

**Modelos paramétricos para séries temporais
de contagem**

Igor André Milhorança

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO
DE
MESTRE EM CIÊNCIAS

Programa: Estatística

Orientadora: Profa. Dra. Airlane Pereira Alencar

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da CNPq

São Paulo, maio de 2014

Modelos paramétricos para séries temporais de contagem

Esta versão da dissertação/tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 14/05/2014. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof^a. Dr^a. Airlane Pereira Alencar (orientadora) - IME-USP
- Prof^a Dr^a Clelia Maria de Castro Toloi - IME-USP
- Prof^a Dr^a Adriana Bruscato Bortoluzzo - INSPER

Agradecimentos

Agradeço à minha família pelo apoio, às professoras Adriana Bruscato Bortoluzzo, Chang Chiann e Clelia Maria de Castro Toloí pela atenção dedicada à correção da minha dissertação, aos meus amigos Aline Santos Damascena, Elivane da Silva Victor e Orlando Yesid Esparza Albarracin pela ajuda durante todo o período do mestrado e, principalmente, à professora Airlane Pereira Alencar, por estar sempre presente e pela excelente orientação desde a época da graduação.

Resumo

MILHORANÇA, I. A. **Modelos paramétricos para séries temporais de contagem**. 2014. 120 f. Dissertação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2014.

Diversas situações práticas exigem a análise de series temporais de contagem, que podem apresentar tendência, sazonalidade e efeitos de variáveis explicativas. A motivação de nosso trabalho é a análise de internações diárias por doenças respiratórias para pessoas com mais que 65 anos residentes no município de São Paulo. O efeito de variáveis climáticas e concentrações de poluentes foram incluídos nos modelos e foram usadas as funções seno e cosseno com periodicidade de um ano para explicar o padrão sazonal e obter os efeitos das variáveis climáticas e poluentes controlando essa sazonalidade. Outro aspecto a ser considerado é a inclusão da população nas análises de modo que a interpretação dos efeitos seja para as taxas diárias de internações.

Diferentes modelos paramétricos foram propostos para as internações. O mais simples é o modelo de regressão linear para o logaritmo das taxas. Foram ajustados os modelos lineares generalizados (MLG) para as internações com função de ligação logaritmo e com a população como offset, por este modelo permitir o uso das distribuições Poisson e Binomial Negativa, usadas para dados de contagem. Devido à heteroscedasticidade extra, foram propostos modelos GAMLSS incluindo variáveis para explicar o desvio padrão. Foram ajustados modelos ARMA e GARMA, por incluírem uma estrutura de correlação serial. O objetivo desse trabalho é comparar as estimativas, os erros padrões, a cobertura dos intervalos de confiança e o erro quadrático médio para o valor predito segundo os vários modelos e a escolha do modelo mais apropriado, que depende da completa análise de resíduos, geralmente omitida na literatura. O modelo GARMA com distribuição Binomial Negativa apresentou melhor ajuste, pois os erros parecem seguir a distribuição proposta e tem baixa autocorrelação, além de ter tido uma boa cobertura pelo intervalo de confiança e um baixo erro quadrático médio.

Também foi analisado o efeito da autocorrelação dos dados nas estimativas nos vários modelos baseado em dados simulados.

Palavras-chave: GARMA, GAMLSS, modelos lineares generalizados.

Abstract

MILHORANÇA, I. A. **Parametric models for count time series**. 2014. 120 f. Thesis (Master) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2014.

Many practical situations require the analysis of time series of counts, which may present trend, seasonality and effects of covariates. The motivation of this work is the analysis of daily hospital admissions for respiratory diseases in people over 65 living in the city of São Paulo. The effect of climatic variables and concentrations of pollutants were included in the models and the sine and cosine functions with annual period were included to explain the seasonal pattern and obtain the effects of pollutants and climatic variables partially controlled by this seasonality. Another aspect to be considered is the inclusion of the population in the analyses in order to interpret the effects based on daily hospitalization rates .

Different parametric models have been proposed for hospitalizations. The simplest is the linear regression model for the logarithm of the hospitalization rate. The generalized linear models (GLM) were adjusted for daily admissions with logarithmic link function and the population as offset to consider the Poisson and Negative Binomial distributions for counting data. Due to the extra heteroscedasticity, GAMLSS models were proposed including variables to explain the standard error. Moreover, the ARMA and GARMA models were fitted to include the serial correlation structure. The aim of this work is to compare estimates, standard errors, coverage of confidence intervals and mean squared error of predicted value for the various models and choose the most appropriate model, which depends on a complete analysis of residuals, usually omitted in the literature. The GARMA model with Negative Binomial distribution was the best fit since the errors seem to follow the proposed distribution and they have small values of autocorrelation. Besides, this model had low mean squared error and a good coverage of confidence interval.

The effect of autocorrelation of data in the estimates was also analyzed in the setting of several models based on simulated data.

Keywords: GARMA, GAMLSS, generalized linear models.

Sumário

Lista de Abreviaturas	ix
Lista de Símbolos	xi
Lista de Figuras	xiii
Lista de Tabelas	xv
1 Introdução	1
2 Metodologia	5
2.1 Regressão linear	7
2.2 Modelo autorregressivo e de média móvel - ARMA	8
2.3 Modelos Lineares Generalizados	9
2.4 Modelos generalizados autorregressivos de médias móveis - GARMA	11
2.5 Modelos aditivos generalizados para posição, escala e forma - GAMLSS	12
2.6 Resíduos	14
2.7 Erro quadrático médio e intervalo de confiança	15
2.8 Imputação	16
3 Aplicação	17
3.1 Dados utilizados	17
3.2 Análise descritiva	17
3.3 Análise Inferencial	22
3.3.1 Ajuste dos modelos	23
3.3.2 Análise de resíduos	26
3.3.3 Ausência de correlação serial dos resíduos	26
3.3.4 Homocedasticidade	27
3.3.5 Normalidade	28
3.3.6 Erro quadrático médio e intervalo de confiança	28
3.4 Escolha e ajuste do modelo final	29
4 Simulação	33
5 Conclusões e perspectivas	49

6 Apêndice	51
6.1 Gráficos do núcleo da densidade	51
6.2 Figuras	52
6.3 Tabelas	70
Referências Bibliográficas	75

Lista de Abreviaturas

ARMA	Modelo autorregressivo e média móvel
COmax	Gás carbônico máximo
COmed	Gás carbônico médio
corr	Correlação linear de Pearson
EQM	Erro quadrático médio
GAMLSS	Modelo generalizado aditivo para localização, escala e forma
GARMA	Modelo generalizado autorregressivo e média móvel
MCMC	Monte Carlo via Cadeia de Markov
MLG	Modelo linear generalizado
O3max	Gás ozônio máximo
O3med	Gás ozônio médio
PM10max	Material particulado máximo
PM10med	Material particulado médio
SE	Erro padrão
Tmax	Temperatura máxima
Tmed	Temperatura média
Tmin	Temperatura mínima
Umax	Umidade relativa do ar máxima
Umed	Umidade relativa do ar média
Umin	Umidade relativa do ar mínima

