \ln c(\theta, \mu) - n_1 \theta \mu + \frac{n_1}{2} \ln \theta \\ &\quad + (\theta - 1) \sum_{i=1}^n y_i + \sum_{i=1}^n y_i \ln y_i - \sum_{i=1}^n \ln(y_i!) \\ &\quad + \theta \sum_{i=1}^n y_i \ln \mu - \theta \sum_{i=1}^n y_i \ln y_i, \end{aligned}$$

em que

$$n_0 = \sum_{i=1}^n (1 - h_i)$$

é o número de observações iguais a zero, e $n_1 = n - n_0$ é o número de observações diferentes de zero. Os estimadores de máxima verossimilhança são obtidos pela maximização da função log-verossimilhança, com relação a θ, μ e β , utilizando-se métodos iterativos, como o algoritmo de Newton-Raphson. Considerando-se $c(\theta, \mu) = 1$, as primeiras derivadas de $\ln L(\theta, \mu, \beta)$, com relação a θ, μ e β são dadas, respectivamente, por:

$$\begin{aligned} \frac{\partial \ln L(\theta, \mu, \beta)}{\partial \theta} &= n_1 \left(\frac{1}{2\theta} - \mu \right) - \frac{n_0 e^{-\theta\mu} (2\theta\mu - 1)}{2(e^\beta \sqrt{\theta} + \theta e^{-\theta\mu})} + \\ &+ \sum_{i=1}^n y_i + \sum_{i=1}^n y_i \ln \mu + \sum_{i=1}^n y_i \ln y_i, \end{aligned}$$

$$\frac{\partial \ln L(\theta, \mu, \beta)}{\partial \mu} = \theta \left(\frac{1}{\mu} \sum_{i=1}^n y_i - n_1 \right) - \frac{\theta^{\frac{3}{2}} n_0 e^{-\theta\mu}}{e^\beta + e^{-\theta\mu} \sqrt{\theta}}$$

e

$$\frac{\partial \ln L(\theta, \mu, \beta)}{\partial \beta} = e^\beta \left(\frac{n_0}{e^\beta + e^{-\theta\mu} \sqrt{\theta}} - \frac{n}{e^\beta + 1} \right).$$

Para a análise Bayesiana, devem-se propor distribuições a priori para θ, μ e β . Pode-se considerar uma reparametrização de μ e θ , dada por $\log \mu = \lambda_1$ e $\log \theta = \lambda_2$ e, portanto, λ_1 e λ_2 seguem, a priori, uma distribuição priori Normal, ou seja, $\lambda_1 \sim N(a_1, b_1)$ e $\lambda_2 \sim N(a_2, b_2)$, em que $a_1, a_2, b_1 > 0$ e $b_2 > 0$ são hiperparâmetros conhecidos, assumindo-se independência a priori entre λ_1, λ_2 e β . Considera-se, também, uma Normal, com hiperparâmetros conhecidos, como distribuição a priori para β . Combinando-se $L(\mu, \theta)$ com a distribuição conjunta a priori e aplicando-se o teorema de Bayes, obtêm-se as distribuições conjuntas a posteriori para os parâmetros de interesse.

4.2.6 Comparações entre os Modelos

A comparação entre os ajustes dos modelos baseados na estimação por máxima verossimilhança pode ser feita utilizando o Critério de Informação de Akaike (AIC), dado por:

$$AIC = -2 \log L(\hat{\psi}_{MLE} | \mathbf{y}) + 2k.$$

Nessa expressão, $L(\hat{\boldsymbol{\psi}}_{MLE}|\mathbf{y})$ é a função log-verossimilhança dos parâmetros $\boldsymbol{\psi}$ do modelo em questão, substituída pelas estimativas de máxima verossimilhança ($\hat{\boldsymbol{\psi}}_{MLE}$), e k é o número de parâmetros do modelo. Quanto menor o valor do AIC , melhor o ajuste do modelo aos dados.

O *Deviance Information Criterion* (DIC) é a generalização do AIC na comparação de modelos ajustados sob a abordagem Bayesiana. O valor DIC , proposto por Spiegelhalter et al. (2002), é dado por:

$$DIC = D(\hat{\boldsymbol{\psi}}) + 2p_D,$$

em que $D(\hat{\boldsymbol{\psi}}) = -2 \log L(\mathbf{D}|\hat{\boldsymbol{\psi}}) + C$ é o deviance avaliado por meio das médias a posteriori dos parâmetros $\boldsymbol{\psi}$, $\mathbf{D}$ é o conjunto de dados observado, C é uma constante não necessária quando se comparam os modelos, e p_D é o número efetivo de parâmetros do modelo dado por $p_D = \overline{D}(\theta) - D(\hat{\boldsymbol{\psi}})$, com $\overline{D}(\theta)$ sendo a média a posteriori do deviance. Baixos valores para o DIC sugerem um melhor ajuste do modelo aos dados. O DIC pode ser facilmente obtido pelo ajuste de modelos Bayesianos utilizando-se métodos MCMC (Markov Chain Monte Carlo), disponíveis no software OpenBUGS.

Um outro critério conhecido de seleção de modelos, quando na modelagem Bayesiana, é a preditiva ordenada condicional (CPO) (GELFAND et al., 1992). Para a i –ésima observação, a CPO_i é dada por:

$$f(D_i | y_{[i]}) = \int f(D_i | \Theta) f(\Theta | D_{[i]}) d\Theta,$$

em que Θ é o vetor de parâmetros; D_i é cada observação do vetor de dados completo D ; $D_{[i]}$ é o vetor de dados D , sem a observação D_i , e $f(\Theta | D_{[i]})$ é a densidade a posteriori para Θ dado $D_{[i]}$ ($i = 1, \dots, n$). Assim, a estatística CPO expressa a probabilidade a posteriori de se observar um valor, ou um conjunto de valores de D_i , quando o modelo é ajustado sem se considerar D_i . Uma aproximação da CPO, via MCMC (CHEN et al., 2001) é dada por:

$$\widehat{CPO}_i = \left[\frac{1}{B} \sum_{b=1}^B \frac{1}{f(D_i | \Theta_b)} \right]^{-1}$$

em que B é o número de iterações do procedimento MCMC, após o período *burn-in*, e Θ_b é o vetor de amostras obtidas na b –ésima iteração. Assumindo-se uma aproximação normal, valores inversos da $\widehat{CPO}_i$ (ICPO) acima de 40 podem ser considerados possíveis *outliers*, e valores acima de 70 são considerados extremos (NTZOUFRAS, 2009). A log-pseudo verossimilhança marginal (LPML) é uma medida Bayesiana de adequação do modelo, calculada a partir da CPO, sendo que

$$\widehat{LPML} = \sum_{i=1}^n \log(\widehat{CPO}_i),$$

e quanto maior a LPML, melhor o ajuste do modelo (GEISSER e EDDY, 1979).

a priori não informativas, $Gama(0, 5; 0, 0001)$, para os parâmetros θ e μ . Os resultados desses ajustes encontram-se na Tabela 4

Tabela 4: Resultados do estudo de simulação - casos (a), (b) e (c).

n	Parâmetro	Modelo 1			Modelo 2			Modelo 3		
		Estimativa	EP	IC 95%	Estimativa	EP	IC 95%	Estimativa	DVP	ICr 95%
$n = 10$	μ	1,70	0,4271	(0,8578; 2,5422)	1,6941	0,4015	(0,9025; 2,4858)	1,746	0,4416	(1,0270; 2,7570)
	θ	0,9319	0,4167	(0,1101; 1,7537)	1,0635	0,4068	(0,2613; 1,8656)	1,036	0,3924	(0,4471; 1,9770)
	μ/θ	1,8243	0,9358	(-0,0297; 3,6695)	1,5930	0,7421	(0,1297; 3,0564)	1,991	1,2	(0,7495; 5,0280)
	$c(\theta; \mu)$	1	-	-	1,0046	0,0273	(0,9598; 1,0584)	0,988	0,041	(0,8803; 1,0380)
$n = 50$	μ	2,02	0,1778	(1,6694; 2,3706)	2,0056	0,1718	(1,6669; 2,3443)	2,01	0,1775	(1,6770; 2,3740)
	θ	1,2782	0,2556	(0,7741; 1,7822)	1,3743	0,2518	(0,8778; 1,8707)	1,357	0,2478	(0,9228; 1,880)
	μ/θ	1,5804	0,3453	(0,8994; 2,2614)	1,4594	0,2974	(0,8730; 2,0458)	1,533	0,3270	(1,0270; 2,3110)
	$c(\theta; \mu)$	1	-	-	1,0157	0,0072	(1,0014; 1,0300)	1,014	0,0081	(0,9948; 1,0270)
$n = 100$	μ	2,04	0,129187	(1,7855; 2,2945)	2,0266	0,1245	(1,7812; 2,2720)	2,029	0,1256	(1,788; 2,282)
	θ	1,2248	0,1732	(0,8832; 1,5664)	1,3215	0,1705	(0,9853; 1,6577)	1,316	0,1667	(1,012; 1,666)
	μ/θ	1,6656	0,2580	(1,1567; 2,1744)	1,5335	0,2212	(1,0974; 1,9697)	1,568	0,2273	(1,183; 2,074)
	$c(\theta; \mu)$	1	-	-	1,0139	0,0053	(1,0034; 1,0244)	1,013	0,0055	(1,001; 1,022)

Os intervalos de confiança 95% para μ/θ , considerando-se o Modelo 1 e uma amostra de tamanho $n = 10$, mostram que a abordagem assintótica pode resultar em intervalos que abrangem valores negativos para a variância.

Nos três conjuntos de dados anteriores, os valores para as médias e variâncias foram próximos. Na Tabela 5, têm-se os resultados baseados nas mesmas modelagens propostas na Tabela 4, mas considerando-se outras duas amostras ($n = 100$), geradas a partir da distribuição Poisson Dupla, para se verificar como o modelo se comporta em situações com variâncias muito pequenas (d) ou muito grandes (e), em relação à média.

(d) Para $\mu = 10$ e $\theta = 4$ seja o vetor de observações $\mathbf{y} = (11, 8, 9, 11, 11, 9, 10, 11, 7, 10, 8, 11, 7, 10, 12, 12, 6, 11, 9, 9, 10, 9, 10, 9, 11, 9, 7, 9, 12, 11, 8, 8, 13, 10, 9, 10, 9, 11, 13, 8, 10, 11, 11, 9, 12, 10, 11, 12, 9, 10, 11, 9, 11, 11, 7, 12, 12, 10, 10, 13, 11, 10, 10, 11, 8, 10, 9, 13, 10, 9, 8, 11, 10, 8, 13, 11, 8, 10, 11, 10, 10, 11, 10, 11, 13, 10, 13, 9, 10, 9, 10, 9, 6, 8, 12, 10, 10, 10, 12, 10)'$. Nesse caso, a média amostral é $\bar{y} = 10,03$, e a variância amostral é igual a 2,494.

(e) Para $\mu = 10$ e $\theta = 0.5$ considere o vetor de observações $\mathbf{y} = (10, 14, 6, 5, 18, 6, 9, 18, 9, 12, 10, 8, 3, 14, 16, 12, 3, 7, 7, 11, 9, 15, 6, 7, 6, 18, 12, 9, 6, 5, 6, 4, 6, 7, 8, 16, 18, 18, 11, 12, 5, 9, 6, 12, 6, 16, 8, 9, 12, 12, 4, 16, 5, 14, 9, 19, 10, 13, 13, 5, 15, 12, 5, 9, 14, 4, 14, 13, 8, 18, 10, 5, 8, 9, 9, 11, 8, 14, 15, 3, 11, 10, 7, 10, 4, 11, 6, 7, 6, 10, 12, 8, 15, 10, 4, 11, 12, 16, 12, 19)'$. Nesse caso, a média amostral é $\bar{y} = 10,05$, e a variância amostral é igual a 17,866.

Tabela 5: Resultados do estudo de simulação - casos (d) e (e).

n	Parâmetro	Modelo 1			Modelo 2			Modelo 3		
		Estimativa	EP	IC 95%	Estimativa	EP	IC 95%	Estimativa	DVP	ICr 95%
$n = 100$	μ	10,03	0,1589	(9,7166; 10,3434)	10,0284	0,1586	(9,7156; 10,3411)	10,03	0,1619	(9,7140; 10,340)
	θ	3,9715	0,5617	(2,8640; 5,0790)	3,9873	0,5616	(2,8799; 5,0947)	3,948	0,5526	(2,945; 5,110)
	μ/θ	2,5255	0,3594	(1,8168; 3,2342)	2,5151	0,3564	(1,8122; 3,2179)	2,591	0,3714	(1,958; 3,415)
	$c(\theta; \mu)$	1	-	-	1,0064	0,0003	(1,0058; 1,0070)	1,006	0,0003	(1,006; 1,007)
$n = 100$	μ	10,05	0,4242	(9,2134; 10,8866)	10,0639	0,4169	(9,2418; 10,8861)	10,06	0,4271	(9,244; 10,92)
	θ	0,5584	0,0789	(0,4027; 0,7141)	0,5781	0,0786	(0,4230; 0,7333)	0,574	0,0777	(0,433; 0,7376)
	μ/θ	17,9982	2,6563	(12,7603; 23,2362)	17,4073	2,4878	(12,5016; 22,3130)	17,86	2,603	(13,42; 23,73)
	$c(\theta; \mu)$	1	-	-	0,9930	0,0024	(0,9882; 0,9977)	0,9925	0,0025	(0,9867; 0,9967)

5.2 Exemplos com Dados Reais

Nesta Seção, serão considerados diferentes conjuntos de dados reais para se ilustrar o uso do modelo baseado na distribuição Poisson Dupla.

5.2.1 Registros de Picadas de Escorpiões em Ribeirão Preto, SP, Brasil

Os dados a seguir são referentes ao número total mensal de picadas de escorpiões registradas em Ribeirão Preto, SP, entre os anos de 2010 e 2013. Seguem: 13, 11, 8, 7, 11, 13, 13, 17, 12, 21, 10, 10, 5, 9, 7, 2, 13, 6, 10, 11, 12, 17, 24, 21, 14, 15, 12, 10, 11, 7, 14, 19, 19, 12, 9, 16, 16, 10, 11, 8, 10, 15, 7, 14, 16, 15, 9 e 11. Esses dados foram obtidos do Sistema de Informação de Agravos de Notificação (SINAN). A média amostral é igual a 12,15, e a variância amostral é 19,53, indicando uma superdispersão.

A Tabela 6 mostra estimadores de máxima verossimilhança (EMV) para os parâmetros da distribuição Poisson Dupla e seus respectivos intervalos de confiança 95%. O Modelo 1 considera que $c(\theta, \mu) = 1$, e o Modelo 2 assume que $c(\theta, \mu)$ é dada pela expressão (5). Ambos os modelos fornecem resultados similares para as estimativas dos parâmetros e valores próximos para o AIC . Pode-se afirmar, também, que, nesse exemplo, a aproximação $c(\theta, \mu) = 1$ é satisfatória e as estimativas para a média μ e para a variância μ/θ são próximas àquelas obtidas diretamente dos dados.

Por meio do ajuste de um modelo de regressão de Poisson a esses dados, obtêm-se um valor de 284,7 para o AIC . Apesar de esse valor ser bastante próximo ao reportado na Tabela 6, a distribuição Poisson Dupla parece ajustar melhor os dados, pois a variância μ/θ estimada é mais próxima daquela obtida diretamente dos dados.

Na Figura 1 é mostrado o gráfico de contorno do logaritmo natural da função de verossimilhança para esses dados, incluindo-se $c(\theta, \mu)$, dada pela expressão 5. O gráfico de contorno indica que o máximo da função localiza-se na coordenada (0,6352;12,1536), como mostrado na Tabela 6 (Modelo 2).

Tabela 6: Picadas de escorpião em Ribeirão Preto, SP - EMV.

Parâmetro	Estimativa	EP	IC 95%	AIC
Modelo 1				280,2
μ	12,1458	0,6391	(10,8856 ; 13,4061)	
θ	0,6195	0,1265	(0,3702 ; 0,8689)	
μ/θ	19,6053	4,1328	(11,4560 ; 27,7547)	
Modelo 2				280,6
μ	12,1536	0,6312	(10,9090 ; 13,3982)	
θ	0,6352	0,1262	(0,3864 ; 0,8840)	
μ/θ	19,1340	3,9382	(11,3683 ; 26,8996)	
$c(\theta; \mu)$	0,9956	0,0025	(0,9906 ; 1,0005)

Figura 1: Gráfico do contorno da função log-verossimilhança, considerando-se $c(\theta, \mu)$ para os dados de picadas de escorpiões em Ribeirão Preto, SP.

