



JOSÉ ANDRÉ MOTA DE QUEIROZ

**Modelos Bayesianos com ponto de mudança e  
termos autorregressivos aplicados em estudos  
de séries temporais de dados epidemiológicos**

Dissertação de Mestrado

Maringá - Paraná  
2016

JOSÉ ANDRÉ MOTA DE QUEIROZ

**Modelos Bayesianos com ponto de mudança e termos  
autorregressivos aplicados em estudos de séries  
temporais de dados epidemiológicos**

Dissertação apresentada ao Programa de  
Pós-Graduação em Bioestatística do Centro  
de Ciências Exatas da Universidade Estadual  
de Maringá como requisito parcial para obten-  
ção do título de Mestre em Bioestatística.  
Orientador: Prof Dr Edson Z. Martinez  
Coorientadora: Profa Dra Isolde T. S. Previdelli

Maringá - Paraná

2016

Dados Internacionais de Catalogação na Publicação (CIP)  
(Biblioteca Central - UEM, Maringá, PR, Brasil)

Q3m      Queiroz, José André Mota de  
Modelos Bayesianos com ponto de mudança e termos autorregressivos aplicados em estudos de séries temporais de dados epidemiológicos / José André Mota de Queiroz. -- Maringá, 2016.  
76 f. : figs., tabs.

Orientador: Prof. Dr. Edson Zangiacomi Martinez.  
Dissertação (mestrado) - Universidade Estadual de Maringá, Centro Ciências Exatas, Programa de Pós-Graduação em Bioestatística, 2016.

1. Série temporal. 2. Modelos de ponto de mudança. 3. Métodos bayesianos. 4. Modelos de regressão. 5. Dados epidemiológicos. 6. Bioestatística - Interdisciplinariedade. I. Martinez, Edson Zangiacomi, orient. II. Universidade Estadual de Maringá. Centro Ciências Exatas. Programa de Pós-Graduação em Bioestatística. III. Título.

CDD 23.ed. 570.15195

GVS-003671

*A minha futura esposa Carol.*

---

# Agradecimentos

---

Agradeço a minha família por trabalhar muito enquanto só me preocupava em estudar, meu pai Izaias e minha mãe Tereza me ensinaram o valor do estudo e fizeram entender que ele seria o caminho mais fácil para uma mudança de vida. Minhas irmãs Cida, Mariana e Wilma sempre me trataram como a seus filhos e trabalharam duro para ajudar na minha educação.

Agradeço a meu orientador Edson Martinez pela paciência e conhecimento compartilhado, além de ser um pesquisador e professor acima da média é um exemplo de pai e cidadão preocupado com o bem do próximo.

Tenho muito a agradecer às professoras Isolde e Eunice por terem aberto a mim as portas do Mestrado em Bioestatística da UEM, a todos professores do programa por seus ensinamentos e conselhos e à CAPES pelo apoio financeiro.

A todos colegas do mestrado, pois os professores devem ensinar e orientar devido a função que escolheram para vida, é claro que alguns os fazem com mais paixão, porém os colegas de sala o fazem exclusivamente por amor ao ensino. Não posso deixar de agradecer em especial ao Tiago, Marcos, Diego, Ricardo, Márcia, Edilenia, Omar, Oilson e Jardel, companheiros que se transformaram em grandes amigos.

Por todos os momentos que passamos juntos não posso me esquecer da minha Carol, a pessoa que mais apostou em mim, a grande incentivadora e inspiração para meus sonhos. Muito obrigado meu bem.

Agradeço também a Deus por ter colocado pessoas tão abençoados no meu caminho.

*"O mundo não é um mar de rosas; é um lugar sujo, um lugar cruel, que não quer saber o quanto você é durão. Vai botar você de joelhos e você vai ficar de joelhos para sempre se você deixar. Você, eu, ninguém vai bater tão forte como a vida, mas não se trata de bater forte. Se trata de quanto você aguenta apanhar e seguir em frente, o quanto você é capaz de aguentar e continuar tentando. É assim que se consegue vencer. Agora se você sabe do teu valor, então vá atrás do que você merece, mas tem que estar preparado para apanhar. E nada de apontar dedos, dizer que você não consegue por causa dele ou dela, ou de quem quer que seja. Só covardes fazem isso e você não é covarde, você é melhor que isso."*

*Rocky Balboa*

# Resumo

Neste trabalho, foram introduzidos modelos Bayesianos para análise de séries temporais com funções seno e cosseno para explicar a sazonalidade dos dados na presença de um ou mais pontos de mudança aleatórios e termos autorregressivos. O desenvolvimento deste modelo é motivado por duas coleções de dados, sendo uma relacionada ao número mensal de dispensação de medicamentos para asma pelos serviços de saúde pública de Ribeirão Preto, Sudeste do Brasil, de 1999 a 2011. A outra coleção de dados refere-se às contagens mensais de notificações de acidentes por picadas de cobras no Estado de São Paulo entre janeiro de 2007 e dezembro de 2014. Para obter estimativas para os parâmetros de interesse, foi utilizado procedimento Bayesiano de simulação em Monte Carlo em Cadeias de Markov (MCMC - *Markov Chain Monte Carlo*), baseado no algoritmo de amostradores de Gibbs. O modelo conseguiu capturar a sazonalidade nas contagens de acidentes por picadas de cobras no Estado de São Paulo e pode ser usado em uma variedade de dados onde há a suposição de sazonalidade. No segundo banco de dados, o Modelo Bayesiano com termos autorregressivos de ordem 2 e componente sazonal ajustou-se bem aos dados e permitiu estimar o ponto de mudança em novembro de 2007, provavelmente refletindo os resultados das novas políticas e programas de saúde voltados a pacientes com asma introduzidas até o ano de 2006. Os resultados do modelo com três pontos de mudança sugerem que o presente modelo é útil para analisar o número mensal de medicamentos para a asma dispensados e que pode ser usado para descrever uma grande e ampla variedade de séries temporais epidemiológicas, onde há um ou mais pontos de mudança.

**Palavras-chave:** Série temporal, modelos de ponto de mudança, métodos bayesianos, modelos de regressão, dados epidemiológicos.

# Abstract

In this paper, Bayesian models have been introduced for analysis of time series with sine and cosine functions to explain the seasonality of the data in the presence of one or more random change-point and autoregressive terms. The development of this model is driven by two data collections, one related to the monthly number of dispensing drugs to asthma by public health services in Ribeirão Preto, Southeastern Brazil, 1999-2011. The other collection of data refer to monthly counts of accident notifications snake bites in the state of São Paulo between January 2007 and December 2014. For obtain estimates for the parameters of interest, a Bayesian Markov Chain Monte Carlo (MCMC) simulation procedure using the Gibbs sampler algorithm was developed. The model was able to capture the seasonality in accident counts for snake bites in the state of São Paulo and can be used in a variety of data where there is seasonality assumption. In the second database, the Bayesian model with autoregressive terms of order 2 and seasonal component set and the data and allowed us to estimate the change-point in November 2007, probably reflecting the results of new health policies and programs aimed at patients with asthma introduced until 2006. The model results with three change-point suggest that this model is useful to analyze the monthly number of drugs for dispensed asthma that can be used to describe a large variety of epidemiologic time series where there are one or more change-point.

**Key words:** Time series, change-point model, Bayesian methods, regression models, epidemiological data.



---

## Lista de ilustrações

---

|  |    |
|--|----|
| Figura 1 – Gráfico de $S_t$ , considerando $\eta_1 = \eta_2 = 1$ e período $T = 12$ . A amplitude é denotada por $\alpha$ . . . . .  | 16 |
| Figura 2 – Gráficos de $S_t$ para valores diferentes de $\eta_{11}$ , $\eta_{12}$ , $\eta_{21}$ e $\eta_{22}$ . . . . .  | 17 |
| Figura 3 – Séries cronológicas do número de acidentes por (a) <i>Bothrops</i> , (b) <i>Crotalus</i> e (c) Cobras não venenosas. Estado de São Paulo, Brasil, Janeiro de 2007 a Dezembro de 2014. . . . .   | 18 |
| Figura 4 – Comparação entre as observações da série temporal (linha preta) e os valores preditos obtidos nos Modelos 1 com $J = 1$ (linha azul) e 2 com $J = 2$ (linha vermelha) com termos autorregressivos de ordem 1. . . . .   | 22 |
| Figura 5 – Função de autocorrelação (ACF) e autocorrelação parcial (PACF) para os resíduos dos Modelos 1 e 2 com termos autorregressivos de ordem 1. . . . .   | 23 |
| Figura 6 – Número mensal de medicamentos dispensados (comprimido de aminofilina, 100 mg) pelos serviços públicos de saúde de Ribeirão Preto, Brasil, de fevereiro de 1999 a dezembro de 2011. . . . .  | 32 |
| Figura 7 – Descrição da convergência e histogramas das distribuições do parâmetro $\alpha$ do Modelo 3 com resíduos que seguem uma distribuição normal. . . . .  | 34 |
| Figura 8 – Descrição da convergência e histogramas das distribuições dos parâmetros do Modelo 3 com resíduos que seguem uma distribuição normal: $\beta_1$ , $\beta_2$ , $\gamma_1$ e $\gamma_2$ . . . . .   | 35 |
| Figura 9 – Descrição da convergência e histogramas das distribuições dos parâmetros do Modelo 3 com resíduos que seguem uma distribuição normal: $\sigma_1^2$ , $\sigma_2^2$ , $\eta_1$ e $\eta_2$ . . . . .   | 36 |
| Figura 10 – Descrição da convergência e histogramas das distribuições dos parâmetros $\theta$ do Modelo 3 com resíduos que seguem uma distribuição normal, mais descrição da convergência e histograma da distribuição do parâmetro $\theta$ do Modelo 3 com resíduos que seguem uma distribuição t-Student. . . . . | 37 |

|   |    |
|---|----|
| Figura 11 – Descrição da convergência e histogramas das distribuições dos parâmetros do Modelo 3 com resíduos que seguem uma distribuição t-Student:<br>$\beta_1, \beta_2, \gamma_1$ e $\gamma_2$ .                                 | 38 |
| Figura 12 – Descrição da convergência e histogramas das distribuições dos parâmetros do Modelo 3 com resíduos que seguem uma distribuição t-Student:<br>$\alpha, v(df), \sigma_1^2, \sigma_2^2$ .                                   | 39 |
| Figura 13 – Descrição da convergência e histogramas das distribuições dos parâmetros do Modelo 3 com resíduos que seguem uma distribuição t-Student:<br>$\sigma_1^2 v(v-2)^{-1}, \sigma_2^2 v(v-2)^{-1}, \eta_1$ e $\eta_2$ .       | 40 |
| Figura 14 – Comparação entre a série temporal observada e os valores estimados do Modelo 1, com resíduos que seguem uma distribuição normal e distribuição t-Student.   | 42 |
| Figura 15 – Comparação entre a série temporal observada e os valores estimados do Modelo 2, com resíduos que seguem uma distribuição normal e distribuição t-Student.   | 43 |
| Figura 16 – Comparação entre a série temporal observada e os valores estimados do Modelo 3, com resíduos que seguem uma distribuição normal e distribuição t-Student.   | 44 |
| Figura 17 – Comparação entre os valores inversos para CPO obtidos do Modelo 3 com resíduos que seguem uma distribuição normal e outra que segue uma distribuição t-Student.   | 45 |
| Figura 18 – Gráficos de autocorrelações, histograma e resíduos do Modelo 3 com ponto de mudança e termos autorregressivos de ordem 1, com uma componente sazonal $S_t$ e resíduos que seguem uma distribuição normal.               | 46 |
| Figura 19 – Gráficos de autocorrelações, histograma e resíduos do Modelo 5 com ponto de mudança e termos autorregressivos de ordem 2, com uma componente sazonal $S_t$ e resíduos que seguem uma distribuição t-student.            | 47 |
| Figura 20 – Comparação entre a série temporal observada e o ajuste obtido do Modelo 5 com ponto de mudança e termos autorregressivos de ordem 2, com uma componente sazonal $S_t$ e resíduos que seguem uma distribuição t-student. | 49 |
| Figura 21 – Número mensal de medicamentos dispensados (xarope de salbutamol, 120 ml) pelos serviços públicos de saúde de Ribeirão Preto - SP, Brasil, de fevereiro de 1999 a dezembro de 2011.                                      | 54 |
| Figura 22 – Padrão sazonal mensal do número de dispensação do xarope de salbutamol ajustados pelo (a) Modelo 3, (b) e (c) Modelo 4 com covariáveis.   | 61 |
| Figura 23 – Comparação entre a série observada e o ajuste calculado a partir do Modelo 1 - com três pontos de mudança.  | 62 |

|   |    |
|---|----|
| Figura 24 – Comparação entre a série observada e o ajuste calculado a partir do Modelo 2 - com três pontos de mudança e termos autorregressivos de ordem 1. . . . .                     | 63 |
| Figura 25 – Comparação entre a série observada e o ajuste calculado a partir do Modelo 3 - com três pontos de mudança, termos autorregressivos de ordem 1 e componente sazonal. . . . . | 64 |
| Figura 26 – Gráficos de autocorrelações dos resíduos que seguem uma distribuição normal dos Modelos 1, 2 e 3. . . . .   | 65 |
| Figura 27 – Comparação entre os valores inversos para CPO obtidos com os Modelos 1, 2 e 3 com resíduos que seguem uma distribuição normal. . . . .                                      | 66 |

---

## Lista de tabelas

---

|  |    |
|--|----|
| Tabela 1 – Resultados bayesianos dos modelos ajustados 1 e 2, considerando os dados sobre os acidentes por picadas de serpentes venenosas do gênero <i>Bothrops</i> e <i>Crotalus</i> , e cobras não venenosas. . . . .  | 20 |
| Tabela 2 – Resultados bayesianos dos modelos ajustados 1 e 2 mas sem AR(1), agora nomeados de Modelo 3 e 4 considerando os dados sobre os acidentes por picadas de serpentes venenosas do gênero <i>Bothrops</i> e <i>Crotalus</i> , e cobras não venenosas. . . . . | 21 |
| Tabela 3 – Os dados mensais sobre o número de medicamentos dispensados (comprimidos de aminofilina de 100 mg) em Ribeirão Preto, Brasil, de fevereiro de 1999 a dezembro de 2011. . . . .  | 27 |
| Tabela 4 – Os resultados dos modelos estatísticos de mudança de ponto - resíduos normais. . . . .  | 33 |
| Tabela 5 – Os resultados dos modelos estatísticos com mudança de ponto - resíduos t-Student. . . . .   | 33 |
| Tabela 6 – Resultados dos modelos de regressão que consideram a temperatura máxima e mínima mensal absoluta como variáveis independentes. . . .  | 41 |
| Tabela 7 – Resultados dos modelos que consideram a temperatura média mensal e precipitação como variáveis independentes. . . . .   | 41 |
| Tabela 8 – Os resultados do Modelo 5 com ponto de mudança bayesiano com termos auto regressivos de ordem 2 e resíduos que seguem uma distribuição normal. . . . .  | 48 |
| Tabela 9 – Dados mensais sobre o número de medicamentos dispensados (Xarope de Sabultamol) em Ribeirão Preto - SP, Brasil, de fevereiro de 1999 a dezembro de 2011. . . . .  | 53 |
| Tabela 10 – Resultados dos modelos estatísticos de mudança de ponto, com sequência de resíduos que seguem uma distribuição normal. . . . .   | 58 |
| Tabela 11 – Resultados dos modelos de regressão que consideram a temperatura máxima e mínima mensal absoluta como variáveis independentes. . . .   | 59 |

|   |    |
|---|----|
| Tabela 12 – Resultados dos modelos de regressão que consideram a temperatura média mensal e precipitações como variáveis independentes. . . . . | 60 |
|---|----|

---

# Sumário

---

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introdução</b>   | <b>14</b> |
| <b>2</b> | <b>Modelo Bayesiano com componente sazonal aplicado a estudos de séries temporais de dados epidemiológicos</b>                              | <b>15</b> |
| 2.1      | Introdução  | 15        |
| 2.2      | Métodos   | 16        |
| 2.2.1    | O campo de estudo e o conjunto de dados   | 16        |
| 2.2.2    | Modelo estatístico  | 18        |
| 2.3      | Resultados e Discussões   | 20        |
| 2.4      | Conclusão   | 24        |
| <b>3</b> | <b>Modelo Bayesiano com um ponto de mudança e termos autorregressivos aplicados em estudos de séries temporais de dados epidemiológicos</b> | <b>25</b> |
| 3.1      | Introdução  | 25        |
| 3.2      | Métodos   | 26        |
| 3.2.1    | O campo de estudo e o conjunto de dados   | 26        |
| 3.2.2    | Modelo estatístico  | 27        |
| 3.2.3    | Especificações dos Modelos  | 30        |
| 3.2.4    | Seleção do Modelo   | 31        |
| 3.3      | Resultados  | 32        |
| 3.3.1    | Modelos com termos autorregressivos de ordem 1  | 32        |
| 3.3.2    | Modelo com termos autorregressivos de ordem 2   | 43        |
| 3.4      | Discussão   | 46        |
| 3.5      | Conclusão   | 50        |
| <b>4</b> | <b>Modelo com mais de um ponto de mudança</b>   | <b>51</b> |
| 4.1      | Introdução  | 51        |
| 4.2      | Métodos   | 53        |
| 4.2.1    | O campo de estudo e o conjunto de dados   | 53        |
| 4.2.2    | Modelo estatístico  | 55        |
| 4.2.3    | Especificações dos Modelos  | 57        |
| 4.2.4    | Seleção do Modelo   | 57        |
| 4.3      | Resultados  | 58        |
| 4.4      | Discussão   | 64        |
| 4.5      | Conclusão   | 65        |

**5 Comentários Gerais . . . . . 67**  
**Referências . . . . . 69**

---

## Capítulo 1

---

---

# Introdução

---

Em muitas situações são utilizados dados epidemiológicos sob a forma de séries temporais. Notificações de morbidades, hospitalizações devidas a uma doença específica e taxas de mortalidade observados em um intervalo de tempo são exemplos de medidas que podem ser estudadas como séries temporais. Os modelos estatísticos são úteis para descrever padrões destas séries, como as tendências temporais e flutuações sazonais. Estes modelos podem também ser usados para prever futuras observações após observar uma série temporal, fornecendo informações para ajudar na vigilância e controle de eventos de interesse em saúde pública.

O primeiro estudo, apresentado no Capítulo 2, descreve um modelo bayesiano com funções seno e cosseno para explicar a sazonalidade e termos autorregressivos. O modelo foi aplicado às contagens mensais de notificações de acidentes por picadas de cobras no Estado de São Paulo entre janeiro de 2007 e dezembro de 2014.

O segundo estudo, apresentado no Capítulo 3, também descreve um modelo bayesiano com uma função que captura a sazonalidade e termos autorregressivos, mas incluem ponto de mudança ([RUGGERI et al., 2007](#)). O desenvolvimento deste modelo é motivado por um conjunto de dados do número mensal de dispensação de medicamentos para asma por parte dos serviços de saúde pública de Ribeirão Preto, Sudeste do Brasil, de 1999 a 2011 (cápsula de aminofilina 100mg).

O terceiro estudo, apresentado no Capítulo 4, introduz um modelo mais completo, pois além de contemplar todas as funções já descritas nos modelos anteriores, insere mais de um ponto de mudança no modelo. Esse modelo bayesiano descreve também um conjunto de dados do número mensal de dispensação de medicamentos para asma por parte dos serviços de saúde pública de Ribeirão Preto, Sudeste do Brasil, de 1999 a 2011 (frasco de xarope de salbutamol 120ml).



## Capítulo 2

# Modelo Bayesiano com componente sazonal aplicado a estudos de séries temporais de dados epidemiológicos

## 2.1 Introdução

Variações sazonais são comuns em estudos epidemiológicos. Por exemplo, sabe-se que a dengue no Brasil é mais incidente na estação chuvosa (MARTINEZ; SILVA, 2011) e óbitos por síndrome de morte súbita ocorrem predominantemente nos meses frios (NUNES et al., 2001). Modelos com termos autorregressivos integrados a médias móveis com sazonalidade (SARIMA - *Seasonal autoregressive integrated moving average*) (NOBRE et al., 2001) são frequentemente utilizados para analisar dados de séries temporais contendo componentes sazonais para monitorar a incidência de doenças infecciosas (BOWIE; PROTHERO, 1981). Modelos SARIMA incluem componentes sazonais e consideraram a série temporal como uma combinação linear de suas próprias observações passadas e valores atuais de um termo de erro.

Alternativamente, podemos modelar a série de tempo, atribuindo aos dados uma equação que inclui um termo que descreve o padrão sazonal sobre unidades de tempo  $T$ , na forma:

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

em que  $\eta_1$  e  $\eta_2$  são valores constantes (STOLWIJK et al., 1999). Considerando, por exemplo, uma série temporal de observações mensais com um padrão de sazonalidade anual ( $T = 12$ ), podemos ter um gráfico (Figura 1) que ilustra a  $S_t$  com  $\eta_1 = \eta_2 = 1$ . Podemos notar que o gráfico de  $S_t$  é uma onda com período  $T$  e amplitude dada por:

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

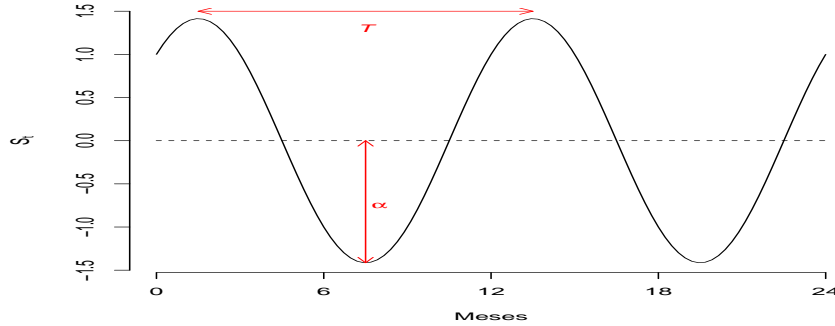


Figura 1 – Gráfico de  $S_t$ , considerando  $\eta_1 = \eta_2 = 1$  e período  $T = 12$ . A amplitude é denotada por  $\alpha$ .

Nota-se que se  $t = 0$ , então  $S_t = \eta_2$ . Na Figura 1, observamos que  $S_t = 1$  quando  $t = 0$ .

Uma expansão de  $S_t$  que pode ser útil para modelar séries temporais mais complexas é dada por

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

em que  $J$  é um número inteiro,  $J \geq 2$ . Considerando  $J = 2$  e  $T = 12$ , a Figura 2 ilustra os gráficos de  $S_t$  para valores diferentes de  $\eta_{11}$ ,  $\eta_{12}$ ,  $\eta_{21}$  e  $\eta_{22}$ .

Assim, o objetivo deste capítulo é apresentar um modelo estatístico para investigar as propriedades das séries temporais epidemiológicas, onde  $S_t$  será incluído para os possíveis efeitos sazonais. Dados de séries temporais para o número de acidentes por picadas de cobras serão considerados para ilustrar a utilidade do modelo proposto. O método de estimativa de parâmetros é baseada em inferência Bayesiana, que utiliza métodos Markov Chain Monte Carlo (MCMC).

## 2.2 Métodos

### 2.2.1 O campo de estudo e o conjunto de dados

O Estado de São Paulo está localizado na parte Sudeste do Brasil. As cobras venenosas que ocorrem nesta região pertencem a três gêneros: *Bothrops*, *Crotalus* e *Micrurus* (RIBEIRO et al., 1998). No Estado de São Paulo, não há cobras do gênero *Lachesis* na natureza (RIBEIRO; JORGE, 1997). Sabe-se que as cobras do gênero *Bothrops* são as mais causadoras de acidentes por picadas no Brasil (BERTOLOZZI et al., 2015). O

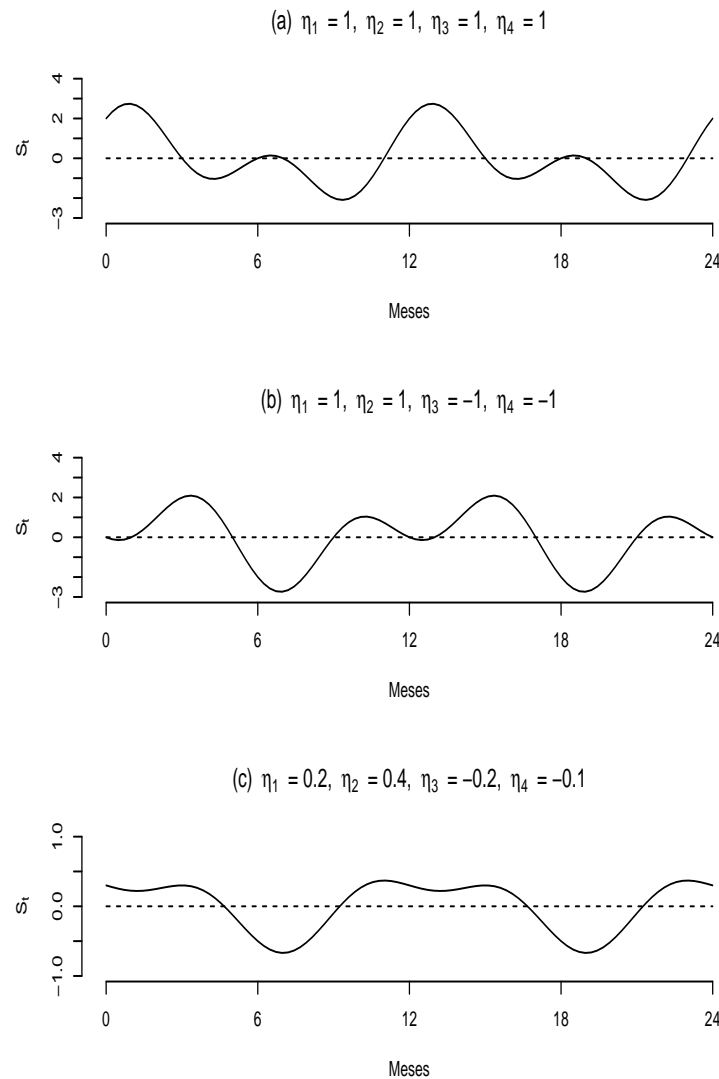


Figura 2 – Gráficos de  $S_t$  para valores diferentes de  $\eta_{11}, \eta_{12}, \eta_{21}$  e  $\eta_{22}$ .

presente estudo utilizará um conjunto de dados dos registros mensais de acidentes por serpentes que ocorreram no Estado de São Paulo, entre janeiro de 2007 e dezembro de 2014. Consideramos os registros de picadas por serpentes do gênero *Bothrops* e *Crotalus*, dado que há poucos acidentes causados por serpentes *Micrurus* notificados no Estado de São Paulo. Além disso, consideramos os registros de acidentes causados por serpentes não venenosas neste período. Estes dados são de domínio público e livremente disponível no site do Departamento de Informática do Sistema Único de Saúde do Brasil (DATASUS, <http://tabnet.datasus.gov.br>).

A Figura 3 mostra os gráficos da série temporal do número de acidentes por *Bothrops*, *Crotalus* e cobras não venenosas. Estes gráficos mostram uma sazonalidade clara da série, com uma tendência de um maior número de acidentes durante os meses quentes, entre novembro e março de cada ano.

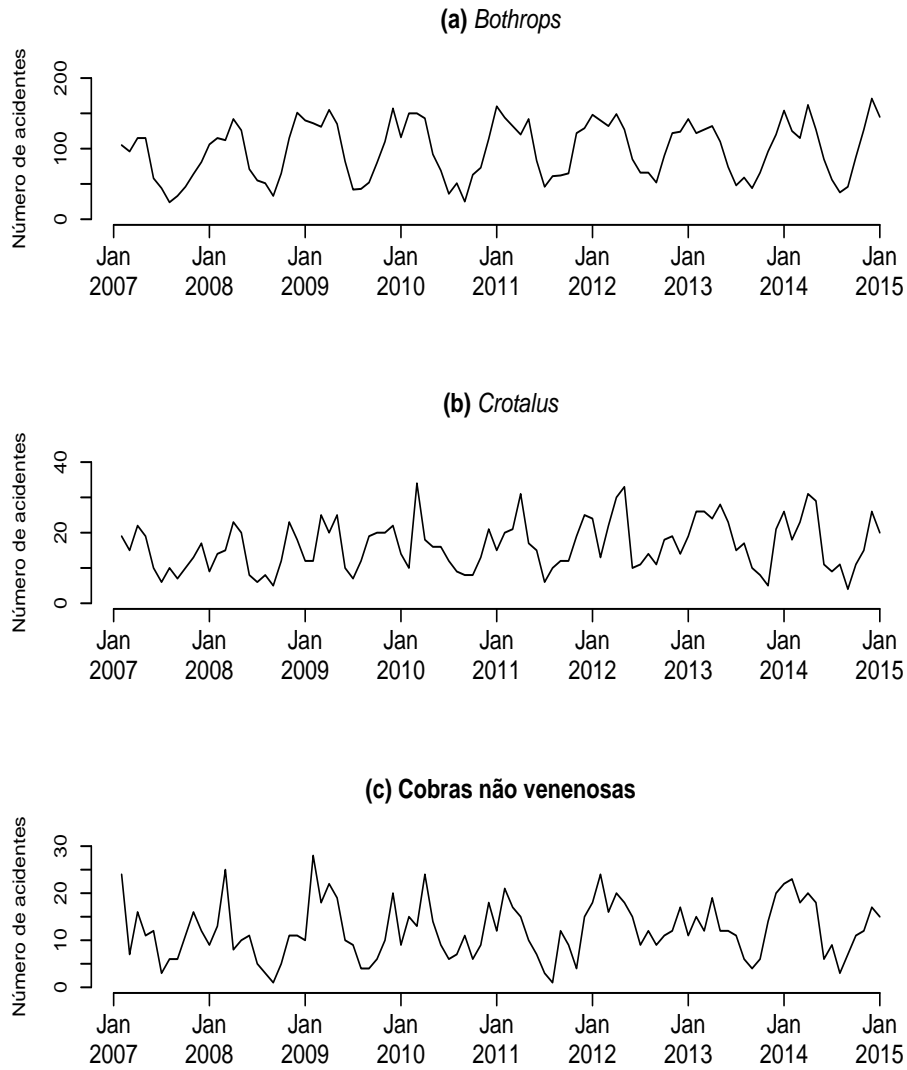


Figura 3 – Séries cronológicas do número de acidentes por (a) *Bothrops*, (b) *Crotalus* e (c) Cobras não venenosas. Estado de São Paulo, Brasil, Janeiro de 2007 a Dezembro de 2014.