Lista de Símbolos

y_t	contagem observada no dia t
δ_t	variável utilizada como offset
Pop_t	população no instante t
tx_t	taxa correspondente às contagens, como por exemplo, $tx_t = \frac{y_t}{Pop_t/100000}$
T	número de observações, sendo que o índice $t = 1, \dots, T$
\mathbf{x}_t	vetor $m \times 1$ composto pelas p variáveis explicativas observadas no instante t
\mathbf{X}	matriz $m \times T$ composta pelas p variáveis explicativas para todas as T observações
β	vetor $m \times 1$ de parâmetros correspondentes às variáveis explicativas
μ_t	valor esperado da variável resposta
σ_t	Variância da variável resposta
ϑ_t	parâmetro canônico
ϑ_2	o parâmetro de dispersão
η	preditor linear
\mathbf{H}_t	vetor de informação passada $\mathbf{H}_t = \{\mathbf{x}_t, \dots, \mathbf{x}_1, y_t, \dots, y_1, \mu_t, \dots, \mu_1\}$
ϕ	vetor de parâmetros autorregressivos
θ	vetor de parâmetros de médias móveis
B	operador backshift, $B^k y_t = y_{t-k}$

Lista de Figuras

3.1	Séries diárias	19
3.2	Box-plot da variável internação dessazonalizada segundo dia da semana	20
3.3	Box-plot da variável internação dessazonalizada segundo feriado	21
3.4	Análise dos resíduos padronizados do modelo reduzido	31
3.5	FAC dos resíduos e dos resíduos ao quadrado do modelo reduzido.	32
3.6	Intervalo de confiança do modelo reduzido.	32
6.1	Gráficos de dispersão entre a variável logaritmo da taxa de internações e as variáveis de poluição e climáticas.	52
6.2	Análise dos resíduos padronizados do modelo regressão linear.	53
6.3	Análise dos resíduos padronizados do modelo ARMA.	53
6.4	Análise dos resíduos padronizados do modelo MLG Normal.	54
6.5	Análise dos resíduos padronizados do modelo MLG Poisson.	54
6.6	Análise dos resíduos padronizados do modelo MLG Binomial Negativa.	55
6.7	Análise dos resíduos padronizados do modelo GARMA Normal.	55
6.8	Análise dos resíduos padronizados do modelo GARMA Poisson.	56
6.9	Análise dos resíduos padronizados do modelo GARMA Binomial Negativa.	56
6.10	Análise dos resíduos padronizados do modelo 1 - GAMLSS Normal.	57
6.11	Análise dos resíduos padronizados do modelo 2 - GAMLSS Normal.	57
6.12	Análise dos resíduos padronizados do modelo 3 - GAMLSS Normal.	58
6.13	Análise dos resíduos padronizados do modelo 4 - GAMLSS Binomial Negativa.	58
6.14	Análise dos resíduos padronizados do modelo 5 - GAMLSS Binomial Negativa.	59
6.15	Análise dos resíduos padronizados do modelo 6 - GAMLSS Binomial Negativa.	59
6.16	FAC dos resíduos padronizados dos modelos regressão linear e ARMA.	60
6.17	FAC dos resíduos padronizados dos modelos MLG.	60
6.18	FAC dos resíduos padronizados dos modelos GARMA.	61
6.19	FAC dos resíduos padronizados dos modelos GAMLSS com distribuição Normal.	61
6.20	FAC dos resíduos padronizados dos modelos GAMLSS com distribuição Binomial Negativa.	62
6.21	FAC dos resíduos ao quadrado dos modelos regressão linear e ARMA.	62
6.22	FAC dos resíduos ao quadrado dos modelos MLG.	63
6.23	FAC dos resíduos ao quadrado dos modelos GARMA.	63
6.24	FAC dos resíduos ao quadrado dos modelos GAMLSS com distribuição Normal.	64

6.25	FAC dos resíduos ao quadrado dos modelos GAMLSS com distribuição Binomial Negativa.	64
6.26	Intervalo de confiança dos modelos regressão linear e ARMA.	65
6.27	Intervalo de confiança dos modelos MLG.	65
6.28	Intervalo de confiança dos modelos GARMA.	66
6.29	Intervalo de confiança dos modelos GAMLSS com distribuição Normal.	66
6.30	Intervalo de confiança dos modelos GAMLSS com distribuição Binomial Negativa.	67
6.31	Desvio padrão dos modelos regressão linear e ARMA.	67
6.32	Desvio padrão dos modelos MLG.	68
6.33	Desvio padrão dos modelos GARMA.	68
6.34	Desvio padrão dos modelos GAMLSS com distribuição Normal.	69
6.35	Desvio padrão dos modelos GAMLSS com distribuição Binomial Negativa.	69

Lista de Tabelas

3.1	Medidas resumo do número de internações segundo dias da semana e feriado	20
3.2	Estimativas dos modelos de regressão linear e ARMA.	23
3.3	Estimativas dos modelos lineares generalizados.	24
3.4	Estimativas dos modelos GARMA.	24
3.5	Estimativas dos modelos GAMLSS com distribuição Normal.	25
3.6	Estimativas dos modelos GAMLSS com distribuição Binomial Negativa.	25
3.7	Variáveis significativas na regressão auxiliar do resíduo ao quadrado.	27
3.8	Estimativas do modelo final reduzido.	30
4.1	Estimativas do modelo GARMA(1,1) com distribuição Binomial Negativa da Tabela 3.4.	34
4.2	Número de modelos com erro na estimação.	34
4.3	Medidas resumo das estimativas das simulações com distribuição Normal e $\phi = 0, 0$. .	37
4.4	Medidas resumo dos erros padrões das simulações com distribuição Normal e $\phi = 0, 0$.	38
4.5	Medidas resumo das estimativas das simulações com distribuição Binomial Negativa e $\phi = 0, 0$	39
4.6	Medidas resumo dos erros padrões das simulações com distribuição Binomial Negativa e $\phi = 0, 0$	40
4.7	Medidas resumo das estimativas das simulações com distribuição Normal e $\phi = 0, 3$. .	41
4.8	Medidas resumo dos erros padrões das simulações com distribuição Normal e $\phi = 0, 3$.	42
4.9	Medidas resumo das estimativas das simulações com distribuição Binomial Negativa e $\phi = 0, 3$	43
4.10	Medidas resumo dos erros padrões das simulações com distribuição Binomial Negativa e $\phi = 0, 3$	44
4.11	Medidas resumo das estimativas das simulações com distribuição Normal e $\phi = 0, 5$. .	45
4.12	Medidas resumo dos erros padrões das simulações com distribuição Normal e $\phi = 0, 5$.	46
4.13	Medidas resumo das estimativas das simulações com distribuição Binomial Negativa e $\phi = 0, 5$	47
4.14	Medidas resumo dos erros padrões das simulações com distribuição Binomial Negativa e $\phi = 0, 5$	48
6.1	Correlação de Pearson para as variáveis de poluição, climáticas e logaritmo da taxa de internações.	70
6.2	Correlação de Pearson para as variáveis sazonais e Feriados.	71
6.3	Tabela descritiva de internação e variáveis de poluição e climáticas.	71