Alternativamente, ICPVs, para μ e θ , podem ser expressos por meio de gráficos. Considerando-se os EMV obtidos no Modelo 2, os gráficos mostrados na Figura 2 descrevem o perfil da função log-verossimilhança $\ln L(\mu, \hat{\theta}_{ML})$ e $\ln L(\hat{\mu}_{ML}, \theta)$. As linhas horizontais pontilhadas representam o valor

$$\ln L(\hat{\mu}_{ML} = 12.1536, \hat{\theta}_{ML} = 0.6352) - \frac{1}{2}\chi_1^2(.95) = -140.2452.$$

Em cada gráfico, os pontos em que a curva log-verossimilhança intercepta as linhas pontilhadas definem a extensão do intervalo de confiança para cada parâmetro, respectivamente. No painel (a) da Figura 2, a linha horizontal intercepta a curva log-verossimilhança em $\mu = 10.958$ e $\mu = 13.433$. Portanto, o ICPV 95% para μ é $(10, 958; 13, 433)$. Analogamente, o Painel (b) mostra que o ICPV 95% para θ é $(0, 420; 0, 917)$.

Figura 2: Gráfico do perfil da log-verossimilhança para μ (painel (a)) e θ (painel (b)), considerando-se os dados de picadas de escorpiões em Ribeirão Preto, SP.

A Tabela 7 mostra as estimativas Bayesianas para os parâmetros da distribuição Poisson Dupla, considerando-se diferentes distribuições a priori para os mesmos. Assim, $N(a; b)$ denota uma distribuição Normal com média a e variância b ; $Gama(c; d)$ denota uma distribuição Gama com hiperparâmetros c e d , e $U(e; f)$ denota uma distribuição Uniforme com hiperparâmetros e e f . Nas simulações, baseadas no algoritmo MCMC, foram geradas 1.000.000 de amostras, considerando um *burn-in period* de 1000 amostras, para eliminação da influência dos valores iniciais, além de saltos de tamanho 100, ou seja, considerando-se as 100^a , 200^a , 300^a , ... iterações, para obtenção das estimativas a posteriori, evitando-se possíveis autocorrelações entre elas.

Tabela 7: Picadas de escorpião em Ribeirão Preto, SP - Estimativas Bayesianas.

Distribuições a priori	Parâmetro	Estimativa	<i>DVP</i>	<i>ICr</i> 95%	<i>DIC</i>
$\log(\mu) \sim N(0; 10^3)$ $\log(\theta) \sim N(0; 10^3)$	Modelo 1				280,2
	μ	12,13	0,664	(10,88 ; 13,48)	
	θ	0,5976	0,125	(0,3835 ; 0,8779)	
	μ/θ	20,22	4,695	(13,73 ; 32,08)	
	Modelo 2				280,7
	μ	12,15	0,6503	(10,92 ; 13,48)	
	θ	0,6166	0,1256	(0,4031 ; 0,8917)	
	μ/θ	19,69	4,478	(13,49 ; 30,75)	
	$c(\theta; \mu)$	0,9952	0,0029	(0,9878 ; 0,9991)	
$\mu \sim Gama(0, 5; 10^{-4})$ $\theta \sim Gama(0, 5; 10^{-4})$	Modelo 1				280,2
	μ	12,16	0,6572	(10,89; 13,48)	
	θ	0,6178	0,1245	(0,4035; 0,8876)	
	μ/θ	20,50	4,454	(13,58; 30,77)	
	Modelo 2				280,7
	μ	12,17	0,6483	(10,94; 13,47)	
	θ	0,6363	0,126	(0,4137; 0,9076)	
	μ/θ	19,91	4,283	(13,24; 29,90)	
	$c(\theta; \mu)$	0,995	0,0027	(0,9884; 0,9992)	
$\mu \sim U(0; 10^3)$ $\theta \sim U(0; 10^3)$	Modelo 1				280,2
	μ	12,17	0,6505	(10,94; 13,50)	
	θ	0,6309	0,1274	(0,4048; 0,9028)	
	μ/θ	20,11	4,414	(13,37; 30,62)	
	Modelo 2				280,6
	μ	12,19	0,6311	(10,98; 13,49)	
	θ	0,6479	0,1286	(0,4224; 0,9222)	
	μ/θ	19,60	4,226	(13,00; 29,34)	
	$c(\theta; \mu)$	0,9953	0,0027	(0,9888; 0,9994)	

5.2.2 Registros de Picadas de Cobras em Benjamin Constant, AM, Brasil

Os dados a seguir referem-se a registros mensais de picadas de cobras no município de Benjamin Constant, Amazonas, Brasil, entre os anos de 2009 e 2013, obtidos na base de dados do SINAN. Os dados são: 1, 1, 1, 3, 4, 2, 0, 4, 2, 2, 1, 1, 2, 4, 4, 3, 2, 0, 2, 0, 0, 1, 0, 4, 5, 2, 3, 2, 1, 4, 3, 2, 1, 4, 3, 2, 4, 1, 3, 3, 1, 1, 1, 1, 5, 3, 2, 4, 4, 3, 1, 4, 2, 3, 1, 3, 1, 2, 1 e 4. Nota-se que existe uma leve subdispersão, com média 2,23 e variância de 1,88. A Tabela 8 mostra as estimativas de máxima verossimilhança para os parâmetros da distribuição Poisson Dupla. O Modelo 1 considera $c(\theta, \mu) = 1$, e o Modelo 2 inclui $c(\theta, \mu)$, dada pela expressão (5). Em ambos os casos, os valores para o AIC foram bastante similares. Entretanto, o segundo modelo estimou um valor para μ/θ mais próximo da variância observada diretamente da amostra. Quando se ajusta um modelo aos mesmos dados, a partir da distribuição Poisson, tem-se um valor para o AIC igual a 207,5.

5.2.3 Número de Visitas ao Médico durante o Primeiro Trimestre da Gravidez

Para ilustrar o uso do modelo Poisson Dupla com excesso de zeros, considere-se o conjunto de dados reais, introduzido por Hosmer et al. (2013), coletado no Baystate Medical Center, Springfield, Massachusetts, em 1986. O objetivo original do estudo foi identificar fatores associados ao baixo peso, ao nascer, em bebês de 189 gestantes, mas, neste caso, considerou-se o número de visitas ao médico durante o primeiro trimestre da gestação. Nas análises, foram considerados os modelos DP (Poisson Dupla) e ZIDP (Modelo Poisson Dupla para excesso de zeros) e os resultados foram comparados, também, com os obtidos pelo método de máxima verossimilhança (EMV). Ambos os Modelos, 1 e 2, consideram a distribuição Poisson Dupla e $c(\theta, \mu) = 1$ e $c(\theta, \mu)$, dada pela expressão (5), respectivamente. O Modelo 3 é baseado na distribuição Poisson Dupla para excesso de zeros e assume $c(\theta, \mu)$, dada pela expressão (5). A Tabela 9 mostra comparações das contagens observadas com as previstas pelos três modelos, além de apresentar $\log L(\hat{\psi}_{MLE}|\mathbf{y})$ e valores de AIC . Nota-se que o modelo ZIDP (Modelo 3) ajusta melhor os dados, dado que os valores preditos estão muito próximos dos observados e que o valor do AIC , para esse modelo, é o menor, dentre os comparados.

O Painel 1 da Tabela 10 mostra estimativas de máxima verossimilhança (EMV) e Bayesianas para o Modelo 3, baseado no modelo Poisson Dupla com excesso de zeros. Erros padrão e $IC95\%$ assintóticos para as estimativas de p e $c(\theta, \mu)$ foram obtidos pelo método delta. O Painel 2 dessa tabela mostra as estimativas de máxima verossimilhança e Bayesianas, para o mesmo modelo, considerando-se a etnia como variável indicadora *dummy*, tal que $x_1 = 1$ e $x_2 = 0$, se negro; $x_1 = 0$ e $x_2 = 1$,

Tabela 8: Picadas de cobras em Benjamin Constant, AM, Brasil - Estimativas Bayesianas.

Distribuições a priori	Parâmetro	Estimativa	<i>DVP</i>	<i>ICr95%</i>	<i>DIC</i>
$\log(\mu) \sim (0; 10^3)$ $\log(\theta) \sim N(0; 10^3)$	Modelo 1				209,5
	μ	2,25	2,231	(1,867; 2,63)	
	θ	0,9983	1,009	(0,6793; 1,402)	
	μ/θ	2,221	2,286	(1,537; 3,397)	
	Modelo 2				209,1
	μ	2,2285	0,1851	(1,884; 2,610)	
	θ	1,097	0,1840	(0,7851; 1,4960)	
	μ/θ	2,014	0,4022	(1,432; 2,998)	
	$c(\theta; \mu)$	1,005	0,0084	(0,9843; 1,017)	
$\mu \sim Gama(0, 5; 10^{-4})$ $\theta \sim Gama(0, 5; 10^{-4})$	Modelo 1				209,5
	μ	2,239	0,1935	(1,876; 2,639)	
	θ	1,027	0,1859	(0,6992; 1,415)	
	μ/θ	2,254	0,4696	(1,523; 3,339)	
	Modelo 2				209,0
	μ	2,238	0,1842	(1,89; 2,617)	
	θ	1,123	0,184	(0,7942; 1,522)	
	μ/θ	2,048	0,3959	(1,421; 2,972)	
	$c(\theta; \mu)$	1,004	0,0082	(0,9855; 1,017)	
$\mu \sim U(0; 10^3)$ $\theta \sim U(0; 10^3)$	Modelo 1				209,5
	μ	2,249	0,1945	(1,887; 2,643)	
	θ	1,046	0,1897	(0,7102; 1,452)	
	μ/θ	2,224	0,4688	(1,496; 3,305)	
	Modelo 2				209,0
	μ	2,246	0,1847	(1,90; 2,66)	
	θ	1,138	0,1837	(0,8079; 1,536)	
	μ/θ	2,029	0,3863	(1,411; 2,920)	
	$c(\theta; \mu)$	1,005	0,0078	(0,9866; 1,017)	

Tabela 9: Contagens observadas e previstas para o número de visitas ao médico no primeiro trimestre da gestação.

	Número de visitas ao médico									$\log L(\hat{\psi}_{MLE} \mathbf{y})$	<i>AIC</i>
	0	1	2	3	4	5	6	7	8		
Contagens observadas	100	47	30	7	4	0	1	0	0		
Contagens previstas, Modelo 1	86,42	69,41	26,43	6,57	1,21	0,18	0,02	0	0	-231,30	466,61
Contagens previstas, Modelo 2	81,20	51,74	24,49	8,97	2,72	0,71	0,16	0,03	0,01	-237,16	478,32
Contagens previstas, Modelo 3	100	47,95	27,72	10,23	2,76	0,58	0,10	0,01	0	-41,57	89,14

a 2,509. Embora a comparação dos valores do AIC não sugiram um melhor modelo, pode-se concluir que o modelo Poisson Dupla é preferível, pois captura a subdispersão dos dados, diferentemente do modelo Poisson padrão. As estimativas bayesianas para o modelo Poisson Dupla, considerando-se distribuições a priori não informativas, são bem próximas das de máxima verossimilhança.

Para ilustrar o uso de um modelo de regressão baseado na distribuição Poisson Dupla, as seguintes covariáveis são consideradas:

- (a) x_1 é a cor da pele da mulher, em que $x_1 = 1$ se branca ou 0, caso contrário;
- (b) x_2 é a idade da menarca, em anos completos;
- (c) x_3 é o índice de massa corpórea (IMC), no momento da entrevista, dado por peso/altura².

Assim, a regressão em μ é dada por:

$$\log(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}, \quad (11)$$

$i = 1, \dots, 55$, em que $\beta_0, \beta_1, \beta_2$ e β_3 são parâmetros desconhecidos.

O Modelo 1 considera a regressão em μ , descrita pela expressão (11) e, também, que θ é um parâmetro desconhecido. O Modelo 2 considera a regressão em μ , como descrita em (11) e, também, uma regressão em θ dada por:

$$\log(\theta_i) = \gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i} + \gamma_3 x_{3i},$$

$i = 1, \dots, 55$, em que $\gamma_0, \gamma_1, \gamma_2$ e γ_3 são parâmetros desconhecidos. A Tabela 11 mostra os resultados baseados nos métodos de máxima verossimilhança e Bayesiano. Nas simulações, baseadas no algoritmo MCMC, foram geradas 1.000.000 de amostras, considerando-se um *burn-in period* de 1000 amostras, para eliminação da influência dos valores iniciais, além de saltos de tamanho 100, ou seja, considerando-se as 100^a, 200^a, 300^a, ... iterações para a obtenção das estimativas a posteriori, evitando-se possíveis autocorrelações entre elas. Nota-se que as estimativas fornecidas pelas duas abordagens são bem próximas. Também, parece que não há evidências de que as variáveis consideradas estejam associadas ao número de filhos, nessas mulheres, dado que os intervalos de confiança (frequentistas) e os de credibilidade (bayesianos) 95% para β_m e γ_m , $m = 0, \dots, 3$, contêm o valor zero.

Tabela 11: Estimadores de máxima verossimilhança e bayesianas para o modelo de regressão Poisson dupla, considerando-se dados referentes a 55 mulheres do estudo apresentado por Conde et al. (2012).

Parâmetro	EMVs				Estimativas Bayesianas			
	Estimativa	EP	IC95%	AIC	Estimativa	DVP	ICr95%	DIC
Modelo 1				188,1				188,1
β_0	-0,3136	0,7260	(-17,452 ; 11,180)		-0,3113	0,764	(-1,848 ; 1,171)	
β_1	-0,1088	0,1619	(-0,4280 ; 0,2104)		-0,1041	0,1705	(-0,4292 ; 0,2373)	
β_2	0,0595	0,0417	(-0,0226 ; 0,1418)		0,0590	0,0437	(-0,0291 ; 0,1447)	
β_3	0,0197	0,0187	(-0,0171 ; 0,0565)		0,0196	0,0196	(-0,0187 ; 0,0582)	
θ	1,5237	0,2738	(0,9838 ; 2,0637)		1,418	0,261	(0,9656 ; 1,986)	
Modelo 2				192,2				192,2
β_0	-0,3795	0,7279	(-18,149 ; 10,560)		-0,3542	0,7819	(-1,898 ; 1,176)	
β_1	-0,0576	0,1521	(-0,3575 ; 0,2423)		-0,0643	0,1701	(-0,3975 ; 0,2719)	
β_2	0,0368	0,0402	(-0,0425 ; 0,1160)		0,0386	0,0444	(-0,0446 ; 0,1297)	
β_3	0,0319	0,0216	(-0,0106 ; 0,0744)		0,0296	0,0226	(-0,0153 ; 0,0733)	
γ_0	0,1942	1,9578	(-3,6663 ; 4,0548)		0,2405	1,93	(-3,75 ; 3,769)	
γ_1	-0,3375	0,4382	(-1,2015 ; 0,5265)		-0,2207	0,4488	(-1,052 ; 0,6927)	
γ_2	0,1403	0,1275	(-0,1111 ; 0,3917)		0,1152	0,1256	(-0,1244 ; 0,367)	
γ_3	-0,0480	0,0531	(-0,1527 ; 0,0566)		-0,0435	0,0528	(-0,1462 ; 0,0603)	

5.3 Modelos de Séries Temporais de Dados de Contagem Baseados na Distribuição Poisson Dupla

5.3.1 Registros de Acidentes com Cobras no Estado de São Paulo, entre 2007 e 2014

Acidentes com cobras são um problema de saúde pública em diversas áreas do Brasil (CUPO, 2015). Dados oficiais afirmam que ocorrem, aproximadamente, 26 mil casos de picadas de cobras, a cada ano, no Brasil, mas esse número pode estar subestimado devido à ineficiência na coleta de dados e no número de notificações. Sistemas de informações de boa performance são necessários para uma vigilância adequada dos acidentes, por região geográfica, espécies de cobras e, conseqüentemente, os registros das picadas (MACHADO et al., 2012). Dificuldades de acesso aos serviços de saúde também contribuem para a subnotificação dos acidentes com cobras, sendo que, se fossem registrados, poderiam promover educação em saúde, reduzindo a gravidade e frequência dos acidentes, bem como as sequelas e letalidade dos mesmos. (BOCHNER et al., 2014; FEITOSA et al., 2015)

O estado de São Paulo é o mais populoso do Brasil, com grande concentração de atividades industriais, comerciais e agrícolas. As cobras encontradas nessa região são dos gêneros *Bothrops*, *Crotalus* e *Micrurus*, sendo que as primeiras são as mais comuns, dada a sua capacidade de adaptação aos variados ambientes. Cobras do gênero *Lachesis* não são encontradas no estado, e acidentes com corais (gênero *Micrurus*) são pouco comuns (FEITOSA et al., 2015; RIBEIRO, et al., 1998, RIBEIRO e JORGE, 1997;

BUCARETCHI et al., 2016).

Nesta Seção, serão analisados 14.419 casos de picadas de cobras, obtidos por meio do DATASUS/SINAN (<http://tabnet.datasus.gov.br>), que ocorreram no estado de São Paulo, Sudeste do Brasil, entre janeiro de 2007 e dezembro de 2014. No período, foram registrados 9.414 acidentes com cobras *Bothrops*, 1.570 com *Crotalus*, 177 com *Micrurus* e 1.167 com cobras não peçonhentas. Entretanto, foram registrados 2.268 acidentes com cobras não identificadas. Os acidentes com cobras *Micrurus* não foram considerados nas análises devido ao número pequeno de registros. Os resultados dos modelos ajustados, mostrados nesta Seção, foram publicados por Aragon et al. (2016).