### 2.2.2 Modelo estatístico

Seja  $y_t$  o número de eventos de interesse observado no tempo  $t$ , com  $t = 1, \dots, n$ . Vamos supor que  $y_t$  é uma observação de uma variável aleatória que segue uma distribuição de Poisson com parâmetro  $\mu_t$ , em que:

$$\log(\mu_t) = \beta_0 + \beta_1 t + S_t + \sum_{k=1}^p \gamma_k (y_{t-k} - \bar{y}),$$

Nesta equação,  $\beta_0$  é uma constante,  $\beta_1$  é o termo linear,  $S_t$  é a componente de efeito sazonal,  $\gamma_1, \dots, \gamma_p$  são termos autorregressivos e  $\bar{y}$  denota a média de  $y_1, \dots, y_n$ . A função de log-verossimilhança é dada por:

$$\ell(\boldsymbol{\theta}|\mathbf{Y}) = \log L(\boldsymbol{\theta}|\mathbf{Y}) = \sum_{t=1}^n y_t \log(\mu_t) - \sum_{t=1}^n \lambda_t - \sum_{t=1}^n \log(y_t!),$$

em que  $\boldsymbol{\theta} = (\beta_0, \beta_1, \eta_{11}, \dots, \eta_{J1}, \eta_{12}, \dots, \eta_{J2}, \gamma_1, \dots, \gamma_p)'$  é o vetor dos parâmetros,  $\mathbf{Y}$  denota o conjunto de dados e  $L(\boldsymbol{\theta}|\mathbf{Y})$  é a função de verossimilhança. Na análise Bayesiana, foram consideradas as seguintes distribuições *a priori*:

$$\begin{aligned}\beta_0 &\sim N(0, 10^5), \\ \beta_1 &\sim N(0, 10^5), \\ \eta_{j1} &\sim N(0, 10^5), \quad j = 1, \dots, J, \\ \eta_{j2} &\sim N(0, 10^5), \quad j = 1, \dots, J,\end{aligned}$$

e

$$\gamma_k \sim \text{Uniform}(-1, 1), \quad k = 1, \dots, p,$$

Desta forma, assumimos distribuições não informativas com média 0 e variância relativamente grande para os parâmetros  $\beta_0, \beta_1, \eta_{j1}$  e  $\eta_{j2}$ ,  $j = 1, \dots, J$ . Em adição, vamos usar uma distribuição uniforme *a priori* para os parâmetros autorregressivos, uma vez que assumimos  $|\gamma_k| < 1$ , mas transformações alternativas podem ser propostas. Vamos assumir independência entre as distribuições *a priori*.

A distribuição conjunta *a posteriori* para os parâmetros do modelo foi obtida via teorema de Bayes, combinando a distribuição conjunta *a priori* com a função de verossimilhança  $L(\boldsymbol{\theta}|\mathbf{Y})$  para o vetor dos parâmetros (GELMAN et al., 2014). Assim, as amostras da distribuição conjunta *a posteriori* serão geradas utilizando-se métodos de simulação MCMC. O OpenBUGS é um programa computacional de código aberto para a inferência Bayesiana, que fornece uma grande vantagem neste tipo de simulação, uma vez que a implementação do modelo depende apenas da função de probabilidade e as distribuições *a priori* para os parâmetros de interesse (FEITOSA et al., 2015). O algoritmo foi executado para 500.000 iterações com uma fase de descarte ("burn-in samples") de 1.000, para eliminar os efeitos dos valores iniciais, sendo amostradas a cada 10 simulações para assegurar a independência entre as amostras sucessivas.

A convergência foi verificada examinando gráficos de diagnóstico. A identificação do melhor ajuste para o modelo de séries temporais foi baseado no DIC (*Deviance Information Criterion*) (SPIEGELHALTER et al., 2002) e por meio de uma inspeção dos resíduos do modelo para autocorrelação usando função de autocorrelação (ACF) e autocorrelações parciais ACF (FACP). Modelos com valores menores para a DIC são considerados como aqueles que melhor se ajustam aos dados.

Tabela 1 – Resultados bayesianos dos modelos ajustados 1 e 2, considerando os dados sobre os acidentes por picadas de serpentes venenosas do gênero *Bothrops* e *Crotalus*, e cobras não venenosas.

| Gênero               | Parâmetro   | Modelo 1 ( $J = 1$ ) |                             | Modelo 2 ( $J = 2$ ) |                             |
|----------------------|-------------|----------------------|-----------------------------|----------------------|-----------------------------|
|                      |             | Estimação Bayesiana  | 95% intervalo credibilidade | Estimação Bayesiana  | 95% intervalo credibilidade |
| <i>Bothrops</i>      | $\beta_0$   | 4,444                | (4,3980 ; 4,490)            | 4,415                | (4,369 ; 4,462)             |
|                      | $\beta_1$   | 0,0015               | (0,0007 ; 0,0023)           | 0,0017               | (0,0009 ; 0,0026)           |
|                      | $\eta_{11}$ | 0,1654               | (0,1126 ; 0,2187)           | 0,2186               | (0,1553 ; 0,2819)           |
|                      | $\eta_{12}$ | 0,3507               | (0,3138 ; 0,3874)           | 0,4160               | (0,3761 ; 0,4559)           |
|                      | $\eta_{21}$ | -                    | -                           | -0,1653              | (-0,1982 ; -0,1323)         |
|                      | $\eta_{22}$ | -                    | -                           | -0,1005              | (-0,1331 ; 0,0687)          |
|                      | $\gamma_1$  | 0,0034               | (0,0023 ; 0,0045)           | 0,0026               | (0,0014 ; 0,0039)           |
|                      | DIC         |                      | 987,4                       |                      | 834,2                       |
| <i>Crotalus</i>      | $\beta_0$   | 2,594                | (2,481 ; 2,705)             | 2,554                | (2,440 ; 2,666)             |
|                      | $\beta_1$   | 0,0032               | (0,0013 ; 0,0052)           | 0,0037               | (0,0017 ; 0,0057)           |
|                      | $\eta_{11}$ | 0,2596               | (0,1676 ; 0,3509)           | 0,2859               | (0,1886 ; 0,3840)           |
|                      | $\eta_{12}$ | 0,226                | (0,1547 ; 0,2966)           | 0,2670               | (0,1908 ; 0,3440)           |
|                      | $\eta_{21}$ | -                    | -                           | -0,1350              | (-0,2151 ; -0,0548)         |
|                      | $\eta_{22}$ | -                    | -                           | -0,1688              | (-0,2406 ; -0,0961)         |
|                      | $\gamma_1$  | 0,0096               | (0,0055 ; 0,0188)           | 0,0039               | (-0,0061 ; 0,0141)          |
|                      | DIC         |                      | 603,4                       |                      | 574,7                       |
| Cobras não venenosas | $\beta_0$   | 2,26                 | (2,13 ; 2,387)              | 2,2447               | (2,116 ; 2,376)             |
|                      | $\beta_1$   | 0,0033               | (0,0011 ; 0,0055)           | 0,0033               | (0,0011 ; 0,0055)           |
|                      | $\eta_{11}$ | 0,3763               | (0,2596 ; 0,4922)           | 0,3987               | (0,2793 ; 0,5181)           |
|                      | $\eta_{12}$ | 0,3844               | (0,2951 ; 0,4725)           | 0,4233               | (0,3309 ; 0,5179)           |
|                      | $\eta_{21}$ | -                    | -                           | -0,1231              | (-0,2084 ; -0,0377)         |
|                      | $\eta_{22}$ | -                    | -                           | -0,0725              | (-0,1570 ; 0,0123)          |
|                      | $\gamma_1$  | -0,0026              | (-0,0163 ; 0,011)           | -0,0045              | (-0,0184 ; 0,0093)          |
|                      | DIC         |                      | 541,9                       |                      | 535,2                       |

## 2.3 Resultados e Discussões

A Tabela 1 mostra os resultados a partir do modelo proposto, considerando o termo autorregressivo de ordem  $p = 1$  e padrão sazonal anual ( $T = 12$ ) com  $J = 1$  (denotado por Modelo 1) e  $J = 2$  (Modelo 2). Para estes modelos, os respectivos parâmetros foram estimados pela média das amostras *a posteriori* e os intervalos de credibilidade de 95% são dados pelos percentis 2.5 e 97.5. Os modelos foram ajustados separadamente para os dados sobre os acidentes por picadas de serpentes venenosas dos gênero *Bothrops* e *Crotalus* e cobras não venenosas.

Considerando os dados sobre os acidentes causados por picadas de serpentes do gênero *Bothrops*, podemos notar que o Modelo 2 produz um melhor ajuste para os dados em termos de valor DIC (Tabela 1). Os gráficos na Figura 4 compara os valores preditivos com base nos Modelos 1 e 2 com termos autorregressivos de ordem 1 da série histórica observada. O painel (a) da Figura 4 sugere que os valores previstos obtidos a partir do Modelo 2 estão mais próximos dos valores reais do que as do Modelo 1. Além

Tabela 2 – Resultados bayesianos dos modelos ajustados 1 e 2 mas sem AR(1), agora nomeados de Modelo 3 e 4 considerando os dados sobre os acidentes por picadas de serpentes venenosas do gênero *Bothrops* e *Crotalus*, e cobras não venenosas.

| Gênero               | Parâmetro   | Modelo 3 ( $J = 1$ ) |                             | Modelo 4 ( $J = 2$ ) |                             |
|----------------------|-------------|----------------------|-----------------------------|----------------------|-----------------------------|
|                      |             | Estimação Bayesiana  | 95% intervalo credibilidade | Estimação Bayesiana  | 95% intervalo credibilidade |
| <i>Bothrops</i>      | $\beta_0$   | 4,399                | (4,355 ; 4,443)             | 4,382                | (4,338 ; 4,426)             |
|                      | $\beta_1$   | 0,0025               | (0,0001 ; 0,0032)           | 0,0025               | (0,0017 ; 0,0032)           |
|                      | $\eta_{11}$ | 0,3083               | (0,2784 ; 0,3379)           | 0,3395               | (0,3092 ; 0,3697)           |
|                      | $\eta_{12}$ | 0,4211               | (0,3911 ; 0,4511)           | 0,4668               | (0,4351 ; 0,4988)           |
|                      | $\eta_{21}$ | -                    | -                           | -0,1927              | (-0,223 ; -0,163)           |
|                      | $\eta_{22}$ | -                    | -                           | -0,0720              | (-0,1018 ; -0,0424)         |
|                      | DIC         |                      | 1027                        |                      | 850,5                       |
| <i>Crotalus</i>      | $\beta_0$   | 2,562                | (2,453 ; 2,67)              | 2,542                | (2,433 ; 2,65)              |
|                      | $\beta_1$   | 0,0039               | (0,0021 ; 0,0057)           | 0,0039               | (0,0021 ; 0,0057)           |
|                      | $\eta_{11}$ | 0,3182               | (0,2455 ; 0,3902)           | 0,312                | (0,2413 ; 0,3825)           |
|                      | $\eta_{12}$ | 0,2268               | (0,1548 ; 0,2987)           | 0,2681               | (0,1916 ; 0,3447)           |
|                      | $\eta_{21}$ | -                    | -                           | -0,1483              | (-0,2207 ; -0,0767)         |
|                      | $\eta_{22}$ | -                    | -                           | -0,1665              | (-0,239 ; -0,0946)          |
|                      | DIC         |                      | 605,7                       |                      | 573,3                       |
| Cobras não venenosas | $\beta_0$   | 2,267                | (2,139 ; 2,392)             | 2,255                | (2,128 ; 2,382)             |
|                      | $\beta_1$   | 0,0032               | (0,0010 ; 0,0053)           | 0,0031               | (0,0010 ; 0,0053)           |
|                      | $\eta_{11}$ | 0,3611               | (0,2759 ; 0,4459)           | 0,372                | (0,2857 ; 0,4588)           |
|                      | $\eta_{12}$ | 0,3795               | (0,2943 ; 0,4648)           | 0,4153               | (0,3252 ; 0,5054)           |
|                      | $\eta_{21}$ | -                    | -                           | -0,12                | (-0,2055 ; -0,0356)         |
|                      | $\eta_{22}$ | -                    | -                           | -0,0735              | (-0,1588 ; 0,0111)          |
|                      | DIC         |                      | 540,1                       |                      | 533,6                       |

disso, a Figura 5 mostra os gráficos ACF e PACF para os resíduos brutos ( $y_t - \hat{y}_t$ ) dos Modelos 1 e 2. A Figura 5 mostra que não houve significativa auto-correlação entre os resíduos nos diferentes intervalos de tempo (defasagens) no Modelo 2, o que indica novamente que o Modelo 2 é melhor do que o Modelo 1. No Modelo 2 temos evidência de um leve aumento nos registros de acidentes causados por serpentes do gênero *Bothrops* ( $\hat{\beta}_1 = 0,0017$ ) durante o período considerado e o termo autorregressivo de ordem 1 mostrou-se significativo ( $\hat{\gamma}_1 = 0,0026$ , com um respectivo intervalo de credibilidade de 95% que não inclui o valor 0).

Um outro modelo, incluindo um termo autorregressivo de ordem 2 foi ajustado, mas o seu ajuste para os dados não sugeriu uma melhora. Além disso, os modelos que não incluem termos autorregressivos, considerando  $J = 1$  e  $J = 2$ , também foram ajustados aos dados sobre os acidentes causados por picadas de serpentes do gênero *Bothrops* e *Crotalus*, e cobras não venenosas (Tabela 2). Percebemos também que mesmo sem AR(1) os modelos com  $J = 2$  se mostram melhores do que os com  $J = 1$ . Mais uma vez, não foi obtida um melhor ajuste para os dados, em comparação com os resultados mostrados na Tabela 1.

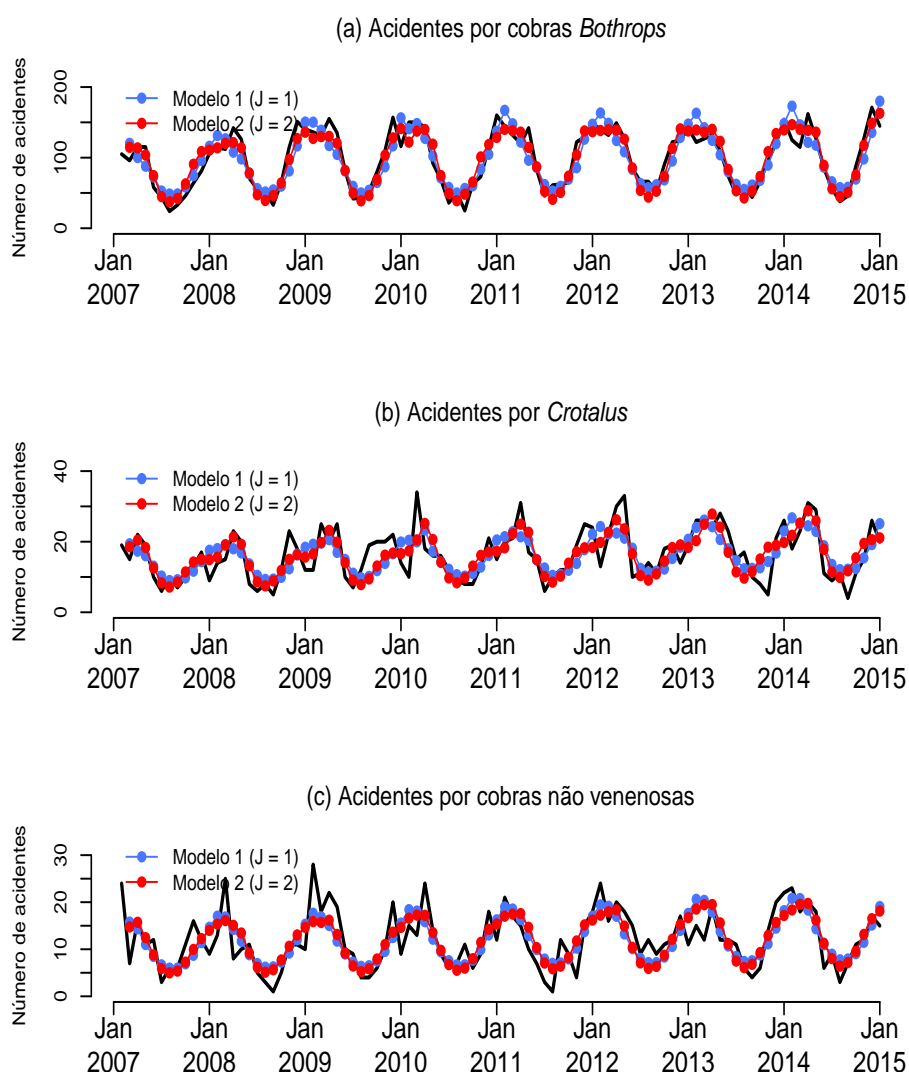


Figura 4 – Comparação entre as observações da série temporal (linha preta) e os valores preditos obtidos nos Modelos 1 com  $J = 1$  (linha azul) e 2 com  $J = 2$  (linha vermelha) com termos autorregressivos de ordem 1.

Considerando os dados sobre os acidentes por serpentes *Crotalus*, o Modelo 2 também proporciona um melhor ajuste para os dados em termos do valor DIC (Tabela 1), mas o respectivo intervalo de credibilidade 95% para  $\gamma_1$  inclui o valor zero, o que sugere um papel não significativo do termo autorregressivo na modelagem desses dados. O ajuste de um modelo que não inclui o termo  $\gamma_1$  resulta em um valor DIC semelhante (573,3 - Tabela 2). O Modelo 2 também evidencia um leve aumento de tendência nos registros de acidentes causados por picadas de serpentes *Crotalus* ( $\hat{\beta}_1 = 0,0037$ ) no período considerado. Gráficos de ACF e PACF para os resíduos brutos descrevem que uma ausência de auto-correlação significativa entre os resíduos nos diferentes intervalos desses resíduos no Modelo 2 (gráficos não mostrado). O painel (b) da Figura 4 compara os valores previstos obtidos a partir dos modelos com  $J = 1$  e  $J = 2$  e termo autorregressivo de ordem



1.

Os resultados do ajuste dos Modelos 1 e 2 para os dados sobre os acidentes causados por picadas de serpentes não venenosas também são apresentados na Tabela 1, onde observamos que o Modelo 2 produz um melhor ajuste para os dados em termos do valor DIC. Este modelo também evidenciou um leve aumento na tendência dos registros de acidentes ( $\hat{\beta}_1 = 0,0033$ ) durante o período considerado e um termo autorregressivo não significativo ( $\hat{\gamma}_1 = -0,0045$ , com um respectivo intervalo de credibilidade de 95%, que inclui o valor zero). O ajuste de um modelo que não inclui o termo  $\gamma_1$  resulta em um valor DIC semelhante (533,6 - Tabela 2). O painel (c) da Figura 4 compara os valores previstos obtidos a partir dos modelos com  $J = 1$  e  $J = 2$  e termo auto regressivo de ordem 1.

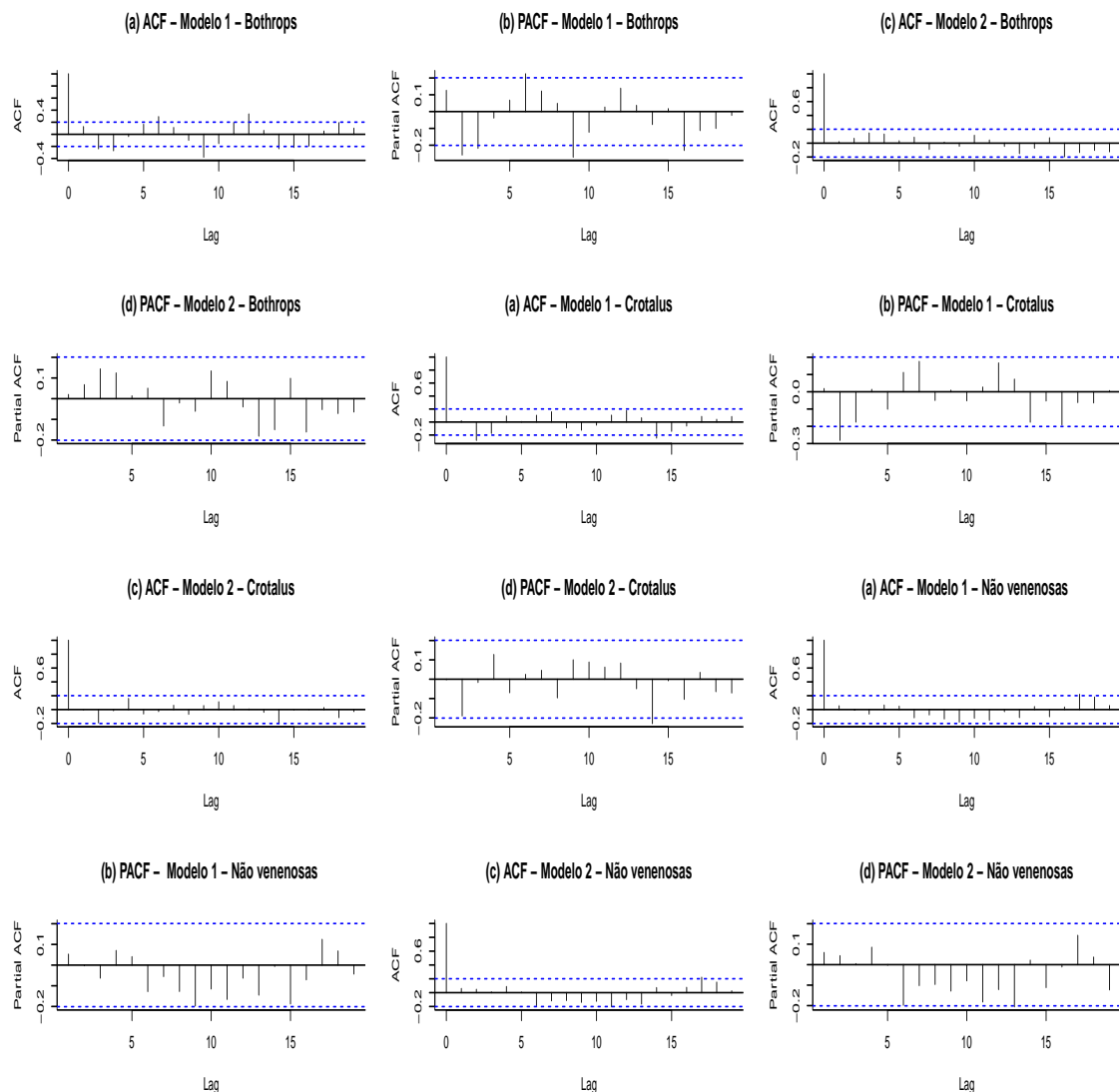


Figura 5 – Função de autocorrelação (ACF) e autocorrelação parcial (PACF) para os resíduos dos Modelos 1 e 2 com termos autorregressivos de ordem 1.

Acidentes com serpentes são descritos como importantes problemas de saúde pública e ambiental, principalmente, em comunidades rurais de países pobres e em desen-

volvimento (CUPO, 2015). Feitosa et al. (2015) descrevem que, desde 1986, profissionais de saúde do Brasil são obrigadas a notificar as autoridades sobre envenenamentos por serpentes. Estes autores consideram que esses registros são úteis para reduzir a incidência deste tipo de acidentes, promovendo a educação para a saúde, e para reduzir sua gravidade, a frequência de sequelas e a letalidade. No entanto, Bochner et al. (2014) relataram que dificuldades no acesso aos serviços de saúde podem contribuir para uma não notificação dos acidentes ofídicos, e isso pode ser uma limitação importante do presente estudo, uma vez que o número real de acidentes por serpentes pode ser maior do que a obtida a partir do DATASUS.

Apesar deste problema, o presente artigo mostra que o modelo estatístico proposto é útil para descrever séries temporais epidemiológicas na presença de termos de sazonalidade e autorregressivos. O ajuste do modelo aos dados descritos mostrou uma sazonalidade anual dos acidentes por serpentes no Estado de São Paulo, onde o maior número de eventos são observados nas estações mais quentes do ano, entre novembro e março.

## 2.4 Conclusão

Um modelo estatístico para as séries temporais epidemiológicas com base em funções seno e co-seno, por meio de abordagem Bayesiana mostrou ser uma ferramenta útil para estimar os parâmetros de interesse do modelo proposto. As comparações entre os valores estimados a partir do modelo e os dados observados mostram um ajuste razoavelmente bom. O modelo apresentado pode ser útil no estudo de séries temporais em outros contextos diferentes, quando se observa um padrão sazonal nos dados e componentes autorregressivos.

---

## Capítulo 3

---

# Modelo Bayesiano com um ponto de mudança e termos autorregressivos aplicados em estudos de séries temporais de dados epidemiológicos

---

### 3.1 Introdução

Modelos com ponto de mudança ([RUGGERI et al., 2007](#); [LEE, 2010](#)) têm sido cada vez mais utilizados em um amplo espectro de aplicações, como em Econometria, Medicina ([GHOSH; VAIDA, 2007](#)) e Estudos Ambientais ([ACHCAR et al., 2010](#); [ACHCAR et al., 2011](#)). Essas ferramentas estatísticas são usadas em problemas práticos em que uma variável aleatória indexada pelo tempo muda seu comportamento em um ou mais momentos. Assim, estes modelos são úteis, por exemplo, quando o interesse do pesquisador consiste em determinar se a série de tempo observada é homogênea ao longo do intervalo de tempo. Como um exemplo, [Achcar et al. \(2008\)](#) consideraram uma análise de ponto de mudança para a incidência de casos de tuberculose em Nova York de 1970 a 2000, quando o número de casos da doença apresentou diferentes comportamentos em três períodos distintos. Num primeiro período, a tendência de incidência estava em declínio, provavelmente relacionado a bons programas de controle da doença. Num segundo período, houve um aumento nas taxas de incidência; e no terceiro período, houve um novo declínio nas taxas de incidência. Métodos Bayesianos modernos de inferência utilizam simulação em Cadeias de Markov (MCMC - *Markov Chain Monte Carlo*). Essas técnicas têm sido utilizadas para modelar dados de séries temporais na presença de um ou mais pontos de mudança ([ACHCAR et al., 2008](#); [BARRY; HARTIGAN, 1993](#); [CARLIN et al., 1992](#); [DEY; PURKAYASTHA, 1997](#); [LAVIELLE; LEBARBIER, 2001](#)), e incluem vários modelos de pontos de mudança em que o número de

pontos é desconhecido (CHIB, 1998; FEARNHEAD, 2006). No presente estudo, o objetivo é introduzir um modelo estatístico Bayesiano com único ponto de mudança, incluindo termos autorregressivos, aplicado ao número mensal de medicamentos para asma dispensados pelos serviços públicos de saúde de Ribeirão Preto, Sudeste do Brasil. Variáveis climáticas estão incluídas como variáveis independentes. Este modelo é uma extensão do modelo apresentado no capítulo anterior, sendo introduzido um ponto de mudança e a inclusão de covariáveis.

## 3.2 Métodos

### 3.2.1 O campo de estudo e o conjunto de dados

O presente estudo insere-se em uma pesquisa maior sobre a dispensação de medicamentos para tratar doenças pulmonares nos serviços públicos de saúde de Ribeirão Preto, município localizado na região noroeste do Estado de São Paulo, Brasil. Ribeirão Preto é classificado como o oitavo maior município do Estado, com cerca de 600 mil habitantes (dados do censo do IBGE, disponíveis em [www.censo2010.ibge.gov.br/sinopse/](http://www.censo2010.ibge.gov.br/sinopse/)). O município pertence à área de cobertura de saúde do XIII Departamento Regional da Secretaria da Saúde do Estado de São Paulo, e é considerado um centro regional de referência na área da saúde com intervenções de média e alta complexidades, atendendo mais de 1,2 milhões de habitantes, dos quais cerca de 62% dependem exclusivamente do Sistema Único de Saúde (SUS) (BITTAR et al., 2011). A rede pública de serviços de saúde em Ribeirão Preto é composta por serviços municipais, estaduais e filantrópicos, tendo 36 farmácias que prestam assistência farmacêutica de acordo com as diretrizes da Política Nacional de Medicamentos (Portaria GM 3.916 de 30 de Novembro, 1998) e atualmente fornecem para a população mais de 260 medicamentos indicados para o tratamento de várias doenças, incluindo a asma.

Os dados sobre o número de medicamentos dispensados foram fornecidos pelo Sistema Hygia Web, um sistema de informação em saúde utilizado pela Secretaria Municipal de Saúde de Ribeirão Preto desde 1992. Este sistema permite o registro de informações sobre os cuidados de saúde em toda a rede pública municipal de saúde. Em 1998, a implementação de um módulo de gestão de medicamentos no Hygia Web permitiu o registro de dados de dispensação de medicamentos e assistência farmacêutica. Assim, foi possível recuperar informações secundárias sobre a dispensação dos principais medicamentos para o tratamento da asma que cobre todo o município, a partir de 1999.

Para efeitos do presente estudo, os dados de apenas um medicamento usado para aliviar os sintomas da asma foram considerados no desenvolvimento do modelo estatístico, os comprimidos de aminofilina de 100 mg. A aminofilina é uma medicação com ação antiinflamatória e broncodilatadora (CRUZ; PONTE, 2003), mas considerada uma

Tabela 3 – Os dados mensais sobre o número de medicamentos dispensados (comprimidos de aminofilina de 100 mg) em Ribeirão Preto, Brasil, de fevereiro de 1999 a dezembro de 2011.

| ANO\MÊS | JAN    | FEV    | MAR    | ABR    | MAI    | JUN    | JUL    | AGO    | SET    | OUT    | NOV    | DEZ    |
|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1999    | -      | 5.409  | 7.455  | 14.766 | 19.430 | 23.704 | 26.408 | 27.840 | 19.365 | 26.998 | 22.854 | 20.976 |
| 2000    | 22.107 | 28.574 | 34.598 | 30.269 | 38.301 | 33.110 | 31.475 | 33.727 | 31.986 | 31.334 | 30.695 | 34.814 |
| 2001    | 38.886 | 31.585 | 31.734 | 35.491 | 40.044 | 37.546 | 43.186 | 45.343 | 35.847 | 41.003 | 40.734 | 32.591 |
| 2002    | 38.236 | 34.783 | 38.996 | 39.382 | 41.758 | 38.750 | 44.330 | 43.815 | 40.823 | 44.162 | 40.461 | 37.671 |
| 2003    | 43.055 | 21.452 | 29.513 | 36.305 | 39.258 | 38.015 | 51.559 | 35.279 | 34.930 | 37.197 | 42.306 | 41.742 |
| 2004    | 43.774 | 43.547 | 62.281 | 58.343 | 54.102 | 53.283 | 61.681 | 65.380 | 56.163 | 53.098 | 57.890 | 54.866 |
| 2005    | 50.786 | 44.233 | 60.625 | 58.377 | 61.750 | 63.355 | 57.444 | 67.948 | 57.358 | 55.572 | 58.325 | 62.312 |
| 2006    | 58.473 | 52.563 | 71.813 | 61.630 | 77.719 | 66.508 | 62.040 | 68.141 | 59.367 | 66.398 | 63.548 | 59.999 |
| 2007    | 63.637 | 58.074 | 64.708 | 60.228 | 74.773 | 55.869 | 68.219 | 68.580 | 52.919 | 63.627 | 52.965 | 49.376 |
| 2008    | 57.650 | 51.590 | 58.503 | 59.372 | 55.746 | 57.484 | 55.301 | 51.615 | 53.762 | 48.998 | 43.559 | 45.231 |
| 2009    | 44.045 | 43.653 | 46.932 | 46.475 | 45.409 | 47.917 | 44.326 | 40.203 | 42.418 | 37.270 | 35.944 | 35.402 |
| 2010    | 34.348 | 34.731 | 39.384 | 39.835 | 38.974 | 42.536 | 35.969 | 40.271 | 39.235 | 34.603 | 34.334 | 35.034 |
| 2011    | 34.107 | 34.663 | 34.006 | 30.365 | 35.415 | 36.884 | 30.394 | 29.596 | 26.854 | 23.614 | 25.663 | 15.262 |

opção secundária para alívio imediato dos sintomas da asma por seus conhecidos efeitos colaterais, como o risco aumentado de náuseas, dor de cabeça, arritmias cardíacas e convulsões. Atualmente, essa medicação tem indicação restrita ao manejo de pacientes em estado de mal asmático ou doença refratária (DALCIN et al., 2000). Os dados completos sobre o número de medicamentos dispensados, a partir de fevereiro de 1999 a dezembro de 2011, estão disponíveis na Tabela 3. Além disso, os dados sobre temperatura e precipitação no município de Ribeirão Preto foram obtidos a partir do Centro Integrado de Informações Agrometeorológicas do Instituto Agrônomo (CIIAGRO).