6.4	Estimativa pontual, erro padrão e valor p dos efeitos das variáveis do modelos de regressão linear simples para os resíduos ao quadrado dos modelos de regressão linear e ARMA.	71
6.5	Estimativa pontual, erro padrão e valor p dos efeitos das variáveis do modelos de regressão linear simples para os resíduos ao quadrado dos modelos MLG.	72
6.6	Estimativa pontual, erro padrão e valor p dos efeitos das variáveis do modelos de regressão linear simples para os resíduos ao quadrado dos modelos GARMA.	72
6.7	Estimativa pontual, erro padrão e valor p dos efeitos das variáveis do modelos de regressão linear simples para os resíduos ao quadrado dos modelos GAMLSS com distribuição Normal.	72
6.8	Estimativa pontual, erro padrão e valor p dos efeitos das variáveis do modelos de regressão linear simples para os resíduos ao quadrado dos modelos GAMLSS com distribuição Binomial Negativa.	73
6.9	Medidas de diagnóstico.	73
6.10	Erro quadrático médio.	73
6.11	Tabela descritiva para o desvio padrão.	74
6.12	Proporção de pontos fora do intervalo de confiança.	74

Capítulo 1

Introdução

Na área epidemiológica, é importante analisar mortalidade e morbidade por várias causas ao longo do tempo. Em geral, tais eventos são medidos por taxas de mortalidade, internações ou outros procedimentos hospitalares ou ambulatoriais. Em todos esses casos, o evento de interesse consiste em uma contagem ao longo do tempo e é utilizada a população de interesse para o cálculo das taxas. Por se tratar de uma série temporal, tais taxas podem, por exemplo, apresentar padrões sazonais, tendências e mudança de nível.

Outra questão muito discutida na literatura é a associação entre variáveis climáticas e de poluição e as taxas de mortalidade e morbidade. Diversos estudos concluem que quanto maior a concentração de poluentes e menores a temperatura e a umidade relativa do ar, maiores são as taxas de internação. No nosso trabalho analisaremos a taxa diária de internação por doenças respiratórias de pessoas com 65 anos e mais residentes no município de São Paulo de 2006 a 2011. Os dados foram obtidos pela Secretaria de Saúde do Município de São Paulo.

A princípio, as áreas mais interessadas nesse tipo de análise são profissionais de saúde pública, como secretarias de saúde. Nessas áreas, quanto mais simples os modelos que consigam levar a conclusões corretas, melhor. A priori, poderia ser proposto um modelo de regressão linear múltiplo para as taxas. Entretanto, espera-se que a suposição de homoscedasticidade não seja válida, já que quanto maiores as médias das internações, maior deve ser a variabilidade. Para tal comportamento, é usual propor a distribuição de Poisson para contagens. Para variáveis com tal distribuição, a variância é igual a média, o que já é indício de heteroscedasticidade. Em modelos de regressão, é possível, utilizando os modelos lineares generalizados (MLG), assumir a distribuição Poisson para os dados. Tais modelos permitem propor que uma função da média dependa da função linear dos parâmetros para distribuições da família exponencial. Também é possível incluir variáveis com coeficientes conhecidos, chamados de offset, que permitem, por exemplo, a modelagem de taxas esperadas a partir das médias de contagens. Diversos estudos analisam dados de mortalidade e morbidade utilizando a distribuição Poisson, como [Katsouyanni *et al.* \[1995\]](#).

É muito usual na prática observar situações em que a variância da variável resposta é maior que a média, tal fato é chamado de superdispersão. Assim, é necessária a inclusão de outro parâmetro que permita considerar essa variância extra. Para os dados de contagem foi proposto o modelo com distribuição Binomial Negativa, como extensão do modelo Poisson, de modo que se o parâmetro de dispersão for igual a 0, temos o caso particular da distribuição Poisson. Muitos trabalhos ajustam modelos com distribuição binomial negativa, como [Van den Bergh *et al.* \[2008\]](#) Maiores informações sobre o ajuste de modelos lineares generalizados às distribuições Normal, Poisson e Binomial Negativa podem ser encontradas em [Paula \[2013\]](#).

Outros modelos muito utilizados para analisar taxa de internações em função de variáveis climáticas e de poluição, como em [Conceição *et al.* \[2001\]](#), são os modelos aditivos generalizados (MAG) ([Hastie e Tibshirani \[1990\]](#)). Nesses modelos é possível propor que a média da taxa seja uma função das variáveis explicativas. Tais funções podem ser, por exemplo, splines. Assim, o efeito do acréscimo de uma variável explicativa na taxa média de internação pode variar dependendo do valor da variável explicativa. Os modelos aditivos generalizados são chamados de modelos não paramétricos,

já que não dependem somente da estimação de parâmetros, mas de funções inteiras. Isso complica a interpretação dos efeitos e, por isso, nosso trabalho se restringe aos modelos paramétricos.

Ao utilizarmos modelos lineares generalizados para a média, temos que a variância da variável resposta pode ser função da média, dependendo da distribuição utilizada. Muitas vezes tal dependência não é suficiente para explicar a variabilidade e pode ser necessário propor um modelo para a variância em função de variáveis explicativas. Nesse contexto, o modelo semiparamétrico aditivo generalizado para localização, escala e forma (GAMLSS=Generalized Additive Models for Location, Scale and Shape) foi proposto por [Rigby e Stasinopoulos \[2005\]](#). O GAMLSS permite a modelagem não só da média, mas de todos os parâmetros da distribuição, como a variância, assimetria e curtose, usando variáveis explicativas. Isso é interessante tanto para os modelos com distribuição Poisson e Binomial Negativa, mas principalmente para a Normal, que, em geral, é proposta sob a suposição de homocedasticidade.

É necessário avaliar a correlação serial ao ajustar modelos para as taxas ao longo do tempo, já que é esperado que a taxa de um dia seja correlacionada com taxas passadas. Os modelos autorregressivos e médias móveis (ARMA) propostos por [Box e Jenkins \[1976\]](#) incluem observações e ruídos passados no modelo de modo que dados em instantes mais próximos tenham autocorrelação mais alta, que dados mais distantes no tempo. É possível incluir variáveis explicativas para modelar a média da variável resposta e é comum utilizar também a denominação ARMAX ([Shumway e Stoffer \[2000\]](#)).