A Tabela 12 mostra a distribuição dos acidentes com cada tipo de cobra, de acordo com sexo e faixa etária da vítimas. Nota-se que os acidentes são mais prováveis em homens, com idades entre 20 e 59 anos, o que corrobora estudos que mostram que as ocorrências de picadas são mais comuns em áreas agrícolas, em que as pessoas afetadas são homens adultos, trabalhadores em atividades rurais (HUI WEN et al., 2015).

Tabela 12: Distribuição dos acidentes, por gênero das cobras, de acordo com sexo e faixa etária das vítimas - Estado de São Paulo, entre 2007 e 2014.

	Cobras				
	<i>Bothrops</i>	<i>Crotalus</i>	<i>Micrurus</i>	Não peçonhentas	Não identificadas
Sexo					
Masculino	78,4%	81,6%	74,2%	68,9%	76,1%
Feminino	21,6%	18,4%	25,8%	31,1%	23,9%
Faixa etária (anos)					
Menor que 1	0,9%	0,6%	0,6%	1,5%	1,4%
1 - 4	1,7%	1,5%	4,5%	4,1%	3,2%
5 - 9	4,4%	2,6%	5,6%	6,3%	3,8%
10 - 14	6,6%	5,3%	4,5%	7,9%	8,7%
15 - 19	7,3%	5,1%	9,0%	7,7%	8,6%
20 - 39	32,1%	35,9%	41,2%	35,6%	34,3%
40 - 59	33,5%	35,1%	28,8%	28,7%	28,0%
60 - 64	5,5%	6,0%	4,0%	3,9%	4,9%
65 - 69	3,6%	3,7%	0,0%	1,9%	3,0%
70 - 79	3,6%	3,3%	1,1%	1,9%	3,3%
80 ou mais	0,9%	0,9%	0,6%	0,5%	0,8%

A Figura 3 apresenta as séries com os registros dos acidentes com cobras, no Estado de São Paulo, entre 2007 e 2014.

Observa-se um comportamento sazonal no número de registros, indicando um aumento de casos entre os meses de novembro e abril, caracterizando uma temporada mais quente. Isso corrobora o fato de que é sabido que as cobras são animais ectotérmicos, ou seja, precisam de uma alta temperatura ambiental para realizarem termorregulação corporal e, conseqüentemente, uma atividade metabólica satisfatória. Assim, os acidentes em épocas mais quentes são mais comuns, pois os animais esta-

Figura 3: Séries temporais de registros de acidentes com cobras, no Estado de São Paulo, entre 2007 e 2014 - Fonte: DATASUS/SINAN.

rão mais ativos no meio, devido à maior disponibilidade de presas (MARQUES et al., 2000).

Para se modelarem dados que apresentam um comportamento sazonal, Stolwijk et al. (1999) propuseram o uso de uma equação, envolvendo funções seno e cosseno, que descreve o padrão sazonal por T unidades de tempo, dada por:

$$S_t = \eta_1 \sin\left(\frac{2\pi t}{T}\right) + \eta_2 \cos\left(\frac{2\pi t}{T}\right), \quad (12)$$

em que η_1 e η_2 são constantes. Assim, considerando-se, por exemplo, uma série temporal com sazonalidade anual, ou seja, $T = 12$, a Figura 4 mostra o comportamento de S_t para $\eta_1 = 1$ e $\eta_2 = 1$.

Figura 4: Gráfico de S_t .

Portanto, S_t é uma onda com período igual a T e amplitude dada por

$$\alpha = \sqrt{\eta_1^2 + \eta_2^2}.$$

Uma expansão de S_t para a modelagem de séries de dados com comportamentos mais incomuns pode ser escrita como:

$$S_t = \sum_{j=1}^J \left[\eta_{j1} \sin \left(\frac{2j\pi t}{T} \right) + \eta_{j2} \cos \left(\frac{2j\pi t}{T} \right) \right],$$

em que J é um número inteiro e $J \geq 2$. Considerando-se $J = 2$ e $T = 12$, a Figura 5 mostra o comportamento de S_t para diferentes valores de η_{11} , η_{12} , η_{21} e η_{22} .

Figura 5: Comportamento de S_t em diferentes situações.

Assim, percebe-se que o emprego de uma soma de funções S_t traz maior flexibilidade no ajuste de dados em séries com sazonalidades.

Nas modelagens a seguir, serão levados em conta modelos de regressão baseados na distribuição Poisson Dupla e de Poisson, com e sem adição de termos autorregressivos de primeira e segunda ordens e, ainda, com $J = 1$ e $J = 2$, definindo um ou dois pares de funções seno e cosseno para S_t , respectivamente. Portanto, a média (μ_t), para ambas as distribuições, será expressa da seguinte forma:

$$\log(\mu_t) = \alpha + \beta t + S_t + \sum_{j=1}^r \gamma_j (y_{t-j} - \bar{y}), \quad (13)$$

em que α é uma constante; β é a tendência linear ao longo do tempo; S_t é a função periódica que representa todo o ciclo sazonal ao longo das T unidades de tempo; γ_j são os termos autorregressivos de $r - \text{ésima}$ ordem e $\bar{y}$ é a média do número de picadas de cobra no período. Levando-se em conta uma sazonalidade anual, toma-se $T = 12$. Como distribuições a priori não informativas para os parâmetros α , β , θ , η_{11} , η_{12} , η_{21} e η_{22} , foram consideradas distribuições $N(0; 1000)$. Já, para γ_1 e γ_2 , foram propostas distribuições $U(-1; 1)$.

Nas simulações, baseadas no algoritmo MCMC, foram geradas 1.000.000 de amostras, considerando-se um *burn-in period* de 1000 amostras, para eliminação da influência dos valores iniciais, além de saltos de tamanho 100, ou seja, tomando-se as 100^a , 200^a , 300^a , ... iterações, para a obtenção das estimativas a posteriori, evitando-se as possíveis autocorrelações entre elas.

Considerando-se os modelos baseados na distribuição Poisson Dupla, a Tabela 13 mostra os resultados dos ajustes do modelo dado por (13), para o número de acidentes com cobras no Estado de São Paulo, sem o termo autorregressivo. A Tabela 14 mostra os resultados dos ajustes do modelo que leva em conta o termo autorregressivo de primeira ordem, e a Tabela 15 traz os resultados para o modelo que considera termos autoregressivos de primeira e segunda ordens.

As Tabelas 16, 17 e 18 mostram os resultados dos ajustes dos modelos de Poisson, para o número de acidentes com cobras no Estado de São Paulo, sem o termo autorregressivo, somente com termo autorregressivo de primeira ordem e com termos autorregressivos de primeira e segunda ordens, respectivamente.

As Figuras 6, 7 e 8 mostram gráficos das ICPO, valores preditos e observados e modelos ajustados para as comparações dos resultados obtidos a partir das modelagens baseadas na distribuição Poisson Dupla e na de Poisson, levando-se em conta os dados dos acidentes com cobras do gênero *Bothrops*. As linhas pontilhadas, nos gráficos da ICPO, representam os limites que definem possíveis *outliers* (limite =

Tabela 13: Resultados dos ajustes dos modelos de regressão Poisson dupla para os acidentes com cobras no Estado de São Paulo, entre 2007 e 2014 - Modelos sem termo autorregressivo.

Gênero	Parâmetro	Modelo 1 ($J = 1$)		Modelo 2 ($J = 2$)	
		Estimativa	ICr 95%	Estimativa	ICr 95%
<i>Bothrops</i>	α	4,380	(4,290; 4,475)	4,375	(4,304; 4,444)
	β	0,003	(0,001; 0,004)	0,002	(0,001; 0,004)
	θ	0,220	(0,163; 0,288)	0,384	(0,281; 0,501)
	η_{11}	0,307	(0,243; 0,370)	0,339	(0,291; 0,389)
	η_{12}	0,416	(0,354; 0,480)	0,467	(0,415; 0,513)
	η_{21}	*	*	-0,195	(-0,243; -0,147)
	η_{22}	*	*	-0,073	(-0,121; -0,025)
	<i>DIC</i>	857,6		806,1	
<i>Crotalus</i>	α	2,561	(2,149; 2,704)	2,550	(2,420; 2,675)
	β	0,004	(0,001; 0,006)	0,004	(0,002; 0,006)
	θ	0,564	(0,414; 0,739)	0,702	(0,517; 0,922)
	η_{11}	0,317	(0,221; 0,415)	0,313	(0,227; 0,398)
	η_{12}	0,227	(0,130; 0,323)	0,273	(0,182; 0,363)
	η_{21}	*	*	-0,144	(-0,231; -0,056)
	η_{22}	*	*	-0,164	(-0,248; 0,077)
	<i>DIC</i>	596,4		577,2	
Não peçonhentas	α	2,292	(2,142; 2,435)	2,287	(2,141; 2,432)
	β	0,003	(0,002; 0,005)	0,003	(0,002; 0,005)
	θ	0,685	(0,500; 0,894)	0,717	(0,523; 0,937)
	η_{11}	0,364	(0,262; 0,465)	0,376	(0,276; 0,481)
	η_{12}	0,391	(0,287; 0,494)	0,425	(0,317; 0,532)
	η_{21}	*	*	-0,109	(-0,210; -0,007)
	η_{22}	*	*	-0,064	(-0,164; -0,035)
	<i>DIC</i>	544,6		542,5	
Não identificadas	α	2,916	(2,810; 3,020)	2,910	(2,810; 3,009)
	β	0,004	(0,002; 0,006)	0,003	(0,002; 0,005)
	θ	0,700	(0,515; 0,912)	0,805	(0,592; 1,047)
	η_{11}	0,277	(0,206; 0,349)	0,294	(0,225; 0,362)
	η_{12}	0,376	(0,304; 0,449)	0,406	(0,336; 0,479)
	η_{21}	*	*	-0,124	(-0,192; -0,055)
	η_{22}	*	*	-0,062	(-0,129; 0,005)
	<i>DIC</i>	611,1		599,7	

Tabela 14: Resultados dos ajustes dos modelos de regressão Poisson dupla para os acidentes com cobras no Estado de São Paulo, entre 2007 e 2014 - Modelos com termo autorregressivo de primeira ordem.

Gênero	Parâmetro	Modelo 1 ($J = 1$)		Modelo 2 ($J = 2$)	
		Estimativa	ICr 95%	Estimativa	ICr 95%
<i>Bothrops</i>	α	4,440	(4,346; 4,537)	4,417	(4,340; 4,493)
	β	0,002	(-0,000002; 0,003)	0,002	(0,0002; 0,003)
	θ	0,244	(0,178; 0,317)	0,495	(0,301; 0,536)
	γ_1	0,003	(0,001; 0,006)	0,003	(0,001; 0,006)
	η_{11}	0,165	(0,059; 0,269)	0,216	(0,116; 0,312)
	η_{12}	0,353	(0,279; 0,429)	0,415	(0,352; 0,477)
	η_{21}	*	*	-0,165	(-0,219; -0,119)
	η_{22}	*	*	-0,100	(-0,153; -0,048)
	<i>DIC</i>		839,8		793,0
<i>Crotalus</i>	α	2,594	(2,444; 2,743)	2,551	(2,417; 2,689)
	β	0,003	(0,0005; 0,0057)	0,004	(0,001; 0,006)
	θ	0,567	(0,418; 0,738)	0,697	(0,512; 0,904)
	γ_1	0,010	(-0,002; 0,022)	0,004	(-0,008; 0,016)
	η_{11}	0,258	(0,134; 0,381)	0,288	(0,172; 0,402)
	η_{12}	0,226	(0,129; 0,323)	0,266	(0,174; 0,355)
	η_{21}	*	*	-0,136	(-0,233; -0,042)
	η_{22}	*	*	-0,167	(-0,252; -0,081)
	<i>DIC</i>		590,5		573,0
Não peçonhentas	α	2,259	(2,102; 2,416)	2,245	(2,087; 2,392)
	β	0,003	(0,0006; 0,006)	0,003	(0,0007; 0,006)
	θ	0,688	(0,505; 0,893)	0,728	(0,532; 0,954)
	γ_1	-0,002	(-0,019; 0,014)	-0,005	(-0,021; 0,012)
	η_{11}	0,374	(0,235; 0,514)	0,400	(0,259; 0,543)
	η_{12}	0,386	(0,278; 0,495)	0,425	(0,316; 0,535)
	η_{21}	*	*	-0,124	(-0,229; -0,019)
	η_{22}	*	*	-0,071	(-0,169; 0,029)
	<i>DIC</i>		539,1		535,3
Não identificadas	α	2,909	(2,792; 3,026)	2,888	(2,774; 2,994)
	β	0,003	(0,001; 0,006)	0,004	(0,002; 0,006)
	θ	0,685	(0,499; 0,894)	0,802	(0,585; 1,044)
	γ_1	0,001	(-0,008; 0,011)	-0,001	(-0,010; 0,008)
	η_{11}	0,262	(0,144; 0,382)	0,303	(0,188; 0,416)
	η_{12}	0,366	(0,280; 0,450)	0,406	(0,324; 0,489)
	η_{21}	*	*	-0,130	(-0,200; -0,059)
	η_{22}	*	*	-0,063	(-0,132; 0,006)
	<i>DIC</i>		606,6		593,9

Tabela 15: Resultados dos ajustes dos modelos de regressão Poisson dupla para os acidentes com cobras no Estado de São Paulo, entre 2007 e 2014 - Modelos com termos autorregressivos de primeira e segunda ordens.

Gênero	Parâmetro	Modelo 1 ($J = 1$)		Modelo 1 ($J = 2$)	
		Estimativa	ICr 95%	Estimativa	ICr 95%
<i>Bothrops</i>	α	4,426	(4,328; 4,517)	4,420	(4,345; 4,499)
	β	0,002	(0,0003; 0,004)	0,002	(0,0002; 0,003)
	θ	0,260	(0,184; 0,342)	0,398	(0,252; 0,544)
	γ_1	0,004	(0,002; 0,006)	0,002	(0,0004; 0,005)
	γ_2	-0,003	(-0,005; -0,681)	0,0003	(-0,002; 0,003)
	η_{11}	0,276	(0,140; 0,401)	0,220	(0,088; 0,334)
	η_{12}	0,327	(0,246; 0,410)	0,422	(0,348; 0,498)
	η_{21}	*	*	-0,162	(-0,221; -0,102)
	η_{22}	*	*	-0,101	(-0,160; -0,043)
	<i>DIC</i>		827,1		787,0
<i>Crotalus</i>	α	2,567	(2,414; 2,720)	2,538	(2,359; 2,673)
	β	0,004	(0,001; 0,006)	0,004	(0,002; 0,006)
	θ	0,601	(0,439; 0,787)	0,715	(0,520; 0,939)
	γ_1	0,011	(-0,0001; 0,023)	0,005	(-0,007; 0,017)
	γ_2	-0,015	(-0,027; -0,003)	-0,010	(-0,022; 0,002)
	η_{11}	0,329	(0,199; 0,458)	0,336	(0,208; 0,462)
	η_{12}	0,189	(0,089; 0,290)	0,238	(0,138; 0,340)
	η_{21}	*	*	-0,139	(-0,228; -0,048)
	η_{22}	*	*	-0,140	(-0,237; -0,047)
	<i>DIC</i>		580,1		566,5
Não peçonhentas	α	2,152	(2,152; 2,467)	2,288	(2,128; 2,451)
	β	0,002	(-0,0004; 0,005)	0,003	(-0,0002; 0,005)
	θ	0,707	(0,513; 0,932)	0,748	(0,547; 0,981)
	γ_1	0,001	(-0,016; 0,018)	-0,0006	(-0,016; 0,016)
	γ_2	0,003	(-0,014; 0,019)	0,004	(-0,013; 0,020)
	η_{11}	0,346	(0,171; 0,526)	0,360	(0,173; 0,546)
	η_{12}	0,390	(0,282; 0,496)	0,427	(0,314; 0,541)
	η_{21}	*	*	-0,108	(-0,214; -0,004)
	η_{22}	*	*	-0,079	(-0,180; 0,022)
	<i>DIC</i>		532,0		529,2
Não identificadas	α	2,872	(2,738; 3,005)	2,856	(2,733; 2,974)
	β	0,005	(0,002; 0,007)	0,005	(0,003; 0,007)
	θ	0,695	(0,507; 0,918)	0,836	(0,600; 1,095)
	γ_1	-0,0001	(-0,010; 0,010)	-0,003	(-0,012; 0,006)
	γ_2	-0,003	(-0,013; 0,008)	0,00002	(-0,009; 0,009)
	η_{11}	0,296	(0,132; 0,454)	0,312	(0,159; 0,463)
	η_{12}	0,367	(0,283; 0,450)	0,407	(0,327; 0,489)
	η_{21}	*	*	-0,144	(-0,214; -0,075)
	η_{22}	*	*	-0,057	(-0,127; 0,012)
	<i>DIC</i>		599,7		584,4

Tabela 16: Resultados dos ajustes dos modelos de regressão de Poisson para os acidentes com cobras no Estado de São Paulo, entre 2007 e 2014 - Modelos sem termo autorregressivo.