O Comitê de Ética em Pesquisa da instituição dos pesquisadores responsáveis pelo projeto maior aprovou o estudo (CEP/CSE/FMRP/USP, protocolo número 453) e a permissão para acessar e usar os registros do Sistema de Informação Hygia Web foi fornecido pelo representante local do Departamento de Saúde (documento 248/11 GS RAS/ras).

### 3.2.2 Modelo estatístico

Sendo  $y_t$  o número de medicamentos dispensados no mês  $t$ ,  $t = 1, \dots, n$ , em que  $n$  é o número de meses na série temporal. O modelo proposto é dado de uma forma geral pela equação:

$$y_t = \alpha + g_1 I_{[1, \theta]}(t) + g_2 I_{(\theta, n]}(t) + S_t + \varepsilon_t \quad (3.1)$$

em que,

$$g_k = \beta_k(t - \theta) + \sum_{r=1}^R \psi_{kr}(x_{rt} - \bar{x}_r) + \sum_{j=1}^p \gamma_{kj}(y_{t-j} - \bar{y}) \quad (3.2)$$

com  $k = 1, 2$ .

Nas equações (3.1) e (3.2),  $\alpha$  é um intercepto, e  $I_{\{A\}}(t)$  denota uma função indicadora de tal forma que  $I_{\{A\}}(t) = 1$  se  $t \in \{A\}$ , e 0 caso contrário,  $\theta$  é o ponto de mudança

a ser estimado de tal modo que  $\theta$  é um número inteiro no intervalo  $[1, n]$ ,  $x_{1t}, x_{2t}, \dots, x_{Rt}$  são observações de  $R$  covariáveis no mês  $t$ ,  $\bar{x}_r$  denota a média de  $x_{r1}, x_{r2}, \dots, x_{rn}$ ,  $r = 1, \dots, R$ ,  $\psi_{1r}$  e  $\psi_{2r}$  são os efeitos da covariável  $x_{rt}$  sobre  $y_t$  antes e depois de  $\theta$ , respectivamente,  $\bar{y}$  indica a média de  $y_1, \dots, y_n$ , os termos  $\gamma_{11}, \dots, \gamma_{1p}, \gamma_{21}, \dots, \gamma_{2p}$  são parâmetros autorregressivos de ordem  $p$  a serem estimados e  $\varepsilon_t$  é o erro aleatório.

Além disso,

$$S_t = \eta_1 \sin\left(\frac{2\pi t}{12}\right) + \eta_2 \cos\left(\frac{2\pi t}{12}\right) \quad (3.3)$$

é uma função periódica mensal para estimar padrões sazonais, em que  $\eta_1$  e  $\eta_2$  são parâmetros que assumem números reais. Utilizando este modelo, assume-se que os termos  $\varepsilon_t, t = 1, \dots, n$  são independentes e seguem uma distribuição normal com média 0 e variância que depende do ponto de mudança  $\theta$ , ou melhor:

$$\varepsilon_t \sim N\left(0, \sigma_1^2 I_{[1, \theta]}(t) + \sigma_2^2 I_{(\theta, n]}(t)\right) \quad (3.4)$$

Assim,  $\sigma_1^2$  e  $\sigma_2^2$  são as variâncias de  $\varepsilon_t$  antes e depois do ponto de mudança, respectivamente. Esta formulação do modelo corresponde à seguinte função de verossimilhança:

$$f(y|\theta, \xi) = \prod_{t=1}^n (2\pi\lambda_t)^{-\frac{1}{2}} \exp\left[-\sum_{t=1}^n \frac{(y_t - \mu_t)^2}{2\lambda_t}\right] \quad (3.5)$$

em que

$$\lambda_t = \sigma_1^2 I_{[1, \theta]}(t) + \sigma_2^2 I_{(\theta, n]}(t), \quad (3.6)$$

$$\mu_t = \alpha + g_1 I_{[1, \theta]}(t) + g_2 I_{(\theta, n]}(t) + S_t, \quad (3.7)$$

tal que  $y = (y_1, y_2, \dots, y_n)$  e

$\xi = (\alpha, \beta_1, \beta_2, \gamma_{1,1}, \dots, \gamma_{1,p}, \gamma_{2,1}, \dots, \gamma_{2,p}, \psi_{1,1}, \dots, \psi_{1,R}, \psi_{2,1}, \dots, \psi_{2,R}, \sigma_1^2, \sigma_2^2, \eta_1, \eta_2)$  é o vetor dos parâmetros. Por definição,  $f(y|\theta, \xi)$  denota a função de densidade de probabilidade conjunta da amostra  $Y = (Y_1, Y_2, \dots, Y_n)$ . Na análise Bayesiana, presume-se que os parâmetros do vetor  $\xi$  e  $\theta$  seguem distribuições baseadas em conhecimento anterior (as distribuições *a priori*), a qual são atualizadas com os dados (representada por  $f(y|\theta, \xi)$ ) para produzir as distribuições *a posteriori*. Isto é formalizado pelo teorema de Bayes, dado por  $f(\theta, \xi|y) \propto f(y|\theta, \xi)p(\theta, \xi)$ , em que  $p(\theta, \xi)$  é a distribuição conjunta *a priori* e  $f(\theta, \xi|y)$  é a distribuição conjunta *a posteriori*. As distribuições *a priori* podem ser "não-informativas", com um pequeno efeito sobre a distribuição *a posteriori*. Assim, as seguintes distribuições *a priori* para os parâmetros do vetor  $\xi$  são consideradas:

- $\alpha \sim N(0, c_1)$ ,

- $\beta_1 \sim N(0, c_2), \beta_2 \sim N(0, c_3),$
- $\gamma_{1,j} \sim N(0, c_{4,j}), \gamma_{2,j} \sim N(0, c_{5,j}), j = 1, \dots, p,$
- $\eta_1 \sim N(0, c_6), \eta_2 \sim N(0, c_7),$
- $\psi_{1,r} \sim N(0, c_{8,r}), \psi_{2,r} \sim N(0, c_{9,r}), r = 1, \dots, R,$
- $\sigma_1^2 \sim IG(c_{10}, c_{11})$  e  $\sigma_2^2 \sim IG(c_{12}, c_{13}).$

em que  $c_1, \dots, c_{13}$  são valores conhecidos para os hiperparâmetros das distribuições *a priori*,  $N(0, c)$  indica uma distribuição normal com média 0 e variância  $c$ , e  $IG(h_1, h_2)$  denota uma distribuição Gama Inversa com média  $\frac{h_2}{h_1-1}$  e variância  $\frac{h_2^2}{[(h_1-1)^2(h_1-2)]}$ . Valores relativamente grandes de  $c_1, \dots, c_{13}$  produzem distribuições *a priori* não informativas para os respectivos parâmetros. Assumiu-se ainda independência entre estes parâmetros.

Além disso, é assumido uma distribuição uniforme discreta *a priori* para o ponto de mudança  $\theta$  de tal modo que as probabilidades prévias dos valores  $1, 2, \dots, n$  são assumidos como iguais a  $\frac{1}{n}$ .

Como alternativa, pode-se considerar que os termos  $\varepsilon_t$  seguem uma distribuição t-Student não padronizada com  $v$  graus de liberdade, um parâmetro de locação  $\mu_t$ , um parâmetro de escala  $\lambda_t$  e variância  $\lambda_t^2 v(v-2)^{-1}$  para  $v > 2$ . Neste caso, a formulação do modelo corresponde à seguinte função de verossimilhança:

$$f(y|\theta, v, \xi) = \prod_{t=1}^n \left\{ \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right) \lambda_t \sqrt{\pi v}} \left[ 1 + \frac{1}{v} \left( \frac{y_t - \mu_t}{\lambda_t} \right)^2 \right]^{-\frac{v}{2}} \right\} \quad (3.8)$$

em que  $\Gamma(\cdot)$  é a função gama, e  $\lambda_t$  e  $\mu_t$  são dados por (3.6) e (3.7) respectivamente. Para a análise Bayesiana, pode-se considerar as mesmas distribuições *a priori* assumidas para o modelo anterior e uma uniforme contínua *a priori* para  $v$ , ou seja,  $v \sim U(2, c_v)$ , onde  $c_v$  é um hiperparâmetro conhecido ( $c_v > 2$ ). Devido a problemas de programação no OpenBugs quando deixamos os graus de liberdade da distribuição t-Student aleatórios (LANGE et al., 1989), o OpenBugs tende a encontrar um valor para  $v$  próximo ao centro do intervalo estabelecido. Portanto para comparações serão ajustados outros modelos, fixando  $v$  em valores maiores que 2, para verificarmos se o modelo encontrado com  $v$  aleatório corresponde a um melhor ajuste.

O procedimento Bayesiano de simulação Monte Carlo em Cadeias de Markov (MCMC) com o algoritmo de amostragem de Gibbs (CASELLA; GEORGE, 1992), foi usado para estimar as distribuições *a posteriori* dos parâmetros de interesse e os componentes de variância (CARLIN; LOUIS, 1997). O algoritmo de amostragem de Gibbs foi executado para 510.000 iterações e amostradas em cada 10 simulações. Para eliminar os efeitos dos valores iniciais, as primeiras 10.000 iterações foram descartadas ("burn-in samples"). Deste modo,

50.000 amostras de Gibbs finais foram usadas na inferência. O intervalo de credibilidade a 95% (IC 95%) foi obtido a partir dos percentis 2,5% e 97,5% dos valores simulados para as distribuições a posteriori dos parâmetros de interesse. O intervalo de credibilidade de 95% bayesiano é análogo ao tradicional intervalo de confiança 95%, expressando 95% do intervalo central de valores que são credíveis para o respectivo parâmetro.

Habituais métodos de diagnóstico existentes foram empregados para verificar a convergência dos cálculos MCMC (CARLIN; LOUIS, 1997). Depois do ajuste do modelo, a suposição de independência entre os sucessivos termos dos erros aleatórios  $\varepsilon_t$  foram verificados graficamente traçando as respectivas autocorrelações e funções de autocorrelações parciais em relação a diferentes defasagens. A estimação bayesiana dos parâmetros foi realizada utilizando o algoritmo MCMC implementado no programa livre OpenBUGS (LUNN et al., 2000). O código OpenBUGS utilizado para esta análise é dado no Apêndice.

### 3.2.3 Especificações dos Modelos

Na ausência de covariáveis, foram ajustados três modelos diferentes para os dados, denotados como os Modelos 1 a 3, tal como descrito a seguir:

- Modelo 1: Neste modelo, os termos autorregressivos  $\gamma_{11}, \dots, \gamma_{1p}, \gamma_{21}, \dots, \gamma_{2p}$  foram descartados, e conseqüentemente,  $g_k = \beta_k(t - \theta)$ , para  $k=1,2$ . Este modelo não considera a presença de covariáveis e a função periódica mensal  $S_t$ ;
- Modelo 2: Este modelo é similar ao Modelo 1, mas considera os termos autorregressivos  $\gamma_{11}, \dots, \gamma_{1p}, \gamma_{21}, \dots, \gamma_{2p}$  de ordem 1;
- Modelo 3: Similar ao Modelo 2, mas considera a função mensal periódica  $S_t$ .

Os modelos 1 a 3 assumem que os resíduos  $\varepsilon_t$  seguem uma distribuição normal ou distribuição t-Student. Além disso, os modelos 4 e 5 são definidos como se segue:

- Modelo 4: Semelhante ao Modelo 3, mas inclui uma variável independente. As seguintes variáveis foram consideradas: a temperatura média mensal ( $^{\circ}\text{C}$ ), temperatura máxima e mínima mensal ( $^{\circ}\text{C}$ ) e precipitação média mensal (mm) em uma escala logarítmica. De acordo com esta formulação, quatro modelos diferentes foram ajustadas aos dados, um para cada variável independente, evitando assim problemas de colinearidade entre as variáveis;
- Modelo 5: Este modelo possui termos autorregressivos  $\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}$  de ordem 2, com ponto de mudança e considera a função mensal periódica  $S_t$ , com a premissa dos resíduos seguirem uma distribuição normal.



### 3.2.4 Seleção do Modelo

O critério de informação deviance (DIC) (SPIEGELHALTER et al., 2002) é amplamente utilizado para a comparação de modelo Bayesiano. No entanto, o modelo proposto é interpretado pelo OpenBUGS como um modelo de mistura, e este programa não é capaz de calcular o valor do DIC nesta situação. Um outro critério para a seleção do modelo é derivado das ordenadas preditivas condicionais (CPO) (GELFAND et al., 1992). Para a  $i$ -ésima observação, a  $CPO_i$  é dada por:

$$f(\mathcal{D}_i|y_{[i]}) = \int f(\mathcal{D}_i|\Theta)f(\Theta|\mathcal{D}_{[i]})d\Theta \quad (3.9)$$

em que  $\Theta$  é o vetor completo de parâmetros,  $\mathcal{D}_i$  é o conjunto de dados completos  $\mathcal{D}$ ,  $\mathcal{D}_{[i]}$  é  $\mathcal{D}$  sem a observação corrente  $i$  e  $f(\Theta|\mathcal{D}_i)$  é a densidade a posteriori de  $\Theta$  dado  $\mathcal{D}_i, i = 1, \dots, n$ . Assim, as estatísticas de CPO expressam a probabilidade a posteriori de se observar o valor ou um conjunto de valores de  $\mathcal{D}_i$  quando o modelo é ajustado para todos os dados exceto  $\mathcal{D}_i$ . Um valor muito baixo para  $CPO_i$  sugere que  $\mathcal{D}_i$  é um valor atípico ("outlier") e uma observação influente. Uma aproximação de  $CPO_i$  obtida pelo método MCMC (CHEN et al., 2012) é dada por:

$$\widehat{CPO}_i = \left[ \frac{1}{B} \sum_{b=1}^B \frac{1}{f(\mathcal{D}_i|\Theta_b)} \right]^{-1}, \quad (3.10)$$

em que  $B$  é o número de iterações durante a implementação do procedimento MCMC após o período de depuração e  $\Theta_b$  é o vetor das amostras obtidas na  $b$ -ésima iteração. Assim, as estatísticas aproximadas CPO podem ser calculadas diretamente no OpenBUGS pela definição das amostras simuladas para  $f(\Theta|\mathcal{D}_i)^{-1}$ . Assumindo normalidade aproximada, os valores inversos para  $\widehat{CPO}_i$  maiores que 40 podem ser consideradas como possíveis outliers e maior do que 70 como valores extremos, conforme Ntzoufras (2011). A pseudo log verossimilhança marginal (LPML - *log pseudo marginal likelihood*) é uma medida Bayesiana de ajuste ou adequação definida com base nas estatísticas da CPO (GEISSER; EDDY, 1979). Para um determinado modelo, o valor LPML é estimado por  $\widehat{LPML} = \sum_{i=1}^n \log \widehat{CPO}_i$ .

Quanto maior for o valor do LPML, melhor é o ajuste do modelo. É equivalente ao cálculo do pseudofator de Bayes (PBF - *pseudo Bayes factor*), que compara os modelos  $m$  e  $m'$  dado por:

$$PBF_{mm'} = \exp(\widehat{LPML}_m - \widehat{LPML}_{m'}).$$

Além disso, a discrepância entre os dados e um modelo de estimativa pode ser medida pela soma dos quadrados dos resíduos (SSR) dados por:

$$SSR = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \hat{\mu}_t)^2, \quad (3.11)$$

onde  $\hat{\mu}_t$  é obtido por meio da substituição dos parâmetros em (3.7) pelas suas respectivas estimativas. Para ajuste de modelos diferentes para um dado conjunto de dados, um valor menor da SSR indica um melhor ajuste para os dados.

### 3.3 Resultados

#### 3.3.1 Modelos com termos autorregressivos de ordem 1

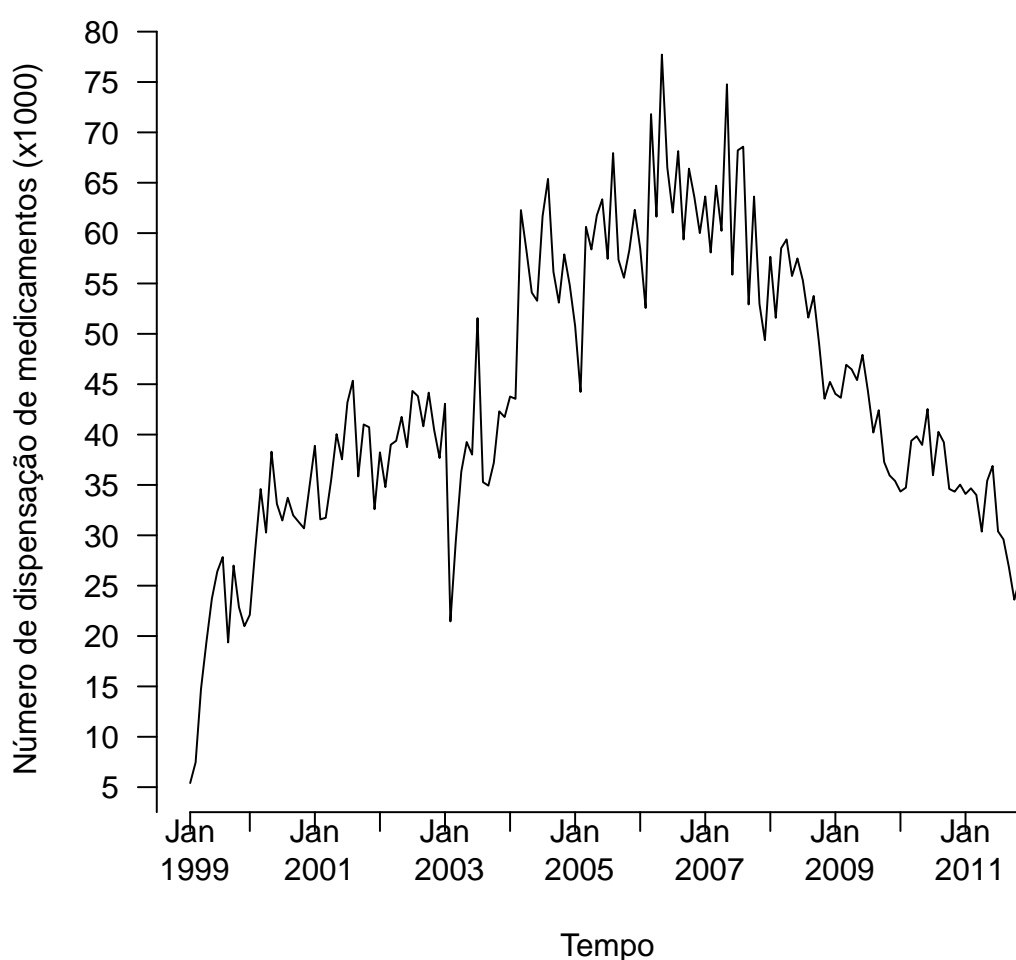


Figura 6 – Número mensal de medicamentos dispensados (comprimido de aminofilina, 100 mg) pelos serviços públicos de saúde de Ribeirão Preto, Brasil, de fevereiro de 1999 a dezembro de 2011.

O gráfico na Figura 6 mostra uma série temporal do número de dispensações mensais do aminofilina a partir de fevereiro de 1999 a dezembro de 2011. Este gráfico evidencia a presença de um ponto de mudança para a série temporal. Observa-se um aumento no número de dispensações da medicação no início do período considerado e

Tabela 4 – Os resultados dos modelos estatísticos de mudança de ponto - resíduos normais.

| Parâmetro      | Modelo 1             |                                | Modelo 2             |                                | Modelo 3             |                                |
|----------------|----------------------|--------------------------------|----------------------|--------------------------------|----------------------|--------------------------------|
|                | Estimativa Bayesiana | Intervalo de credibilidade 95% | Estimativa Bayesiana | Intervalo de credibilidade 95% | Estimativa Bayesiana | Intervalo de credibilidade 95% |
| $\alpha$       | 65,93                | (64,12 ; 67,74)                | 55,45                | (50,18 ; 60,04)                | 58,64                | (53,65 ; 63,52)                |
| $\beta_1$      | 0,5031               | (0,4511 ; 0,5566)              | 0,2176               | (0,1087 ; 0,3267)              | 0,2782               | (0,1747 ; 0,3893)              |
| $\beta_2$      | -0,6477              | (-0,7185 ; -0,5821)            | -0,4945              | (-0,647 ; -0,331)              | -0,6214              | (-0,766 ; -0,4688)             |
| $\theta$       | 92,6                 | (89 ; 96)                      | 102,9                | (93 ; 109)                     | 102,8                | (98 ; 107)                     |
| $\gamma_{1,1}$ | -                    | -                              | 0,5084               | (0,3298 ; 0,692)               | 0,3965               | (0,2113 ; 0,5793)              |
| $\gamma_{2,1}$ | -                    | -                              | 0,2462               | (-0,0051 ; 0,5247)             | 0,02651              | (-0,201 ; 0,2684)              |
| $\sigma_1^2$   | 45,04                | (33,55 ; 60,38)                | 37,86                | (27,93 ; 51,09)                | 35,25                | (26,02 ; 47,34)                |
| $\sigma_2^2$   | 20,78                | (14,51 ; 29,74)                | 15,38                | (8,939 ; 24,97)                | 10,8                 | (6,398 ; 17,58)                |
| $\eta_1$       | -                    | -                              | -                    | -                              | 1,563                | (0,5025 ; 2,593)               |
| $\eta_2$       | -                    | -                              | -                    | -                              | -2,515               | (-3,732 ; -1,32)               |
| LPML           | -493,8               |                                | -484,1               |                                | -473,2               |                                |
| SSR            | 5212,4               |                                | 4403,8               |                                | 3861,8               |                                |

Tabela 5 – Os resultados dos modelos estatísticos com mudança de ponto - resíduos t-Student.

| Parâmetro                | Modelo 1             |                                | Modelo 2             |                                | Modelo 3             |                                |
|--------------------------|----------------------|--------------------------------|----------------------|--------------------------------|----------------------|--------------------------------|
|                          | Estimativa Bayesiana | Intervalo de credibilidade 95% | Estimativa Bayesiana | Intervalo de credibilidade 95% | Estimativa Bayesiana | Intervalo de credibilidade 95% |
| $\alpha$                 | 65,7                 | (63,88 ; 67,52)                | 54,65                | (49,67 ; 59,18)                | 58,02                | (53,43 ; 62,75)                |
| $\beta_1$                | 0,4981               | (0,447 ; 0,5513)               | 0,2032               | (0,1005 ; 0,3094)              | 0,2649               | (0,1708 ; 0,3711)              |
| $\beta_2$                | -0,6358              | (-0,7071 ; -0,5709)            | -0,4628              | (-0,6186 ; -0,3023)            | -0,6014              | (-0,7487 ; -0,4491)            |
| $\theta$                 | 92,25                | (89,0 ; 96,0)                  | 102,9                | (93 ; 110)                     | 102,9                | (97 ; 106)                     |
| $v(df)$                  | 23,27                | (4,757 ; 48,29)                | 15,73                | (3,401 ; 46,14)                | 20,38                | (4,338 ; 47,68)                |
| $\gamma_{1,1}$           | -                    | -                              | 0,5256               | (0,3522 ; 0,6984)              | 0,4203               | (0,2382 ; 0,591)               |
| $\gamma_{2,1}$           | -                    | -                              | 0,2856               | (-0,02728 ; 0,5507)            | 0,04623              | (-0,182 ; 0,2843)              |
| $\sigma_1^2$             | 38,95                | (25,28 ; 54,92)                | 28,62                | (16,63 ; 43,0)                 | 28,67                | (17,82 ; 41,25)                |
| $\sigma_2^2$             | 18,21                | (11,37 ; 27,18)                | 12,87                | (6,508 ; 22,17)                | 9,602                | (5,424 ; 16,03)                |
| $\sigma_1^2 v(v-2)^{-1}$ | 45,18                | (32,37 ; 63,36)                | 37,36                | (25,57 ; 55,36)                | 34,2                 | (24,3 ; 47,99)                 |
| $\sigma_2^2 v(v-2)^{-1}$ | 21,16                | (14,13 ; 31,81)                | 17,01                | (8,97 ; 30,67)                 | 11,55                | (6,623 ; 19,76)                |
| $\eta_1$                 | -                    | -                              | -                    | -                              | 1,491                | (0,4625 ; 2,495)               |
| $\eta_2$                 | -                    | -                              | -                    | -                              | -2,397               | (-3,586 ; -1,232)              |
| LPML                     | -691,70              |                                | -671,34              |                                | -640,9               |                                |
| SSR                      | 5222,3               |                                | 4447,5               |                                | 3882,7               |                                |

uma grande redução em fevereiro de 2003, provavelmente devido a um período em que a droga estava faltando, seguido de um aumento do número de dispensas até meados de maio de 2006. Após este período, observou-se uma redução do número de dispensações do medicamento, dado que ele foi progressivamente substituído por outros, considerando seus conhecidos efeitos adversos.

Na análise Bayesiana, foram consideradas distribuições a priori não informativas para todos os parâmetros do modelo. Deste modo, considerou-se  $c_1 = c_2 = c_3 = c_{4,j} = c_{5,j} = c_6 = c_7 = c_{8,r} = c_{9,r} = 10^6$ ,  $j = 1, \dots, p$ ,  $r = 1, \dots, R$ , nas distribuições a priori para  $\alpha, \beta_1, \beta_2, \gamma_{1,j}, \gamma_{2,j}, \psi_{1,r}$  e  $\psi_{2,r}$ , e  $c_{10} = c_{11} = c_{12} = c_{13} = 0, 1$  nas distribuições a priori para  $\sigma_1^2$  e  $\sigma_2^2$ . No caso do modelo com erros que seguem uma distribuição t-Student, é considerado  $c_v = 50$ , ou seja,  $v \sim U(2, 50)$  e fixamos  $v$  igual a 3, 5, 15, 20, 25, 30 para comparações entre os ajustes dos modelos. O número de medicamentos dispensados mensais foi dividido por 1000, a fim de facilitar a convergência do algoritmo computacional.

As Tabelas 4 e 5 mostram os resultados para os Modelos 1 a 3 obtidos pelo programa OpenBUGS. Os resultados da Tabela 4 considera que os resíduos dos modelos seguem uma distribuição normal, enquanto que a Tabela 5 mostra os resultados dos modelos com

resíduos que seguem uma distribuição t-Student com  $v$  graus de liberdade.

No caso dos Modelos 2 e 3, foram ajustados considerando termos autorregressivos de ordem 1 para todos os casos. A Tabela 4 mostra que os resultados do Modelo 3 tem o menor valor SSR e o valor mais alto LPML, sugerindo que ele proporciona o melhor ajuste para os dados. Em adição, o valor PBF comparando os Modelos 3 e 2 com resíduos que seguem uma distribuição normal é 5476 (final da seção 2.2.4). Em todos os modelos ajustados, as estimativas para  $\beta_1$  são positivas e as estimativas para  $\beta_2$  são negativas, descrevendo que o número de medicamentos dispensados está aumentando ao longo do tempo até o ponto de mudança  $\theta$ , depois começa a diminuir a partir desse valor. O intervalo de credibilidade 95% para  $\gamma_{1,1}$  não contém valor zero, evidenciando a importância do parâmetro autorregressivo de ordem  $p = 1$ . Os resultados do Modelo 3 também mostram que o intervalo de credibilidade de 95% para  $\eta_1$  e  $\eta_2$  não contém o valor zero, o que sugere a evidência de um padrão sazonal anual na série.

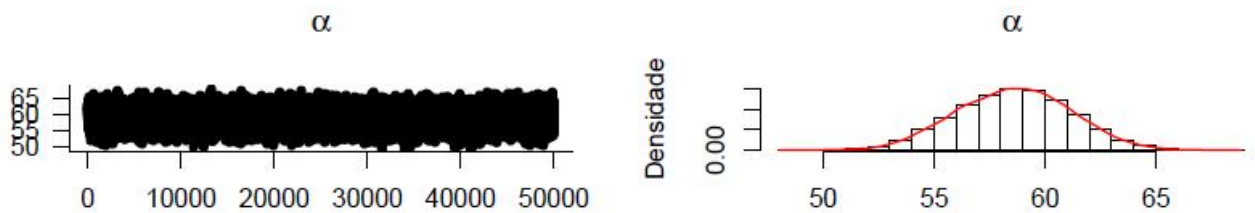


Figura 7 – Descrição da convergência e histogramas das distribuições do parâmetro  $\alpha$  do Modelo 3 com resíduos que seguem uma distribuição normal.

Os resultados da Tabela 5 indicam que as estimativas obtidas a partir dos modelos com resíduos que seguem distribuição t-Student estão próximas às encontradas quando se considera uma distribuição normal (Tabela 4). A partir da equação (3.6), as variâncias de  $\varepsilon_t$  antes e depois do ponto de mudança são dadas por  $\sigma_1^2 v(v-2)^{-1}$  e  $\sigma_2^2 v(v-2)^{-1}$ , respectivamente. Estimadores para estas quantidades também são apresentados na Tabela 5, podendo notar que eles são muito semelhantes aos de  $\sigma_1^2$  e  $\sigma_2^2$  obtidos a partir do ajuste dos modelos baseados na distribuição normal (Tabela 4).

Os gráficos das Figuras 7 a 13 descrevem a convergência dos parâmetros do Modelo 3, considerando os modelos com resíduos que seguem distribuição normal e distribuição t-Student. Na Figura 10 podemos comparar a distribuição a posteriori para o valor de  $\theta$  obtido do modelo normal com aquele obtido do modelo t-Student, em gráficos com a mesma escala. As distribuições estão muito próximas, o que já era esperado por suas médias serem  $\theta_{normal} = 102,8$  e  $\theta_{t-student} = 102,9$  e com o intervalo de credibilidade semelhantes (Tabelas 4 e 5).