Outra proposta que contempla a possível presença de correlação serial são os modelos generalizados autorregressivos e médias móveis (GARMA) ([Benjamin *et al.* \[2003\]](#)). Tal modelo une a flexibilidade do modelo linear generalizado, que permite assumir distribuições como Normal, Poisson e Binomial Negativa, e tem a vantagem de modelar a autocorrelação, problema comum em séries temporais.

Muitos modelos podem ser propostos e obteremos estimadores não viesados, mas ao não considerarmos tanto a heteroscedasticidade, quanto a autocorrelação, os erros padrões do modelo não são estimados corretamente, podendo ser super ou subestimados ([Wooldridge \[2012\]](#)). Logo, uma preocupação do nosso trabalho é comparar não somente as estimativas dos parâmetros, mas, também os erros padrões e verificar se os modelos diferem muito entre si com relação a tais erros.

No nosso estudo, como esperado, as taxas de internação apresentam maiores valores no inverno e menores no verão. O padrão sazonal pode ser em parte explicado utilizando funções seno e cosseno como em [Serfling \[1963\]](#). O objetivo de se incluir tal padrão sazonal no modelo não é somente para explicar a sazonalidade, mas evitar confundimentos do efeito da sazonalidade com o efeito de covariáveis com padrão sazonal e, assim, possamos analisar o excesso de internações com relação ao modelo de referência com seno e cosseno. Diversos modelos para dados ao longo do tempo utilizam o modelo de Serfling como em [Souza *et al.* \[2013\]](#) e [Linares e Díaz \[2008\]](#).

O principal objetivo do nosso trabalho é propor e comparar diferentes modelos paramétricos para séries temporais de internações (contagem). Todos os modelos devem ser equivalentes no sentido de permitir a interpretação dos parâmetros como variações percentuais nas taxas de internação em função do acréscimo de uma unidade de cada variável explicativa. Tal comparação não se restringe somente às estimativas dos modelos mas também aos erros padrões destes. Além disso, também será apresentada uma análise do efeito da autocorrelação sobre os resultados dos modelos com base em simulações de modelos com coeficientes fixados de correlação entre as observações de determinado dia e do dia anterior.

O trabalho está dividido em 4 capítulos. No Capítulo 2, apresentamos e detalhamos toda a metodologia utilizada, incluindo os modelos ajustados nesse trabalho e suas propriedades e limitações. O Capítulo 3, foi dividido em três partes. Na Seção 3.1, apresentamos o banco de dados reais referentes ao número diário de hospitalizações por problemas respiratórios de pessoas com mais de 65 anos na cidade de São Paulo. Na Seção 3.2, estão análise descritiva desse conjunto de dados. Na Seção 3.3, mostramos os resultados da estimação dos modelos propostos no capítulo anterior e a análise de resíduos para estes modelos. No Capítulo 4, os problemas causados pela autocorrelação são mais explorados, com base na simulação de séries com características semelhantes às do banco de dados reais, incluindo autocorrelação, utilizando o modelo GARMA(1,0) com diferentes parâme-

tros autorregressivos. A partir dos dados simulados, foram comparadas as estimativas com relação ao valor verdadeiro e também comparamos os erros padrões obtidos no ajuste de modelos MLG e GARMA com distribuições Normal e Binomial Negativa.

Capítulo 2

Metodologia

Neste capítulo são apresentados os modelos propostos para séries temporais de contagem incluindo variáveis explicativas. As distribuições investigadas são: Normal, por ser mais simples, e Poisson e Binomial Negativa, por se tratarem de distribuições para dados de contagem.

Na prática, é usual analisar a taxa de ocorrências por unidade de medida. Em estudos em que a variável resposta é o número de óbitos e internações observado ao longo do tempo, é imprescindível levar em conta que a população não é constante, então é preferível analisar a taxa de óbitos ou internações do que o número de ocorrências. Uma primeira abordagem consiste em analisar diretamente a taxa de óbitos ou internações por 100 mil habitantes como variável resposta.

Outra questão levantada é que ao se aumentar uma unidade em cada variável explicativa; por exemplo, a variação de uma unidade em uma variável de poluição, talvez não acarrete diretamente em uma variação de alguns pontos na taxa, mas, sim, em uma variação percentual da taxa. Isso é mais usual já que para valores altos da taxa, os aumentos são maiores do que para valores mais baixos. Esse conceito de elasticidade, muito utilizado na economia, é também bastante usado em modelos para taxas de morbimortalidade (Wooldridge [2012]). Portanto, os modelos de regressão linear podem ter o logaritmo da taxa como variável resposta e, nesse caso, o modelo explica a esperança (média) do logaritmo da taxa de internação, mas é importante ressaltar que os efeitos das variáveis explicativas serão variações percentuais das taxa e não do logaritmo da taxa.

Uma alternativa à modelagem direta das taxas é propor modelos para contagem e incluir a população com coeficiente constante (offset) e função de ligação logaritmo. Os modelos lineares generalizados apresentam essa possibilidade, de modo que seja possível estudar o logaritmo da taxa esperada de internação. Assim, utilizando a função exponencial também é possível medir a variação percentual da taxa devido ao acréscimo de uma unidade da variável explicativa. A distribuição de Poisson é muito utilizada para esses tipos de dados e por ser uma distribuição da família exponencial, pode ser usada nos modelos lineares generalizados. Ela assume que a variância é igual à média, o que já é indicio de heterocedasticidade. Porém, caso a variância seja maior que a média, situação conhecida como superdispersão dos dados, é mais adequado usar a distribuição Binomial Negativa. A distribuição Binomial Negativa permite que haja essa superdispersão dos dados e também faz parte das distribuições utilizadas em modelos lineares generalizados.

Como os dados foram medidos ao longo do tempo, a usual suposição de independência pode ser violada e, assim, é importante incluir no modelo, componentes que considerem a correlação serial. Um primeiro modo de realizar essa inclusão é ajustar os modelos ARMA ou ARIMA propostos por Box e Jenkins [1976]. Em geral, esses modelos supõem que os erros são não correlacionados e gaussianos. Outro modo é utilizar modelos para variáveis resposta que consistem em dados de contagem, adotando modelos lineares generalizados GARMA, que incluem termos autorregressivos e médias móveis, semelhantes ao modelo ARMA.

Outra classe de modelos é o GAMLSS que não considera a autocorrelação, porém tem a possibilidade de modelar a variância em função de variáveis explicativas e permite o uso das distribuições Poisson e Binomial Negativa, além da Normal.