Gênero	Parâmetro	Modelo 1 ($J = 1$)		Modelo 2 ($J = 2$)	
		Estimativa	ICr 95%	Estimativa	ICr 95%
<i>Bothrops</i>	α	4,386	(4,343; 4,429)	4,377	(4,334; 4,420)
	β	0,0027	(0,002; 0,0034)	0,0025	(0,0018; 0,0032)
	η_{11}	0,306	(0,277; 0,336)	0,339	(0,309; 0,369)
	η_{12}	0,416	(0,386; 0,445)	0,465	(0,434; 0,497)
	η_{21}	*	*	-0,195	(-0,225; -0,165)
	η_{22}	*	*	-0,073	(-0,103; -0,043)
	<i>DIC</i>	1041,0		858,3	
<i>Crotalus</i>	α	2,563	(2,458; 2,665)	2,552	(2,445; 2,654)
	β	0,004	(0,002; 0,005)	0,003	(0,002; 0,005)
	η_{11}	0,318	(0,244; 0,389)	0,312	(0,242; 0,383)
	η_{12}	0,227	(0,156; 0,300)	0,271	(0,199; 0,345)
	η_{21}	*	*	-0,144	(-0,214; -0,073)
	η_{22}	*	*	-0,163	(-0,233; -0,092)
	<i>DIC</i>	610,5		578,6	
Não peçonhentas	α	2,292	(2,171; 2,415)	2,288	(2,167; 2,403)
	β	0,003	(0,0006; 0,005)	0,002	(0,0005; 0,004)
	η_{11}	0,363	(0,279; 0,448)	0,375	(0,287; 0,462)
	η_{12}	0,390	(0,306; 0,476)	0,424	(0,334; 0,514)
	η_{21}	*	*	-0,109	(-0,196; -0,030)
	η_{22}	*	*	-0,063	(-0,148; 0,021)
	<i>DIC</i>	548,1		543,4	
Não identificadas	α	2,917	(2,829; 3,005)	2,912	(2,820; 2,997)
	β	0,004	(0,002; 0,005)	0,003	(0,002; 0,005)
	η_{11}	0,277	(0,221; 0,336)	0,292	(0,228; 0,353)
	η_{12}	0,375	(0,315; 0,434)	0,406	(0,343; 0,469)
	η_{21}	*	*	-0,124	(-0,184; -0,065)
	η_{22}	*	*	-0,062	(-0,126; -0,003)
	<i>DIC</i>	614,0		598,0	

Tabela 17: Resultados dos ajustes dos modelos de regressão de Poisson para os acidentes com cobras no Estado de São Paulo, entre 2007 e 2014 - Modelos com termo autorregressivo de primeira ordem.

Gênero	Parâmetro	Modelo 1 ($J = 1$)		Modelo 2 ($J = 2$)	
		Estimativa	ICr 95%	Estimativa	ICr 95%
<i>Bothrops</i>	α	4,444	(4,397; 4,494)	4,415	(4,367; 4,462)
	β	0,001	(0,0006; 0,002)	0,001	(0,0009; 0,002)
	γ_1	0,003	(0,002; 0,004)	0,003	(0,001; 0,004)
	η_{11}	0,162	(0,101; 0,216)	0,213	(0,133; 0,278)
	η_{12}	0,348	(0,309; 0,388)	0,412	(0,350; 0,453)
	η_{21}	*	*	-0,163	(-0,196; -0,129)
	η_{22}	*	*	-0,100	(-0,138; -0,067)
	<i>DIC</i>	988,3		838,4	
<i>Crotalus</i>	α	2,595	(2,482; 2,712)	2,554	(2,440; 2,664)
	β	0,003	(0,001; 0,005)	0,003	(0,002; 0,005)
	γ_1	0,009	(0,0002; 0,002)	0,004	(-0,005; 0,014)
	η_{11}	0,260	(0,169; 0,351)	0,285	(0,192; 0,383)
	η_{12}	0,226	(0,155; 0,301)	0,265	(0,188; 0,341)
	η_{21}	*	*	-0,135	(-0,212; -0,053)
	η_{22}	*	*	-0,169	(-0,244; -0,096)
	<i>DIC</i>	603,5		574,6	
Não peçonhentas	α	2,260	(2,128; 2,385)	2,248	(2,125; 2,374)
	β	0,003	(0,001; 0,006)	0,003	(0,001; 0,005)
	γ_1	-0,003	(-0,016; 0,011)	-0,004	(-0,018; 0,009)
	η_{11}	0,376	(0,258; 0,489)	0,398	(0,283; 0,515)
	η_{12}	0,384	(0,293; 0,473)	0,424	(0,337; 0,517)
	η_{21}	*	*	-0,125	(-0,213; -0,041)
	η_{22}	*	*	-0,072	(-0,158; 0,015)
	<i>DIC</i>	541,9		535,2	
Não identificadas		0,384			
	α	2,909	(2,812; 3,005)	2,891	(2,797; 2,992)
	β	0,004	(0,002; 0,006)	0,004	(0,002; 0,006)
	γ_1	0,001	(-0,007; 0,009)	-0,001	(-0,009; 0,008)
	η_{11}	0,262	(0,165; 0,358)	0,299	(0,191; 0,403)
	η_{12}	0,368	(0,299; 0,437)	0,406	(0,330; 0,477)
	η_{21}	*	*	-0,129	(-0,193; -0,068)
	η_{22}	*	*	-0,065	(-0,125; -0,004)
	<i>DIC</i>	609,5		591,8	

Tabela 18: Resultados dos ajustes dos modelos de regressão de Poisson para os acidentes com cobras no Estado de São Paulo, entre 2007 e 2014 - Modelos com termos autorregressivos de primeira e segunda ordens.

Gênero	Parâmetro	Modelo 1 ($J = 1$)		Modelo 2 ($J = 2$)	
		Estimativa	ICr 95%	Estimativa	ICr 95%
<i>Bothrops</i>	α	4,404	(4,374; 4,457)	4,420	(4,346; 4,477)
	β	0,002	(0,0015; 0,0033)	0,002	(0,001; 0,003)
	γ_1	-0,001	(-0,002; 0,0001)	0,001	(-0,0002; 0,006)
	γ_2	0,0002	(-0,001; 0,002)	0,0005	(-0,007; 0,003)
	η_{11}	0,360	(0,278; 0,433)	0,257	(0,035; 0,305)
	η_{12}	0,430	(0,364; 0,502)	0,484	(0,052; 0,600)
	η_{21}	*	*	-0,189	(-0,259; -0,008)
	η_{22}	*	*	-0,088	(-0,244; -0,039)
	<i>DIC</i>		1008,0		838,4
<i>Crotalus</i>	α	2,525	(2,406; 2,644)	2,544	(2,427; 2,662)
	β	0,005	(0,003; 0,007)	0,004	(0,002; 0,006)
	γ_1	-0,014	(-0,024; -0,004)	-0,009	(-0,019; 0,002)
	γ_2	-0,001	(-0,010; 0,008)	0,008	(-0,003; 0,019)
	η_{11}	0,397	(0,308; 0,487)	0,354	(0,263; 0,448)
	η_{12}	0,182	(0,082; 0,282)	0,294	(0,185; 0,403)
	η_{21}	*	*	-0,178	(-0,254; -0,098)
	η_{22}	*	*	-0,137	(-0,212; -0,061)
	<i>DIC</i>		594,8		565,3
Não peçonhentas	α	2,310	(2,182; 2,435)	2,292	(2,161; 2,422)
	β	0,002	(0,0001; 0,005)	0,002	(0,0002; 0,005)
	γ_1	0,003	(-0,011; 0,017)	0,003	(-0,011; 0,017)
	γ_2	0,007	(-0,007; 0,021)	0,008	(-0,007; 0,022)
	η_{11}	0,337	(0,212; 0,458)	0,342	(0,216; 0,464)
	η_{12}	0,434	(0,311; 0,559)	0,472	(0,349; 0,599)
	η_{21}	*	*	-0,111	(-0,197; -0,026)
	η_{22}	*	*	-0,078	(-0,161; 0,010)
	<i>DIC</i>		532,2		527,0
Não identificadas	α	2,854	(2,754; 2,952)	2,859	(2,751; 2,965)
	β	0,005	(0,003; 0,007)	0,005	(0,003; 0,006)
	γ_1	-0,002	(-0,010; 0,006)	0,0004	(-0,008; 0,009)
	γ_2	-0,006	(-0,014; 0,003)	-0,005	(-0,014; 0,004)
	η_{11}	0,318	(0,202; 0,427)	0,299	(0,186; 0,416)
	η_{12}	0,305	(0,200; 0,406)	0,349	(0,240; 0,455)
	η_{21}	*	*	-0,132	(-0,195; -0,064)
	η_{22}	*	*	-0,066	(-0,137; 0,002)
	<i>DIC</i>		599,5		582,2

40), e os considerados extremos (limite = 70) (NTZOUFRAS, 2009). Devido à presença de valores muito discrepantes, os valores da ICPO foram considerados na forma de logaritmo natural.

Com base nas Figuras 6, 7 e 8, nota-se que os modelos que consideram dois pares de funções seno e cosseno ($J = 2$) ajustam melhor os dados, tanto quando se utiliza a distribuição de Poisson quanto a Poisson Dupla, sendo esta a melhor opção para o ajuste, levando-se em conta os valores para ICPO, DIC e LPML. Com relação aos termos autorregressivos, o modelo que considera apenas o termo de primeira ordem e dois pares de funções seno e cosseno ($J = 2$), ajusta melhor os dados, sendo que parece haver uma correlação entre o número de picadas registradas em um mês e o das do mês anterior (ver gráficos de autocorrelações no Apêndice 1). Também, nota-se que a série tem um comportamento crescente ($\beta > 0$) ao longo do período, indicando que existe um aumento do número de casos ou um interesse maior na notificação dos mesmos.

As Figuras 9, 10 e 11 mostram gráficos das ICPO, valores preditos e observados e modelos ajustados para as comparações dos resultados obtidos a partir das modelagens baseadas na distribuição Poisson Dupla e na de Poisson, considerando-se os dados dos acidentes com cobras do gênero *Crotalus*.

Observa-se nas Figuras 9, 10 e 11, que os modelos que consideram dois pares de funções seno e cosseno ($J = 2$) ajustam melhor os dados, tanto quando se utiliza a distribuição de Poisson quanto a Poisson Dupla. Existe um ganho no ajuste quando se considera ($J = 2$) com relação aos modelos que consideram ($J = 1$), mas não quando se utiliza uma distribuição ou outra na modelagem. Também, parece não existir autocorrelação entre as picadas ocorridas em um mês e as do mês anterior (ver gráficos de autocorrelações no Apêndice 1). Nota-se, ainda, que a série tem um comportamento crescente ($\beta > 0$) ao longo do período, indicando que existe um aumento do número de casos ou um interesse maior na notificação dos mesmos.

As Figuras 12, 13 e 14 mostram gráficos das ICPO, valores preditos e observados e modelos ajustados para as comparações dos resultados obtidos a partir das modelagens baseadas na distribuição Poisson Dupla e na de Poisson, considerando-se os dados dos acidentes com cobras não peçonhentas.

Nas Figuras 12, 13 e 14, observa-se que não existe um ganho significativo nas modelagens que consideram dois pares de funções seno e cosseno, quando comparadas às que levam em conta apenas um par. Também, o uso da distribuição Poisson Dupla não trouxe maior ganho quando comparado com os modelos de Poisson. Quanto aos parâmetros autorregressivos, parece não existir autocorrelação entre as picadas ocorridas em um mês e as do mês anterior (ver gráficos de autocorrelações no Apêndice

(a) DP - Sem termo autorregressivo

(b) Poisson - Sem termo autorregressivo

(c) DP - AR(1)

(d) Poisson - AR(1)

(e) DP - AR(2)

(f) Poisson - AR(2)

Figura 6: Gráficos da ICPO para os ajustes dos modelos Poisson Dupla e de Poisson - cobras do gênero *Bothrops*.

(a) DP - Sem termo autorregressivo

(b) Poisson - Sem termo autorregressivo

(c) DP - AR(1)

(d) Poisson - AR(1)

(e) DP - AR(2)

(f) Poisson - AR(2)

Figura 7: Gráficos dos valores preditos e observados para os ajustes dos modelos Poisson Dupla e de Poisson - cobras do gênero *Bothrops*.

(a) DP - Sem termo autorregressivo

(b) Poisson - Sem termo autorregressivo

(c) DP - AR(1)

(d) Poisson - AR(1)

(e) DP - AR(2)

(f) Poisson - AR(2)

Figura 8: Gráficos de séries dos registros de acidentes e curvas ajustadas pelos modelos Poisson Dupla e de Poisson - cobras do gênero *Bothrops*.

(a) DP - Sem termo autorregressivo

(b) Poisson - Sem termo autorregressivo

(c) DP - AR(1)

(d) Poisson - AR(1)

(e) DP - AR(2)

(f) Poisson - AR(2)

Figura 9: Gráficos da ICPO para os ajustes dos modelos Poisson Dupla e de Poisson - cobras do gênero *Crotalus*.

(a) DP - Sem termo autorregressivo

(b) Poisson - Sem termo autorregressivo

(c) DP - AR(1)

(d) Poisson - AR(1)

(e) DP - AR(2)

(f) Poisson - AR(2)

Figura 10: Gráficos dos valores preditos e observados para os ajustes dos modelos Poisson Dupla e de Poisson - cobras do gênero *Crotalus*.

(a) DP - Sem termo autorregressivo

(b) Poisson - Sem termo autorregressivo

(c) DP - AR(1)

(d) Poisson - AR(1)

(e) DP - AR(2)

(f) Poisson - AR(2)

Figura 11: Gráficos de séries dos registros de acidentes e curvas ajustadas pelos modelos Poisson Dupla e de Poisson - cobras do gênero *Crotalus*.

(a) DP - Sem termo autorregressivo

(b) Poisson - Sem termo autorregressivo

(c) DP - AR(1)

(d) Poisson - AR(1)

(e) DP - AR(2)

(f) Poisson - AR(2)

Figura 12: Gráficos da ICPO para os ajustes dos modelos Poisson Dupla e de Poisson - cobras não peçonhentas.

(a) DP - Sem termo autorregressivo

(b) Poisson - Sem termo autorregressivo

(c) DP - AR(1)

(d) Poisson - AR(1)

(e) DP - AR(2)

(f) Poisson - AR(2)

Figura 13: Gráficos dos valores preditos e observados para os ajustes dos modelos Poisson Dupla e de Poisson - cobras não peçonhentas.

(a) DP - Sem termo autorregressivo

(b) Poisson - Sem termo autorregressivo

(c) DP - AR(1)

(d) Poisson - AR(1)

(e) DP - AR(2)

(f) Poisson - AR(2)

Figura 14: Gráficos de séries dos registros de acidentes e curvas ajustadas pelos modelos Poisson Dupla e de Poisson - cobras não peçonhentas.

1). Nota-se, ainda, que a série tem um comportamento crescente ($\beta > 0$) ao longo do período, indicando que existe um aumento do número de casos ou um interesse maior na notificação dos mesmos.

As Figuras 15, 16 e 17 mostram gráficos das ICPO, valores preditos e observados e modelos ajustados para as comparações dos resultados obtidos a partir das modelagens baseadas na distribuição Poisson Dupla e na de Poisson, considerando-se os dados dos acidentes com cobras não identificadas.

Nas Figuras 12, 13 e 14, nota-se que não existe um ganho significativo nas modelagens que consideram dois pares de funções seno e cosseno, quando comparadas às que levam em conta apenas um par. Também, o uso da distribuição Poisson Dupla não trouxe maior ganho quando comparado com os modelos de Poisson. Quanto aos parâmetros autorregressivos, parece não existir autocorrelação entre as picadas ocorridas em um mês e as do mês anterior (ver gráficos de autocorrelações no Apêndice 1). Nota-se que, ainda, que a série tem um comportamento crescente ($\beta > 0$) ao longo do período, indicando que existe um aumento do número de casos ou um interesse maior na notificação dos mesmos.

De forma geral, houve um ganho significativo referente ao uso da distribuição Poisson Dupla nas modelagens dos dados de picadas de cobra dos gêneros *Bothrops* e *Crotalus*. Isso pode ser devido ao fato de serem casos em que a variabilidade dos dados é maior do que o que ocorre com os dados dos acidentes com cobras não peçonhentas e não identificadas. O uso de dois pares de funções seno e cosseno, para contemplar a sazonalidade dos dados, parece ter contribuído bastante para um melhor ajuste dos modelos baseados na distribuição Poisson Dupla e de Poisson. Pelo menos um dos pares foi significativo, de acordo com os resultados mostrados nas tabelas (ver parâmetros η_{11} , η_{12} , η_{21} e η_{22}). A inserção do parâmetro autorregressivo de primeira ordem (γ_1) trouxe um melhor ajuste dos modelos aos dados de picadas de cobras do gênero *Bothrops*. Os resultados também corroboraram o fato de os acidentes com cobras ocorrerem em períodos quentes, mais especificamente entre novembro e abril, pois esses animais necessitam de temperaturas mais altas para realizarem um metabolismo satisfatório.