As amostras de Gibbs para os parâmetros  $\alpha$ ,  $\beta_1$ ,  $\beta_2$ ,  $\gamma_1$ ,  $\gamma_2$ ,  $\eta_1$  e  $\eta_2$  e descritas nas Figuras 7 a 9 do modelo em que os resíduos seguem uma distribuição normal, têm suas

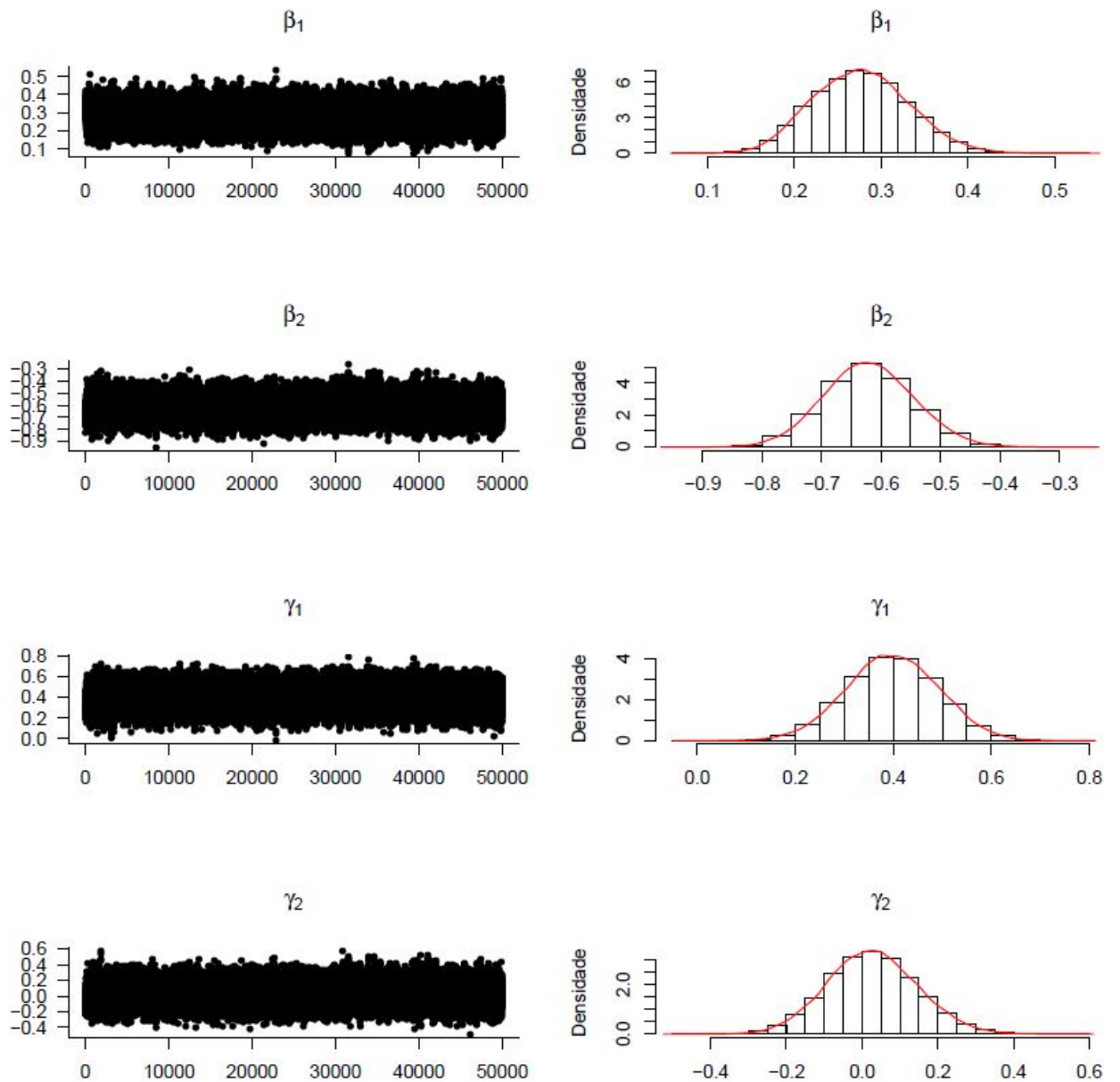


Figura 8 – Descrição da convergência e histogramas das distribuições dos parâmetros do Modelo 3 com resíduos que seguem uma distribuição normal:  $\beta_1$ ,  $\beta_2$ ,  $\gamma_1$  e  $\gamma_2$ .

distribuições simétricas como mostrado em seus respectivos histogramas. Por sua vez, os parâmetros  $\sigma_1^2$  e  $\sigma_2^2$ , apresentam assimetria à direita (Figura 9).

No modelo com resíduos que seguem uma distribuição t-Student, as amostras simuladas para os parâmetros  $\beta_1$ ,  $\beta_2$ ,  $\gamma_1$ ,  $\gamma_2$ ,  $\alpha$ ,  $\eta_1$ ,  $\eta_2$  descritas nas Figuras 11 a 13, têm suas distribuições simétricas como mostrado em seus histogramas. Já os parâmetros  $v(df)$ ,  $\sigma_2^2$ ,  $\sigma_1^2 v(v-2)^{-1}$  e  $\sigma_2^2 v(v-2)^{-1}$  apresentam uma assimetria à direita, enquanto  $\sigma_1^2$  aparenta ter uma distribuição mais simétrica (Figuras 12 e 13).

O gráfico representado na Figura 14 mostra o número observado de medicamentos

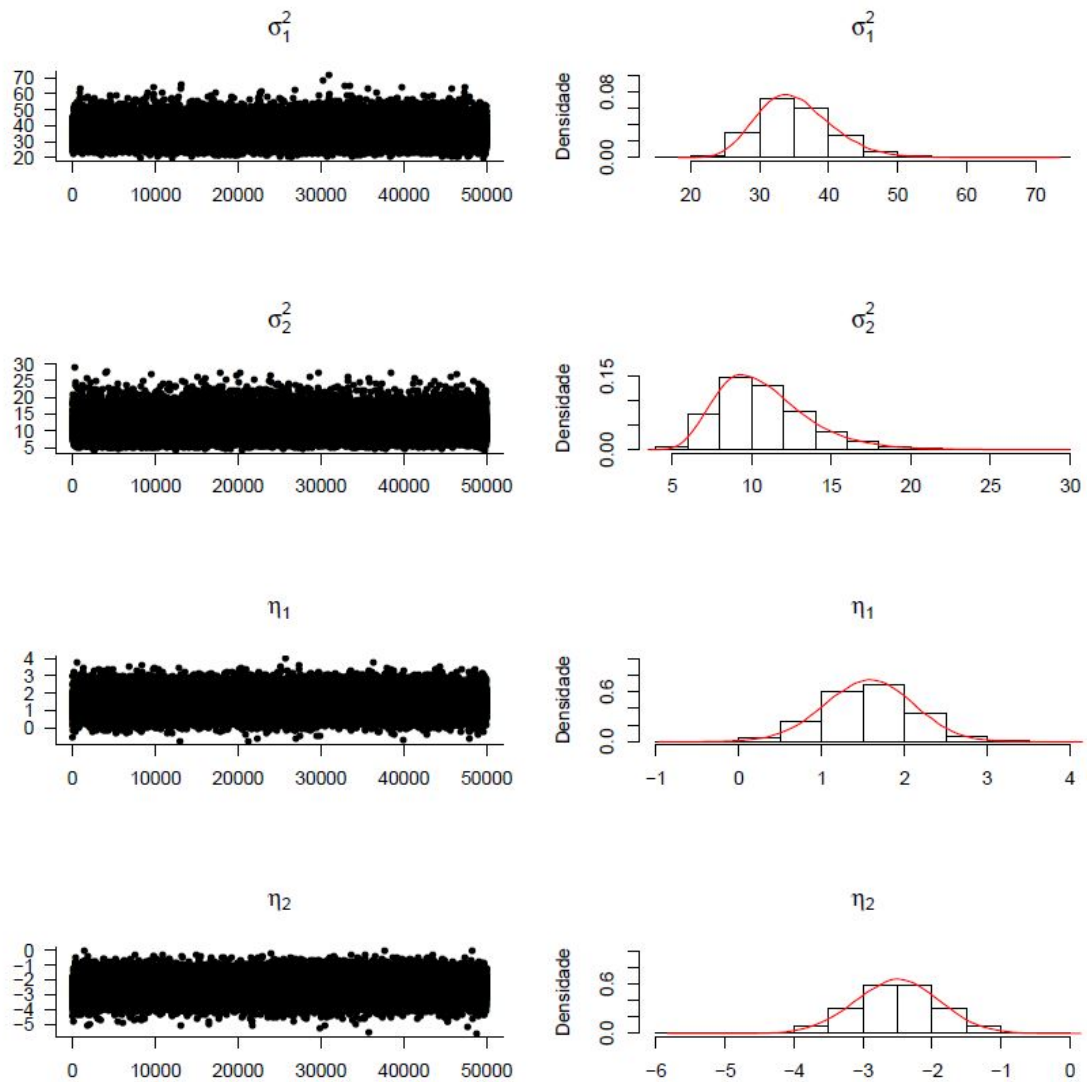


Figura 9 – Descrição da convergência e histogramas das distribuições dos parâmetros do Modelo 3 com resíduos que seguem uma distribuição normal:  $\sigma_1^2$ ,  $\sigma_2^2$ ,  $\eta_1$  e  $\eta_2$ .

dispensados e os valores previstos obtidos a partir do Modelo 1. Considerando os resultados do Modelo 1 com resíduos que seguem uma distribuição normal, os valores preditos linearmente aumentam até o ponto de mudança  $\theta$  estimado por  $t = 92,6$  (Tabela 4), correspondente ao mês de outubro de 2006, com um intervalo de credibilidade de 95% variando de junho de 2006 a janeiro de 2007.

Após este ponto de mudança, os valores preditos diminuiriam de forma linear como é verificado pelo coeficiente de  $\beta_2$  estimado por  $-0,6477$ . Considerando o Modelo 1 com resíduos que seguem uma distribuição t-Student, o ponto de mudança  $\theta$  é estimado por  $t = 92,25$  (Tabela 5), correspondente ao mês de setembro de 2006, com um intervalo de credibilidade de 95% de junho de 2006 a janeiro de 2007. No entanto, gráficos de autocorrelações parciais (não mostrados) para os resíduos do Modelo 1 mostram evidências de correlação serial significativa entre os valores sucessivos do  $\varepsilon_t$ , ou seja, a suposição de

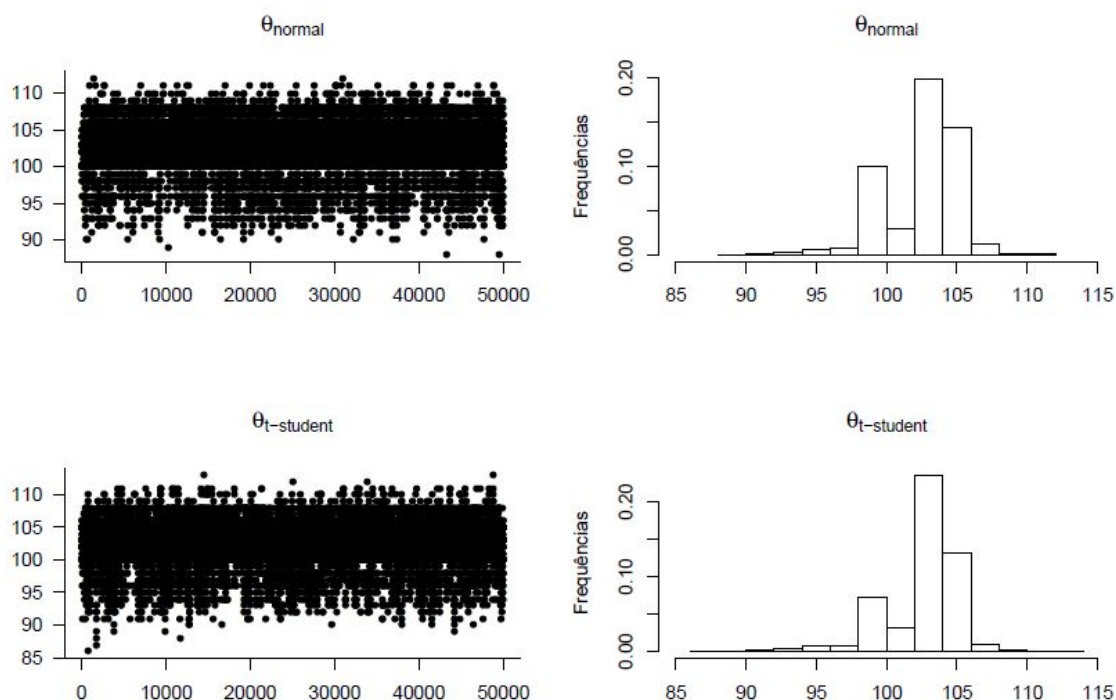


Figura 10 – Descrição da convergência e histogramas das distribuições dos parâmetros  $\theta$  do Modelo 3 com resíduos que seguem uma distribuição normal, mais descrição da convergência e histograma da distribuição do parâmetro  $\theta$  do Modelo 3 com resíduos que seguem uma distribuição t-Student.

independência entre os resíduos não foi atingida.

Portanto, Modelo 1 é útil para descrever a tendência linear da série de tempo antes e depois do ponto de mudança (Figura 14), mas inferências para os parâmetros do modelo podem ser prejudicadas pela falta de independência entre os resíduos. As estimativas dos parâmetros ajustados segundo os Modelos 2 e 3, considerando termos autorregressivos de ordem 1, são apresentados nas Tabelas 4 e 5 e visualizadas nas Figuras 15 e 16.

Considerando o ajuste do Modelo 2 com resíduos que seguem uma distribuição normal, o ponto de mudança é estimado em  $t = 102,9$  (Tabela 4), correspondente ao mês de agosto de 2007, com um intervalo de credibilidade de 95% variando de outubro de 2006 a fevereiro de 2008.

Os Modelos 2 e 3 apresentaram uma estimativa para a variância  $\sigma_1^2$  maior do que a estimativa para  $\sigma_2^2$ , sugerindo uma maior dispersão do número de medicamentos dispensados antes do ponto de mudança.

O ajuste do Modelo 3 representado na Figura 16 com resíduos que seguem uma distribuição normal, apresenta ponto de mudança estimado em  $t = 102,8$  (Tabela 4), correspondente ao mês de agosto de 2007, com um intervalo de credibilidade de 95% de março de 2007 a dezembro de 2007, sete meses menor que o do Modelo 2 com resíduos que seguem uma distribuição normal.



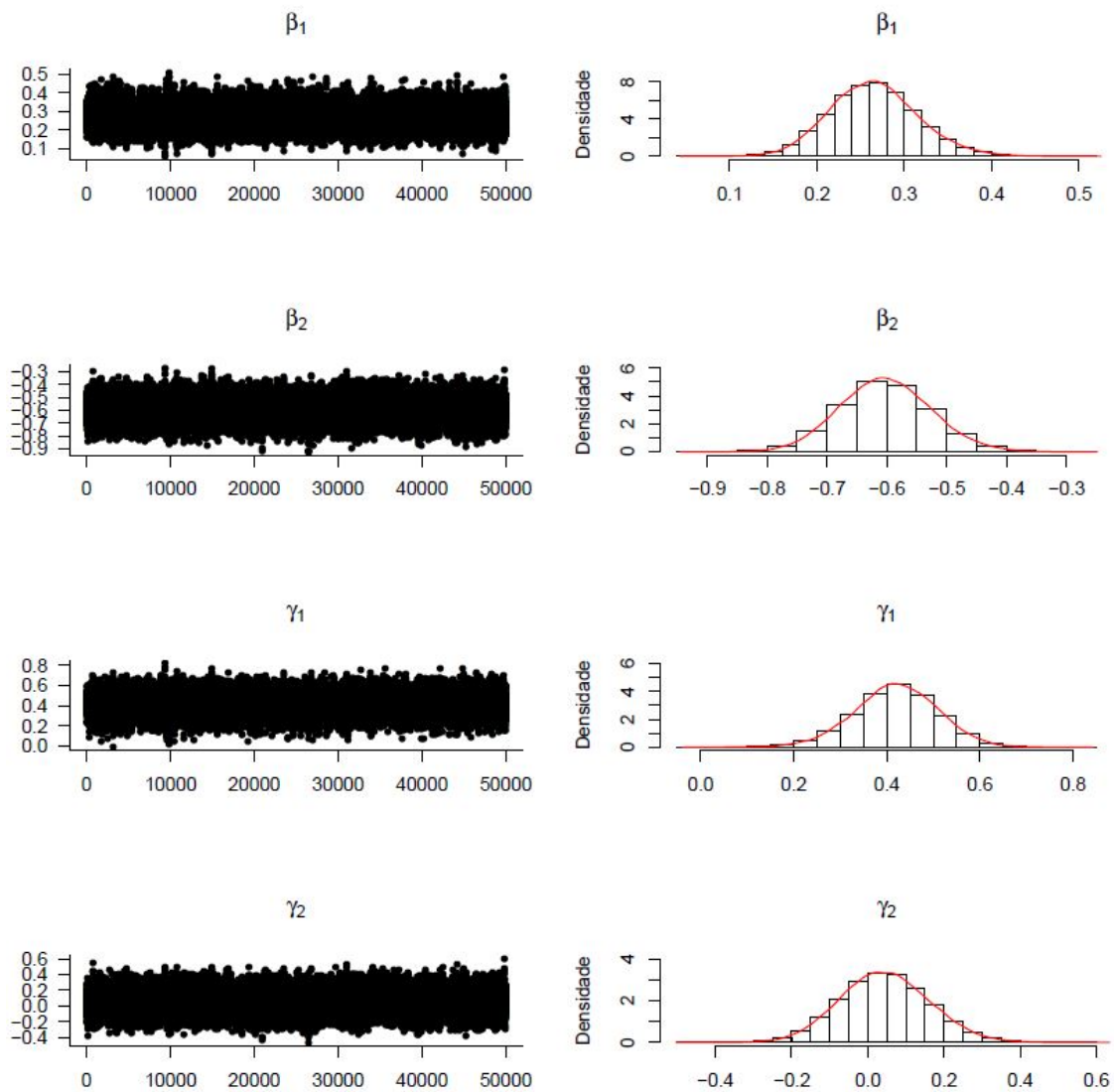


Figura 11 – Descrição da convergência e histogramas das distribuições dos parâmetros do Modelo 3 com resíduos que seguem uma distribuição t-Student:  $\beta_1$ ,  $\beta_2$ ,  $\gamma_1$  e  $\gamma_2$ .

O Modelo 3 com resíduos que seguem uma distribuição t-Student foi ajustado além dos resultados que são encontrados na Tabela 5 fixando os valores para  $\nu$  em 3, 5, 15, 20, 25, 30 tendo como resposta para LPML os valores respectivos  $-757, 57$ ;  $-660, 73$ ;  $-636, 847$ ;  $-636, 297$ ;  $-636, 05$ ;  $-635, 87$  e para SSR 3924, 61; 3909, 10; 3885, 27; 3874, 05; 3871, 47; 3868, 60. Notamos que o Modelo 3 em que consideramos uma distribuição uniforme contínua *a priori* para  $\nu$ , ou seja,  $\nu \sim U(2, c_\nu)$  (resultados na Tabela 5), não teve diferença significativa nos parâmetros estimados para os modelos com  $\nu$  fixos.

Observamos que a soma dos quadrados dos resíduos foi diminuindo à medida que aumentamos o valor de  $\nu$ . Isso já era esperado, pois quando se aumenta o  $\nu$  os resíduos vão se aproximar de uma distribuição normal, portanto a soma dos quadrados dos resíduos diminuem ficando próximo do modelo que ofereceu o melhor ajuste (Modelo 3 com resíduos que seguem uma distribuição normal). A comparação com a LPML não se mostrou



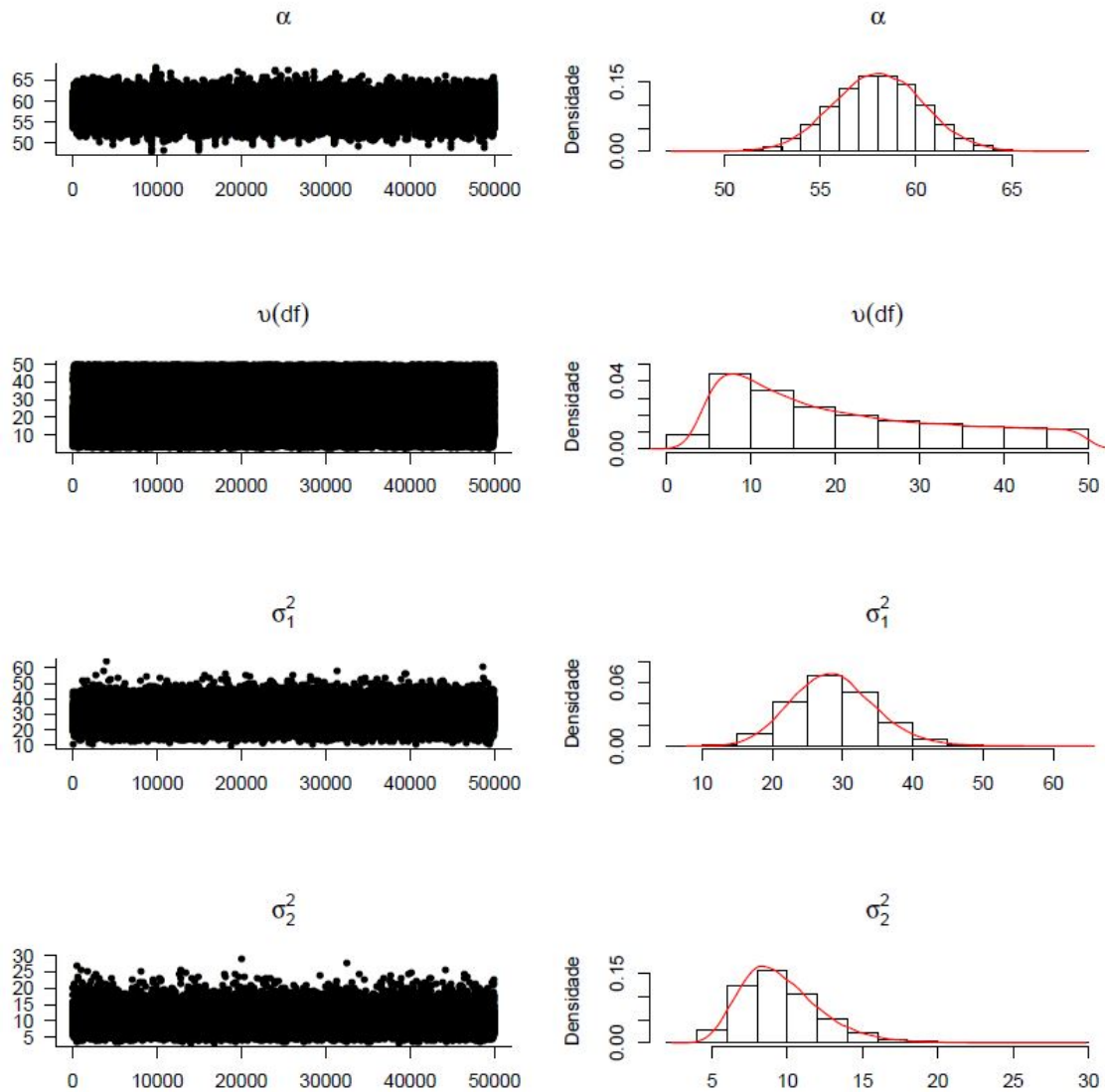


Figura 12 – Descrição da convergência e histogramas das distribuições dos parâmetros do Modelo 3 com resíduos que seguem uma distribuição t-Student:  $\alpha$ ,  $\nu(df)$ ,  $\sigma_1^2$ ,  $\sigma_2^2$ .

significante, pois o modelo com  $\nu$  aleatório apresentou valor igual a  $-640,90$  e a sequência dos valores obtidos de LPML que considerou  $\nu$  com valores fixos já tinha encontrado um "platô" de aproximadamente  $-635$ , ou seja, poderíamos aumentar muito o valor do  $\nu$  que não teríamos um aumento considerável na LPML. Por essas evidências o Modelo 3 com  $\nu \sim U(2, c_\nu)$ , dentre os modelos que os resíduos seguem uma distribuição t-Student, foi o modelo que melhor se ajustou aos dados. Porém, ele não se mostra maior que os valores obtidos do Modelo 3 com resíduos que seguem uma distribuição normal (comparar Tabelas 4 e 5).

Um exame da sequência de valores inversos da CPO (Figura 17) pode identificar possíveis valores extremos no modelo ajustado e permitir comparações entre os modelos, considerando cada ponto de observação.

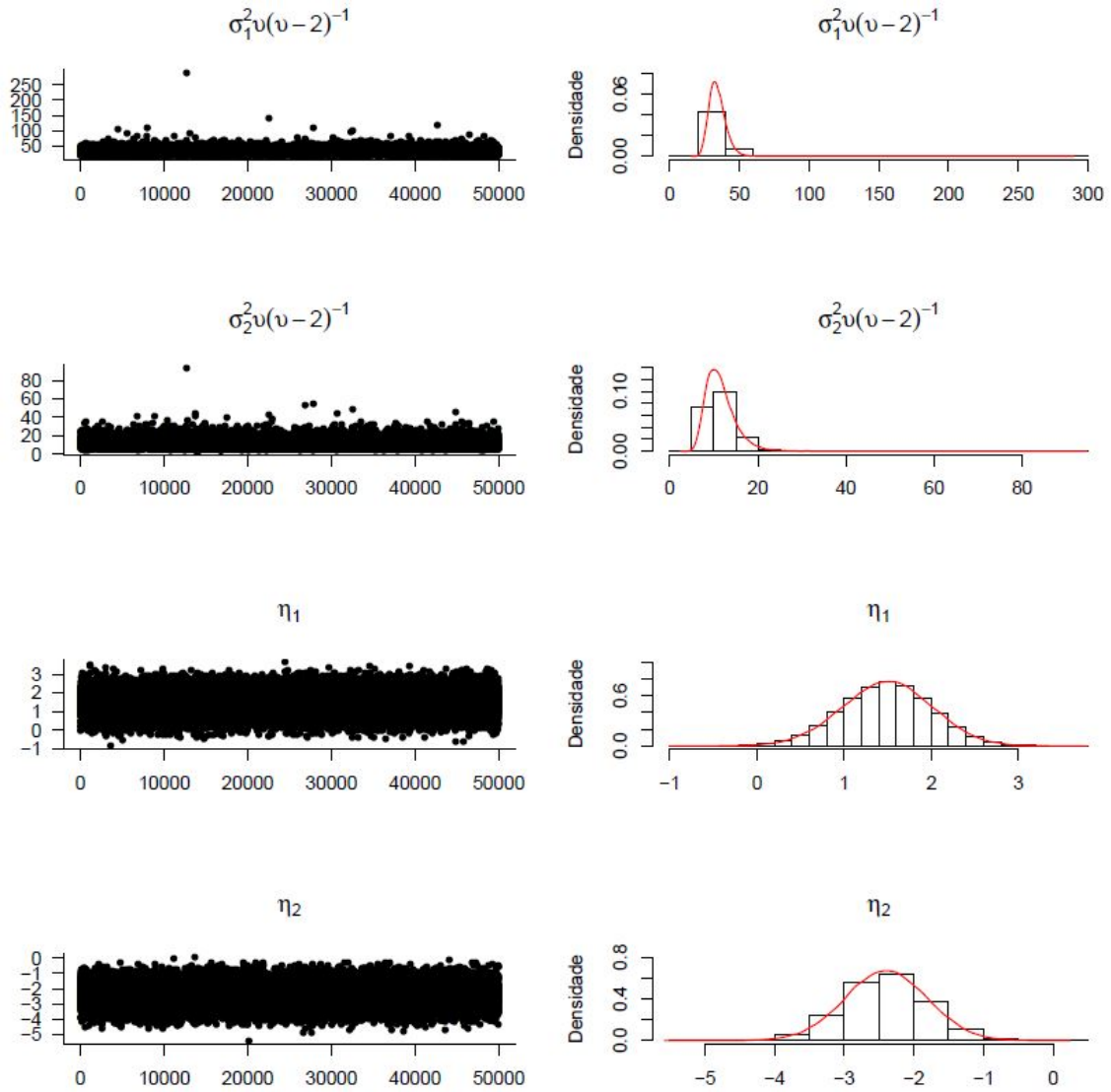


Figura 13 – Descrição da convergência e histogramas das distribuições dos parâmetros do Modelo 3 com resíduos que seguem uma distribuição t-Student:  $\sigma_1^2 v(v-2)^{-1}$ ,  $\sigma_2^2 v(v-2)^{-1}$ ,  $\eta_1$  e  $\eta_2$ .

Considerando os resultados do Modelo 3, o gráfico da Figura 17 apresenta os valores inversos para CPO, onde as duas linhas tracejadas horizontais na figura passam por meio dos valores de 40 e 70, com o objetivo de identificar possíveis *outliers* e valores extremos, respectivamente (ver subseção 2.2.4). Este gráfico compara os valores inversos para CPO obtidos a partir do Modelo 3 com resíduos de uma distribuição normal e distribuição t-Student com  $v$  graus de liberdade.

Podemos notar que o número de valores extremos é maior quando se considera o modelo com resíduos que assumem distribuição t-Student, reforçando que o modelo com erros provenientes de uma distribuição normal é o modelo que melhor se ajusta aos dados (Figura 17).