Fixando a notação utilizada nesse capítulo, considere:

- y_t : contagem observada no dia t ;
- δ_t : variável utilizada como offset;
- Pop_t : população no instante t ;
- tx_t : taxa correspondente às contagens, como por exemplo, $tx_t = \frac{y_t}{Pop_t/100000}$;
- T : número de observações, sendo que o índice $t = 1, \dots, T$;
- \mathbf{x}_t : vetor $m \times 1$ composto pelas p variáveis explicativas observadas no instante t ;
- \mathbf{X} : matriz $m \times T$ composta pelas p variáveis explicativas para todas as T observações;
- β : vetor $m \times 1$ de parâmetros correspondentes às variáveis explicativas;
- μ_t : valor esperado da variável resposta;
- σ_t Variância da variável resposta;
- ϑ_t parâmetro canônico;
- ϑ_2 : parâmetro de dispersão;
- η : preditor linear;
- \mathbf{H}_t : vetor de informação passada $\mathbf{H}_t = \{\mathbf{x}_t, \dots, \mathbf{x}_1, y_t, \dots, y_1, \mu_t, \dots, \mu_1\}$;
- ϕ : vetor de parâmetros autorregressivos;
- θ : vetor de parâmetros de médias móveis;
- B : operador backshift, $B^k y_t = y_{t-k}$.

Serão apresentados os modelos:

1. Regressão Linear;
2. Autorregressivo e de Média Móvel (ARMA);
3. Lineares Generalizados (MLG);
4. Generalizados Autorregressivos de Médias Móveis (GARMA);
5. Aditivos Generalizados para Posição, Escala e forma (GAMLSS).

Nesse trabalho, os parâmetros são estimados por máxima verossimilhança (EMV). Este estimador tem propriedade de suficiência (se o parâmetro tiver alguma estatística suficiente), eficiência, invariância e é não viesado assintoticamente (Mood [1950]).

Respeitando as condições de regularidade, nenhum outro estimador assintoticamente normal tem menor variância assintótica que o EMV, assim este é assintoticamente eficiente. Por fim, sendo $u(\beta)$ uma função que associa β a um único e definido valor (u não é necessariamente uma função injetora) e $\hat{\beta}$ é o EMV de β , então pela propriedade de invariância $u(\hat{\beta})$ é EMV para $u(\beta)$.

Para o modelo de regressão linear, o estimador de mínimos quadrados é simples, tem expressão analítica e coincide com o estimador de máxima verossimilhança sob distribuição normal dos erros. Para maximizar o logaritmo da função de verossimilhança dos modelos MLG e GAMLSS aplica-se o algoritmo Rigby and Stasinopoulos (RS) (Rigby e Stasinopoulos [2005]), que é uma generalização do algoritmo usado por Rigby e Stasinopoulos [1996] e é baseado no algoritmo Newton-Raphson ou Escore de Fisher. Ele utiliza a primeira derivada e o valor esperado ou aproximado da segunda derivada do logaritmo da função de verossimilhança com relação aos parâmetros. Os ajustes dos

modelos ARMA e GARMA são realizados por procedimento iterativo por mínimos quadrados reponderados (Green [1984]) e é usado o algoritmo Escore de Fisher.

Após a estimação dos parâmetros β , calcula-se o efeito das variáveis explicativas como a variação percentual da taxa média devido ao aumento de uma unidade de cada variável explicativa. Tal variação percentual é igual a $[\exp(\beta) - 1]100$. Pela propriedade da invariância, essa variação tem estimador de máxima verossimilhança igual a:

$$\widehat{u}(\widehat{\beta}) = [\exp(\widehat{\beta}) - 1]100 \quad (2.1)$$

com variância assintótica obtida pelo método de Delta (Sen *et al.* [2010]):

$$\text{var}(\widehat{u}(\widehat{\beta})) \approx \text{var}(\widehat{\beta}) \times \left(\frac{\delta u(\widehat{\beta})}{\delta \beta} \right)^2 \approx \text{var}(\widehat{\beta}) \times \exp(\widehat{\beta})^2 \times 100^2 \quad (2.2)$$

As significâncias dos parâmetros de todos os modelos serão testadas por meio do teste de razão de verossimilhanças explicado a seguir.

Considere o modelo completo M_1 e o modelo reduzido M_0 sob a restrição de que alguns parâmetros são nulos. Assim, a hipótese nula é H_0 : alguns parâmetros específicos do vetor β são nulos. Nesse caso, o modelo M_0 é dito encaixado no modelo M_1 . Denotam-se os valores do logaritmo da máxima verossimilhança dos modelos M_1 e M_0 por l_0 e l_1 respectivamente e os números de parâmetros por κ_0 e κ_1 . Para testar a hipótese H_0 é usada a estatística da razão de verossimilhanças, $\Lambda = -2(l_0 - l_1)$, que sob H_0 tem distribuição assintótica χ^2 com $(\kappa_1 - \kappa_0)$ graus de liberdade. O parâmetro ϑ_2 , se desconhecido, é estimado pelo estimador de máxima verossimilhança (EMV) $\widehat{\vartheta}_{2j}$ para cada modelo $j = 0, 1$ (Rigby e Stasinopoulos [2005]).

Todas as análises foram realizadas no programa R versão 3.0.1. Para os modelos regressão linear, MLG e GAMLSS, foi usada a função `gamlss` da biblioteca `gamlss` (Stasinopoulos e Rigby [2007]) e para os modelos ARMA e GARMA, foi usada a função `garmaFit` da biblioteca `gamlss.util` (Stasinopoulos *et al.* [2013]). Para a análise inferencial, utilizou-se um nível de significância de 5%.

2.1 Regressão linear

O modelo de regressão linear é o mais utilizado na prática para explicar a variação de uma variável resposta em função de variáveis explicativas. O modelo é dito linear porque o valor esperado da variável resposta é uma função linear dos parâmetros (Bussab e Morettin [2010]).

Este modelo é dado por:

$$tx_t = \mathbf{x}'_t \beta + e_t, \quad (2.3)$$

e tem as seguintes suposições:

1. Há uma relação linear direta ou de identidade entre o preditor linear e os valores esperados do modelo: $\mu_t = \mathbf{x}'_t \beta$;
2. Os erros e_t têm média 0, o que implica que a taxa tenha valor esperado $E(tx_t) = \mathbf{x}'_t \beta$;
3. Os erros e_t são independentes;
4. Os erros e_t têm variância σ^2 constante para todo t , o que implica que tx_t também seja homocedástico;
5. Os erros e_t têm distribuição Normal.