Uma limitação nessas conclusões pode estar no fato de, muitas vezes, não ser possível a identificação da cobra causadora do acidente, podendo haver uma subnotificação do número de picadas de acordo com o gênero conhecido. Essa subnotificação pode ser ainda maior, dado que, no Brasil, existe um grande número de acidentes que não são notificados (MACHADO et al., 2012).

(a) DP - Sem termo autorregressivo

(b) Poisson - Sem termo autorregressivo

(c) DP - AR(1)

(d) Poisson - AR(1)

(e) DP - AR(2)

(f) Poisson - AR(2)

Figura 15: Gráficos da ICPO para os ajustes dos modelos Poisson Dupla e de Poisson - cobras não identificadas.

(a) DP - Sem termo autorregressivo

(b) Poisson - Sem termo autorregressivo

(c) DP - AR(1)

(d) Poisson - AR(1)

(e) DP - AR(2)

(f) Poisson - AR(2)

Figura 16: Gráficos dos valores preditos e observados para os ajustes dos modelos Poisson Dupla e de Poisson - cobras não identificadas.

(a) DP - Sem termo autorregressivo

(b) Poisson - Sem termo autorregressivo

(c) DP - AR(1)

(d) Poisson - AR(1)

(e) DP - AR(2)

(f) Poisson - AR(2)

Figura 17: Gráficos de séries dos registros de acidentes e curvas ajustadas pelos modelos Poisson Dupla e de Poisson - cobras não identificadas.

5.3.2 Registros de Acidentes com Escorpiões em Ribeirão Preto, São Paulo, entre 2007 e 2014

No Brasil, acidentes com picadas de escorpiões são um problema de saúde pública, tanto pela alta incidência em algumas regiões quanto pelas complicações causadas pela toxicidade do veneno, principalmente em crianças. Os registros desse tipo de acidente começaram a ser feitos desde 1988, pelo Ministério da Saúde, mas somente a partir de 1997 os registros passaram a ser responsabilidade do SINAN, fonte oficial de notificações compulsórias (RECKZIEGEL e PINTO, 2014). No Estado de São Paulo, as espécies mais comumente responsáveis pelos acidentes são *T. bahiensis*, conhecido como escorpião marrom, e *T. serrulatus*, conhecido como escorpião amarelo, sendo este o que possui o veneno mais potente e, conseqüentemente, o que causa a grande parte de agravos à saúde (CUPO, 2015). Segundo dados do DATASUS/SINAN (<http://tabnet.datasus.gov.br>), no período entre 2007 e 2014, as maiores vítimas dos escorpiões foram indivíduos do sexo masculino (60,6%) e pessoas com idade entre 20 e 59 anos (61,2%). Os acidentes com crianças até 9 anos representaram 11,2% do total. As duas espécies de escorpiões (marrom e amarelo), ao contrário de muitas outras, não possuem um padrão quanto à sua distribuição, sendo muito bem adaptadas a ambientes modificados pelo homem. Sua atividade é maior entre os meses de setembro e março, período mais quente e com alta precipitação, o que facilita sua dispersão para novos abrigos. As temperaturas entre 14 °C e 38 °C são consideradas ótimas para que as espécies possam sobreviver, mas elas podem suportar temperaturas menores. Em contrapartida, reagem negativamente às temperaturas maiores, sendo esse fator o mais fortemente associado às atividades dos escorpiões (HOSHINO et al., 2006; CLOUDSLEY-THOMPSON, 1975; CHOWELL et al., 2005).

A cidade de Ribeirão Preto está localizada na região Nordeste do Estado de São Paulo e é considerada um local de grande incidência de acidentes com escorpiões, devido ao clima favorável, que é o tropical semiúmido (Aw na classificação climática de Köppen-Geiger). Considerando-se os últimos 25 anos, as temperaturas médias mensais, máxima e mínima, foram, respectivamente, 16,8 °C e 29,1 °C, e a média mensal de precipitação foi de 118,1 mm (<http://www.ciiagro.sp.gov.br/>). Entre os anos de 2007 e 2014, 55,2% dos acidentes notificados na cidade foram com indivíduos do sexo masculino e 47,9% com aqueles com idade entre 20 e 59 anos. No que diz respeito aos acidentes com crianças de até 9 anos, a cidade possui taxas bem maiores que as do Estado de São Paulo, com 23,5% dos acidentes.

A Tabela 19 apresenta o número de registros anuais de acidentes com picadas de escorpiões, obtidos na base de dados do DATASUS/SINAN, notificados no Brasil, Estado de São Paulo e Ribeirão Preto - SP, entre os anos de 2007 e 2014.

Tabela 19: Registros anuais de acidentes com escorpiões no Brasil, no Estado de São Paulo e em Ribeirão Preto, SP, entre 2007 e 2014 - Fonte: DATASUS/SINAN.

Ano	Brasil	Estado de SP	Ribeirão Preto
2007	37340	4351	250
2008	40320	5328	341
2009	50383	6452	269
2010	51600	7128	169
2011	59200	7017	154
2012	64936	9251	183
2013	79705	11479	184
2014	88247	12592	162

A Figura 18 apresenta dados de acidentes mensais com picadas de escorpiões, obtidos na base de dados do DATASUS/SINAN, notificados no Brasil, Estado de São Paulo e Ribeirão Preto - SP, entre os anos de 2007 e 2014.

A Figura 19 apresenta dados de acidentes mensais com picadas de escorpiões, notificados em Ribeirão Preto, bem como os dados climáticos, como temperaturas médias, máxima e mínima e precipitações médias mensais, em Ribeirão Preto - SP, entre os anos de 2007 e 2014.

O objetivo das modelagens a seguir é verificar a possível associação entre a incidência das picadas de escorpiões e as variáveis climáticas, considerando-se modelos de regressão de Poisson e baseados na distribuição Poisson Dupla, com e sem adição de termos autorregressivos de primeira e segunda ordens e, ainda, com $J = 1$ e $J = 2$, definindo um ou dois pares de funções seno e cosseno para S_t , respectivamente. Portanto, a média (μ_t), para ambas as distribuições, é expressa da seguinte forma:

$$\log(\mu_t) = \alpha + \beta t + S_t + \sum_{j=1}^r \gamma_j (y_{t-j} - \bar{y}) + \lambda_1 T_{\max} + \lambda_2 T_{\min} + \lambda_3 \text{Prec}, \quad (14)$$

em que α é uma constante; β é a tendência linear ao longo do tempo; S_t é a função periódica que representa todo o ciclo sazonal ao longo das T unidades de tempo ($T = 12$, considerando-se a sazonalidade anual); γ_j são os termos autorregressivos de r -ésima ordem; $\bar{y}$ é a média do número de picadas de cobra no período; λ_1 é o parâmetro referente à temperatura máxima média mensal ($T_{\max}$); λ_2 é o parâmetro referente à temperatura mínima média mensal ($T_{\min}$) e λ_3 é o parâmetro referente à precipitação média mensal (Prec). Como distribuições a priori, não informativas, para os parâmetros $\alpha, \beta, \theta, \eta_{11}, \eta_{12}, \eta_{21}, \eta_{22}, \lambda_1, \lambda_2$ e λ_3 foram consideradas distribuições $N(0; 1000)$ e, para γ_1 e γ_2 , distribuições $U(-1; 1)$. Nas simulações, baseadas no algoritmo MCMC, foram

Figura 18: Séries temporais de registros de acidentes com escorpiões no Brasil, no Estado de São Paulo e em Ribeirão Preto, SP, entre 2007 e 2014 - Fonte: DATASUS/-SINAN.

Figura 19: Séries temporais de registros de acidentes com escorpiões, temperaturas máximas e mínimas médias mensais e precipitação mensal, em Ribeirão Preto, SP, entre 2007 e 2014 - Fonte: DATASUS/SINAN.

geradas 1.000.000 de amostras, considerando-se um *burn-in period* de 1000 amostras, para eliminação da influência dos valores iniciais, além de saltos de tamanho 100, ou seja, tomando-se as 100^a, 200^a, 300^a, ... iterações para a obtenção das estimativas a posteriori, evitando-se possíveis autocorrelações entre elas.

A Tabela 20 mostra os resultados dos ajustes do modelo de regressão Poisson Dupla, para o número de acidentes com escorpiões na cidade de Ribeirão Preto, SP, considerando-se o modelo dado por (14), sem o termo autorregressivo. A Tabela 21 mostra os resultados dos ajustes do modelo de regressão Poisson Dupla, levando-se em conta um termo autoregressivo de primeira ordem. A Tabela 22 apresenta os resultados do modelo que considera termos autorregressivos de primeira e segunda ordens. O modelo M1 não contém covariáveis; o M2 considera a temperatura máxima média mensal (Tmax); o M3 contempla a temperatura mínima média mensal (Tmin); o M4 considera a precipitação média mensal (Prec), e o M5 contempla as três covariáveis no mesmo ajuste. Além disso, foram ajustados modelos com um ($J = 1$) e dois ($J = 2$) pares de funções seno e cosseno. Por uma questão de problemas de convergência do algoritmo, não foi possível se obterem as estimativas do modelo que considera dois pares de funções seno e cosseno ($J = 2$) e as covariáveis Tmax, Tmin e Precipitação, no mesmo ajuste, para as duas situações em que termos autorregressivos foram inseridos no modelo. Portanto, nesses casos, os resultados apresentados consideram apenas um par dessas funções ($J = 1$).

Nas Figuras 20, 23 e 26, observam-se os gráficos da ICPO para os ajustes dos modelos de regressão baseados na distribuição Poisson Dupla, considerando-se termos autorregressivos e um ou dois pares de funções seno e cosseno. As Figuras 21, 24 e 27 apresentam os gráficos dos valores preditos e observados para as mesmas situações. As Figuras 22, 25 e 28 mostram os gráficos dos ajustes dos modelos.

A Tabela 23 contém os resultados dos ajustes do modelo de regressão de Poisson, para o número de acidentes com escorpiões na cidade de Ribeirão Preto, SP, considerando-se o modelo dado por (14), sem o termo autorregressivo. A Tabela 24 mostra os resultados dos ajustes do modelo de regressão de Poisson, levando-se em conta um termo autoregressivo de primeira ordem. A Tabela 25 apresenta os resultados do modelo que considera termos autorregressivos de primeira e segunda ordens. O modelo M1 não contém covariáveis; o M2 considera a temperatura máxima média mensal (Tmax); o M3 contempla a temperatura mínima média mensal (Tmin); o M4 considera a precipitação média mensal (Prec), e o M5 contempla as três covariáveis no mesmo ajuste. Além disso, foram ajustados modelos com um ($J = 1$) e dois ($J = 2$) pares de funções seno e cosseno. Por uma questão de problemas de convergência do algoritmo, não foi possível se obterem as estimativas do modelo que considera dois pares de funções seno e cosseno ($J = 2$) e as covariáveis Tmax, Tmin e Precipitação,

Tabela 20: Resultados dos ajustes dos modelos de regressão Poisson Dupla para os acidentes com escorpiões em Ribeirão Preto - SP, entre 2007 e 2014 - Modelos sem termo autorregressivo.

Parâmetro	Modelo 1 ($J = 1$)		Modelo 2 ($J = 2$)	
	Estimativa	ICr 95%	Estimativa	ICr 95%
M1				
α	3,152	(3,014; 3,288)	3,148	(3,012; 3,284)
β	-0,009	(-0,012; -0,006)	-0,009	(-0,012; -0,006)
θ	0,493	(0,360; 0,649)	0,497	(0,368; 0,658)
η_{11}	-0,113	(-0,215; -0,013)	-0,111	(-0,218; -0,008)
η_{12}	0,156	(0,051; 0,259)	0,164	(0,059; 0,275)
η_{21}	*	*	0,007	(-0,096; 0,111)
η_{22}	*	*	-0,076	(-0,177; 0,025)
<i>DIC</i>		603,6		605,2
M2				
α	3,159	(1,597; 4,724)	4,397	(0,723; 6,644)
β	-0,009	(-0,012; -0,006)	-0,009	(-0,012; -0,006)
θ	0,490	(0,358; 0,645)	0,495	(0,346; 0,659)
λ_1	-0,0001	(-0,053; 0,053)	-0,042	(-0,118; 0,084)
η_{11}	-0,113	(-0,214; -0,013)	-0,117	(-0,217; -0,006)
η_{12}	0,158	(0,014; 0,300)	0,246	(0,029; 0,428)
η_{21}	*	*	0,002	(-0,101; 0,111)
η_{22}	*	*	-0,130	(-0,271; 0,014)
<i>DIC</i>		605,5		605,9
M3				
α	2,961	(1,790; 4,108)	3,608	(2,145; 5,104)
β	-0,009	(-0,012; -0,006)	-0,009	(-0,012; -0,006)
θ	0,491	(0,362; 0,641)	0,490	(0,356; 0,640)
λ_2	0,011	(-0,056; 0,079)	-0,027	(-0,115; 0,058)
η_{11}	-0,127	(-0,257; -0,001)	-0,074	(-0,255; 0,079)
η_{12}	0,122	(-0,118; 0,355)	0,253	(-0,048; 0,550)
η_{21}	*	*	-0,002	(-0,109; 0,103)
η_{22}	*	*	-0,100	(-0,231; 0,036)
<i>DIC</i>		605,3		607,1
M4				
α	3,181	(3,001; 3,370)	3,159	(2,979; 3,348)
β	-0,009	(-0,012; -0,006)	-0,009	(-0,012; -0,006)
θ	0,489	(0,362; 0,638)	0,492	(0,366; 0,639)
λ_3	-0,002	(-0,013; 0,008)	-0,001	(-0,011; 0,009)
η_{11}	-0,103	(-0,215; 0,012)	-0,104	(-0,215; 0,008)
η_{12}	0,182	(0,032; 0,326)	0,173	(0,026; 0,315)
η_{21}	*	*	0,013	(-0,089; 0,117)
η_{22}	*	*	-0,070	(-0,175; 0,033)
<i>DIC</i>		605,4		606,9
M5				
α	3,633	(1,537; 5,789)	5,459	(2,860; 8,228)
β	-0,009	(-0,012; -0,006)	-0,009	(-0,012; -0,006)
θ	0,481	(0,355; 0,635)	0,490	(0,362; 0,642)
λ_1	-0,035	(-0,127; 0,059)	-0,088	(-0,200; 0,018)
λ_2	0,034	(-0,063; 0,131)	0,022	(-0,083; 0,130)
λ_3	-0,005	(-0,020; 0,008)	-0,009	(-0,023; 0,005)
η_{11}	-0,134	(-0,288; 0,034)	-0,114	(-0,283; 0,049)
η_{12}	0,166	(-0,129; 0,471)	0,348	(0,023; 0,668)
η_{21}	*	*	0,027	(-0,083; 0,138)
η_{22}	*	*	-0,149	(-0,303; -0,009)
<i>DIC</i>		609,1		608,1

(a) Sem covariável

(b) Temperatura Máxima (Tmax)

(c) Temperatura Mínima (Tmin)

(d) Precipitação

(e) Tmax, Tmin e Precipitação

Figura 20: Gráficos da ICPO para os ajustes dos modelos Poisson Dupla, sem termo autorregressivo.

(a) Sem covariável

(b) Temperatura Máxima (Tmax)

(c) Temperatura Mínima (Tmin)

(d) Precipitação

(e) Tmax, Tmin e Precipitação

Figura 21: Gráficos dos valores preditos e observados para os ajustes dos modelos Poisson Dupla, sem termo autorregressivo.

(a) Sem covariável

(b) Temperatura Máxima (Tmax)

(c) Temperatura Mínima (Tmin)

(d) Precipitação

(e) Tmax, Tmin e Precipitação

Figura 22: Gráficos de séries temporais dos dados de acidentes com escorpiões e dos ajustes dos modelos Poisson Dupla, sem termo autorregressivo.

Tabela 21: Resultados dos ajustes dos modelos de regressão Poisson Dupla para os acidentes com escorpiões em Ribeirão Preto - SP, entre 2007 e 2014 - Modelos com termo autorregressivo de primeira ordem.