As Tabelas 6 e 7 mostram os resultados de modelos de regressão (Modelo 4)

Tabela 6 – Resultados dos modelos de regressão que consideram a temperatura máxima e mínima mensal absoluta como variáveis independentes.

| Parâmetro      | Temperatura máxima absoluta<br>(°C) |                                | Temperatura mínima absoluta<br>(°C) |                                |
|----------------|-------------------------------------|--------------------------------|-------------------------------------|--------------------------------|
|                | Estimativa Bayesiana                | Intervalo de credibilidade 95% | Estimativa Bayesiana                | Intervalo de credibilidade 95% |
| $\alpha$       | 58,92                               | (53,76 ; 63,87)                | 58,33                               | (53,53 ; 63,27)                |
| $\beta_1$      | 0,2847                              | (0,1768 ; 0,3981)              | 0,2695                              | (0,1714 ; 0,3806)              |
| $\beta_2$      | -0,6255                             | (-0,7718 ; -0,4713)            | -0,6248                             | (-0,7741 ; -0,47)              |
| $\theta$       | 102,5                               | (97 ; 106)                     | 103,4                               | (99 ; 107)                     |
| $\gamma_{1,1}$ | 0,3885                              | (0,2015 ; 0,5759)              | 0,4109                              | (0,2236 ; 0,5894)              |
| $\gamma_{2,1}$ | 0,02175                             | (-0,2068 ; 0,2653)             | 0,02721                             | (-0,2079 ; 0,2723)             |
| $\sigma_1^2$   | 35,22                               | (25,84 ; 47,52)                | 35,93                               | (26,58 ; 48,11)                |
| $\sigma_2^2$   | 11,35                               | (6,583 ; 18,64)                | 10,4                                | (6,215 ; 17,19)                |
| $\eta_1$       | 1,387                               | (-0,08801 ; 2,799)             | 1,566                               | (0,4995 ; 2,589)               |
| $\eta_2$       | -2,438                              | (-3,772 ; -1,109)              | -1,787                              | (-4,024 ; 0,381)               |
| $\psi_1$       | -0,1034                             | (-0,742 ; 0,5323)              | -0,04794                            | (-0,4267 ; 0,333)              |
| $\psi_2$       | -0,06951                            | (-0,6551 ; 0,5018)             | -0,1673                             | (-0,517 ; 0,1923)              |
| LPML           |                                     | -475,6                         |                                     | -474,6                         |
| SSR            |                                     | 3820                           |                                     | 3885,6                         |

Tabela 7 – Resultados dos modelos que consideram a temperatura média mensal e precipitação como variáveis independentes.

| Parâmetro      | Temperatura média mensal<br>(°C) |                                | Precipitação<br>(log mm) |                                |
|----------------|----------------------------------|--------------------------------|--------------------------|--------------------------------|
|                | Estimativa Bayesiana             | Intervalo de credibilidade 95% | Estimativa Bayesiana     | Intervalo de credibilidade 95% |
| $\alpha$       | 58,52                            | (53,67 ; 63,53)                | 59,14                    | (53,97 ; 64,14)                |
| $\beta_1$      | 0,2741                           | (0,1743 ; 0,3891)              | 0,2888                   | (0,1802 ; 0,4017)              |
| $\beta_2$      | -0,6274                          | (-0,7748 ; -0,4741)            | -0,6326                  | (-0,7785 ; -0,4783)            |
| $\theta$       | 103,1                            | (98 ; 106)                     | 102,5                    | (98 ; 106)                     |
| $\gamma_{1,1}$ | 0,4036                           | (0,2121 ; 0,5837)              | 0,381                    | (0,1903 ; 0,5714)              |
| $\gamma_{2,1}$ | 0,02009                          | (-0,2111 ; 0,2619)             | 0,007956                 | (-0,2228 ; 0,2529)             |
| $\sigma_1^2$   | 35,47                            | (26,21 ; 47,55)                | 35,2                     | (25,91 ; 47,37)                |
| $\sigma_2^2$   | 10,7                             | (6,307 ; 17,64)                | 11,13                    | (6,51 ; 18,06)                 |
| $\eta_1$       | 1,293                            | (0,1076 ; 2,438)               | 1,539                    | (0,4631 ; 2,581)               |
| $\eta_2$       | -1,666                           | (-3,663 ; 0,3239)              | -2,19                    | (-3,613 ; -0,768)              |
| $\psi_1$       | -0,205                           | (-0,9775 ; 0,5674)             | -0,08521                 | (-0,6223 ; 0,4457)             |
| $\psi_2$       | -0,4258                          | (-1,139 ; 0,2802)              | -0,2251                  | (-0,6852 ; 0,2353)             |
| LPML           |                                  | -474,8                         |                          | -474,8                         |
| SSR            |                                  | 3839,4                         |                          | 3816,5                         |

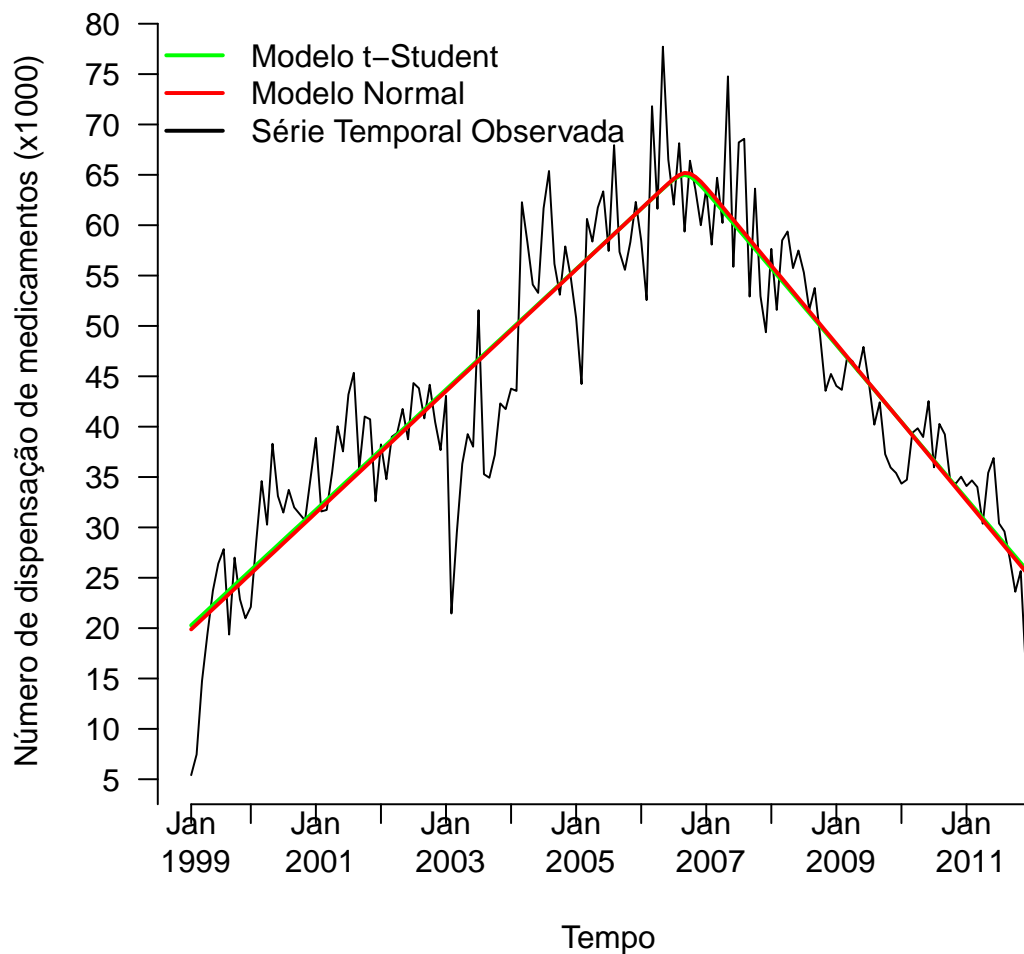


Figura 14 – Comparação entre a série temporal observada e os valores estimados do Modelo 1, com resíduos que seguem uma distribuição normal e distribuição t-Student.

que consideram a temperatura máxima mensal e a mínima mensal em valor absoluto, a temperatura média mensal e a precipitação média mensal como variáveis independentes. Estes modelos supõem que os resíduos seguem uma distribuição normal. Em todos os modelos que consideram estas variáveis independentes, podemos observar que o intervalo de credibilidade 95% para os parâmetros  $\psi_1$  e  $\psi_2$  incluem o valor zero. Isto implica que não temos evidências de que estas variáveis climáticas estão associados com o número mensal de medicamentos dispensados. Estes resultados são observados mesmo em modelos similares que não incluem a componente sazonal  $St$  (resultados não mostrados).

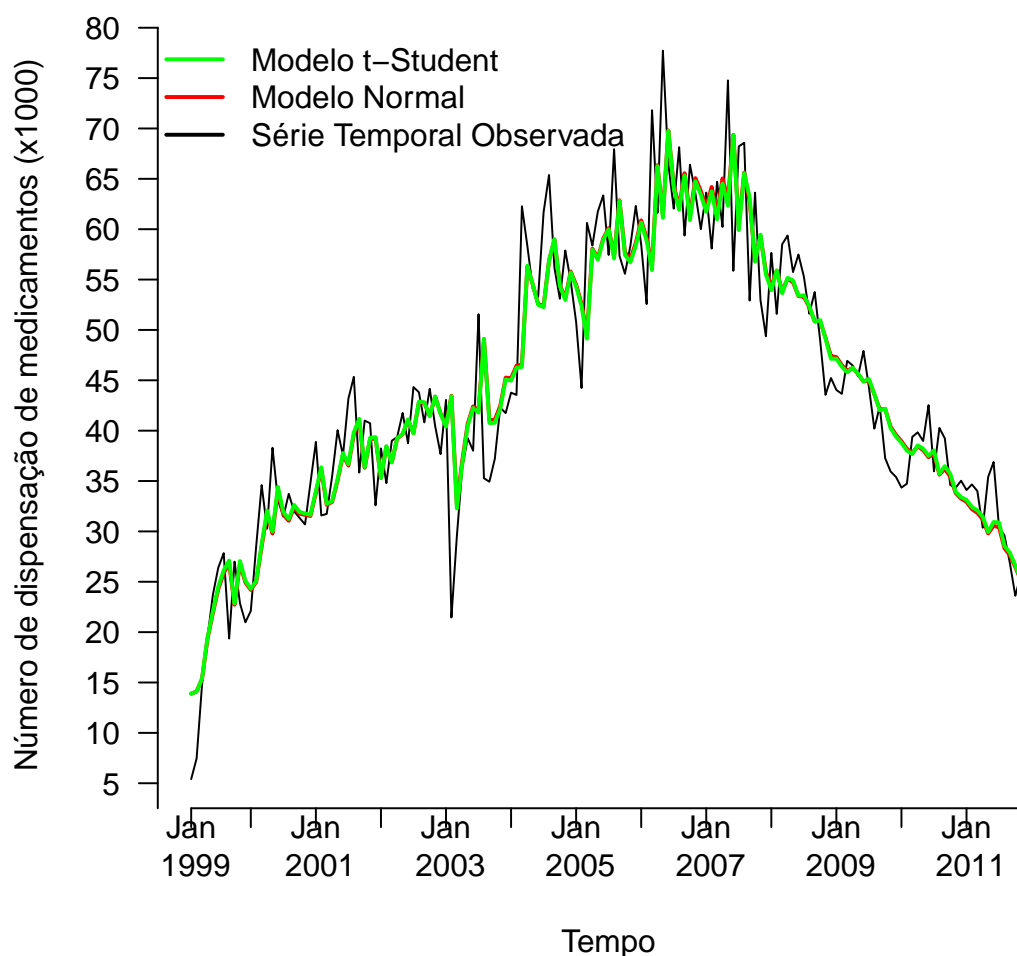


Figura 15 – Comparação entre a série temporal observada e os valores estimados do Modelo 2, com resíduos que seguem uma distribuição normal e distribuição t-Student.

### 3.3.2 Modelo com termos autorregressivos de ordem 2

O Modelo 3 com resíduos que seguem uma distribuição normal, até então o melhor modelo encontrado, mostrou-se deficiente na análise dos gráficos de autocorrelação para os resíduos (Figura 18). Observou-se evidência de correlação serial significativa entre os valores sucessivos de  $\varepsilon_t$ , ou seja, a suposição de independência entre os resíduos pode não ter sido atingida indicando um mal ajuste.

Tentamos encontrar um novo modelo (Modelo 5), onde os gráficos de auto correlação ACF e PACF fossem melhor ajustados para independência dos resíduos e mantivesse um bom valor para LPML e SSR. Então ajustamos um modelo com ponto de mudança e resíduos que seguem uma distribuição normal, porém com termos autorregressivos de ordem 2 e com a componente  $St$ . Os gráficos de autocorrelação (Figura 19) para os resíduos do

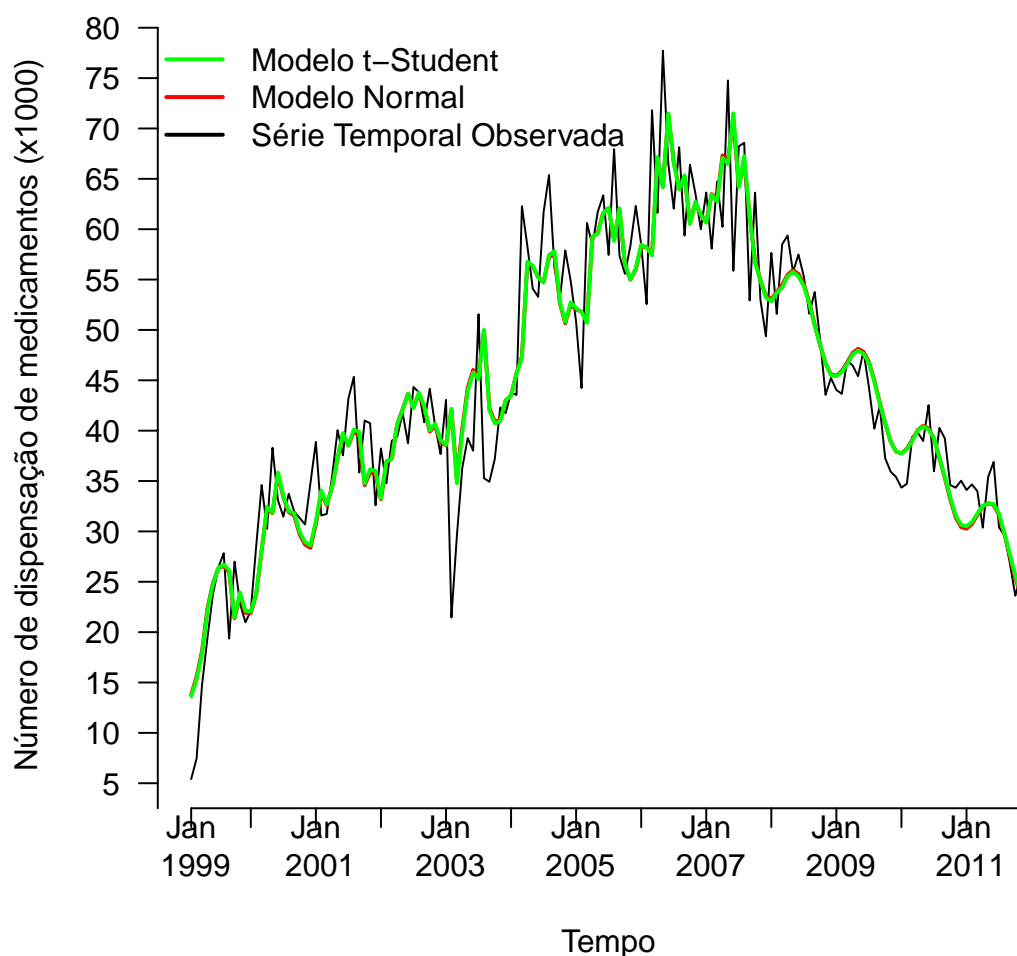


Figura 16 – Comparação entre a série temporal observada e os valores estimados do Modelo 3, com resíduos que seguem uma distribuição normal e distribuição t-Student.

Modelo 5, se mostraram melhor ajustados que do Modelo 3, e não evidenciou-se correlação serial significativa entre os valores sucessivos de  $\varepsilon_t$ , indicando um melhor ajuste do modelo aos dados.

A LPML do Modelo 5 apresentou o valor maior dentre todos os modelos,  $-466,2$  (Tabela 8). O valor do PBF comparando o Modelo 5 e o Modelo 3 (com resíduos que seguem uma distribuição normal) foi  $1096,6$ , também com menor  $SSR = 3618,3$  (Tabela 8). Esses indicadores nos leva a eleger o Modelo 5, como aquele de maior evidência de ajustar melhor a série temporal estudada (mais parcimonioso).

Os gráficos de autocorrelações e resíduos do Modelo 5 podem ser comparados com os do Modelo 3, analisando as Figuras 18 e 19. Os dados do Modelo 5 estão melhores comportados na faixa de independência entre os resíduos (Figuras 19a e 19b). A estrutura

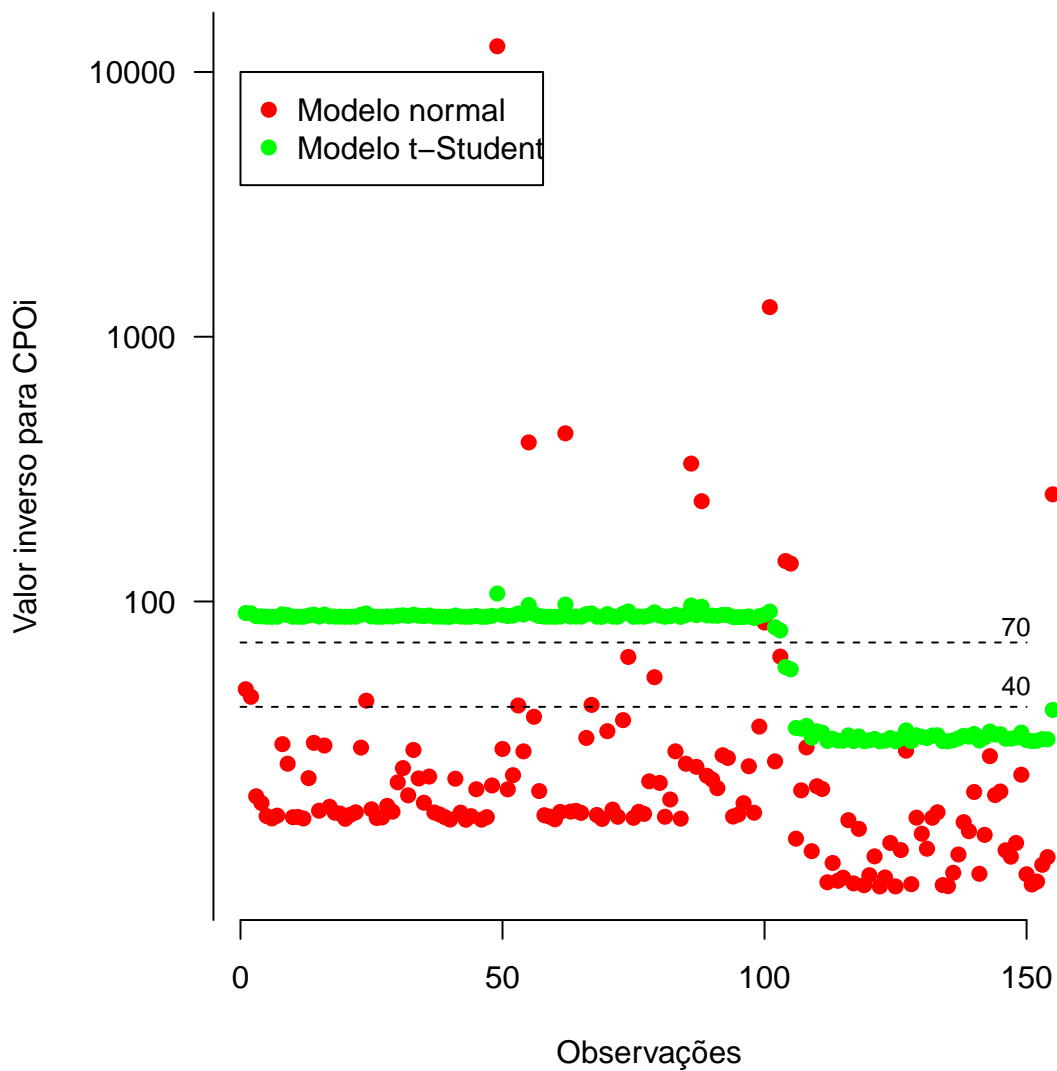


Figura 17 – Comparação entre os valores inversos para CPO obtidos do Modelo 3 com resíduos que seguem uma distribuição normal e outra que segue uma distribuição t-Student.

do Modelo 5 com termos autorregressivos de ordem 2 parece acomodar melhor os resíduos (Figura 19d), apesar que os intervalos de credibilidade 95% para as estimativas de  $\gamma_{12}$  e  $\gamma_{22}$  conter o valor zero (Tabela 6).

Quando plotamos os dados do Modelo 5 com a série temporal estudada, temos a aparência de um bom ajuste (Figura 20), com ponto de mudança estimado para  $t = 105.7$  (Tabela 8), correspondente ao mês de novembro de 2007, com um intervalo de credibilidade de 95% de maio de 2007 a abril de 2008.

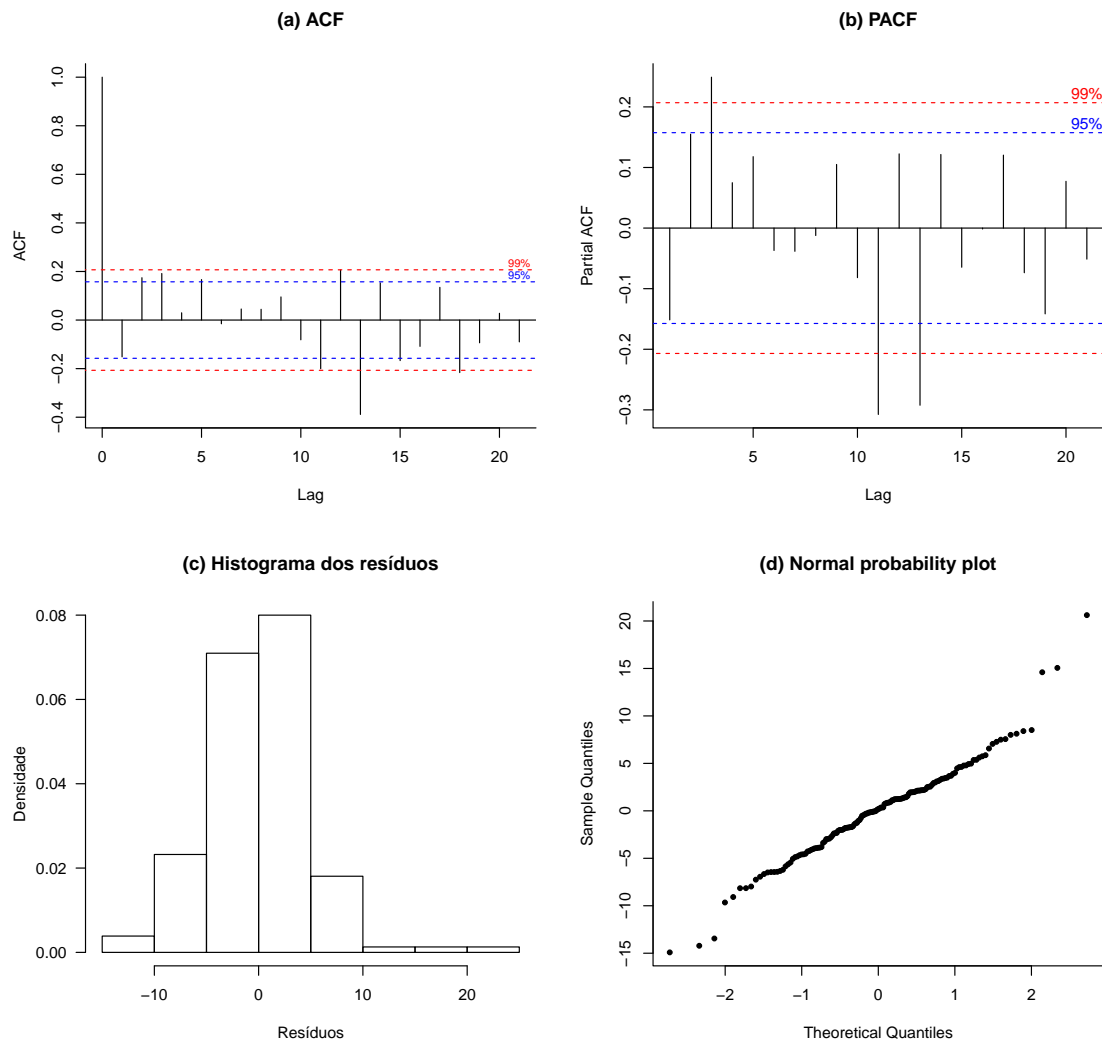


Figura 18 – Gráficos de autocorrelações, histograma e resíduos do Modelo 3 com ponto de mudança e termos autorregressivos de ordem 1, com uma componente sazonal  $S_t$  e resíduos que seguem uma distribuição normal.

### 3.4 Discussão

Os gráficos de ajuste dos modelos descrevem como o comportamento da série histórica do número de dispensações mensais dos comprimidos de aminofilina de 100 mg, de 1999 a 2011 é interpretado em termos da presença de um ponto de mudança. A Política Nacional Brasileira de Medicamentos introduziu em 30 de outubro de 1998 novas diretrizes para a assistência farmacêutica na saúde pública, definindo entre outras coisas, a lista de medicamentos essenciais, considerando os problemas de saúde mais comuns relatados na população. Posteriormente, e com o processo de descentralização da distribuição de drogas para os Estados e Municípios (Portaria GM 176 de 8 de Março de 1999), foi possível ampliar a oferta de medicamentos na rede pública de saúde (BOTEGA et al., 2007), o que explica a aumento do número de dispensas da medicação no início do ano de 1999 (Figura 6). Flutuações no número mensal de medicamentos dispensados são observadas na Figura



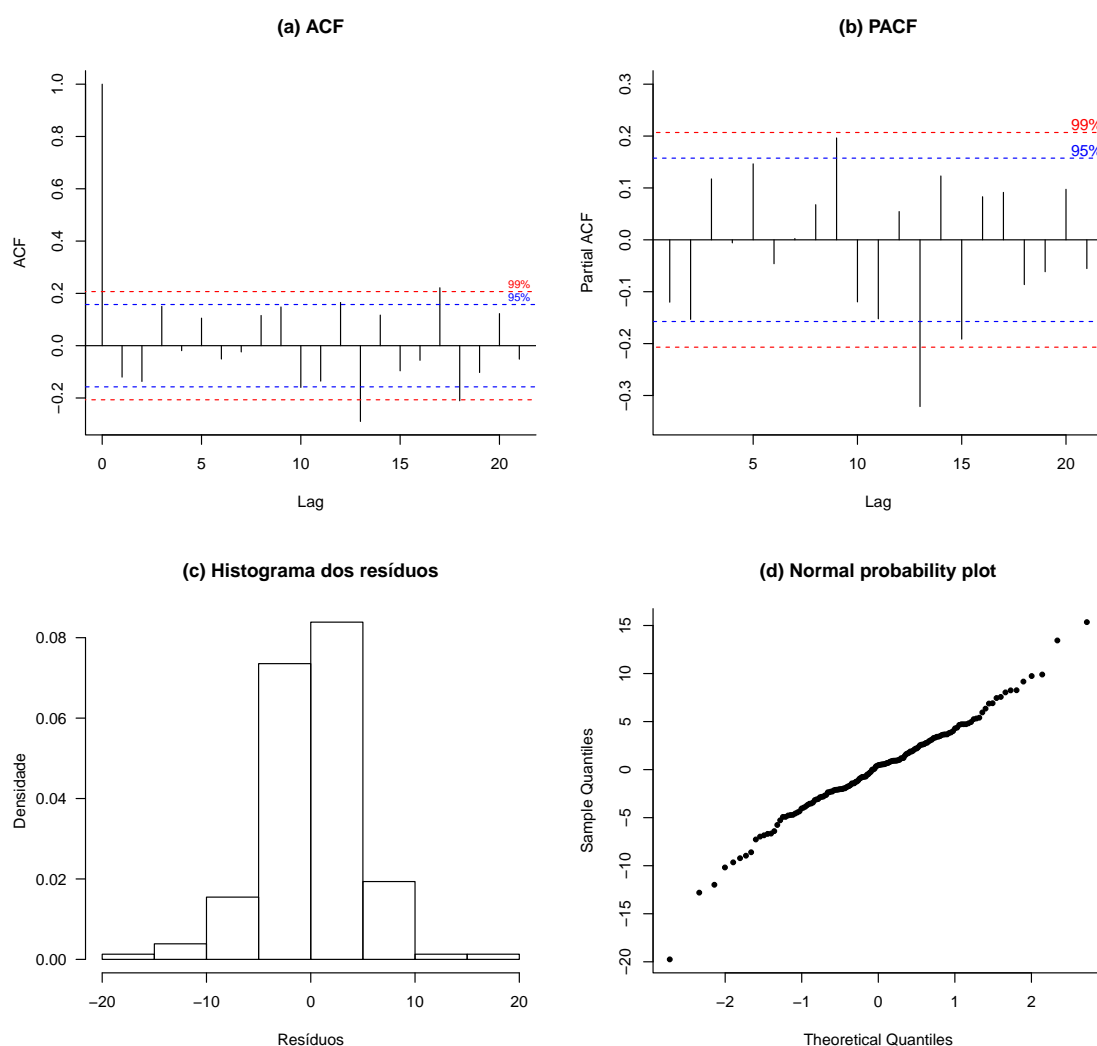


Figura 19 – Gráficos de autocorrelações, histograma e resíduos do Modelo 5 com ponto de mudança e termos autorregressivos de ordem 2, com uma componente sazonal  $S_t$  e resíduos que seguem uma distribuição t-student.

6, com alguma sazonalidade. Levanta-se a hipótese que o máximo de pontos locais nas séries temporais são coincidentes com os períodos mais frios e secos, quando as infecções das vias aéreas e episódios de broncoespasmo ocorrem com mais frequência, criando uma maior demanda para o uso do medicamento em períodos específicos do ano (THOMAZELLI et al., 2007; PETERSON et al., 2012). No entanto, os modelos de regressão utilizados neste estudo não mostram uma associação significativa entre variáveis climáticas e o número de medicamentos dispensados (Tabela 6 e 7).

A Figura 6 mostra uma grande redução na dispensação de aminofilina em fevereiro e março de 2003, sugerindo um período de descontinuidade no fornecimento do medicamento. Apesar dos bons resultados da política de descentralização da assistência farmacêutica, o fornecimento de medicamentos essenciais e medicamentos para algumas situações especiais (medicamentos que fazem parte do componente especializado da assistência

Tabela 8 – Os resultados do Modelo 5 com ponto de mudança bayesiano com termos autorregressivos de ordem 2 e resíduos que seguem uma distribuição normal.

| Parâmetro     | Estimativa Bayesiana | Intervalo de credibilidade 95% |
|---------------|----------------------|--------------------------------|
| $\alpha$      | 52,35                | (47,35 ; 57,85)                |
| $\beta_1$     | 0,1486               | (0,05126 ; 0,267)              |
| $\beta_2$     | -0,427               | (-0,6082 ; -0,2443)            |
| $\theta$      | 105,7                | (100 ; 111)                    |
| $\eta_1$      | 2,171                | (1,173 ; 3,167)                |
| $\eta_2$      | -1,761               | (-2,931 ; -0,6003)             |
| $\gamma_{11}$ | 0,2787               | (0,09527 ; 0,4609)             |
| $\gamma_{12}$ | 0,09419              | (-0,2007 ; 0,3985)             |
| $\gamma_{21}$ | 0,3647               | (0,1694 ; 0,5477)              |
| $\gamma_{22}$ | 0,233                | (-0,05044 ; 0,5112)            |
| $\sigma_1$    | 33,06                | (24,81 ; 43,7)                 |
| $\sigma_2$    | 8,618                | (5,218 ; 14,33)                |
| LPML          |                      | -466,2                         |
| SSR           |                      | 3618,3                         |

farmacêutica) dependia em grande parte das importações do Sistema Nacional de Saúde (BOTEGA et al., 2007), o que levou a indisponibilidade do medicamento para a população. A redução em março de 2003, foi seguida por um aumento do número de dispensações, observada até o mês de maio de 2007. Em 2004, como parte da Política Nacional de Atenção Integral à Saúde das Pessoas com Doenças Respiratórias, os municípios com serviços de saúde de atenção primária começaram a receber do Ministério da Saúde a beclometasona 250 mcg spray oral, beclometasona 50 mcg spray nasal e salbutamol 100 mcg spray para o tratamento de asma e rinite alérgica, considerando a necessidade de tratamento das duas doenças, uma vez que asma e rinite alérgica muitas vezes coexistem no mesmo indivíduo e o controle de uma dessas doenças favorece o controle de outra, contribuindo assim para a implementação das melhores práticas em saúde na asma. Naquele momento, os comprimidos de aminofilina 100 mg anteriormente fornecidos pelo Sistema Nacional de Saúde já não eram fornecidos e foi substituído pelo spray de salbutamol (BOTEGA et al., 2007). De 2007 até o final de 2011, as Figuras 14,15,16 e 20 mostram uma redução no número de dispensas de aminofilina caracterizando um ponto de mudança que provavelmente reflete a melhoria da assistência prestada ao paciente com asma, quando os novos medicamentos foram introduzidos.