A validade das suposições anteriormente citadas deve ser checada após o ajuste do modelo e sua validade implica várias propriedades dos estimadores. Algumas dessas propriedades serão comentadas para motivar a necessidade da utilização de outros modelos quando algumas dessas suposições forem violadas. No modelo de regressão linear, é possível utilizar o estimador de mínimos

quadrados, que simplesmente minimiza $\sum_{t=1}^T (tx_t - \mathbf{x}'_t \boldsymbol{\beta})^2$. O estimador de mínimos quadrados é não viesado, valendo as suposições 1 e 2. Sob as suposições de 1 a 4, os estimadores obtidos serão os melhores estimadores lineares não viesados (BLUE= Best Linear Unbiased Estimator) e a variância de $\hat{\boldsymbol{\beta}}$ é dada por:

$$\text{var}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}' \mathbf{X})^{-1}. \quad (2.4)$$

Incluindo também a suposição 5, tem-se que $\hat{\boldsymbol{\beta}}$ tem distribuição Normal e este resultado, juntamente com estimadores para a variância dos erros σ^2 , pode ser utilizado para a realização de testes e construção de intervalos de confiança. Além disso, o estimador de máxima verossimilhança coincide com o estimador de mínimos quadrados e assim é possível utilizar de várias propriedades dos estimadores de máxima verossimilhança.

O logaritmo da função de verossimilhança é dado por:

$$l(\boldsymbol{\beta}; \mathbf{x}) = -\frac{T}{2} \log(2\pi\sigma) + \sum_{t=1}^T -\frac{(tx_t - \mathbf{x}'_t \boldsymbol{\beta})^2}{2\sigma^2}, \quad (2.5)$$

Na prática, a violação dessas suposições exige modelos mais sofisticados para obter estimadores não viesados de $\boldsymbol{\beta}$ e da variância. Por exemplo, é possível que para valores mais altos das taxas haja maior variabilidade do que para valores mais baixos. Outra situação é não ter a normalidade dos erros e_t , o que pode ser verificado em análises dos resíduos. Nesse caso, é interessante propor outras distribuições para y_t como, por exemplo, a distribuição Poisson.

2.2 Modelo autorregressivo e de média móvel - ARMA

Nesse modelo, proposto por [Box e Jenkins \[1976\]](#), a variável resposta y_t é explicada por variáveis defasadas $y_{t-1}, y_{t-2}, \dots, y_{t-p}$ e por erros defasados $e_{t-1}, e_{t-2}, \dots, e_{t-q}$, de tal modo que $y_t = \phi_0 + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}$.

Este modelo considera que a correlação entre a variável y_t e suas defasagens y_{t-1}, y_{t-2}, \dots é não nula. Detalhes são apresentados em [Morettin e Tolo \[2006\]](#).

O modelo ARMA é uma generalização do modelo de regressão linear e permite o uso de parâmetro para explicar a autocorrelação da variável resposta em relação ao tempo.

O modelo ARMA(p,q) com variáveis explicativas apresentado por [Shumway e Stoffer \[2000\]](#) e [Benjamin *et al.* \[2003\]](#) é dado pela seguinte equação:

$$\phi(B)(tx_t - \mathbf{x}'_t \boldsymbol{\beta}) = \theta(B)e_t, \quad (2.6)$$

em que $\phi(B) = (1 - \sum_{j=1}^p \phi_j B^j)$, $\theta(B) = (1 - \sum_{j=1}^q \theta_j B^j)$ e e_t é ruído branco com distribuição $N(0, \sigma_e^2)$ e B é o operador backshift definido por:

$$B^k tx_t = tx_{t-k} \quad (2.7)$$

A autocovariância pode ser escrita como:

$$\gamma_k = E \left[(tx_t - \mathbf{x}'_t \boldsymbol{\beta})(tx_{t-k} - \mathbf{x}'_{t-k} \boldsymbol{\beta}) \right] - E(tx_t - \mathbf{x}'_t \boldsymbol{\beta})E(tx_{t-k} - \mathbf{x}'_{t-k} \boldsymbol{\beta}) \quad (2.8)$$

E a autocorrelação como:

$$\rho_k = \frac{E \left[(tx_t - \mathbf{x}'_t \boldsymbol{\beta})(tx_{t-k} - \mathbf{x}'_{t-k} \boldsymbol{\beta}) \right] - E(tx_t - \mathbf{x}'_t \boldsymbol{\beta})E(tx_{t-k} - \mathbf{x}'_{t-k} \boldsymbol{\beta})}{\gamma_0} \quad (2.9)$$

O modelo ARMA precisa atender à condição de estacionariedade, ou seja, a série tem que ser distribuída de forma aleatória em torno de uma média e variância constante.

Essa condição é verificada, considerando o operador autorregressivo, $\phi(B)$, como um polinômio em B com p graus de liberdade: $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$. A série será estacionária se todas as raízes de $\phi(B) = 0$ forem, em módulo, maiores do que 1, sendo que podem ser números complexos, ou seja, todas as raízes devem estar fora do círculo unitário.

Essa condição é válida se $tx_t - \mathbf{x}'_t \boldsymbol{\beta}$ tiver média e variância constantes.

Análoga à condição de estacionariedade, o modelo será invertível se o polinômio $\theta(B) = 0$ tiver todas as raízes fora do círculo unitário. Caso isso não ocorra, $tx_t - \mathbf{x}'_t \boldsymbol{\beta}$ irá depender de $tx_{t-1} - \mathbf{x}'_{t-1} \boldsymbol{\beta}, tx_{t-2} - \mathbf{x}'_{t-2} \boldsymbol{\beta}, \dots, tx_{t-k} - \mathbf{x}'_{t-k} \boldsymbol{\beta}$ e o peso irá aumentar a medida que k aumenta (Box *et al.* [2013, p. 50]).

Os parâmetros do modelo podem ser estimados por máxima verossimilhança. O logaritmo da função de verossimilhança condicional às primeiras observações é dado pela expressão:

$$l(\phi, \theta; w_t) = \log(2\pi\sigma_e)^{-\frac{T}{2}} \sum_{t=\max(p,q)}^T \left(-\frac{w_t - \phi_1 w_{t-1} - \dots - \phi_p w_{t-p} - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}}{2\sigma_e^2} \right)^2, \quad (2.10)$$

em que $w_t = tx_t - \mathbf{x}'_t \boldsymbol{\beta}$ tem distribuição ARMA(p, q), e_t são erros não correlacionados com distribuição $N(0, \sigma_e^2)$. Caso todos os ϕ e θ sejam iguais à zero, este modelo se torna um modelo de regressão linear.

O modelo ARMA está implementado no R, na biblioteca `gamlss.util` utilizando a função `garmafit` (Stasinopoulos *et al.* [2013]).

Mais informações sobre este modelo estão em Box *et al.* [2013].