Parâmetro	Modelo 1 ($J = 1$)		Modelo 2 ($J = 2$)	
	Estimativa	ICr 95%	Estimativa	ICr 95%
M1				
α	3,014	(2,850; 3,174)	3,009	(2,841; 3,182)
β	-0,006	(-0,009; -0,003)	-0,006	(-0,010; -0,003)
θ	0,525	(0,384; 0,691)	0,524	(0,378; 0,687)
γ_1	0,016	(0,005; 0,028)	0,016	(0,005; 0,028)
η_{11}	-0,112	(-0,209; -0,013)	-0,109	(-0,206; -0,011)
η_{12}	0,099	(-0,006; 0,209)	0,108	(0,002; 0,220)
η_{21}	*	*	0,021	(-0,078; 0,121)
η_{22}	*	*	-0,068	(-0,171; 0,034)
<i>DIC</i>		591,7		593,5
M2				
α	3,058	(1,470; 4,624)	4,191	(2,078; 6,294)
β	-0,006	(-0,009; -0,003)	-0,006	(-0,010; -0,003)
θ	0,521	(0,380; 0,687)	0,529	(0,383; 0,688)
γ_1	0,016	(0,005; 0,028)	0,016	(0,004; 0,028)
λ_1	-0,001	(-0,055; 0,050)	-0,040	(-0,110; 0,031)
η_{11}	-0,112	(-0,211; -0,013)	-0,120	(-0,219; -0,018)
η_{12}	0,104	(-0,039; 0,253)	0,184	(0,008; 0,360)
η_{21}	*	*	0,018	(-0,087; 0,123)
η_{22}	*	*	-0,123	(-0,263; 0,013)
<i>DIC</i>		593,5		594,1
M3				
α	2,809	(1,721; 3,946)	3,271	(1,796; 4,710)
β	-0,006	(-0,009; -0,003)	-0,007	(-0,010; -0,003)
θ	0,522	(0,383; 0,678)	0,526	(0,388; 0,696)
γ_1	0,017	(0,005; 0,028)	0,016	(0,004; 0,028)
λ_2	0,012	(-0,053; 0,074)	-0,015	(-0,097; 0,072)
η_{11}	-0,125	(-0,252; 0,003)	-0,089	(-0,225; 0,058)
η_{12}	0,064	(-0,175; 0,301)	0,156	(-0,139; 0,453)
η_{21}	*	*	0,015	(-0,094; 0,126)
η_{22}	*	*	-0,084	(-0,212; 0,044)
<i>DIC</i>		593,5		595,4
M4				
α	3,044	(2,831; 3,254)	3,023	(2,810; 3,240)
β	-0,006	(-0,010; -0,004)	-0,006	(-0,009; -0,003)
θ	0,521	(0,378; 0,685)	0,520	(0,381; 0,683)
γ_1	0,016	(0,004; 0,028)	0,017	(0,005; 0,028)
λ_3	-0,002	(-0,014; 0,009)	-0,001	(-0,013; 0,011)
η_{11}	-0,099	(-0,251; 0,012)	-0,104	(-0,220; 0,009)
η_{12}	0,125	(-0,032; 0,284)	0,117	(-0,038; 0,282)
η_{21}	*	*	0,023	(-0,077; 0,217)
η_{22}	*	*	-0,068	(-0,172; 0,037)
<i>DIC</i>		593,7		595,6
M5				
α	3,542	(1,430; 5,709)	*	*
β	-0,006	(-0,009; -0,003)	*	*
θ	0,510	(0,371; 0,678)	*	*
γ_1	0,016	(0,004; 0,028)	*	*
λ_1	-0,036	(-0,133; 0,061)	*	*
λ_2	0,035	(-0,062; 0,134)	*	*
λ_3	-0,006	(-0,022; 0,010)	*	*
η_{11}	-0,132	(-0,288; 0,029)	*	*
η_{12}	0,110	(-0,197; 0,411)	*	*
η_{21}	*	*	*	*
η_{22}	*	*	*	*
<i>DIC</i>		597,4		*

(a) Sem covariável

(b) Temperatura Máxima (Tmax)

(c) Temperatura Mínima (Tmin)

(d) Precipitação

(e) Tmax, Tmin e Precipitação

Figura 23: Gráficos da ICPO para os ajustes dos modelos Poisson Dupla, com termo autorregressivo de primeira ordem.

(a) Sem covariável

(b) Temperatura Máxima (Tmax)

(c) Temperatura Mínima (Tmin)

(d) Precipitação

(e) Tmax, Tmin e Precipitação

Figura 24: Gráficos dos valores preditos e observados para os ajustes dos modelos Poisson Dupla, com termo autorregressivo de primeira ordem.

(a) Sem covariável

(b) Temperatura Máxima (Tmax)

(c) Temperatura Mínima (Tmin)

(d) Precipitação

(e) Tmax, Tmin e Precipitação

Figura 25: Gráficos de séries temporais dos dados de acidentes com escorpiões e dos ajustes dos modelos Poisson Dupla, com termo autorregressivo de primeira ordem.

Tabela 22: Resultados dos ajustes dos modelos de regressão Poisson Dupla para os acidentes com escorpiões em Ribeirão Preto - SP, entre 2007 e 2014 - Modelos com termos autorregressivos de primeira e segunda ordens.

Parâmetro	Modelo 1 ($J = 1$)		Modelo 2 ($J = 2$)	
	Estimativa	ICr 95%	Estimativa	ICr 95%
M1				
α	2,978	(2,801; 3,156)	2,966	(2,793; 3,141)
β	-0,006	(-0,009; -0,003)	-0,005	(-0,009; -0,002)
θ	0,538	(0,393; 0,712)	0,538	(0,392; 0,707)
γ_1	0,015	(0,002; 0,027)	0,014	(0,003; 0,026)
γ_2	0,009	(-0,004; 0,021)	0,010	(-0,002; 0,023)
η_{11}	-0,116	(-0,222; -0,014)	-0,116	(-0,219; -0,015)
η_{12}	0,088	(-0,024; 0,198)	0,092	(-0,015; 0,207)
η_{21}	*	*	0,039	(-0,066; 0,137)
η_{22}	*	*	-0,085	(-0,184; 0,012)
<i>DIC</i>		585,0		585,4
M2				
α	2,904	(1,388; 4,558)	4,240	(2,222; 6,213)
β	-0,005	(-0,009; -0,002)	-0,005	(-0,009; -0,002)
θ	0,526	(0,377; 0,689)	0,541	(0,399; 0,705)
γ_1	0,014	(0,002; 0,027)	0,014	(0,001; 0,026)
γ_2	0,009	(-0,003; 0,022)	0,011	(-0,001; 0,023)
λ_1	0,002	(-0,054; 0,053)	-0,043	(-0,111; 0,025)
η_{11}	-0,118	(-0,228; -0,012)	-0,126	(-0,232; -0,021)
η_{12}	0,081	(-0,072; 0,235)	0,174	(-0,002; 0,337)
η_{21}	*	*	0,034	(-0,061; 0,135)
η_{22}	*	*	0,011	(-0,001; 0,023)
<i>DIC</i>		587,0		585,9
M3				
α	2,833	(1,661; 3,972)	3,446	(2,116; 4,818)
β	-0,005	(-0,009; -0,002)	-0,006	(-0,009; -0,002)
θ	0,525	(0,388; 0,689)	0,543	(0,392; 0,716)
γ_1	0,014	(0,002; 0,027)	0,013	(0,001; 0,025)
γ_2	0,009	(-0,004; 0,021)	0,011	(-0,001; 0,023)
λ_2	0,008	(-0,059; 0,074)	-0,028	(-0,108; 0,051)
η_{11}	-0,125	(-0,257; 0,010)	-0,084	(-0,277; 0,055)
η_{12}	0,061	(-0,182; 0,302)	0,182	(-0,095; 0,468)
η_{21}	*	*	0,027	(-0,076; 0,131)
η_{22}	*	*	-0,108	(-0,230; 0,016)
<i>DIC</i>		587,3		587,0
M4				
α	3,016	(2,783; 3,243)	2,990	(2,753; 3,225)
β	-0,006	(-0,009; -0,002)	-0,005	(-0,009; -0,002)
θ	0,529	(0,384; 0,694)	0,530	(0,385; 0,700)
γ_1	0,014	(0,001; 0,027)	0,014	(0,002; 0,026)
γ_2	0,008	(-0,004; 0,021)	0,010	(-0,002; 0,023)
λ_3	-0,003	(-0,015; 0,009)	-0,002	(-0,014; 0,010)
η_{11}	-0,101	(-0,216; 0,017)	-0,107	(-0,226; 0,008)
η_{12}	0,116	(-0,042; 0,268)	0,112	(-0,061; 0,277)
η_{21}	*	*	0,041	(-0,064; 0,145)
η_{22}	*	*	-0,080	(-0,185; 0,027)
<i>DIC</i>		586,6		587,5
M5				
α	3,403	(1,292; 5,536)	*	*
β	-0,006	(-0,009; -0,002)	*	*
θ	0,512	(0,369; 0,669)	*	*
γ_1	0,014	(0,001; 0,027)	*	*
γ_2	0,008	(-0,003; 0,021)	*	*
λ_1	-0,023	(-0,124; 0,068)	*	*
λ_2	0,020	(-0,075; 0,118)	*	*
λ_3	-0,005	(-0,022; 0,010)	*	*
η_{11}	-0,121	(-0,283; 0,045)	*	*
η_{12}	0,125	(-0,179; 0,418)	*	*
η_{21}	*	*	*	*
η_{22}	*	*	*	*
<i>DIC</i>		590,6		*

(a) Sem covariável

(b) Temperatura Máxima (Tmax)

(c) Temperatura Mínima (Tmin)

(d) Precipitação

(e) Tmax, Tmin e Precipitação

Figura 26: Gráficos da ICPO para os ajustes dos modelos Poisson Dupla, com termos autorregressivos de primeira e segunda ordens.

(a) Sem covariável

(b) Temperatura Máxima (Tmax)

(c) Temperatura Mínima (Tmin)

(d) Precipitação

(e) Tmax, Tmin e Precipitação

Figura 27: Gráficos dos valores preditos e observados para os ajustes dos modelos Poisson Dupla, com termos autorregressivos de primeira e segunda ordens.

(a) Sem covariável

(b) Temperatura Máxima (Tmax)

(c) Temperatura Mínima (Tmin)

(d) Precipitação

(e) Tmax, Tmin e Precipitação

Figura 28: Gráficos de séries temporais dos dados de acidentes com escorpiões e dos ajustes dos modelos Poisson Dupla, com termos autorregressivos de primeira e segunda ordens.

no mesmo ajuste, para a situação em que termos autorregressivos de primeira e segunda ordens foram inseridos no mesmo modelo. Portanto, nesse caso, os resultados apresentados consideram apenas um par dessas funções ($J = 1$).

As Figuras 29, 32 e 35 apresentam os gráficos da ICPO para os ajustes dos modelos de regressão de Poisson, considerando-se termos autorregressivos e um ou dois pares de funções seno e cosseno. Nas Figuras 30, 33 e 36, observam-se os gráficos dos valores preditos e observados para as mesmas situações. As Figuras 31, 34 e 37 mostram os gráficos dos ajustes dos modelos.

A partir dos resultados, nota-se que, de forma geral, a série de dados de picadas de escorpiões, em Ribeirão Preto - SP, vem caindo ($\beta < 0$). Isso foi observado tanto na modelagem baseada na distribuição Poisson Dupla quanto na de Poisson. Além disso, a inclusão do termo autorregressivo de primeira ordem foi essencial para melhoria do ajuste dos modelos aos dados ($\gamma_1 > 0$). A observação dos gráficos dos ajustes dos modelos às séries originais corroboram isso.

Após a inserção do termo autorregressivo de primeira ordem, apesar de alguns correlogramas ainda sugerirem uma correlação entre observações sucessivas, observou-se que, mesmo considerando-se termos autorregressivos de primeira e segunda ordens nos mesmos modelos, não houve ganho nos ajustes e, na grande maioria, tem-se que $\gamma_2 = 0$. Também, não se pode estabelecer uma razão prática para essas autocorrelações remanescentes.

Com relação às covariáveis climáticas, apesar de alguns estudos afirmarem que a temperatura é um fator bastante determinante nas atividades dos escorpiões (HOSHINO et al., 2006; CLOUDSLEY-THOMPSON, 1975; CHOWELL et al., 2005), não se encontrou associação de nenhuma delas (temperaturas mínima e máxima), bem como da precipitação, com a incidência de picadas, no período avaliado.

Na grande parte dos modelos ajustados, pelo menos um dos termos η_{11} , η_{12} , η_{21} e η_{22} , referentes aos pares funções seno e cosseno, foi diferente de zero, mas não houve um ganho maior nos ajustes quando se inseriram dois pares ($J = 2$) dessas funções, quando comparados aos modelos com somente um par ($J = 1$).

Quando se confrontaram os resultados provenientes dos modelos baseados nas distribuições Poisson Dupla e de Poisson, a primeira mostrou ser a melhor opção, como se observa nos gráficos das ICPOs, nas curvas ajustadas e, também, nos valores do DIC e da LPML.

Uma interpretação prática dos resultados pode ser viesada, pois parece que os dados obtidos para Ribeirão Preto estão subnotificados, dado que a cidade possui, reconhecidamente, alta incidência de acidentes com escorpiões (CARRARO et al., 2015). Um fato que, possivelmente, pode confirmar a subnotificação é que a série de dados

Tabela 23: Resultados dos ajustes dos modelos de regressão de Poisson para os acidentes com escorpiões em Ribeirão Preto - SP, entre 2007 e 2014 - Modelos sem termo autorregressivo.

Parâmetro	Modelo 1 ($J = 1$)		Modelo 2 ($J = 2$)	
	Estimativa	ICr 95%	Estimativa	ICr 95%
M1				
α	3,154	(3,061; 3,247)	3,148	(3,055; 3,240)
β	-0,009	(-0,011; -0,007)	-0,009	(-0,011; -0,007)
η_{11}	-0,112	(-0,185; -0,039)	-0,109	(-0,178; -0,039)
η_{12}	0,157	(0,084; 0,233)	0,162	(0,087; 0,234)
η_{21}	*	*	0,008	(-0,062; 0,081)
η_{22}	*	*	-0,074	(-0,151; -0,002)
<i>DIC</i>		628,7		628,2
M2				
α	3,157	(2,060; 4,241)	4,480	(2,956; 5,997)
β	-0,009	(-0,011; -0,007)	-0,009	(-0,011; -0,007)
λ_1	0,001	(-0,037; 0,037)	-0,045	(-0,097; 0,006)
η_{11}	-0,110	(-0,185; -0,036)	-0,117	(-0,190; -0,046)
η_{12}	0,157	(0,052; 0,257)	0,248	(0,123; 0,372)
η_{21}	*	*	0,004	(-0,067; 0,079)
η_{22}	*	*	-0,138	(-0,228; -0,037)
<i>DIC</i>		630,7		627,3
M3				
α	2,958	(2,179; 3,753)	3,589	(2,615; 4,612)
β	-0,009	(-0,011; -0,007)	-0,009	(-0,011; -0,007)
λ_2	0,011	(-0,035; 0,057)	-0,026	(-0,086; 0,031)
η_{11}	-0,125	(-0,215; -0,033)	-0,076	(-0,175; 0,028)
η_{12}	0,119	(-0,042; 0,289)	0,248	(0,039; 0,462)
η_{21}	*	*	-0,002	(-0,077; 0,072)
η_{22}	*	*	-0,099	(-0,188; -0,008)
<i>DIC</i>		630,4		629,5
M4				
α	3,183	(3,055; 3,303)	3,161	(3,027; 3,292)
β	-0,009	(-0,011; -0,007)	-0,009	(-0,011; -0,007)
λ_3	-0,002	(-0,009; 0,005)	-0,001	(-0,008; 0,006)
η_{11}	-0,101	(-0,182; -0,018)	-0,104	(-0,184; -0,026)
η_{12}	0,182	(0,080; 0,280)	0,175	(0,071; 0,281)
η_{21}	*	*	0,011	(-0,060; 0,083)
η_{22}	*	*	-0,072	(-0,145; 0,002)
<i>DIC</i>		630,1		630,2
M5				
α	3,611	(2,270; 5,022)	5,397	(3,417; 7,218)
β	-0,009	(-0,011; -0,007)	-0,009	(-0,011; -0,007)
λ_1	-0,031	(-0,096; 0,029)	-0,087	(-0,161; -0,008)
λ_2	0,031	(-0,033; 0,100)	0,024	(-0,050; 0,096)
λ_3	-0,005	(-0,014; 0,004)	-0,008	(-0,019; 0,001)
η_{11}	-0,131	(-0,242; -0,019)	-0,112	(-0,224; 0,002)
η_{12}	0,161	(-0,029; 0,360)	0,335	(0,109; 0,556)
η_{21}	*	*	0,025	(-0,056; 0,106)
η_{22}	*	*	-0,152	(-0,253; -0,056)
<i>DIC</i>		633,1		627,8

(a) Sem covariável

(b) Temperatura Máxima (Tmax)

(c) Temperatura Mínima (Tmin)

(d) Precipitação

(e) Tmax, Tmin e Precipitação

Figura 29: Gráficos da ICPO para os ajustes dos modelos de Poisson, sem termo autorregressivo.

(a) Sem covariável

(b) Temperatura Máxima (Tmax)

(c) Temperatura Mínima (Tmin)

(d) Precipitação

(e) Tmax, Tmin e Precipitação

Figura 30: Gráficos dos valores preditos e observados para os ajustes dos modelos de Poisson, sem termo autorregressivo.

(a) Sem covariável

(b) Temperatura Máxima (Tmax)

(c) Temperatura Mínima (Tmin)

(d) Precipitação

(e) Tmax, Tmin e Precipitação

Figura 31: Gráficos de séries temporais dos dados de acidentes com escorpiões e dos ajustes dos modelos de Poisson, sem termo autorregressivo.