O modelo escolhido como o mais parcimonioso entre os testados fica evidente ao se observar os resultados. O Modelo 5, devido a seus termos autorregressivos de ordem 2, possuir uma estrutura que acomoda melhor os resíduos (Figura 19d) e mostra o melhor desempenho no resultado da LPML (Tabela 8), além da menor soma dos quadrados dos resíduos. As estimativas para  $\beta_1$  são positivas e as estimativas para  $\beta_2$  são negativas, descrevendo que o número de medicamentos dispensados está aumentando ao longo do

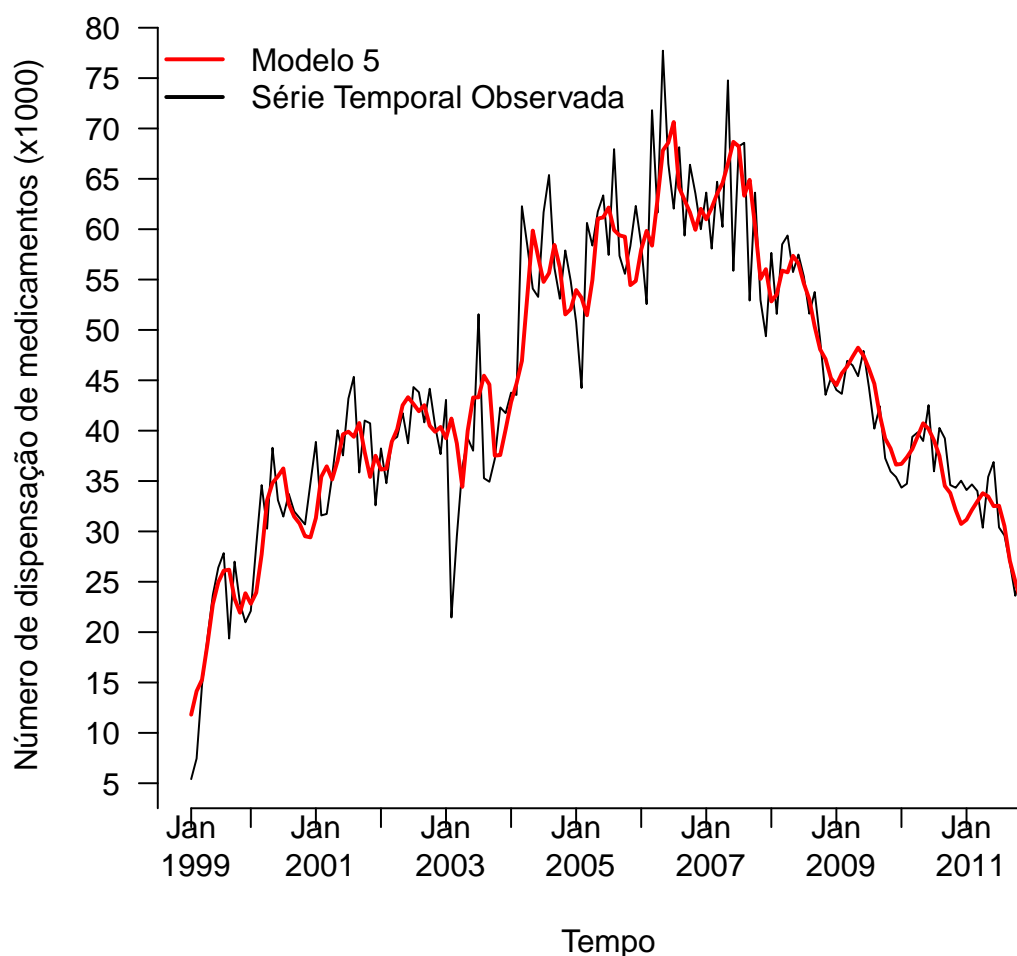


Figura 20 – Comparação entre a série temporal observada e o ajuste obtido do Modelo 5 com ponto de mudança e termos autorregressivos de ordem 2, com uma componente sazonal  $S_t$  e resíduos que seguem uma distribuição t-student.

tempo até o ponto de mudança  $\theta$  (Tabela 8). Depois, começa a diminuir a partir desse valor, o que vem a corroborar com o acerto em modelar o banco de dados com um ponto de mudança. Os intervalos de credibilidade 95% para  $\gamma_{11}$  e  $\gamma_{21}$  não contêm valor zero, evidenciando a importância de se incluir no modelo o parâmetro autorregressivo de ordem  $p = 1$  (Tabela 8). Os resultados do Modelo 5 mostram que o intervalo de credibilidade 95% para  $\eta_1$  e  $\eta_2$  não contém o valor zero, o que sugere a evidência de um padrão sazonal anual na série, ou seja, o componente  $S_t$  é importante para o modelo.

Um dos objetivos do presente estudo foi verificar entre os cinco modelos definidos qual deles ajusta melhor o banco de dados de série temporal de dispensação de aminofilina 100 mg com um ponto de mudança sob enfoque bayesiano. Em estudos futuros podemos testar outros modelos com termos auto regressivos de ordem 2 com condicional da premissa

dos resíduos seguirem uma distribuição t-Student ou ainda usar um modelo bivariado para ajustar essa série temporal de dispensa de aminofilina com outra série temporal de dispensa de outro remédio de asma no mesmo período.

### 3.5 Conclusão

Um modelo estatístico para as séries temporais epidemiológicas com componente sazonal de ordem 1, termos autorregressivos de ordem 1 e 2, covariáveis e um ponto de mudança foi descrita neste capítulo. Observou-se que o modelo proposto foi bem ajustado aos dados do número mensal de medicamentos para asma dispensados pelos serviços públicos de saúde de Ribeirão Preto, Brasil, sendo utilizado para a implementação do modelo um programa livre o OpenBUGS (ver apêndice). O modelo pode ser usado para descrever uma grande variedade de séries de dados epidemiológicos, onde um ponto de mudança está presente. Além de ser flexível, uma vantagem do modelo proposto sob uma abordagem bayesiana é a possibilidade de incluir informações de distribuições anteriores que estão à disposição dos gestores de saúde pública.

---

## Capítulo 4

---

# Modelo com mais de um ponto de mudança

---

### 4.1 Introdução

Dados de séries temporais são originados a partir de estudos em que são reportadas taxas de mortalidade, de hospitalizações, de infecções por alguma doença ou outro evento de interesse, em períodos definidos (dia, semana, mês ou ano), por algum tempo, objetivando-se observar tendências ou sazonalidades ([SHUMWAY; STOFFER, 2006](#)).

A aplicação de modelos de séries temporais com pontos de mudança tem se tornado cada vez mais presente em estudos epidemiológicos, mais especificamente, nos estudos em que existe uma intervenção ou acontecimento natural em um dado instante e, após isto, o comportamento da variável de interesse muda, com relação à tendência anterior. Assim, um ponto de mudança é definido como um momento, no tempo, em que os parâmetros (média, variância) do modelo ou uma tendência, que são utilizados para descrever uma série temporal, mudam abruptamente ([BEAULIEU et al., 2012](#)).

Como exemplo, [Francisco et al. \(2005\)](#) analisaram dados de mortalidade por doenças respiratórias, entre o período de 1980 a 2000 de idosos que viviam no Estado de São Paulo. O objetivo foi verificar uma possível tendência de crescimento ou decréscimo nessas taxas, além de se observar o impacto nas mesmas, do início da vacinação contra influenza, em idosos. Os autores observaram uma redução nas taxas, sugerindo um ponto de mudança na série, talvez, consequente da vacinação.

[Dourado et al. \(2006\)](#) publicaram um estudo retrospectivo com interesse de se ajustar um modelo de séries temporais para o período entre 1990 e 1996, para predizerem as taxas de incidência e de mortalidade por infecção por HIV, em todas as regiões do Brasil, para o período de 1997 a 2003. Com isso, os valores preditos foram comparados com os reais, a fim de se conhecer o possível impacto causado nas taxas de intervenção, por meio do antiretroviral e medidas de prevenção. O modelo ajustado pelos autores não contemplou

um ponto de mudança, mas a estrutura do estudo seria propícia para o ajuste do mesmo, podendo fornecer informações mais detalhadas, comparando-se com o modelo utilizado.

Filho e Rissin (2003) avaliaram a transição nutricional nas cinco regiões brasileiras através de taxas de desnutrição, obesidade e anemia, medidas em apenas três anos, no período entre 1975 e 1996. Os autores consideraram a estrutura de três estudos transversais, mas, provavelmente, esses dados poderiam ter sido analisados de melhor maneira se essas taxas tivessem sido coletadas, ano a ano, e um ponto de mudança fosse considerado na modelagem.

Além do fato da literatura apresentar estudos considerando pontos de mudanças fixos, alguns autores propuseram métodos que o consideram como aleatório, ou seja, podendo ser estimado pelo ajuste do modelo.

Ainda com relação à infecção por HIV, Ghosh e Vaida (2007) publicaram um ensaio clínico em que foram comparadas as contagens de células CD4, de acordo com a introdução de quatro tipos de terapias HAART (Terapia Antiretroviral Altamente Ativa). Assim, dados de contagens de CD4 foram colhidos, dia a dia, e foi ajustado um modelo com ponto de mudança que pudesse mostrar a mudança de tendência para uma terapia em particular.

Tomé e Miranda (2004) publicaram um estudo em que foram ajustados modelos para detecção de pontos de mudança para dados climáticos mundiais (precipitação e temperatura) obtidos através de longos períodos de observação (1880 até 1997). Assim, os autores identificaram três pontos em que essas variáveis tiveram sua tendência alterada, ao longo do período.

Um conceito que deve ser esclarecido é a diferença entre ponto de mudança e sazonalidade. São conceitos bem distintos, mas podem ser confundidos. A sazonalidade refere-se à alternância cíclica de períodos previsíveis de baixas ou altas taxas de uma variável em questão. Portanto, consegue-se verificar um padrão, ao longo do tempo, com relação aos pontos mais altos e mais baixos da série. Na área da saúde, muitas vezes, a sazonalidade é definida pelas estações do ano. Neste contexto, Modarres et al. (2012), publicaram um estudo que objetivou modelar a sazonalidade das incidências de fraturas de quadril, em idosos canadenses, entre 1985 e 2005, relacionando-as com variáveis climáticas. Os autores puderam verificar uma relação entre períodos frios e maiores taxas de incidências de fraturas e, também, um ponto de mudança em 1996 (não estimado diretamente), sugerindo uma série de casos em queda.

Carlin et al. (1992) consideraram dados de alguns estudos já publicados, e propuseram o uso de modelos hierárquicos sob o enfoque bayesiano para a análise com pontos de mudança.

Dey e Purkayastha (1997) propuseram o uso do fator de Bayes para selecionar o modelo com melhor ajuste aos dados (considerando ou não o ponto de mudança). Esse

Tabela 9 – Dados mensais sobre o número de medicamentos dispensados (Xarope de Sabutamol) em Ribeirão Preto - SP, Brasil, de fevereiro de 1999 a dezembro de 2011.

| ANO\MÊS | JAN   | FEV | MAR   | ABR   | MAI   | JUN   | JUL   | AGO   | SET   | OUT   | NOV   | DEZ |
|---------|-------|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----|
| 1998    | -     | -   | -     | -     | -     | -     | -     | -     | -     | -     | -     | 102 |
| 1999    | 394   | 278 | 585   | 358   | 336   | 365   | 269   | 391   | 123   | 107   | 118   | 214 |
| 2000    | 490   | 924 | 1.207 | 1.111 | 1.354 | 1.296 | 1.107 | 1.149 | 957   | 1.073 | 857   | 801 |
| 2001    | 901   | 635 | 1.130 | 1.481 | 1.947 | 2.032 | 1.994 | 1.795 | 1.467 | 1.410 | 1.116 | 690 |
| 2002    | 889   | 930 | 1.569 | 1.759 | 1.735 | 1.615 | 1.794 | 1.530 | 1.136 | 834   | 700   | 980 |
| 2003    | 942   | 819 | 1.437 | 1.600 | 1.944 | 1.550 | 1.778 | 1.007 | 1.089 | 849   | 906   | 817 |
| 2004    | 799   | 851 | 2.051 | 1.843 | 1.906 | 1.713 | 1.488 | 1.773 | 1.274 | 933   | 1061  | 748 |
| 2005    | 703   | 583 | 1.463 | 1.630 | 1.987 | 1.669 | 1.672 | 1.302 | 1.046 | 925   | 674   | 987 |
| 2006    | 1.004 | 939 | 1.112 | 1.426 | 2.407 | 1.832 | 1.266 | 1.120 | 867   | 876   | 879   | 587 |
| 2007    | 490   | 411 | 1.148 | 1.116 | 1.438 | 1.316 | 1.027 | 928   | 663   | 511   | 477   | 474 |
| 2008    | 436   | 577 | 859   | 919   | 1.066 | 835   | 648   | 452   | 570   | 597   | 452   | 489 |
| 2009    | 411   | 561 | 840   | 1144  | 1.138 | 966   | 883   | 802   | 971   | 407   | 405   | 428 |
| 2010    | 344   | 436 | 731   | 742   | 801   | 668   | 571   | 560   | 534   | 436   | 461   | 465 |
| 2011    | 531   | 452 | 682   | 686   | 777   | 792   | 696   | 488   | 341   | 362   | 430   | 239 |

fator resulta em um peso relativo de evidência do modelo com o ponto de mudança sobre o modelo sem o ponto de mudança.

Schütz e Holschneider (2011) propuseram um modelo bayesiano que é capaz de estimar múltiplos pontos de mudança, o que permite verificar não somente uma alteração na tendência da série, mas também, se existe uma mudança com relação à variabilidade da mesma. Chib (1998) e Fearnhead (2006) também propuseram modelos que podem estimar múltiplos pontos de mudança.

Nota-se, na literatura, uma vasta diversidade de propostas de modelos para detecção de pontos de mudanças nos mais variados contextos. O elemento autorregressivo na estimação de parâmetros é um detalhe importante, pois em uma estrutura em que são analisadas medidas ao longo do tempo, é necessário se levar em conta que uma taxa observada em um tempo  $t$  pode ser dependente daquelas medidas nos tempos  $t - 1, t - 2, \dots, t - R$ . Assim, tem-se um parâmetro autorregressivo de  $R$ -ésima ordem.

## 4.2 Métodos

### 4.2.1 O campo de estudo e o conjunto de dados

O presente estudo insere-se em uma pesquisa maior sobre a dispensação de medicamentos para tratar doenças pulmonares nos serviços públicos de saúde de Ribeirão Preto, São Paulo, Brasil. Os dados sobre o número de medicamentos dispensados foram fornecidos pelo Sistema Hygia Web, já comentado na seção 3.2.1.

Para efeitos do presente estudo, os dados de apenas um medicamento usado para aliviar os sintomas da asma foram considerados no desenvolvimento do modelo estatístico. O xarope de salbutamol em frasco de vidro âmbar com volume rotulado 120 ml contém 0,48 mg/ml de sulfato de salbutamol e é largamente utilizado no trato da asma (MARTINS et

al., 2008), sendo mais indicado para crianças e adultos que preferem medicamentos sob a forma líquida. Os dados completos sobre o número de medicamentos dispensados, a partir de dezembro de 1998 a dezembro de 2011, estão disponíveis na Tabela 9. Além disso, os dados sobre temperatura e precipitação no município de Ribeirão Preto foram obtidos a partir do Centro Integrado de Informações Agrometeorológicas do Instituto Agrônomo (CIIAGRO), usados como covariáveis no modelo.

O Comitê de Ética em Pesquisa da instituição dos pesquisadores responsáveis pelo projeto maior aprovou o estudo (CEP/CSE/FMRP/USP, protocolo número 453) e a permissão para acessar e usar os registros do Sistema de Informação Hygia Web foi fornecido pelo representante local do Departamento de Saúde (documento 248/11 GS RAS/ras).

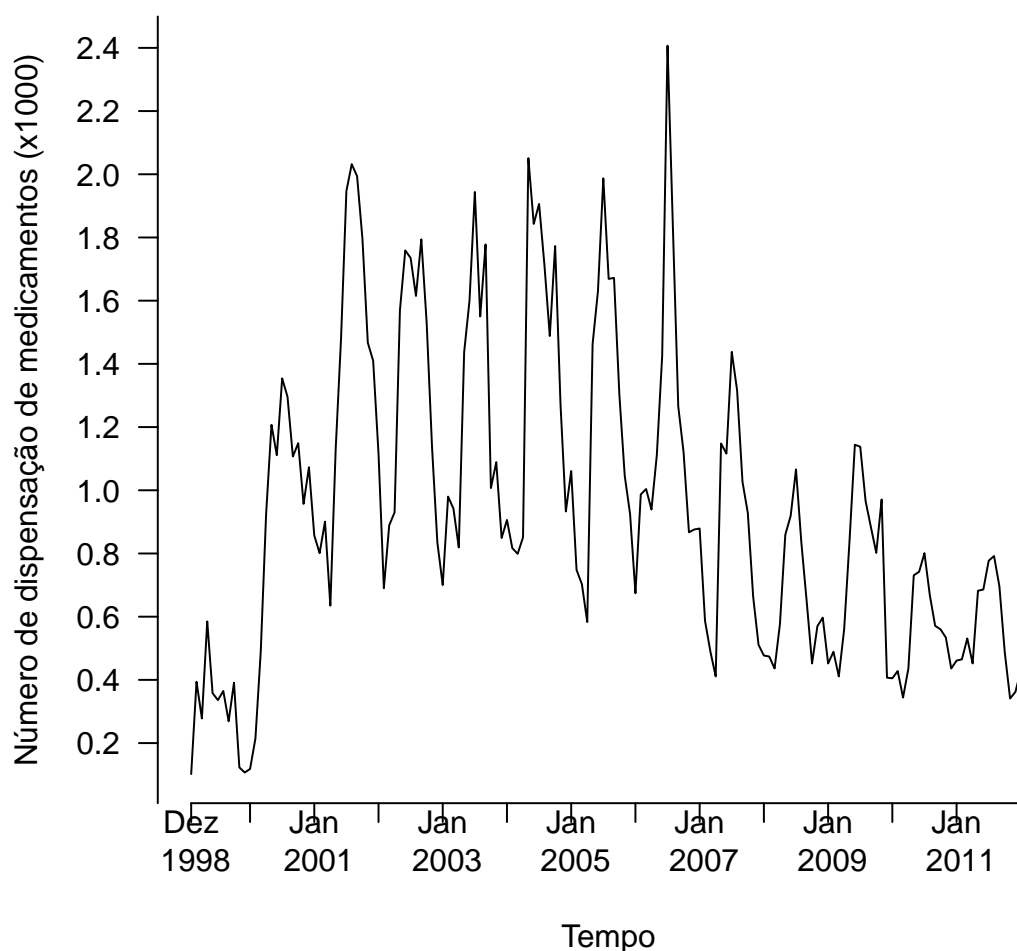


Figura 21 – Número mensal de medicamentos dispensados (xarope de salbutamol, 120 ml) pelos serviços públicos de saúde de Ribeirão Preto - SP, Brasil, de fevereiro de 1999 a dezembro de 2011.



### 4.2.2 Modelo estatístico

O gráfico da Figura 21 mostra a série temporal do número de dispensações mensais do xarope de salbutamol a partir de dezembro de 1998 a dezembro de 2011. Este gráfico dá evidências da presença de mais de um ponto de mudança para a série temporal. Observa-se um aumento no número de dispensações da medicação no início do período considerado até final de 2000. Temos então oscilações constantes até final de 2006, e após observamos uma queda da série até final de 2011. Portanto, vamos supor a presença de três pontos de mudança. Seja  $y_t$  o número de medicamentos dispensados no mês  $t$ ,  $t = 1, \dots, n$ , em que  $n$  é o número de meses na série temporal. O modelo proposto com três pontos de mudança é dado pela equação:

$$y_t = \alpha + g_1 I_{[1, \theta_1]}(t) + g_2 I_{(\theta_1, \theta_2]}(t) + g_3 I_{(\theta_2, \theta_3]}(t) + g_4 I_{(\theta_3, n]}(t) + S_t + \varepsilon_t \quad (4.1)$$

em que,

$$g_k = \beta_k(t - \theta_{k-1}) + \sum_{r=1}^R \psi_{kr}(x_{rt} - \bar{x}_r) + \sum_{j=1}^p \gamma_{kj}(y_{t-j} - \bar{y}) \quad (4.2)$$

com  $k = 1, \dots, 4$  e  $\theta_0 = 0$  para  $t < \theta_1$ .

Nas Equações (4.1) e (4.2),  $\alpha$  é um intercepto, e  $I_{\{A\}}(t)$  denota uma função indicadora de tal forma que  $I_{\{A\}}(t) = 1$  se  $t \in \{A\}$ , e 0 caso contrário,  $\theta_1, \theta_2$  e  $\theta_3$  são os pontos de mudança a serem estimados,  $x_{1t}, x_{2t}, \dots, x_{Rt}$  são observações de  $R$  covariáveis no mês  $t$ ,  $\bar{x}_r$  denota a média de  $x_{r1}, x_{r2}, \dots, x_{rn}$ ,  $r = 1, \dots, R$ ,  $\psi_{1r}, \psi_{2r}, \psi_{3r}$  e  $\psi_{4r}$  são os efeitos da covariável  $x_{rt}$  sobre  $y_t$  nos intervalos de mudança da série  $1, \dots, \theta_1, \dots, \theta_2, \dots, \theta_3, \dots, n$ , respectivamente,  $\bar{y}$  indica a média de  $y_1, \dots, y_n$ , os termos  $\gamma_{11}, \dots, \gamma_{1p}, \gamma_{21}, \dots, \gamma_{2p}, \gamma_{31}, \dots, \gamma_{3p}, \gamma_{41}, \dots, \gamma_{4p}$  são parâmetros autorregressivos de ordem  $p$  a serem estimados e  $\varepsilon_t$  são os termos de erro aleatório.

Além disso,  $S_t$  é dado de forma análoga a Equação (3.3) seção 3.2.2. Utilizando esse modelo assume-se que os termos  $\varepsilon_t$ ,  $t = 1, \dots, n$  são independentes e seguem uma distribuição normal com média 0 e variância que depende do ponto de mudança  $\theta$ , ou melhor:

$$\varepsilon_t \sim N\left(0, \sigma_1^2 I_{[1, \theta_1]}(t) + \sigma_2^2 I_{(\theta_1, \theta_2]}(t) + \sigma_3^2 I_{(\theta_2, \theta_3]}(t) + \sigma_4^2 I_{(\theta_3, n]}(t)\right) \quad (4.3)$$

Assim,  $\sigma_1^2, \sigma_2^2, \sigma_3^2$  e  $\sigma_4^2$  são as variâncias de  $\varepsilon_t$  nos intervalos de mudança da série  $1, \dots, \theta_1, \dots, \theta_2, \dots, \theta_3, \dots, n$ , respectivamente. Esta formulação do modelo corresponde à seguinte função de verossimilhança:

$$f(y|\theta, \xi) = \prod_{t=1}^n (2\pi\lambda_t)^{-\frac{1}{2}} \exp\left[-\sum_{t=1}^n \frac{(y_t - \mu_t)^2}{2\lambda_t}\right] \quad (4.4)$$

em que

$$\lambda_t = \sigma_1^2 I_{[1, \theta_1]}(t) + \sigma_2^2 I_{(\theta_1, \theta_2]}(t) + \sigma_3^2 I_{(\theta_2, \theta_3]}(t) + \sigma_4^2 I_{(\theta_3, n]}(t), \quad (4.5)$$

$$\mu_t = \alpha + g_1 I_{[1, \theta_1]}(t) + g_2 I_{(\theta_1, \theta_2]}(t) + g_3 I_{(\theta_2, \theta_3]}(t) + g_4 I_{(\theta_3, n]}(t) + S_t, \quad (4.6)$$

tal que  $y = (y_1, y_2, \dots, y_n)$  e

$\xi = (\alpha, \beta_1, \beta_2, \beta_3, \beta_4, \gamma_{1,1}, \dots, \gamma_{1,p}, \gamma_{2,1}, \dots, \gamma_{2,p}, \gamma_{3,1}, \dots, \gamma_{3,p}, \gamma_{4,1}, \dots, \gamma_{4,p}, \psi_{1,1}, \dots, \psi_{1,R}, \psi_{2,1}, \dots, \psi_{2,R}, \psi_{3,1}, \dots, \psi_{3,R}, \psi_{4,1}, \dots, \psi_{4,R}, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2, \eta_1, \eta_2)$  é o vetor dos parâmetros.

Por definição,  $f(y|\theta, \xi)$  denota a função de densidade de probabilidade conjunta da amostra  $Y = (Y_1, Y_2, \dots, Y_n)$ . Na análise Bayesiana, presume-se que os parâmetros do vetor  $\xi$  e  $\theta$  seguem distribuições *a priori*, a qual são atualizadas com os dados (representada por  $f(y|\theta, \xi)$ ) para produzir as distribuições *a posteriori*. Isto é formalizado pelo teorema de Bayes, dado por  $f(\theta, \xi|y) \propto f(y|\theta, \xi)p(\theta, \xi)$ , em que  $p(\theta, \xi)$  é a distribuição conjunta *a priori* e  $f(\theta, \xi|y)$  é a distribuição conjunta *a posteriori*. As distribuições *a priori* podem ser "não-informativas", com um pequeno efeito sobre a distribuição *a posteriori*.

Assim, as seguintes distribuições *a priori* para os parâmetros do vetor  $\xi$  são consideradas:

- $\alpha \sim N(0, c_1)$ ,
- $\beta_1 \sim N(0, c_2), \beta_2 \sim N(0, c_3), \beta_3 \sim N(0, c_4), \beta_4 \sim N(0, c_5)$ ,
- $\gamma_{1,j} \sim N(0, c_{6,j}), \gamma_{2,j} \sim N(0, c_{7,j}), \gamma_{3,j} \sim N(0, c_{8,j}), \gamma_{4,j} \sim N(0, c_{9,j}), j = 1, \dots, p$ ,
- $\eta_1 \sim N(0, c_{10}), \eta_2 \sim N(0, c_{11})$ ,
- $\psi_{1,r} \sim N(0, c_{12,r}), \psi_{2,r} \sim N(0, c_{13,r}), \psi_{3,r} \sim N(0, c_{14,r}), \psi_{4,r} \sim N(0, c_{15,r}), r = 1, \dots, R$ ,
- $\sigma_1^2 \sim IG(c_{16}, c_{17})$  e  $\sigma_2^2 \sim IG(c_{18}, c_{19}), \sigma_3^2 \sim IG(c_{20}, c_{21})$  e  $\sigma_4^2 \sim IG(c_{22}, c_{23})$ .

em que  $c_1, \dots, c_{23}$  são valores conhecidos para hiperparâmetros das distribuições *a priori*,  $N(0, c)$  indica uma distribuição normal com média 0 e variância  $c$ , e  $IG(h_1, h_2)$  denota uma distribuição Gama Inversa com média  $\frac{h_2}{h_1-1}$  e variância  $\frac{h_2^2}{[(h_1-1)^2(h_1-2)]}$ . Valores grandes de  $c_1, \dots, c_{23}$  produzem distribuições *a priori* não informativas para os respectivos parâmetros. Assumiu-se ainda independência entre estes parâmetros.

Além disso, é assumido uma distribuição uniforme discreta *a priori* para o ponto de mudança  $\theta$  de tal modo que as probabilidades prévias dos valores  $1, 2, \dots, n$  são assumidos como iguais a  $\frac{1}{n}$ .

O procedimento Bayesiano de simulação em Cadeias de Markov (MCMC - *Markov Chain Monte Carlo*) com o algoritmo de amostragem de Gibbs (CASELLA; GEORGE, 1992) foi usado para estimar as distribuições a posteriori dos parâmetros de interesse e os componentes de variância (CARLIN; LOUIS, 1997) e habituais métodos de diagnóstico existentes foram empregados para verificar a convergência dos cálculos MCMC (CARLIN; LOUIS, 1997), de forma análoga ao utilizado no Capítulo 3 seção 3.2.4 desse trabalho. A estimação bayesiana dos parâmetros foi realizada novamente utilizando o algoritmo MCMC implementado no programa livre OpenBUGS (LUNN et al., 2000). O código OpenBUGS utilizado para esta análise é dado no Apêndice.

### 4.2.3 Especificações dos Modelos

Na ausência de covariáveis, foram ajustados três modelos diferentes para os dados, denotados como os Modelos 1 a 3, tal como descrito a seguir:

- **Modelo 1:** Neste modelo, os termos autorregressivos  $\gamma_{11}, \dots, \gamma_{1p}, \gamma_{21}, \dots, \gamma_{2p}, \gamma_{31}, \dots, \gamma_{3p}, \gamma_{41}, \dots, \gamma_{4p}$  não foram incluídos, e consequentemente,  $g_k = \beta_k(t - \theta_{k-1})$ , para  $k = 1, \dots, 4$ . Este modelo não considera a presença de covariáveis e nem a função periódica mensal  $S_t$ ;
- **Modelo 2:** Este modelo é similar ao Modelo 1, mas considera os termos autorregressivos  $\gamma_{11}, \dots, \gamma_{1p}, \gamma_{21}, \dots, \gamma_{2p}, \gamma_{31}, \dots, \gamma_{3p}, \gamma_{41}, \dots, \gamma_{4p}$  de ordem 1;
- **Modelo 3:** Similar ao Modelo 2, mas considera a função mensal periódica  $S_t$ ;

Os modelos 1 a 3 foram ajustados com base na premissa que os resíduos  $\varepsilon_t$  seguem uma distribuição normal;

- **Modelo 4, 5, 6 e 7:** Semelhantes ao Modelo 3, mas incluem uma variável independente. As seguintes variáveis foram consideradas: a temperatura média mensal ( $^{\circ}\text{C}$ ), temperatura máxima e mínima mensal ( $^{\circ}\text{C}$ ) e precipitação média mensal (mm) em uma escala logarítmica. De acordo com esta formulação, quatro modelos diferentes foram ajustadas aos dados, um para cada variável independente, evitando assim problemas de colinearidade entre as variáveis.