2.3 Modelos Lineares Generalizados

O modelo linear generalizado (MLG) foi desenvolvido por Nelder e Wedderburn [1972] e é uma generalização do modelo de regressão linear, de forma que possa ser usada qualquer distribuição da família exponencial, incluindo as distribuições para contagem Poisson e Binomial Negativa, além da Normal (Hardin e Hilbe [2007]). Outras distribuições pertencentes à família exponencial são: Gama e Binomial. Os principais resultados da teoria e aplicação dos modelos lineares generalizados podem ser encontrados em Paula [2013].

A família exponencial compreende um conjunto de distribuições cujas funções densidade de probabilidade podem ser escritas da seguinte forma:

$$f(y_t) = \exp \left\{ \frac{y_t \vartheta_{1t} - b(\vartheta_{1t})}{\vartheta_2} + d(y_t, \vartheta_2) \right\}, \quad (2.11)$$

em que ϑ_1 é o parâmetro canônico e ϑ_2 é o parâmetro de escala, com $b(\cdot)$ e $d(\cdot)$ sendo funções específicas que definem uma particular família exponencial. Os termos $\mu_t = E(y_t) = b'(\vartheta_{1t})$ e $var(y_t) = \vartheta_2 V(\mu_t) = \vartheta_2 b''(\vartheta_{1t})$, $t = 1, \dots, T$ e $V(\mu_t)$ é a função de variância de μ_t , representam respectivamente a média e a variância de y_t .

No modelo linear generalizado, uma função da média da variável resposta é uma função linear dos parâmetros, dito preditor linear. Assim tem-se que

$$g(\mu_t) = \eta_t = \mathbf{x}'_t \boldsymbol{\beta} + \delta_t. \quad (2.12)$$

Muitas vezes são adotadas funções de ligação g canônicas, que correspondem ao parâmetro canônico igual ao preditor linear, ou seja, $\vartheta_{1t} = g(\mu_t) = \eta_t$.

Para este trabalho, foi considerada a função $g(\mu_t)$ como logaritmo de μ_t . Desse modo, $\mu_t = e^{\eta_t}$ e para todo valor de η_t , a média μ_t é um valor positivo, o que é necessário para variáveis respostas com valores não negativos.

Outra vantagem de se trabalhar com a ligação logaritmo e com o offset da população é a possibilidade de reestruturar a equação para que a interpretação seja relacionada à taxa média (ver

equação 2.13) e não, à média dos logaritmos das taxas.

$$\log(\mu_t) = \mathbf{x}'_t\beta + \delta_t \Rightarrow \log\left(\frac{\mu_t}{\delta_t}\right) = \log[E(tx_t)] = \mathbf{x}'_t\beta \quad (2.13)$$

em que $E(tx_t)$ é a taxa esperada.

Assim como na regressão linear, os parâmetros β podem ser estimados pelo método de máxima verossimilhança. A ligação canônica garante a concavidade de $L(\beta)$ e com isso muitos resultados assintóticos são obtidos mais facilmente, como, por exemplo, a unicidade da estimativa de verossimilhança de β quando essa existe.

Como os dados desse estudo são de contagem, foi trabalhado com os modelos MLG com distribuição Poisson e Binomial Negativa, além da distribuição Normal. Note que o modelo de regressão linear usual é equivalente ao MLG com distribuição Normal e ligação identidade.

A função densidade de probabilidade da distribuição Poisson pode ser escrita como:

$$f(y_t; \mu_t) = \frac{e^{-\mu_t} \mu_t^{y_t}}{y_t!}. \quad (2.14)$$

Reescrevendo esta densidade no formato da função da família exponencial (2.11), tem-se:

$$f(y_t; \mu_t) = \exp\{y_t \log(\mu_t) - \mu_t - \log(\Gamma(y_t + 1))\}. \quad (2.15)$$

O logaritmo da função log-verossimilhança é dada pela expressão:

$$l(\mu; y) = \sum_{t=1}^n \{y_t \log(\mu_t) - \mu_t - \log(\Gamma(y_t + 1))\} \quad (2.16)$$

em que,

$$\log(\mu_t) = \mathbf{x}'_t\beta + \delta_t. \quad (2.17)$$

Na distribuição de Poisson, a variância é igual a média μ_t . É dito que há superdispersão quando é suposto que os dados tenham distribuição Poisson, porém a variância é maior que a média (Paula [2013]). Neste caso, uma alternativa é utilizar a distribuição Binomial Negativa por esta ter variância maior que a média.

A função densidade de probabilidade da distribuição Binomial Negativa pode ser escrita como:

$$f(y_t | \mu_t, \vartheta_2) = \frac{\Gamma(y_t + 1/\vartheta_2)(\mu_t \vartheta_2)^{y_t}}{\Gamma(y_t + 1)\Gamma(1/\vartheta_2)(\mu_t \vartheta_2 + 1)^{y_t + 1/\vartheta_2}} \quad (2.18)$$

O logaritmo da função de verossimilhança é expresso por:

$$l(\mu_t, \vartheta_2; y_t) = \sum_{t=1}^n \left\{ \log(\Gamma(y_t + \frac{1}{\vartheta_2})) + y_t \log(\mu_t \vartheta_2) - \log(\Gamma(y_t + 1)) - (y_t + \frac{1}{\vartheta_2}) \log(\mu_t \vartheta_2 + 1) \right\} - n \log(\Gamma(\frac{1}{\vartheta_2})) \quad (2.19)$$

em que, $\vartheta_2 > 0$ e, assim como na Poisson,

$$\log(\mu_t) = \mathbf{x}'_t\beta + \delta_t. \quad (2.20)$$

A variância é dada pela equação:

$$\text{Var}(Y_t) = \mu_t + \vartheta_2 \mu_t^2, \quad (2.21)$$

e o desvio padrão é dado por:

$$DP(Y_t) = \sqrt{\mu_t + \vartheta_2 \mu_t^2}, \quad (2.22)$$

logo a variância da distribuição Binomial Negativa é sempre maior que a média e portanto é uma distribuição para superdispersão. Para ϑ_2 tendendo a 0, tem-se a distribuição de Poisson.

O modelo MLG está implementado no R, na biblioteca `gamlss.util` utilizando a função `garmafit` (Stasinopoulos *et al.* [2013]).

2.4 Modelos generalizados autorregressivos de médias móveis - GARMA

O modelo GARMA (Benjamin *et al.* [2003]) é uma combinação do modelo ARMA e do modelo linear generalizado, possibilitando a inclusão de termos autorregressivos e de médias móveis em qualquer distribuição da família exponencial.