Tabela 24: Resultados dos ajustes dos modelos de regressão de Poisson para os acidentes com escorpiões em Ribeirão Preto - SP, entre 2007 e 2014 - Modelos com termo autorregressivo de primeira ordem.

Parâmetro	Modelo 1 ($J = 1$)		Modelo 2 ($J = 2$)	
	Estimativa	ICr 95%	Estimativa	ICr 95%
M1				
α	3,020	(2,903; 3,139)	3,102	(2,894; 3,134)
β	-0,006	(-0,009; -0,004)	-0,006	(-0,009; -0,004)
γ_1	0,016	(0,008; 0,024)	0,016	(0,008; 0,025)
η_{11}	-0,111	(-0,185; -0,039)	-0,109	(-0,180; -0,037)
η_{12}	0,102	(0,024; 0,179)	0,107	(0,025; 0,188)
η_{21}	*	*	0,022	(-0,049; 0,094)
η_{22}	*	*	-0,069	(-0,141; 0,004)
<i>DIC</i>		610,1		610,2
M2				
α	2,992	(1,875; 4,086)	4,267	(2,800; 5,765)
β	-0,006	(-0,009; -0,004)	-0,006	(-0,009; -0,004)
γ_1	0,016	(0,008; 0,025)	0,016	(0,008; 0,025)
λ_1	0,001	(-0,037; 0,037)	-0,042	(-0,094; 0,008)
η_{11}	-0,111	(-0,181; -0,039)	-0,119	(-0,192; -0,047)
η_{12}	0,102	(-0,003; 0,204)	0,190	(0,065; 0,317)
η_{21}	*	*	0,016	(-0,057; 0,092)
η_{22}	*	*	-0,122	(-0,218; -0,025)
<i>DIC</i>		612,0		609,6
M3				
α	2,800	(2,022; 3,581)	3,307	(2,271; 4,296)
β	-0,006	(-0,009; -0,004)	-0,006	(-0,009; -0,004)
γ_1	0,016	(0,008; 0,025)	0,016	(0,007; 0,024)
λ_2	0,012	(-0,033; 0,058)	-0,017	(-0,075; 0,042)
η_{11}	-0,124	(-0,210; -0,035)	-0,088	(-0,187; 0,012)
η_{12}	0,060	(-0,106; 0,226)	0,164	(-0,048; 0,374)
η_{21}	*	*	0,015	(-0,055; 0,091)
η_{22}	*	*	-0,086	(-0,175; 0,005)
<i>DIC</i>		611,7		611,7
M4				
α	3,040	(2,896; 3,182)	3,021	(2,873; 3,179)
β	-0,006	(-0,009; -0,004)	-0,006	(-0,008; -0,004)
γ_1	0,016	(0,008; 0,025)	0,016	(0,008; 0,025)
λ_3	-0,002	(-0,010; 0,006)	-0,001	(-0,009; 0,007)
η_{11}	-0,103	(-0,188; -0,022)	-0,106	(-0,188; -0,026)
η_{12}	0,122	(0,010; 0,232)	0,117	(0,008; 0,229)
η_{21}	*	*	0,025	(-0,045; 0,098)
η_{22}	*	*	-0,069	(-0,143; 0,008)
<i>DIC</i>		611,7		612,1
M5				
α	3,470	(2,038; 4,987)	5,264	(3,354; 7,184)
β	-0,006	(-0,009; -0,004)	-0,006	(-0,009; -0,004)
γ_1	0,016	(0,008; 0,024)	0,015	(0,006; 0,023)
λ_1	-0,033	(-0,100; 0,035)	-0,091	(-0,175; -0,009)
λ_2	0,034	(-0,031; 0,103)	0,032	(-0,045; 0,117)
λ_3	-0,005	(-0,016; 0,005)	-0,010	(-0,023; 0,001)
η_{11}	-0,133	(-0,248; -0,020)	-0,120	(-0,248; -0,003)
η_{12}	0,103	(-0,099; 0,307)	0,282	(0,047; 0,525)
η_{21}	*	*	0,041	(-0,040; 0,125)
η_{22}	*	*	-0,135	(-0,233; -0,038)
<i>DIC</i>		614,5		610,4

(a) Sem covariável

(b) Temperatura Máxima (Tmax)

(c) Temperatura Mínima (Tmin)

(d) Precipitação

(e) Tmax, Tmin e Precipitação

Figura 32: Gráficos da ICPO para os ajustes dos modelos de Poisson, com termo autorregressivo de primeira ordem.

(a) Sem covariável

(b) Temperatura Máxima (Tmax)

(c) Temperatura Mínima (Tmin)

(d) Precipitação

(e) Tmax, Tmin e Precipitação

Figura 33: Gráficos dos valores preditos e observados para os ajustes dos modelos de Poisson, com termo autorregressivo de primeira ordem.

(a) Sem covariável

(b) Temperatura Máxima (Tmax)

(c) Temperatura Mínima (Tmin)

(d) Precipitação

(e) Tmax, Tmin e Precipitação

Figura 34: Gráficos de séries temporais dos dados de acidentes com escorpiões e dos ajustes dos modelos de Poisson, com termo autorregressivo de primeira ordem.

Tabela 25: Resultados dos ajustes dos modelos de regressão de Poisson para os acidentes com escorpiões em Ribeirão Preto - SP, entre 2007 e 2014 - Modelos com termos autorregressivos de primeira e segunda ordens.

Parâmetro	Modelo 1 ($J = 1$)		Modelo 2 ($J = 2$)	
	Estimativa	ICr 95%	Estimativa	ICr 95%
M1				
α	2,976	(2,845; 3,106)	2,967	(2,833; 3,095)
β	-0,006	(-0,008; -0,003)	-0,006	(-0,008; -0,003)
γ_1	0,015	(0,006; 0,023)	0,014	(0,005; 0,023)
γ_2	0,009	(0,0005; 0,018)	0,010	(0,001; 0,019)
η_{11}	-0,118	(-0,194; -0,046)	-0,118	(-0,191; -0,047)
η_{12}	0,084	(0,007; 0,166)	0,091	(0,010; 0,173)
η_{21}	*	*	0,039	(-0,035; 0,116)
η_{22}	*	*	-0,081	(-0,152; -0,011)
<i>DIC</i>		601,0		599,0
M2				
α	2,866	(1,781; 3,964)	4,230	(2,739; 5,761)
β	-0,006	(-0,008; -0,003)	-0,006	(-0,008; -0,003)
γ_1	0,014	(0,005; 0,023)	0,014	(0,005; 0,023)
γ_2	0,009	(0,0005; 0,018)	0,011	(0,002; 0,020)
λ_1	0,004	(-0,033; 0,040)	-0,043	(-0,095; 0,008)
η_{11}	-0,116	(-0,191; -0,042)	-0,125	(-0,200; -0,048)
η_{12}	0,080	(-0,025; 0,188)	0,171	(0,040; 0,299)
η_{21}	*	*	0,030	(-0,044; 0,102)
η_{22}	*	*	-0,138	(-0,237; -0,038)
<i>DIC</i>		602,8		598,8
M3				
α	2,846	(2,034; 3,649)	3,427	(2,383; 4,494)
β	-0,006	(-0,008; -0,003)	-0,006	(-0,008; -0,003)
γ_1	0,015	(0,006; 0,023)	0,013	(0,005; 0,022)
γ_2	0,009	(0,0002; 0,018)	0,011	(0,002; 0,020)
λ_2	0,008	(-0,039; 0,054)	-0,027	(-0,088; 0,035)
η_{11}	-0,125	(-0,216; -0,034)	-0,086	(-0,193; 0,020)
η_{12}	0,061	(-0,114; 0,225)	0,178	(-0,031; 0,403)
η_{21}	*	*	0,028	(-0,047; 0,105)
η_{22}	*	*	-0,108	(-0,202; -0,019)
<i>DIC</i>		602,8		600,4
M4				
α	3,016	(2,849; 3,182)	2,991	(2,827; 3,148)
β	-0,006	(-0,008; -0,003)	-0,006	(-0,008; -0,003)
γ_1	0,014	(0,005; 0,023)	0,015	(0,005; 0,022)
γ_2	0,009	(-0,00002; 0,018)	0,010	(0,001; 0,018)
λ_3	-0,003	(-0,011; 0,006)	-0,002	(-0,011; 0,006)
η_{11}	-0,102	(-0,189; -0,016)	-0,107	(-0,191; -0,023)
η_{12}	0,114	(-0,006; 0,233)	0,111	(-0,008; 0,227)
η_{21}	*	*	0,042	(-0,031; 0,115)
η_{22}	*	*	-0,080	(-0,155; -0,003)
<i>DIC</i>		602,5		600,6
M5				
α	3,272	(1,736; 4,877)	*	*
β	-0,006	(-0,008; -0,003)	*	*
γ_1	0,014	(0,005; 0,023)	*	*
γ_2	0,008	(-0,0003; 0,017)	*	*
λ_1	-0,020	(-0,094; 0,049)	*	*
λ_2	0,022	(-0,045; 0,092)	*	*
λ_3	-0,004	(-0,016; 0,006)	*	*
η_{11}	-0,125	(-0,242; -0,015)	*	*
η_{12}	0,099	(-0,099; 0,302)	*	*
η_{21}	*	*	*	*
η_{22}	*	*	*	*
<i>DIC</i>		606,0		*

(a) Sem covariável

(b) Temperatura Máxima (Tmax)

(c) Temperatura Mínima (Tmin)

(d) Precipitação

(e) Tmax, Tmin e Precipitação

Figura 35: Gráficos da ICPO para os ajustes dos modelos de Poisson, com termos autorregressivos de primeira e segunda ordens.

(a) Sem covariável

(b) Temperatura Máxima (Tmax)

(c) Temperatura Mínima (Tmin)

(d) Precipitação

(e) Tmax, Tmin e Precipitação

Figura 36: Gráficos dos valores preditos e observados para os ajustes dos modelos de Poisson, com termos autorregressivos de primeira e segunda ordens.

(a) Sem covariável

(b) Temperatura Máxima (Tmax)

(c) Temperatura Mínima (Tmin)

(d) Precipitação

(e) Tmax, Tmin e Precipitação

Figura 37: Gráficos de séries temporais dos dados de acidentes com escorpiões e dos ajustes dos modelos de Poisson, com termos autorregressivos de primeira e segunda ordens.

segue uma tendência contrária (queda) àquela que observada no Estado de São Paulo e no Brasil (ver Figura 18). Ao mesmo tempo, por se tratar de região de alto índice de acidentes, a população pode ter maior conhecimento sobre como evitar o encontro com os escorpiões.

6 CONCLUSÃO

Neste trabalho, propuseram-se modelos de séries temporais para dados de contagem, sob enfoque Bayesiano, utilizando-se distribuições de probabilidade para variáveis discretas, tais como de Poisson e Poisson Dupla, apresentando-se, para esta última, maiores desenvolvimentos em relação aos estimadores de máxima verossimilhança e seus respectivos erros padrão. Introduziu-se, ainda, um modelo para dados de contagem com excesso de zeros, com base na Poisson Dupla.

A partir de exemplos com dados da literatura, foram ajustados modelos sob as abordagens frequentista e Bayesiana e, nesses casos, não houve um ganho significativo no ajuste, quando comparados os resultados de ambas. Ao mesmo tempo, utilizando-se a abordagem Bayesiana, é possível se incorporarem informações a priori sobre os parâmetros (principalmente a média). Ainda, por meio das distribuições a posteriori, simuladas para os parâmetros, obtiveram-se intervalos de credibilidade livres de definições assintóticas.

Na modelagem dos dados em séries, a abordagem Bayesiana apresentou-se como importante ferramenta, devido à necessidade de se inserirem parâmetros referentes à sazonalidade, bem como às covariáveis climáticas, o que seria um obstáculo na abordagem frequentista dos modelos propostos.

Como aplicações principais, apresentaram-se resultados obtidos do ajuste de modelos para dados de registros de acidentes com picadas de cobras, no Estado de São Paulo, e picadas de escorpiões, na cidade de Ribeirão Preto, SP, entre os anos de 2007 e 2014. Quanto aos acidentes com escorpiões, consideraram-se dados climáticos (temperaturas máximas e mínimas médias mensais e precipitação média mensal).

Os modelos baseados na distribuição Poisson Dupla, apesar de serem pouco encontrados na literatura, surgem como eficaz alternativa na modelagem de dados de contagem cuja variância é maior ou menor que a média. Apesar de essa distribuição possuir dois parâmetros, a interpretação prática dos mesmos é tão imediata quanto a da distribuição de Poisson, em que o único parâmetro é igual à média. Quando utilizados nas séries temporais avaliadas neste trabalho, os modelos baseados na distribuição Poisson Dupla mostraram melhor desempenho no ajuste dos dados, quando comparados aos modelos de Poisson.

Outras distribuições poderiam ser utilizadas nos mesmos casos indicados para o uso da Poisson Dupla, como, por exemplo, distribuição Conway-Maxwell-Poisson

(CONWAY e MAXWELL, 1962) ou a Binomial Negativa, mas a Dupla Poisson pode ser mais vantajosa, na prática, devido à simplicidade em se obterem as estimativas de máxima verossimilhança e Bayesianas, para os parâmetros de interesse.

REFERÊNCIAS BIBLIOGRÁFICAS

ARAGON, D. C., de QUEIROZ, J. A. M., MARTINEZ, E. Z. Incidence of snakebites from 2007 to 2014 in the State of São Paulo, Southeast Brazil, using a Bayesian time series model. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 49, n. 4, p. 515-519, 2016.

ARBEX, M.A., PEREIRA, L.A.A., CARVALHO-OLIVERA, R., SALDIVA, P. H. N., BRAGA, A.L.F. The effect of air pollution on pneumonia-related emergency department visits in a region of extensive sugar cane plantations: a 30-month time-series study. **Journal of Epidemiology and Community Health**, v. 68, p. 669-674, 2014.

BOCHNER, R., FISZON, J.T., MACHADO, C. A profile of snake bites in Brazil, 2001 to 2012. **Journal of Clinical Toxicology**, v. 4, n. 3, 2014.

BOX, G.E.P. , JENKINS, G.M., REINSEL, G.C. **Time series analysis**, 4^a ed., New York: Wiley, 2008, 784p.

BRAS, A. L., GOMES, D., FILIPE, P. A., de SOUSA, B., NUNES, C. Trends, seasonality and forecasts of pulmonary tuberculosis in Portugal. **The International Journal of Tuberculosis and Lung Disease**, v. 18, n. 10, p. 1202-1210(9), 2014.

BRASIL. Secretaria de Vigilância à Saúde. Sistema de Informação de Agravos de Notificação - SINAN. Normas e Rotinas. Brasília: Ministério da Saúde, 2002. Disponível em <<http://sinan.saude.gov.br/sinan>>.

BRUNKARD, J.M., CIFUENTES, E., ROTHENBERG, S.J. Assessing the roles of temperature, precipitation, and ENSO in dengue re-emergence on the Texas-Mexico border region. **Salud Pública de México**, v. 50, n. 3, p. 227-234, 2008.

BUCARETCHI, F., DE CAPITANI, E.M., VIEIRA, R.J., RODRIGUES, C.K., ZANNIN, M., SILVA JR, N. J, CASAIS-E-SILVA, L.L., HYSLOP, S. Coral snake bites (*Micrurus* spp.) in Brazil: a review of literature reports. **Clinical Toxicology**, v. 54, n. 3, p. 222-234, 2016.

CAMERON, A. C., JOHANSSON, P. Count data regression using series expansions: with applications. **Journal of Applied Econometrics**, v. 12, p. 203-223, 1997.

CAMERON, A.C., TRIVEDI, P. K. **Regression analysis of count data**, 2^a ed, Cambridge University Press, 2013, 587p.

CARLIN, B.P., LOUIS, T.A. **Bayes and empirical Bayes methods for data**

analysis, 2^a ed., New York: Chapman and Hall, 2000, 440p.

CARRARO, P. E., BARBOSA, F.O., BEGALI, J.H., RUBIO. M.S., BÜRGER, K.P. Distribuição espacial de acidentes por animais peçonhentos no estado de São Paulo entre os anos de 2007 e 2012. **Ars Veterinária**, v. 31, n. 2, p. 46, 2015.

CASELLA, G., GEORGE, E.I. Explaining the Gibbs sampler. **The American Statistician**, v. 46, n. 3, p. 167-174, 1992.

CHADSUTHI, S., MODCHANG, C., LENBURY, Y., IAMSIRITHAWORN, S., TRIAMPO, W. Modeling seasonal leptospirosis transmission and its association with rainfall and temperature in Thailand using time-series and ARIMAX analyses. **Asian Pacific Journal of Tropical Medicine**, p. 539-546, 2012.

CHEN, M.H., SHAO, Q.M, IBRAHIM, J.G. **Monte Carlo Methods in Bayesian Computation**, 1^a ed., New York: Springer, 2000, 387p.

CHIB, S., GREENBERG, E. Understanding the Metropolis-Hastings Algorithm. **The American Statistician**, v. 49, n. 4, p. 327-335, 1995. 2012.

CHOWELL, G., HYMAN, J.M., DÍAZ-DUEÑAS, P., HENGARTNER, N. W. Predicting scorpion sting incidence in a endemic region using climatological variables. **International Journal of Environmental Health Research**, v. 15, p. 425-435, 2005.