### 4.2.4 Seleção do Modelo

Como já comentado no Capítulo anterior o modelo proposto é interpretado pelo OpenBUGS como um modelo de mistura, e este programa não é capaz de calcular o valor do DIC nesta situação. Um outro critério para a seleção do modelo é derivado das ordenadas preditivas condicionais (CPO) (GELFAND et al., 1992). De forma análoga a análise feita

Tabela 10 – Resultados dos modelos estatísticos de mudança de ponto, com sequência de resíduos que seguem uma distribuição normal.

| Parâmetro      | Modelo 1             |                                | Modelo 2             |                                | Modelo 3             |                                |
|----------------|----------------------|--------------------------------|----------------------|--------------------------------|----------------------|--------------------------------|
|                | Estimativa Bayesiana | Intervalo de credibilidade 95% | Estimativa Bayesiana | Intervalo de credibilidade 95% | Estimativa Bayesiana | Intervalo de credibilidade 95% |
| $\alpha_1$     | 0,023                | (-0,231 ; 0,260)               | 0,732                | (0,403 ; 1,049)                | 0,382                | (0,008 ; 0,709)                |
| $\alpha_2$     | 1,39                 | (1,204 ; 1,571)                | 1,072                | (0,942 ; 1,203)                | 1,149                | (1,026 ; 1,275)                |
| $\alpha_3$     | 1,038                | (0,836 ; 1,222)                | 0,938                | (0,824 ; 1,053)                | 0,984                | (0,872 ; 1,094)                |
| $\alpha_4$     | 0,619                | (0,480 ; 0,771)                | 0,859                | (0,734 ; 0,983)                | 0,700                | (0,554 ; 0,841)                |
| $\beta_1$      | 0,049                | (0,036 ; 0,066)                | 0,013                | (-0,0018 ; 0,031)              | 0,031                | (0,014 ; 0,052)                |
| $\beta_2$      | -0,005               | (-0,010 ; -0,694)              | -0,0018              | (-0,0045 ; 0,752)              | -0,0024              | (-0,0049 ; 0,2095)             |
| $\beta_3$      | -0,012               | (-0,020 ; -0,005)              | -0,0024              | (-0,0085 ; 0,004)              | -0,0077              | (-0,0132 ; -0,0027)            |
| $\beta_4$      | -0,005               | (-0,014 ; 0,005)               | -0,004               | (-0,014 ; 0,0056)              | -0,0055              | (-0,0178 ; 0,0039)             |
| $\theta_1$     | 28,26                | (22 ; 30)                      | 25,54                | (20 ; 30)                      | 25,4                 | (20,0 ; 30,0)                  |
| $\theta_2$     | 94,97                | (90 ; 104)                     | 98,7                 | (90 ; 105)                     | 94,99                | (90,0 ; 104)                   |
| $\theta_3$     | 130,6                | (120 ; 139)                    | 131,2                | (120 ; 140)                    | 132,3                | (120,0 ; 140,0)                |
| $\gamma_{1,1}$ | -                    | -                              | 0,745                | (0,430 ; 1,056)                | 0,391                | (0,043 ; 0,711)                |
| $\gamma_{2,1}$ | -                    | -                              | 0,730                | (0,565 ; 0,902)                | 0,581                | (0,408 ; 0,752)                |
| $\gamma_{3,1}$ | -                    | -                              | 0,749                | (0,487 ; 1,046)                | 0,407                | (0,155 ; 0,668)                |
| $\gamma_{4,1}$ | -                    | -                              | 0,668                | (0,296 ; 1,044)                | 0,146                | (-0,332 ; 0,572)               |
| $\sigma_1^2$   | 0,113                | (0,064 ; 0,197)                | 0,052                | (0,028 ; 0,093)                | 0,045                | (0,023 ; 0,085)                |
| $\sigma_2^2$   | 0,220                | (0,154 ; 0,313)                | 0,116                | (0,082 ; 0,163)                | 0,072                | (0,050 ; 0,103)                |
| $\sigma_3^2$   | 0,098                | (0,056 ; 0,167)                | 0,044                | (0,023 ; 0,077)                | 0,031                | (0,018 ; 0,053)                |
| $\sigma_4^2$   | 0,040                | (0,021 ; 0,074)                | 0,026                | (0,013 ; 0,048)                | 0,022                | (0,011 ; 0,043)                |
| $\eta_1$       | -                    | -                              | -                    | -                              | 0,032                | (-0,031 ; 0,094)               |
| $\eta_2$       | -                    | -                              | -                    | -                              | -0,218               | (-0,286 ; -0,152)              |
| LPML           | -54,54               |                                | -4,24                |                                | 26,45                |                                |
| SSR            | 20,05                |                                | 10,11                |                                | 6,22                 |                                |

nos modelos propostos no Capítulo 2 vamos analisar o melhor ajuste utilizando a LPML (*log pseudo marginal likelihood*) e SSR (soma dos quadrados dos resíduos). Essas medidas estão detalhadas na Seção 3.2.4 desse mesmo trabalho, onde mostramos como comparar o melhor ajuste entre os modelos que usam a abordagem de ponto de mudança.

### 4.3 Resultados

Na análise Bayesiana, foram consideradas distribuições a priori não informativas para todos os parâmetros do modelo. Deste modo, considerou-se  $c_1 = c_2 = c_3 = c_4 = c_5 = c_{6,j} = c_{7,j} = c_{8,j} = c_{9,j} = c_{10} = c_{11} = c_{12,r} = c_{13,r} = c_{14,r} = c_{15,r} = 10^6$ ,  $j = 1, \dots, p$ ,  $r = 1, \dots, R$ , nas distribuições a priori para  $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta_1, \beta_2, \beta_3, \beta_4, \gamma_{1,j}, \gamma_{2,j}, \gamma_{3,j}, \gamma_{4,j}, \psi_{1,r}, \psi_{2,r}, \psi_{3,r}$  e  $\psi_{4,r}$  e  $c_{16} = c_{17} = c_{18} = c_{19} = c_{20} = c_{21} = c_{22} = c_{23} = 0, 1$  nas distribuições a priori para  $\sigma_1^2, \sigma_2^2, \sigma_3^2$  e  $\sigma_4^2$ . O número de medicamentos dispensados mensais foi dividido por 1000, a fim de facilitar a convergência do algoritmo computacional.

A Tabela 10 mostra os resultados para os Modelos 1 a 3 obtidos pelo programa OpenBUGS. Os resultados da Tabela 10 considera que os resíduos dos modelos seguem uma distribuição normal. No caso dos Modelos 2 e 3, foram ajustados considerando termos autorregressivos de ordem 1 para todos os casos.

As Tabelas 11 e 12 mostram os resultados de modelos de regressão (Modelo 4) que consideram a temperatura máxima mensal e a mínima mensal em valor absoluto, a temperatura média mensal e a precipitação média mensal como variáveis independentes.

Tabela 11 – Resultados dos modelos de regressão que consideram a temperatura máxima e mínima mensal absoluta como variáveis independentes.

| Parâmetro      | Temperatura máxima absoluta<br>(°C) |                                | Temperatura mínima absoluta<br>(°C) |                                |
|----------------|-------------------------------------|--------------------------------|-------------------------------------|--------------------------------|
|                | Estimativa Bayesiana                | Intervalo de credibilidade 95% | Estimativa Bayesiana                | Intervalo de credibilidade 95% |
| $\alpha_1$     | 0,4103                              | (0,0399 ; 0,7398)              | 0,4601                              | (0,1238 ; 0,7876)              |
| $\alpha_2$     | 1,151                               | (1,034 ; 1,273)                | 1,181                               | (1,063 ; 1,298)                |
| $\alpha_3$     | 0,9732                              | (0,8626 ; 1,091)               | 1,01                                | (0,893 ; 1,127)                |
| $\alpha_4$     | 0,7396                              | (0,5717 ; 0,8991)              | 0,6793                              | (0,5198 ; 0,8303)              |
| $\beta_1$      | 0,0298                              | (0,0134 ; 0,0498)              | 0,0308                              | (0,01403 ; 0,049)              |
| $\beta_2$      | -0,0025                             | (-0,0049 ; 0,0265)             | -0,0024                             | (-0,0047 ; 0,0214)             |
| $\beta_3$      | -0,0066                             | (-0,0131 ; -0,3063)            | -0,0091                             | (-0,0153 ; -0,0038)            |
| $\beta_4$      | -0,0049                             | (-0,0194 ; 0,0062)             | -0,0054                             | (-0,0199 ; 0,0057)             |
| $\theta_1$     | 25,23                               | (21 ; 30)                      | 23,53                               | (20 ; 28)                      |
| $\theta_2$     | 96,54                               | (90 ; 105)                     | 94,56                               | (90 ; 104)                     |
| $\theta_3$     | 132,10                              | (120 ; 140)                    | 131,4                               | (120 ; 140)                    |
| $\gamma_{1,1}$ | 0,4266                              | (0,0788 ; 0,7506)              | 0,5451                              | (0,2109 ; 0,8774)              |
| $\gamma_{2,1}$ | 0,5665                              | (0,3855 ; 0,7487)              | 0,4863                              | (0,2999 ; 0,6783)              |
| $\gamma_{3,1}$ | 0,48                                | (0,1811 ; 0,8124)              | 0,3434                              | (0,0298 ; 0,6957)              |
| $\gamma_{4,1}$ | 0,2583                              | (-0,3103 ; 0,7861)             | 0,084                               | (-0,461 ; 0,5793)              |
| $\sigma_1^2$   | 0,0426                              | (0,0213 ; 0,0817)              | 0,0363                              | (0,0186 ; 0,06854)             |
| $\sigma_2^2$   | 0,071                               | (0,0497 ; 0,1014)              | 0,0676                              | (0,0472 ; 0,0967)              |
| $\sigma_3^2$   | 0,030                               | (0,0160 ; 0,0533)              | 0,0331                              | (0,0185 ; 0,058)               |
| $\sigma_4^2$   | 0,0257                              | (0,0125 ; 0,0502)              | 0,0237                              | (0,0114 ; 0,0459)              |
| $\eta_1$       | 0,041                               | (-0,024 ; 0,1046)              | 0,025                               | (-0,0613 ; 0,1094)             |
| $\eta_2$       | -0,2441                             | (-0,3219 ; -0,167)             | -0,2328                             | (-0,3108 ; -0,1564)            |
| $\psi_1$       | 0,0326                              | (-0,01437 ; 0,07937)           | 0,02633                             | (0,0023 ; 0,0509)              |
| $\psi_2$       | -0,0036                             | (-0,03802 ; 0,0302)            | -0,0151                             | (-0,0356 ; 0,0059)             |
| $\psi_3$       | 0,01895                             | (-0,01799 ; 0,057)             | -0,0039                             | (-0,0282 ; 0,0236)             |
| $\psi_4$       | 0,01981                             | (-0,02589 ; 0,0667)            | -0,2326                             | (-0,0217 ; 0,0212)             |
| LPML           |                                     | 23,97                          |                                     | 27,77                          |
| SSR            |                                     | 6,02                           |                                     | 5,67                           |

Estes modelos supõem que os resíduos seguem uma distribuição normal. Podemos observar que o intervalo de credibilidade 95% para os parâmetros  $\psi_1$ ,  $\psi_2$ ,  $\psi_3$  e  $\psi_4$  incluem o valor zero em quase todos os modelos que consideram estas variáveis independentes, apenas  $\psi_1 = 0,0263$  no modelo com a covariável temperatura mínima absoluta (Tabela 11) e  $\psi_1 = 0,0613$  no modelo com a covariável temperatura média mensal (Tabela 12) foram significantes para os modelos com covariáveis. Com esses resultados não podemos inferir que há evidências de que estas variáveis climáticas não estão associados com o número mensal de medicamentos dispensados, pois a componente sazonal  $St$  já acomoda possíveis efeitos de variáveis climáticas por serem sazonais (Figura 22), mostrando mais uma vez a eficácia da série de Fourier para captar efeitos periódicos. A Figura 22 mostra que os modelos ajustados com a componente periódica  $S_t$  descreve um aumento no padrão

Tabela 12 – Resultados dos modelos de regressão que consideram a temperatura média mensal e precipitações como variáveis independentes.

| Parâmetro      | Temperatura média mensal<br>(°C) |                                | Precipitação<br>(log mm) |                                |
|----------------|----------------------------------|--------------------------------|--------------------------|--------------------------------|
|                | Estimativa Bayesiana             | Intervalo de credibilidade 95% | Estimativa Bayesiana     | Intervalo de credibilidade 95% |
| $\alpha_1$     | 0,4099                           | (0,0693 ; 0,7501)              | 0,3898                   | (0,0109 ; 0,7336)              |
| $\alpha_2$     | 1,188                            | (1,069 ; 1,306)                | 1,172                    | (1,047 ; 1,3)                  |
| $\alpha_3$     | 1,011                            | (0,8994 ; 1,127)               | 1,001                    | (0,8874 ; 1,114)               |
| $\alpha_4$     | 0,7028                           | (0,537 ; 0,8618)               | 0,6858                   | (0,5317 ; 0,8303)              |
| $\beta_1$      | 0,0327                           | (0,0151 ; 0,0514)              | 0,0314                   | (0,0144 ; 0,0525)              |
| $\beta_2$      | -0,0024                          | (-0,0047 ; -0,0098)            | -0,0024                  | (-0,005 ; 0,1175)              |
| $\beta_3$      | -0,0088                          | (-0,0155 ; -0,0031)            | -0,0086                  | (-0,0146 ; -0,0034)            |
| $\beta_4$      | -0,0053                          | (-0,0208 ; 0,0062)             | -0,0057                  | (-0,0191 ; 0,0044)             |
| $\theta_1$     | 23,95                            | (20 ; 29)                      | 25,33                    | (20 ; 30)                      |
| $\theta_2$     | 96,02                            | (90 ; 104)                     | 95,19                    | (90 ; 105)                     |
| $\theta_3$     | 131,7                            | (120 ; 140)                    | 132,3                    | (120 ; 140)                    |
| $\gamma_{1,1}$ | 0,4832                           | (0,153 ; 0,8165)               | 0,4097                   | (0,0515 ; 0,75)                |
| $\gamma_{2,1}$ | 0,4503                           | (0,2639 ; 0,6438)              | 0,5173                   | (0,3318 ; 0,7026)              |
| $\gamma_{3,1}$ | 0,4344                           | (0,1088 ; 0,8094)              | 0,4055                   | (0,138 ; 0,6908)               |
| $\gamma_{4,1}$ | 0,1427                           | (-0,4214 ; 0,6596)             | 0,0992                   | (-0,4179 ; 0,5554)             |
| $\sigma_1^2$   | 0,0380                           | (0,0195 ; 0,073)               | 0,0488                   | (0,0247 ; 0,0924)              |
| $\sigma_2^2$   | 0,0657                           | (0,046 ; 0,0939)               | 0,0701                   | (0,0486 ; 0,1003)              |
| $\sigma_3^2$   | 0,0326                           | (0,0176 ; 0,0587)              | 0,0313                   | (0,0175 ; 0,0545)              |
| $\sigma_4^2$   | 0,0246                           | (0,0119 ; 0,048)               | 0,0238                   | (0,0116 ; 0,0465)              |
| $\eta_1$       | 0,0126                           | (-0,0633 ; 0,0878)             | 0,0115                   | (-0,0653 ; 0,0877)             |
| $\eta_2$       | -0,2452                          | (-0,329 ; -0,1625)             | -0,2352                  | (-0,3097 ; -0,1617)            |
| $\psi_1$       | 0,0613                           | (0,0061 ; 0,1187)              | 0,0183                   | (-0,017 ; 0,0552)              |
| $\psi_2$       | -0,0329                          | (-0,0754 ; 0,0101)             | -0,0113                  | (-0,0424 ; 0,019)              |
| $\psi_3$       | 0,0159                           | (-0,0336 ; 0,0694)             | 0,015                    | (-0,0127 ; 0,0446)             |
| $\psi_4$       | 0,0103                           | (-0,0389 ; 0,0589)             | 0,006                    | (-0,0213 ; 0,0339)             |
| LPML           | 27,26                            |                                | 23,27                    |                                |
| SSR            | 5,59                             |                                | 5,99                     |                                |

sazonal anual da dispensação do xarope de salbutamol nos meses entre março e setembro, coincidindo com a época de mais frio e seca no ano na região da retirada do banco de dados, quando as infecções das vias aéreas e episódios de broncoespasmo ocorrem com mais frequência, criando uma maior demanda para o uso do medicamento nesse período (THOMAZELLI et al., 2007; PETERSON et al., 2012).

Portanto ajustamos modelos similares que não incluem a componente sazonal  $St$ , mas com termos autorregressivos e as quatro variáveis climáticas separadamente (resultados não mostrados). A LPML desses modelos ficaram entre  $-8$  e  $1$  e a soma do quadrado dos resíduos todas próximas de dez mostrando um bom ajuste. Entretanto, devido ao Modelo 3 ter uma LPML significativamente maior (Tabela 10) e sem covariáveis, ou seja, um modelo mais parcimonioso, com menos parâmetros preferimos ocultar os resultados

dos modelos com covariáveis sem  $S_t$ . Portanto nos próximos parágrafos vamos comparar apenas os resultados dos Modelos 1,2 e 3 contidos na Tabela 10.

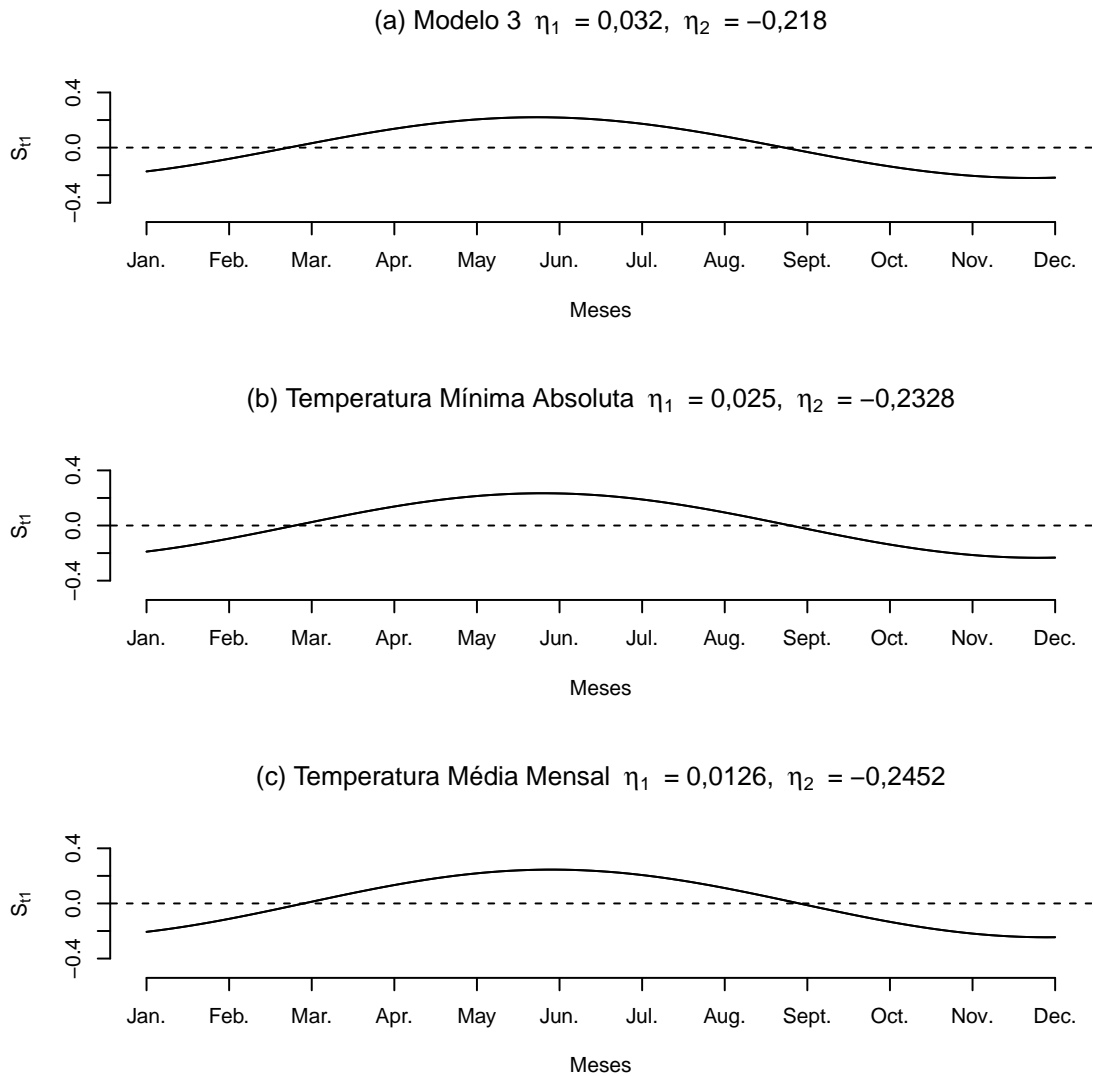


Figura 22 – Padrão sazonal mensal do número de dispensação do xarope de salbutamol ajustados pelo (a) Modelo 3, (b) e (c) Modelo 4 com covariáveis.

Os resultados do Modelo 3 tem um valor baixo para  $SSR = 6,22$  e o valor mais alto para LPML (26,45), sugerindo que ele proporciona o melhor ajuste para os dados entre estes três modelos (Tabela 10). Em todos os modelos ajustados (apenas no Modelo 2 o intervalo de credibilidade mostrou que  $\beta_1$  não era significativo), as estimativas para  $\beta_1$  são positivas, descrevendo que o número de medicamentos dispensados está aumentando ao longo do tempo até o ponto de mudança  $\theta_1$ . O intervalo de credibilidade 95% para  $\gamma_{1,1}$ ,  $\gamma_{2,1}$ ,  $\gamma_{3,1}$  não contém valor zero (Modelo 2 e 3), evidenciando a importância do parâmetro autorregressivo de ordem  $p = 1$ . Apenas  $\gamma_{4,1}$  mostrou-se não significativo no Modelo 3. Os resultados do Modelo 3 também mostram que o intervalo de credibilidade de 95% para  $\eta_2$  não contém o valor zero, o que sugere novamente a evidência de um padrão sazonal

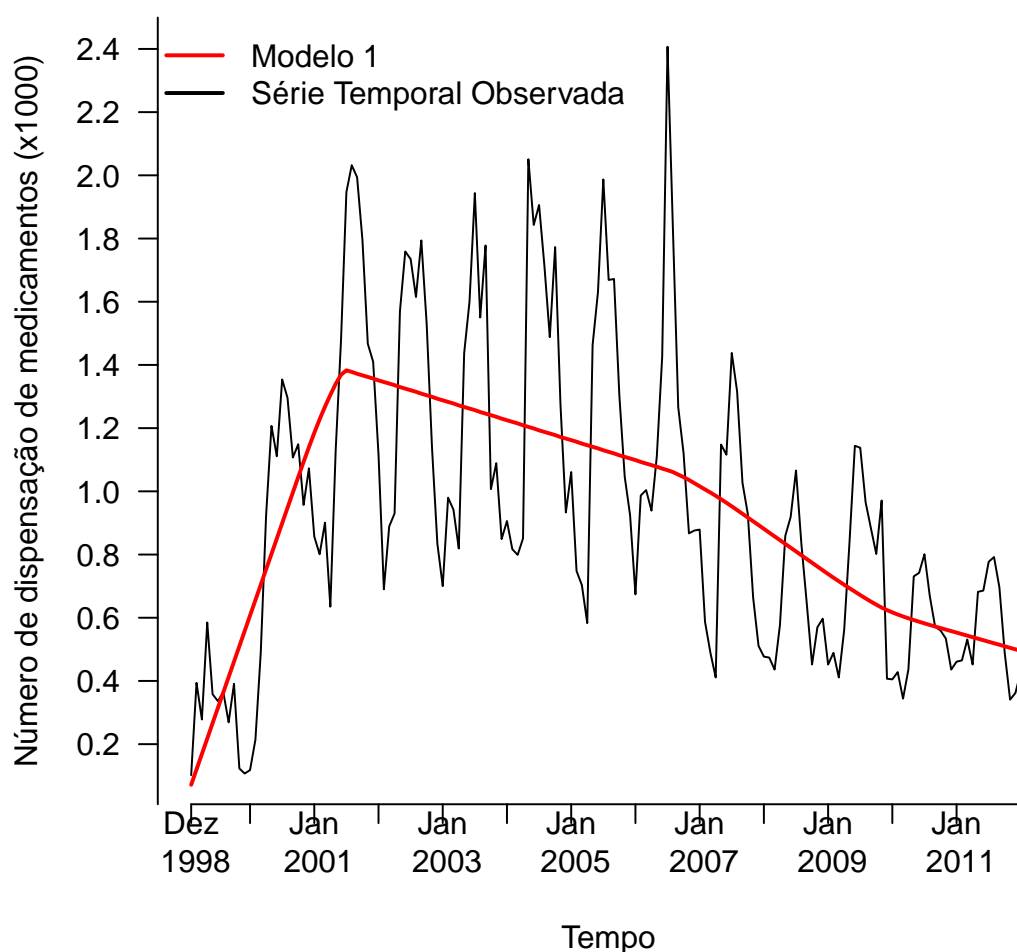


Figura 23 – Comparação entre a série observada e o ajuste calculado a partir do Modelo 1 - com três pontos de mudança.

anual na série. Os valores de  $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_3$  e  $\alpha_4$  são todos significantes para o Modelo 3, evidenciando um intercepto para cada intervalo entre os pontos de mudança.

As Figuras 23, 24 e 25 mostram a comparação entre os dados da série observada de dispensação do medicamento para asma xarope de salbutamol com os ajustes dos Modelos 1, 2 e 3.

A Figura 26 mostra os gráficos da função de autocorrelação e autocorrelação parcial para os resíduos dos Modelos 1, 2 e 3. Observamos que o Modelo 3 tem uma estrutura que acomoda melhor os resíduos em relação aos Modelos 1 e 2, não violando para o Modelo 3 a suposição de independências dos resíduos.

Um exame da sequência de valores inversos da CPO (Figura 27) pode identificar possíveis valores extremos no modelo ajustado e permitir comparações entre os modelos,



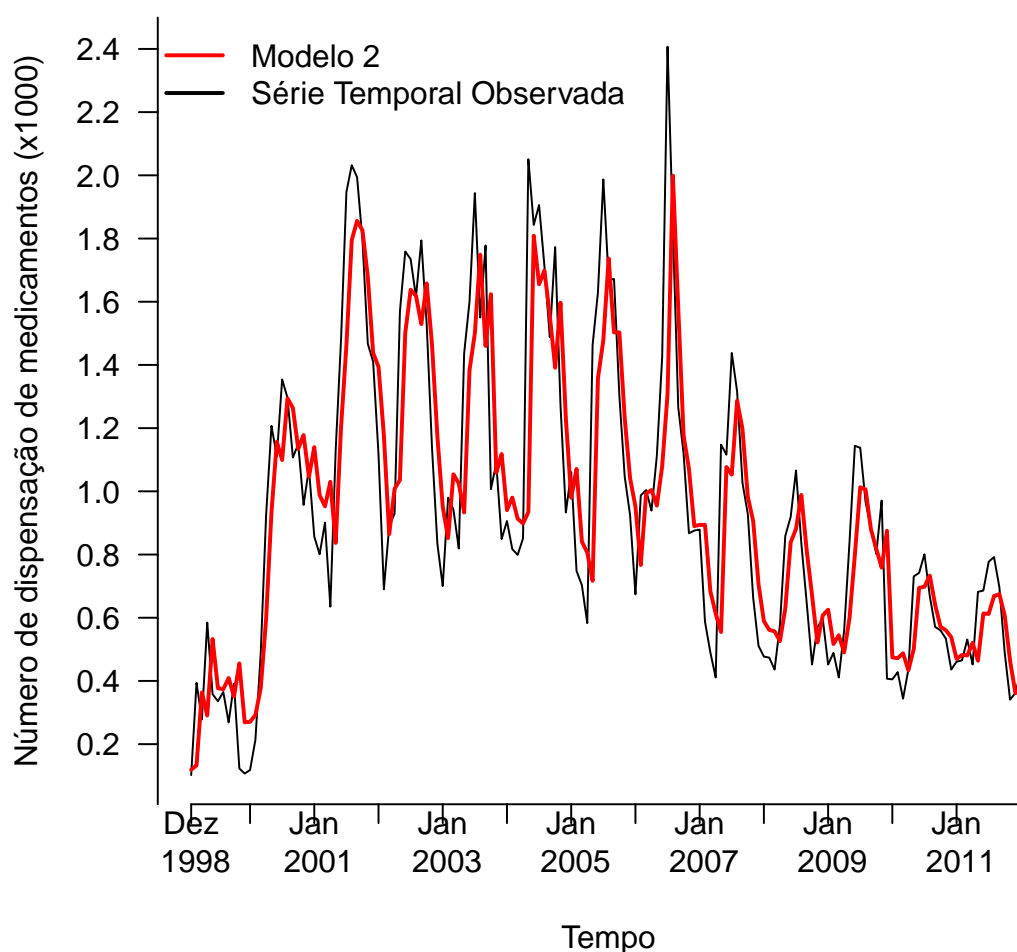


Figura 24 – Comparação entre a série observada e o ajuste calculado a partir do Modelo 2 - com três pontos de mudança e termos autorregressivos de ordem 1.

considerando cada ponto de observação. Considerando os resultados dos Modelos 1, 2 e 3, o gráfico da Figura 27 apresenta os valores inversos para CPO, onde as duas linhas tracejadas horizontais na figura passam por meio dos valores de 40 e 70, com o objetivo de identificar possíveis *outliers* e valores extremos, respectivamente (ver subseção 2.2.4). Apenas no Modelo 1 observamos um valor extremo. No Modelo 3 não há *outliers* e nem valores extremos o que nos dá maior segurança de um bom ajuste por parte desse modelo.