CLOUDSLEY-THOMPSON, J.L. Adaptations of arthropoda to arid environments. **Annual Review of Entomology**, v. 20, p. 261-283, 1975.

CONDE, D. M., COSTA-PAIVA, L., MARTINEZ, E. Z., PINTO-NETO, A. M. Low bone mineral density in middle-aged breast cancer survivors: prevalence and associated factors. **Breast Care**, v. 7, n. 2, p. 121-125, 2012.

CONWAY, R. W., MAXWELL, W. L. A queuing model with state dependent service rates. **Journal of Industrial Engineering**, v. 12, p. 132-136, 1962

CUPO, P. Bites and stings from venomous animals: a neglected Brazilian tropical disease. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 48, n. 6, p. 639-641, 2015.

CUPO, P. Clinical update on scorpion envenoming. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 48, n. 6, p. 642-649, 2015.

EARNEST, A., CHEN, M.I., NG, D., SIN, L.Y. Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore. **BMC Health Services Research**, v. 5, n. 36, 2005.

EFRON, B. Double exponential families and their use in generalized linear regression.

- Journal of the American Statistical Association**, v. 81, n. 395, p. 709-721, 1986.
- EILERS, P. H. C., GAMPE, J., MARX, B. D., RAU, R. Modulation models for seasonal time series and incidence tables. **Statistics in Medicine**, v. 27, p. 3430-3441, 2008.
- FEITOSA, E. S., SAMPAIO, V., SACHETT, J., CASTRO, D. B. D., NORONHA, M. D. D. N., LOZANO, J. L. L., MUNIZ, E., FERREIRA, L. C. L., LACERDA, M. V. G., MONTEIRO, W. M. Snakebites as a largely neglected problem in the Brazilian Amazon: highlights of the epidemiological trends in the State of Amazonas. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 48 (Suppl 1), p. 34-41, 2015.
- GAO, L., ZHANG, Y., DING, G., LIU, Q., ZHOU, M., LI, X., JIANG, B. **The American Journal of Tropical Medicine and Hygiene**, v. 90, n. 4, p. 697-704, 2014.
- GEISSER, S., EDDY, W. A predictive approach to model selection. **Journal of the American Statistical Association**, v. 74, p. 153-160, 1979.
- GELFAND, A.E., DEY, D.K., CHANG, H. Model determination using predictive distributions with implementation via sampling-based methods (with discussion). **Bayesian Statistics**, v. 4, p. 147-167, 1992.
- GELMAN, A., RUBIN, B. D. Inference from iterative simulation using multiple sequences. **Statistical Sciences**, v. 7, n. 4, p. 457-511, 1992.
- GHARBI M., QUENEL P. , GUSTAVE J., CASSADOU S., LA RUCHE G., GIRDARY L., MARRAMA L. Time series analysis of dengue incidence in Guadeloupe, French West Indies: forecasting models using climate variables as predictors. **BMC Infectious Diseases**, v. 11, 2011.
- HII, Y.L., ZHU, H., NG. N. , NG, L.C., ROCKLÖV, J. Forecast of dengue incidence using temperature and rainfall. **PLoS Neglected Tropical Diseases**, v. 6, n. 11, e1908, 2012.
- HOSHINO, K., MOURA, A.T.V., DE PAULA, H.M.G. Selection of enviromental temperature by the yellow scorpion *Tityus serrulatus* Lutz and Mello, 1992 (Scorpiones, Buthidae). **Journal of Venomous Animals and Toxins including Tropical Diseases**, v. 12, p. 59-66, 2006.
- HOSMER, D.W.JR., LEMESHOW, S., STURDIVANT, R. X. **Applied logistic regression**, 3^a ed., Chichester: Wiley, 2013, 528p.
- HUI WEN, F., MONTEIRO, W.M., MOURA DA SILVA, A.M., TAMBOURGI, D.V., MENDONÇA DA SILVA, I., SAMPAIO, V. S., DOS SANTOS, M.C., SACHETT, J., FERREIRA, L.C., KALIL, J., LACERDA, M. Snakebites and scorpion stings in the Brazilian Amazon: identifying research priorities for a largely neglected problem.

PLOS Neglected Tropical Diseases, v. 9, n. 5, e0003701, 2015.

JEFFREYS, H. An Invariant Form for the Prior Probability in Estimation Problems. **Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences**, v. 186, n. 1007, p. 453-461, 1946.

JIANG, H., LIVINGSTON, M., ROOM, R., DIETZE, P. , NORSTRÖM, T., KERR, W.C. Alcohol consumption and liver disease in Australia: A time series analysis of the period 1935-2006. **Alcohol and Alcoholism**, v. 49, n. 3, p. 363-368, 2014.

LAMBERT, D. Zero-inflated Poisson regression, with an application to defects in manufacturing. **Technometrics**, v. 34, n.1, p. 1-14, 1992.

LEE, H. S., HER, M., LEVINE, M., MOORE, G.E. Time series analysis of human and bovine brucellosis in South Korea from 2005 to 2010. **Preventive Veterinary Medicine**, v. 110, p. 190-197, 2013.

LU, L., LIN, H., TIAN, L., YANG, W., SUN, J., LIU, Q. Time series analysis of dengue fever and weather in Guangzhou, China. **BMC Public Health**, v. 9, n. 395, 2009.

LIU, Q., LIU, X., JIANG, B., YANG, W. Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model. **BMC Infectious Diseases**, v. 11, 2011.

LUNN, D., SPIEGELHALTER, D., THOMAS, A., BEST, N. The BUGS project: Evolution, critique, and future directions, **Statistics in Medicine**, v. 28, p. 3049-3067, 2009.

LUZ, P. M., MENDES, B.V., CODEÇO, C.T., STRUCHINER, C.J., GALVANI, A.P. Time series analysis of dengue incidence in Rio de Janeiro, Brazil. **American Journal of Tropical Medicine and Hygiene**, v. 79, n. 6, p. 933-939, 2008.

MACHADO, C., BOCHNER, R., FISZON, J.T. Epidemiological profile of snakebites in Rio de Janeiro, Brazil, 2001-2006. **Journal of Venomous Animals and Toxins including Tropical Diseases**, v. 18, n. 2, p. 217-224, 2012.

MARQUES, O.A.V., ETEROVIC, A., ENDO, W. Seasonal activity of snakes in the Atlantic Forest in southeastern Brazil. **Amphibia-Reptilia**, v. 22, p. 103-111, 2000.

MARTINEZ, E.Z., SILVA, E.A. Predicting the number of cases of dengue infection in Ribeirão Preto, São Paulo State, Brazil, using a SARIMA model. **Cadernos de Saúde Pública**, v. 27, n. 9, p. 1809-1818, 2011.

MARTINEZ, E.Z., SILVA, E.A., FABBRO, A.L. A SARIMA forecasting model to predict the number of cases of dengue in Campinas, State of São Paulo, Brazil. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 44, n. 4, p. 436-440, 2011.

MARTINS, V. E., ALENCAR, C.H., FACÓ, P. E., DUTRA, R.F., ALVES, C.R.,

- PONTES, R.J., GUEDES, M.I. Spatial distribution and breeding site characteristics of *Aedes albopictus* and *Aedes aegypti* in Fortaleza, State of Ceará. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 43, n. 1, p. 73-77, 2010.
- MARTINS, V. E., MARTINS, M.G., de ARAÚJO, J.M., SILVA, L.O., MONTEIRO, H.A., CASTRO, F.C., VASCONCELOS, p. F., GUEDES, M.I. First report of *Aedes (Stegomyia) albopictus* in the state of Ceará, Brazil. **Revista de Saúde Pública**, v. 40, n. 4, p. 737-739, 2006.
- MILLAR, R. B. **Maximum likelihood estimation and inference: with examples in R, SAS and ADMB**, 1^a ed., Chichester: Wiley, 2011, 357p.
- MODARRES, R., OUARDA, T.B.M.J., VANASSE, A., ORZANCO, M.G., GOSSELIN, p. Modeling seasonal variation of hip fracture in Montreal, Canada. **Bone**, v. 50, p. 909-916, 2012.
- NTZOUFRAS, I. **Bayesian modeling using WinBUGS**, 1^a ed., Hoboken: Wiley, 2009, 520p.
- RAMIREZ, A.P., BUITRAGO, J.I.G., GONZÁLEZ, J.P.P., MORÁLEZ, A.H., CARRASQUILLA, G. Frequency and tendency of malaria in Colombia, 1990 to 2011: a descriptive study. **Malaria Journal**, v. 13, n. 202, 2014.
- RAZVODOVSKY, Y. E. The Effect of Beverage Type on Alcoholic Psychoses Rate in Russia, **Alcohol and Alcoholism**, v. 50, n. 2, p. 200-205, 2015.
- RECKZIEGEL, G. C., PINTO JR., V. L. Scorpionism in Brazil in the years 2000 to 2012, **Journal of Venomous Animals and Toxins including Tropical Diseases**, v. 20, n. 46, 2014.
- RIBEIRO, L.A., ALBUQUERQUE, M.J., PIRES DE CAMPOS, V.A.F., KATZ, G., TAKAOKA, N. Y., LEBRÃO, M. L., JORGE, M.T. Obits by venomous snakes in the State of São Paulo: evaluation of 43 cases from 1988 to 1993. **Revista da Associação Médica Brasileira**, v. 44, n. 4, p. 312-318, 1998.
- RIBEIRO, L.A., JORGE, M.T. Acidente por serpentes do gênero *Bothrops*: série de 3.139 casos. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 30, n. 6, p. 475-80, 1997.
- ROCCO, I.M., SILVEIRA, V. R., MAEDA, A.Y., SILVA, S.J., SPENASSATTO, C., BISORDI, I., SUZUKI, A. First isolation of Dengue 4 in the state of São Paulo, Brazil, 2011. **Revista do Instituto de Medicina Tropical de São Paulo**, v. 54, n. 1, p. 49-51, 2012.
- SHUMWAY, R. H., STOFFER, D.S. **Time series analysis and its applications with R examples**. 2^a ed., New York: Springer, 2006, 576p.

STOLWIJK, A. M., STRAATMAN, H., ZIELHUIS, G. A. Studying seasonality by using sine and cosine functions in regression analysis. **Journal of Epidemiology and Community Health**, v. 53, n. 4, p. 235-238, 1999.

PINTO, E., COELHO, M., OLIVER, L., MASSAD, E. The influence of climate variables on dengue in Singapore. **International Journal of Environmental Health Research**, v. 21, n. 6, p. 415-426, 2011.

SILAWAN, T., SINGHASIVANON, P. , KAEWKUNGWAL, J., NIMMANITYA, S., SUWONKERD, W. Temporal patterns and forecast of dengue infection in Northeastern Thailand. **The Southeast Asian Journal of Tropical Medicine and Public Health**, v. 39, n. 1, p. 90-98, 2008.

ZHANG, X., PANG, Y., CUI, M., STALLONES, L., XIANG, H. Forecasting mortality of road traffic injuries in China using seasonal autoregressive integrated moving average model. **Annals of Epidemiology**, v. 25, p. 101-106, 2015.

Apêndices

Apêndice 1

Figura 38: Gráfico das autocorrelações e autocorrelações parciais - Cobra do gênero *Bothrops* - Modelo Poisson Dupla sem termo autorregressivo.

Figura 39: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Cobra do gênero *Crotalus* - Modelo Poisson Dupla sem termo autorregressivo.

Figura 40: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Cobra não peçonhenta - Modelo Poisson Dupla sem termo autorregressivo.

Figura 41: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Cobra não identificada - Modelo Poisson Dupla sem termo autorregressivo.

Figura 42: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Cobra do gênero *Bothrops* - Modelo Poisson Dupla com termo autorregressivo de primeira ordem.

Figura 43: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Cobra do gênero *Crotalus* - Modelo Poisson Dupla com termo autorregressivo de primeira ordem.

Figura 44: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Cobra não peçonhenta - Modelo Poisson Dupla com termo autorregressivo de primeira ordem.

Figura 45: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Cobra não identificada - Modelo Poisson Dupla com termo autorregressivo de primeira ordem.

Figura 46: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Cobra do gênero *Bothrops* - Modelo Poisson Dupla com termos autorregressivos de primeira e segunda ordens.

Figura 47: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Cobra do gênero *Crotalus* - Modelo Poisson Dupla com termos autorregressivos de primeira e segunda ordens.

Figura 48: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Cobra não peçonhenta - Modelo Poisson Dupla com termos autorregressivos de primeira e segunda ordens.

Figura 49: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Cobra não identificada - Modelo Poisson Dupla com termos autorregressivos de primeira e segunda ordens.

Figura 50: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Cobra do gênero *Bothrops* - Modelo Poisson sem termo autorregressivo.

Figura 51: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Cobra do gênero *Crotalus* - Modelo Poisson sem termo autorregressivo.

Figura 52: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Cobra não peçonhenta - Modelo Poisson sem termo autorregressivo.

Figura 53: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Cobra não identificada - Modelo Poisson sem termo autorregressivo.

Figura 54: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Cobra do gênero *Bothrops* - Modelo Poisson com termo autorregressivo de primeira ordem.

Figura 55: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Cobra do gênero *Crotalus* - Modelo Poisson com termo autorregressivo de primeira ordem.

Figura 56: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Cobra não peçonhenta - Modelo Poisson com termo autorregressivo de primeira ordem.

Figura 57: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Cobra não identificada - Modelo Poisson com termo autorregressivo de primeira ordem.

Figura 58: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Cobra do gênero *Bothrops* - Modelo Poisson com termos autorregressivos de primeira e segunda ordens.

Figura 59: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Cobra do gênero *Crotalus* - Modelo Poisson com termos autorregressivos de primeira e segunda ordens.

Figura 60: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Cobra não peçonhenta - Modelo Poisson com termos autorregressivos de primeira e segunda ordens.

Figura 61: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Cobra não identificada - Modelo Poisson com termos autorregressivos de primeira e segunda ordens.

Apêndice 2

Figura 62: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos Poisson Dupla sem termo autorregressivo.

Figura 63: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos Poisson Dupla sem termo autorregressivo - Temperatura Máxima.

Figura 64: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos Poisson Dupla sem termo autorregressivo - Temperatura Mínima.

Figura 65: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos Poisson Dupla sem termo autorregressivo - Precipitação.

Figura 66: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos Poisson Dupla sem termo autorregressivo - Tmax, Tmin e Precipitação.

Figura 67: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos Poisson Dupla com termo autorregressivo de primeira ordem.

Figura 68: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos Poisson Dupla com termo autorregressivo de primeira ordem - Temperatura Máxima.

Figura 69: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos Poisson Dupla com termo autorregressivo de primeira ordem - Temperatura Mínima.

Figura 70: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos Poisson Dupla com termo autorregressivo de primeira ordem - Precipitação.

Figura 71: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos Poisson Dupla com termo autorregressivo de primeira ordem - Tmax, Tmin e Precipitação.

Figura 72: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos Poisson Dupla com termos autorregressivos de primeira e segunda ordens.

Figura 73: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos Poisson Dupla com termos autorregressivos de primeira e segunda ordens - Temperatura Máxima.

Figura 74: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos Poisson Dupla com termos autorregressivos de primeira e segunda ordens - Temperatura Mínima.

Figura 75: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos Poisson Dupla com termos autorregressivos de primeira e segunda ordens - Precipitação.

Figura 76: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos Poisson Dupla com termos autorregressivos de primeira e segunda ordens - Tmax, Tmin e Precipitação.

Figura 77: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos de Poisson sem termo autorregressivo.

Figura 78: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos de Poisson sem termo autorregressivo - Temperatura Máxima.

Figura 79: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos de Poisson sem termo autorregressivo - Temperatura Mínima.

Figura 80: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos de Poisson sem termo autorregressivo - Precipitação.

Figura 81: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos de Poisson sem termo autorregressivo - Tmax, Tmin e Precipitação.

Figura 82: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos de Poisson com termo autorregressivo de primeira ordem.

Figura 83: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos de Poisson com termo autorregressivo de primeira ordem - Temperatura Máxima.

Figura 84: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos de Poisson com termo autorregressivo de primeira ordem - Temperatura Mínima.

Figura 85: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos de Poisson com termo autorregressivo de primeira ordem - Precipitação.

Figura 86: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos de Poisson com termo autorregressivo de primeira ordem - Tmax, Tmin e Precipitação.

Figura 87: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos de Poisson com termos autorregressivos de primeira e segunda ordens.

Figura 88: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos de Poisson com termos autorregressivos de primeira e segunda ordens - Temperatura Máxima.

Figura 89: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos de Poisson com termos autorregressivos de primeira e segunda ordens - Temperatura Mínima.

Figura 90: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos de Poisson com termos autorregressivos de primeira e segunda ordens - Precipitação.

Figura 91: Gráfico das autocorrelações e autocorrelações parciais dos resíduos do modelo - Picadas de escorpiões em Ribeirão Preto - SP - Modelos de Poisson com termos autorregressivos de primeira e segunda ordens - Tmax, Tmin e Precipitação.