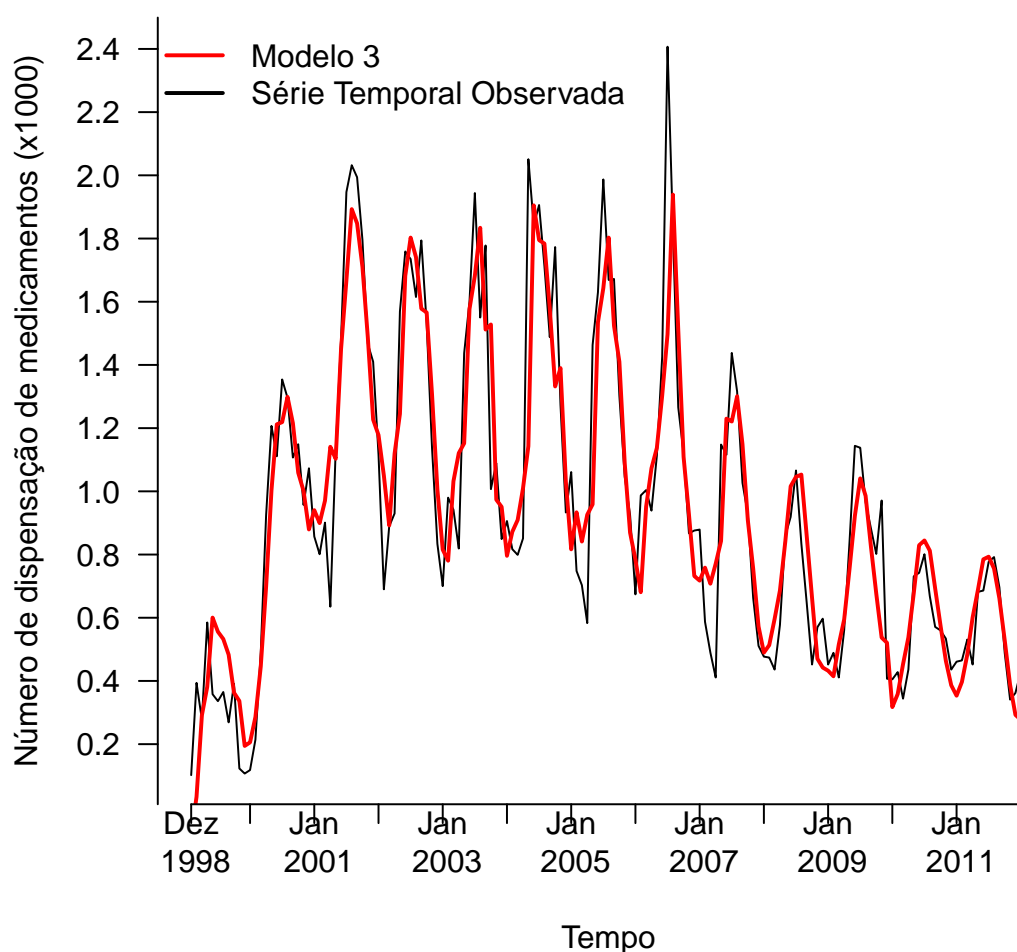


Figura 25 – Comparação entre a série observada e o ajuste calculado a partir do Modelo 3 - com três pontos de mudança, termos autorregressivos de ordem 1 e componente sazonal.

## 4.4 Discussão

Os gráficos de ajuste dos modelos descrevem como o comportamento da série histórica do número de dispensas mensais dos frascos de xarope de salbutamol de 120 ml, de dezembro de 1998 a dezembro de 2011, interpretados em termos da presença de três pontos de mudança. O aumento do número de dispensação da medicação no início do ano de 1999 (Figura 21), pode ser explicado pela mudança da Política Nacional Brasileira de Medicamentos já comentado na Seção 3.4. Com as mudanças na política de distribuição de medicamentos no período estudado, a metodologia de ponto de mudança passa a ser uma ferramenta útil e interessante para caracterizar essas mudanças no modelo.

Mostramos por meio de gráficos e tabelas (seção 4.3) que além dos três pontos de

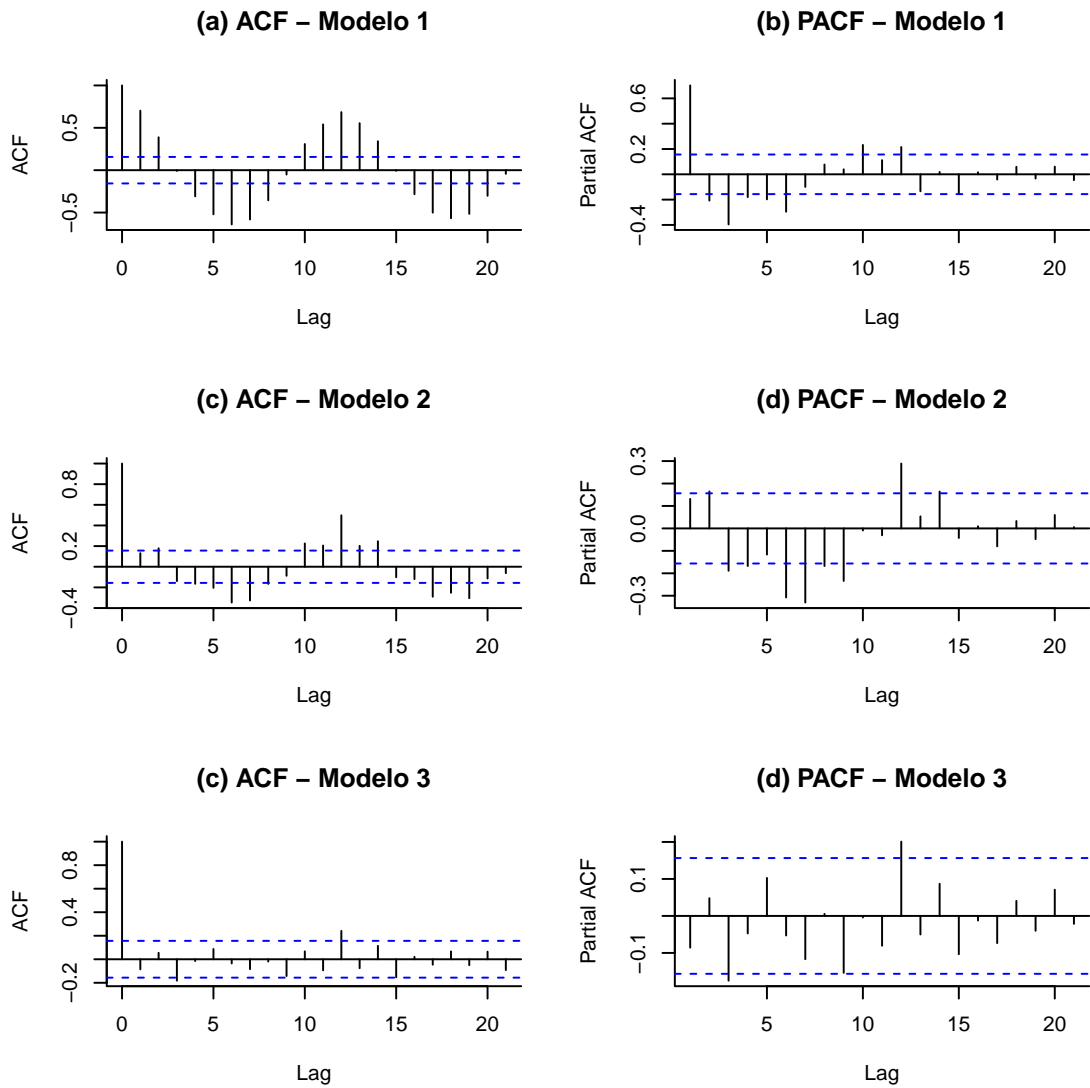


Figura 26 – Gráficos de autocorrelações dos resíduos que seguem uma distribuição normal dos Modelos 1, 2 e 3.

mudança, a importância dos termos autorregressivos de ordem 1 e a componente sazonal  $S_t$  através do intervalo de credibilidade, todas essas componentes estão no Modelo 3. Tendo o maior valor para LPML, menor valor da soma dos quadrados dos resíduos e não violar a suposição de independência dos resíduos e nem apresentar *outliers* ou valores extremos, consideramos dentre os quatro modelos testados, o Modelo 3 como aquele com mais evidências de um melhor ajuste para série observada.

## 4.5 Conclusão

Um modelo estatístico para as séries temporais epidemiológicas com componente sazonal de ordem 1, termos autorregressivos de ordem 1, covariáveis e três pontos de mudança foi descrita neste capítulo. Observou-se que o modelo proposto foi bem ajustado

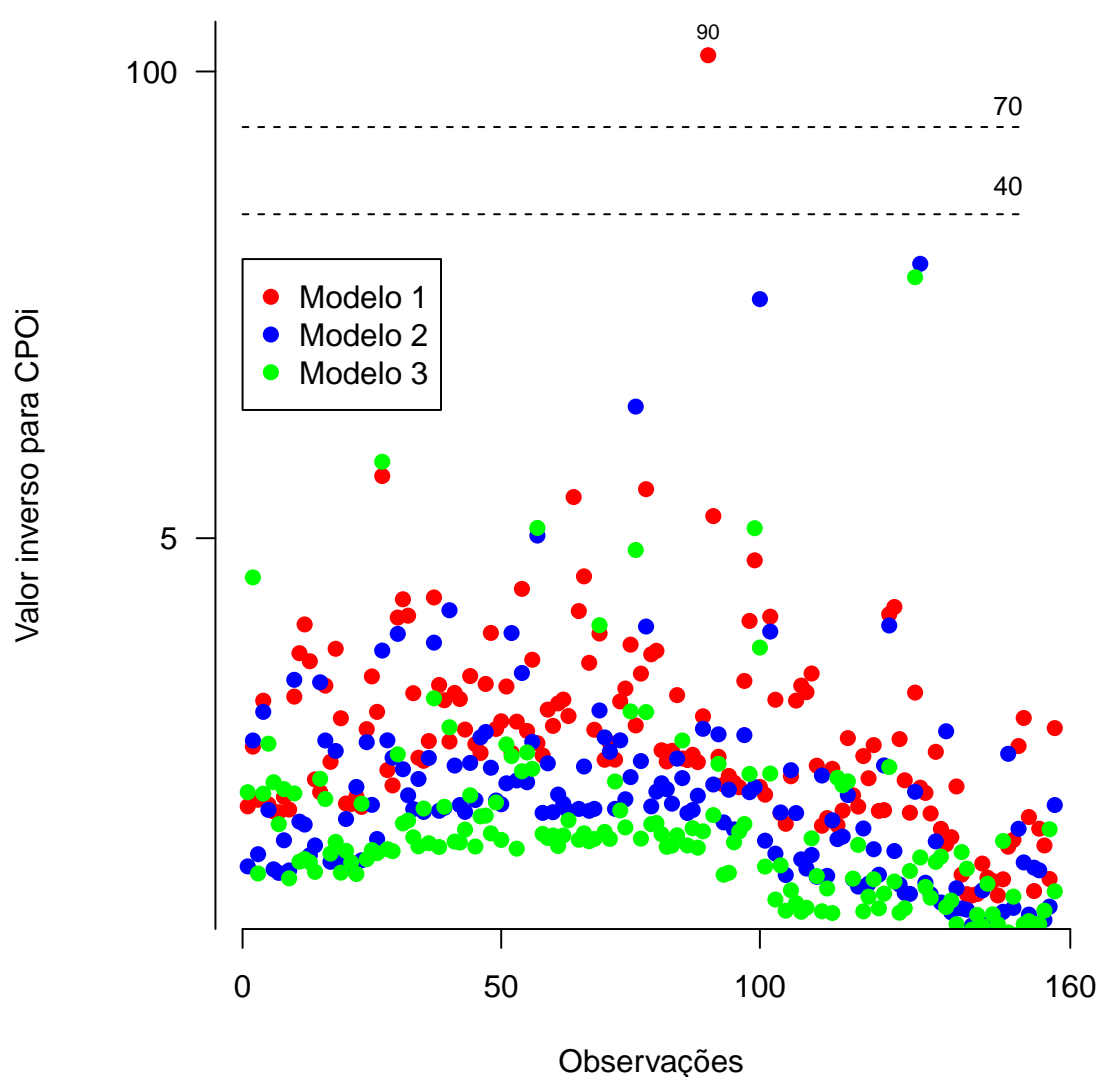


Figura 27 – Comparação entre os valores inversos para CPO obtidos com os Modelos 1, 2 e 3 com resíduos que seguem uma distribuição normal.

(seção 4.3) aos dados do número mensal de medicamentos para asma (xarope salbutamol frasco de 120ml) dispensados pelos serviços públicos de saúde de Ribeirão Preto, São Paulo, Brasil. O modelo pode ser usado para descrever uma grande variedade de séries de dados epidemiológicos, onde mais de um ponto de mudança está presente. Além disso, uma vantagem do modelo proposto sob uma abordagem Bayesiana é a possibilidade de incluir informações *a priori* que estão à disposição dos gestores de saúde pública.

---

## Capítulo 5

---

---

# Comentários Gerais

---

Ao longo do trabalho mostramos métodos voltados ao ajuste de modelos de séries temporais de dados epidemiológicos. Os primeiros modelos (Capítulo 2) aos últimos modelos (Capítulo 4) fomos ampliando o modelo com componentes que apresentavam ajustes melhores para as séries estudadas.

No Capítulo 2 apresentamos a série de Fourier (componente  $S_t$ ) e sua utilidade em ajustar séries que apresentam sazonalidades e comparamos modelos com e sem termos autorregressivos, mostrando a melhora no ajuste. No Capítulo 3 considerando o comportamento do banco de dados, acrescentamos um ponto de mudança, método muito utilizado para ajustar modelos epidemiológicos com mudanças de comportamento da série. Depois de um período crescente a série passa a ficar decrescente ou constante. Apresentamos no Capítulo 3 uma forma alternativa ao DIC (*Deviance Information Criterion*) para comparar ajuste de modelos a LPML (*Log Pseudo Marginal Likelihood*), útil em casos, como em modelos de mistura onde o DIC não é calculado.

No Capítulo 4 foi acrescentado mais pontos de mudança (3 pontos) e apresentado de uma forma que possa auxiliar o leitor a fazer análises com até mais pontos de mudança. Apesar de termos usado covariáveis (variáveis climáticas) nos modelos do Capítulo 3, no Capítulo 4 apresentamos uma discussão mais ampla sobre a componente  $S_t$  acomodar a periodicidade de possíveis covariáveis sazonais deixando o modelo mais parcimonioso e ainda com bom ajuste. Não descartamos o uso de covariáveis junto com a  $S_t$ , que pode encontrar a magnitude da variação além da sazonalidade anual, no caso das variáveis climáticas.

Em todos os modelos ajustados nesse trabalho foi utilizado o procedimento Bayesiano de simulação em Monte Carlo em Cadeias de Markov (MCMC) com o algoritmo de amostragem de Gibbs, para estimar as distribuições *a posteriori* dos parâmetros de interesse e as componentes de variância. Tratando-se de um procedimento muito utilizado

e de simples implementação (Códigos do OpenBugs no Apêndice), a inferência Bayesiana tem sido aplicada com êxito em diversos campos como na análise de séries epidemiológicas. Todos gráficos foram gerados no programa estatístico livre R.

Em trabalhos futuros podemos ajustar um modelo mais complexo para os dados do Capítulo 2 (notificações de acidente ofídicos no Estado de São Paulo, por exemplo), ajustando os três conjuntos de dados no mesmo modelo para verificar a sazonalidade conjunta dos dados. Podemos propor um modelo bivariado para ajustar a série temporal de dispensação de cápsulas aminofilina e xarope de salbutamol (banco de dados dos Capítulos 2 e 3), remédios para asma no mesmo período. Usaremos o mesmo procedimento Bayesiano de simulação (MCMC) com o algoritmo de amostragem de Gibbs para estimar os parâmetros. Vamos implementar um novo código no OpenBugs e analisar a convergência desses novos modelos de forma análoga a explicada nesse trabalho.

O modelo com um ou mais ponto de mudança pode ser usado para descrever uma grande variedade de séries de dados epidemiológicos.

---

## Referências

---

- ACHCAR, J.; MARTINEZ, E.; RUFFINO-NETTO, A.; PAULINO, C.; SOARES, P. A statistical model investigating the prevalence of tuberculosis in New York city using counting processes with two change-points. **Epidemiology and Infection**, Cambridge Univ Press, v. 136, n. 12, p. 1599–1605, 2008.
- ACHCAR, J. A.; RODRIGUES, E. R.; PAULINO, C. D.; SOARES, P. Non-homogeneous Poisson models with a change-point: an application to ozone peaks in Mexico City. **Environmental and Ecological Statistics**, Springer, v. 17, n. 4, p. 521–541, 2010.
- ACHCAR, J. A.; RODRIGUES, E. R.; TZINTZUN, G. Using non-homogeneous Poisson models with multiple change-points to estimate the number of ozone exceedances in Mexico City. **Environmetrics**, Wiley Online Library, v. 22, n. 1, p. 1–12, 2011.
- BARRY, D.; HARTIGAN, J. A. A Bayesian analysis for change point problems. **Journal of the American Statistical Association**, Taylor & Francis, v. 88, n. 421, p. 309–319, 1993.
- BEAULIEU, C.; CHEN, J.; SARMIENTO, J. L. Change-point analysis as a tool to detect abrupt climate variations. **Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences**, The Royal Society, v. 370, n. 1962, p. 1228–1249, 2012.
- BERTOLOZZI, M. R.; SCATENA, C. M. d. C.; FRANÇA, F. O. d. S. Vulnerabilities in snakebites in São Paulo, Brazil. **Revista de Saúde Pública**, SciELO Public Health, v. 49, 2015.
- BITTAR, O.; MENDES, J.; MAGALHÃES, A. Rede hospitalar no estado de São Paulo: mapear para regular. **São Paulo: Secretaria de Estado da Saúde de São Paulo**, 2011.
- BOCHNER, R.; FISZON, J. T.; MACHADO, C. A profile of snake bites in Brazil, 2001 to 2012. **Journal of Clinical Toxicology**, OMICS Group, v. 2014, 2014.
- BOTEGA, A.; SANTOS, M.; NETO, A. C. Descentralização das ações de assistência farmacêutica: asma e rinite. **Asma em saúde pública**, Manole São Paulo, 2007.
- BOWIE, C.; PROTHERO, D. Finding causes of seasonal diseases using time series analysis. **International Journal of Epidemiology**, IEA, v. 10, n. 1, p. 87–92, 1981.
- CARLIN, B. P.; GELFAND, A. E.; SMITH, A. F. Hierarchical Bayesian analysis of changepoint problems. **Applied Statistics**, JSTOR, p. 389–405, 1992.

- CARLIN, B. P.; LOUIS, T. A. Bayes and empirical Bayes methods for data analysis. **Statistics and Computing**, Springer, v. 7, n. 2, p. 153–154, 1997.
- CASELLA, G.; GEORGE, E. I. Explaining the Gibbs sampler. **The American Statistician**, Taylor & Francis, v. 46, n. 3, p. 167–174, 1992.
- CHEN, M.-H.; SHAO, Q.-M.; IBRAHIM, J. G. **Monte Carlo methods in Bayesian computation**. [S.l.]: Springer Science & Business Media, 2012.
- CHIB, S. Estimation and comparison of multiple change-point models. **Journal of Econometrics**, Elsevier, v. 86, n. 2, p. 221–241, 1998.
- CRUZ, A.; PONTE, E. V. Inflamação nas pequenas vias aéreas em asmáticos. **Revista Brasileira Alergia Immunopatologia**, v. 26, p. 25–32, 2003.
- CUPO, P. Bites and stings from venomous animals: a neglected Brazilian tropical disease. **Revista da Sociedade Brasileira de Medicina Tropical**, SciELO Brasil, v. 48, n. 6, p. 639–641, 2015.
- DALCIN, P. d. T. R.; BARRETO, S. S. M.; MEDEIROS, A. C.; SIQUEIRA, M. K.; MALLMANN, F.; LACERDA, M. A.; GAZZANA, M. B. Asma aguda em adultos na sala de emergência: o manejo clínico na primeira hora. **Jornal de Pneumologia**, v. 26, n. 6, p. 297–306, 2000.
- DEY, D. K.; PURKAYASTHA, S. Bayesian approach to change point problems. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 26, n. 8, p. 2035–2047, 1997.
- DOURADO, I.; VERAS, M. A. d. S.; BARREIRA, D.; BRITO, A. M. d. AIDS epidemic trends after the introduction of antiretroviral therapy in brazil. **Revista de Saúde Pública**, SciELO Brasil, v. 40, p. 9–17, 2006.
- FEARNHEAD, P. Exact and efficient Bayesian inference for multiple changepoint problems. **Statistics and Computing**, Springer, v. 16, n. 2, p. 203–213, 2006.
- 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. d. L.; LACERDA, M. V. G. d.; 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**, SciELO Brasil, v. 48, p. 34–41, 2015.
- FILHO, M. B.; RISSIN, A. A transição nutricional no Brasil: tendências regionais e temporais. **Caderno de Saúde Pública**, SciELO Brasil, v. 19, n. Supl 1, p. 181–91, 2003.
- FRANCISCO, P. M. S. B.; DONALISIO, M. R. de C. et al. Impacto da vacinação contra influenza na mortalidade por doenças respiratórias em idosos. **Revista de Saúde Pública**, v. 39, n. 1, p. 75–81, 2005.
- GEISSER, S.; EDDY, W. F. A predictive approach to model selection. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 74, n. 365, p. 153–160, 1979.
- GELFAND, A. E.; DEY, D. K.; CHANG, H. Model determination using predictive distributions with implementation via sampling-based methods. **Department of Statistics - Stanford University**, n. TR-462, 1992.



- GELMAN, A.; CARLIN, J. B.; STERN, H. S.; RUBIN, D. B. **Bayesian data analysis**. [S.l.]: Taylor & Francis, 2014. v. 2.
- GHOSH, P.; VAIDA, F. Random changepoint modelling of HIV immunologic responses. **Statistics in Medicine**, Wiley Online Library, v. 26, n. 9, p. 2074–2087, 2007.
- LANGE, K. L.; LITTLE, R. J.; TAYLOR, J. M. Robust statistical modeling using the t distribution. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 84, n. 408, p. 881–896, 1989.
- LAVIELLE, M.; LEBARBIER, E. An application of MCMC methods for the multiple change-points problem. **Signal Processing**, Elsevier, v. 81, n. 1, p. 39–53, 2001.
- LEE, T.-S. Change-point problems: bibliography and review. **Journal of Statistical Theory and Practice**, Taylor & Francis, v. 4, n. 4, p. 643–662, 2010.
- LUNN, D. J.; THOMAS, A.; BEST, N.; SPIEGELHALTER, D. WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. **Statistics and Computing**, Springer, v. 10, n. 4, p. 325–337, 2000.
- MARTINEZ, E. Z.; SILVA, E. A. S. d. Predicting the number of cases of dengue infection in Ribeirão Preto, São Paulo State, Brazil, using a SARIMA model. **Cadernos de Saúde Pública**, SciELO Public Health, v. 27, n. 9, p. 1809–1818, 2011.
- MARTINS, A. S.; SANTOS, J. K. C.; BATISTA, N. de S.; OLIVEIRA, M. A. O. de; OLIVEIRA, A. T. C.; ALENCAR, J. R. B. de. Monitoramento do processo de envase de salbutamol xarope utilizando controle estatístico. **Revista Brasileira de Farmácia**, v. 89, n. 2, p. 102–108, 2008.
- MODARRES, R.; OUARDA, T. B.; VANASSE, A.; ORZANCO, M. G.; GOSSELIN, P. Modeling seasonal variation of hip fracture in montreal, canada. **Bone**, Elsevier, v. 50, n. 4, p. 909–916, 2012.
- NOBRE, F. F.; MONTEIRO, A. B. S.; TELLES, P. R.; WILLIAMSON, G. D. Dynamic linear model and sarima: a comparison of their forecasting performance in epidemiology. **Statistics in Medicine**, Wiley Online Library, v. 20, n. 20, p. 3051–3069, 2001.
- NTZOUFRAS, I. **Bayesian modeling using WinBUGS**. [S.l.]: John Wiley & Sons, 2011. v. 698.
- NUNES, M. L.; PINHO, A. P. S.; AERTS, D.; SANT'ANNA, A.; MARTINS, M. P.; COSTA, J. C. d. Sudden infant death syndrome: clinical aspects of an underdiagnosed disease. **Jornal de Pediatria**, SciELO Brasil, v. 77, n. 1, p. 29–34, 2001.
- PETERSON, M. G.; GAETA, T. J.; BIRKHAHN, R. H.; FERNÁNDEZ, J. L.; MANCUSO, C. A. History of symptom triggers in patients presenting to the emergency department for asthma. **Journal of Asthma**, Taylor & Francis, v. 49, n. 6, p. 629–636, 2012.
- RIBEIRO, L.; ALBUQUERQUE, M.; CAMPOS, V. Pires de; KATZ, G.; TAKAOKA, N.; LEBRAO, M.; JORGE, M. Obits by venomous snakes in the State of São Paulo: evaluation of 43 cases from 1988 to 1993. **Revista da Associação Médica Brasileira**, SciELO Brasil, 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**, SciELO Brasil, v. 30, n. 6, p. 475–480, 1997.
- RUGGERI, F.; KENETT, R.; FALTIN, F. W. **Encyclopedia of Statistics in Quality and Reliability**. [S.l.]: John Wiley, 2007.
- SCHÜTZ, N.; HOLSCHNEIDER, M. Detection of trend changes in time series using Bayesian inference. **Physical Review E**, APS, v. 84, n. 2, p. 021120, 2011.
- SHUMWAY, R. H.; STOFFER, D. S. Time series regression and exploratory data analysis. **Time Series Analysis and Its Applications: With R Examples**, Springer, p. 48–83, 2006.
- SPIEGELHALTER, D. J.; BEST, N. G.; CARLIN, B. P.; LINDE, A. V. D. Bayesian measures of model complexity and fit. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 64, n. 4, p. 583–639, 2002.
- STOLWIJK, A.; STRAATMAN, H.; ZIELHUIS, G. Studying seasonality by using sine and cosine functions in regression analysis. **Journal of Epidemiology and Community Health**, BMJ Publishing Group Ltd, v. 53, n. 4, p. 235–238, 1999.
- THOMAZELLI, L. M.; VIEIRA, S.; LEAL, A. L.; SOUSA, T. S.; OLIVEIRA, D. B.; GOLONO, M. A.; GILLIO, A. E.; STWIEN, K. E.; ERDMAN, D. D.; DURIGON, E. L. Surveillance of eight respiratory viruses in clinical samples of pediatric patients in Southeast Brazil. **Jornal de Pediatria**, SciELO Brasil, v. 83, n. 5, p. 422–428, 2007.
- TOMÉ, A.; MIRANDA, P. Piecewise linear fitting and trend changing points of climate parameters. **Geophysical Research Letters**, Wiley Online Library, v. 31, n. 2, 2004.

---

## Apêndice

---

O código OpenBugs usado para especificar o modelo estatístico, na sua forma geral e com resíduos que seguem uma distribuição normal é dado abaixo. Observações da variável independente são indicadas por **x[t]**. Além disso, o **cp** indica o ponto de mudança de valor e **N** é o comprimento da série temporal. Na sequência temos o código para os modelos do primeiro, segundo e terceiro capítulos desse trabalho.

```
#Modelo referente ao Capítulo 2
model
{
  for (t in 2:N)
  {
    y[t] ~ dpois(theta[t])
    log(theta[t]) <- alpha + beta*t + S[t] + gama1* (y[t-1] - mean(y[]))
    S[t] <- S1[t] + S2[t]
    S1[t] <- eta[1]*sin(2*pi*t/12) + eta[2]*cos(2*pi*t/12)
    S2[t] <- eta[3]*sin(4*pi*t/12) + eta[4]*cos(4*pi*t/12)
  }

  alpha ~ dnorm(0,0.00001)
  beta ~ dnorm(0,0.00001)
  eta[1] ~ dnorm(0,0.00001)
  eta[2] ~ dnorm(0,0.00001)
  eta[3] ~ dnorm(0,0.00001)
  eta[4] ~ dnorm(0,0.00001)
  gama1 ~ dunif(-1,1)
}

#Modelo referente ao Capítulo 3
model
```

```

{
  for(t in 1:N)
  {
    y[t] ~ dnorm(mu[t], tau[J[t]])
    mu[t] <- alpha + beta[J[t]]* (t-cp)
    + gama[J[t]]* (w[t] - mean(w[])) + St[t]
    + phi[J[t]]*(x[t] - mean(x[]))
    k[t] <- step(t - cp - 0.5)
    J[t] <- 1 + k[t]
    punif[t] <- 1/N
    St[t] <- eta[1]*sin(2*pi*t/12) + eta[2]*cos(2*pi*t/12)
# Likelihood function
    L[t] <- 1/sqrt(2*pi*(pow(sigma[1],1-k[t])
    * pow(sigma[2],k[t])))
    * exp(-(y[t]-mu[t])*(y[t]-mu[t])
    /(2*(pow(sigma[1],1-k[t]) * pow(sigma[2],k[t]))))
# Inverse values for CP0
    P0[t] <- 1/L[t]
  }
  for(i in 2:N)
  {
    w[i] <- y[i-1]
  }
  w[1] <- y[1]
  pi <- 3.14159265359
# Distribuição a priori
  prec <- 1.0E-6
  alpha ~ dnorm(0.0, prec)
  cp ~ dcat(punif[])
  for(j in 1:2)
  {
    beta[j] ~ dnorm(0.0, prec)
    eta[j] ~ dnorm(0.0, prec)
    gama[j] ~ dnorm(0.0, prec)
    phi[j] ~ dnorm(0.0, prec)
    tau[j] ~ dgamma(0.1,0.1)
    sigma[j] <- 1/tau[j]
  }
}

#Modelo referente ao Capítulo 4
model
{
  for(i in 1:N)
  {
    y[i] ~ dnorm(mu[i], tau[J[i]])

```

```

mu[i] <- alpha0[J[i]] + beta[J[i]]*(t[i] - H[J[i]] )
      + gama[J[i]]* (x[i] - mean(x[])) + St[i]
C1[i] <- equals(min(i,cp[1]),i) * equals(min(1,i-1),1)
C2[i] <- equals(min(i,cp[2]),i) * equals(min(cp[1],i-1),cp[1])
C3[i] <- equals(min(i,cp[3]),i) * equals(min(cp[2],i-1),cp[2])
C4[i] <- equals(min(i,N),i)      * equals(min(cp[3],i-1),cp[3])

J[i] <- pow(1,C1[i]) * pow(2,C2[i]) * pow(3,C3[i]) * pow(4,C4[i])

punif1[i] <- (1/(30-20+1)) * equals(min(i,30),i) * equals(min(20,i),20)
punif2[i] <- (1/(105-90+1)) * equals(min(i,105),i) * equals(min(90,i),90)
punif3[i] <- (1/(140-120+1)) * equals(min(i,140),i) * equals(min(120,i),120)

St[i] <- eta[1]*sin(2*pi*t[i]/12) + eta[2]*cos(2*pi*t[i]/12)

L[i] <- 1/sqrt(2*3.14159265*sigma[J[i]] ) *
exp( -(y[i]-mu[i])*(y[i]-mu[i])/(2* sigma[J[i]] ))

PO[i] <- 1/L[i]

}

for(i in 2:N)
{
  x[i] <- y[i-1]
}
x[1] <- y[1]

H[1] <-0
H[2] <- cp[1]
H[3] <- cp[2]
H[4] <- cp[3]

prec<-1.0E-6

for (j in 1:4)
{

  beta[j] ~ dnorm(0.0, prec)
  tau[j] ~dgamma(.1,.1)
  sigma[j] <- 1/tau[j]

}

```

```
pi <- 3.14159265359
alpha0[1] ~ dnorm(0,prec)
alpha0[2] <- alpha0[1] + beta[1]*cp[1]
alpha0[3] <- alpha0[2] + beta[2]*(cp[2]-cp[1])
alpha0[4] <- alpha0[3] + beta[3]*(cp[3]-cp[2])
gama[1] ~ dnorm(0.0, prec)
gama[2] ~ dnorm(0.0, prec)
gama[3] ~ dnorm(0.0, prec)
gama[4] ~ dnorm(0.0, prec)
eta[1] ~ dnorm(0.0, prec)
eta[2] ~ dnorm(0.0, prec)

cp[1] ~ dcat(punif1[])
cp[2] ~ dcat(punif2[])
cp[3] ~ dcat(punif3[])

}
```