No modelo GARMA, a distribuição condicional de cada observação y_t , para $t = 1, \dots, T$, dada toda a informação passada $\mathbf{H}_{t-1} = \{\mathbf{x}_{t-1}, \dots, \mathbf{x}_1, y_{t-1}, \dots, y_1\}$, pertence à família exponencial, ou seja, a densidade condicional é dada por:

$$f(y_t | \mathbf{H}_{t-1}) = \exp \left\{ \frac{y_t \vartheta_{1t} - b(\vartheta_{1t})}{\vartheta_2} + d(y_t, \vartheta_2) \right\}. \quad (2.23)$$

Os termos $\mu_t = E(y_t | \mathbf{H}_{t-1}) = b'(\vartheta_{1t})$ e $var(y_t | \mathbf{H}_{t-1}) = \vartheta_2 V(\mu_t) = \vartheta_2 b''(\vartheta_{1t})$, $t = 1, \dots, T$ e $V(\mu_t)$ é a função de variância de μ_t , representam respectivamente a média e a variância condicionais de y_t dado \mathbf{H}_{t-1} . A notação é a mesma que utilizada para o MLG em (2.11) porém no modelo GARMA são utilizadas distribuições condicionais às informações passadas.

Para dados medidos ao longo do tempo, as observações y_t podem ser correlacionadas com as observações y_{t-1}, y_{t-2}, \dots . O modelo GARMA permite a autocorrelação de y_t incluindo termos autorregressivos e médias móveis no preditor linear para a média:

$$g(\mu_t) = \eta_t = \mathbf{x}'_t \boldsymbol{\beta} + \sum_{j=1}^p \phi_j A(y_{t-j}, x_{t-j}, \boldsymbol{\beta}) + \sum_{j=1}^q \theta_j M(y_{t-j}, \mu_{t-j}), \quad (2.24)$$

em que os termos $\mu_t = b'(\vartheta_{1t}) = E(y_t | \mathbf{H}_{t-1})$ e $var(y_t | \mathbf{H}_{t-1}) = \vartheta_2 v(\mu_t) = \vartheta_2 b''(\vartheta_{1t})$ $t = 1, \dots, n$ representam a média e a variância condicional de y_t dado \mathbf{H}_{t-1} . Os parâmetros autorregressivos são $\boldsymbol{\phi}' = (\phi_1, \dots, \phi_p)$, os parâmetros de média móvel são $\boldsymbol{\theta}' = (\theta_1, \dots, \theta_q)$ e A e M são funções dos termos autorregressivos e de médias móveis, respectivamente.

As funções de ligação dos termos autorregressivo e média móvel, respectivos a A e M , apresentadas anteriormente são muito gerais, então para esse estudo foi considerado o seguinte caso particular para o preditor linear η_t , incluindo o offset δ_t :

$$\begin{aligned} g(\mu_t) = \eta_t &= \mathbf{x}'_t \boldsymbol{\beta} + \delta_t + \sum_{j=1}^p \phi_j \left\{ g(y_{t-j}) - \mathbf{x}'_{t-j} \boldsymbol{\beta} - \delta_t \right\} \\ &+ \sum_{j=1}^q \theta_j \left\{ g(y_{t-j}) - \mu_{t-j} \right\} \end{aligned} \quad (2.25)$$

Neste trabalho, foram considerados os modelos GARMA com distribuições Normal, Poisson e Binomial Negativa.

No caso de função de ligação logaritmo neperiano, é necessário substituir y_{t-j} por y_{t-j}^* quando y_t for igual a 0, pois $\lim_{y \rightarrow 0} \log(y) \rightarrow -\infty$. Nesses casos, pode-se utilizar $y_{t-j}^* = \max(y_{t-j}, c)$ com $0 < c < 1$.

Como no modelo MLG, foi adotada a ligação logaritmo, o modelo estrutural é dado por:

$$\begin{aligned} \log(\mu_t) &= \mathbf{x}'_t \boldsymbol{\beta} + \delta_t + \sum_{j=1}^p \phi_j \left\{ \log(y_{t-j}) - \mathbf{x}'_{t-j} \boldsymbol{\beta} - \delta_t \right\} \\ &\quad + \sum_{j=1}^q \theta_j \left\{ \log(y_{t-j} / \mu_{t-j}) \right\}, \end{aligned} \quad (2.26)$$

Assim como no modelo linear generalizado (ver equação 2.13) é a possível reestruturar a equação 2.26 para que a interpretação seja relacionada à taxa média (ver equação 2.27).

$$\begin{aligned} \log[E(tx_t)] &= \mathbf{x}'_t \boldsymbol{\beta} + \sum_{j=1}^p \phi_j \left\{ \log(y_{t-j}) - \mathbf{x}'_{t-j} \boldsymbol{\beta} - \delta_t \right\} \\ &\quad + \sum_{j=1}^q \theta_j \left\{ \log(y_{t-j} / \mu_{t-j}) \right\} \end{aligned} \quad (2.27)$$

em que $E(tx_t)$ é a taxa esperada.

Analogamente ao modelo linear generalizado, o logaritmo neperiano da função de verossimilhança do modelo GARMA com distribuição Poisson é dado por:

$$l(\mu; y) = \sum_{t=1}^T \left\{ y_t \log(\mu_t) - \mu_t - \log(\Gamma(y_t + 1)) \right\}. \quad (2.28)$$

e com distribuição Binomial Negativa é dado por:

$$\begin{aligned} l(\mu_t, \vartheta_2; y_t) &= \sum_{t=1}^T \left\{ \log(\Gamma(y_t + \frac{1}{\vartheta_2})) + y_t \log(\mu_t \vartheta_2) - \log(\Gamma(y_t + 1)) - (y_t + \frac{1}{\vartheta_2}) \log(\mu_t \vartheta_2 + 1) \right\} \\ &\quad - T \log \left[\Gamma \left(\frac{1}{\vartheta_2} \right) \right] \end{aligned} \quad (2.29)$$

A função de ligação (2.25) é parecida com a função (2.12) do MLG, porém, agora, μ_t depende também de ϕ e θ .

Assim como os resíduos do modelo ARMA, os resíduos padronizados do modelo GARMA precisam atender à suposição de normalidade e estacionariedade.

O modelo GARMA está implementado no R, na biblioteca `gamlss.util` utilizando a função `garmafit` (Stasinopoulos *et al.* [2013]).

2.5 Modelos aditivos generalizados para posição, escala e forma - GAMLSS

O modelo GAMLSS (Rigby e Stasinopoulos [2005]) é uma generalização do MLG e do modelo aditivo generalizado (MAG) (Hastie e Tibshirani [1990]). Para estes dois modelos, a variância e, em geral, a assimetria e a curtose dependem da média e do parâmetro constante de dispersão.

Com o GAMLSS é possível modelar diretamente todos os parâmetros da distribuição condicional de y , e não somente a média (localização). Para a maioria das distribuições, os parâmetros são: média, variância, assimetria e curtose. Isso permite que se explique a heteroscedasticidade por meio de variáveis explicativas, porém não permite ter parâmetros para a autocorrelação. Esta flexibilidade se mostra interessante no exemplo do Capítulo 3. Esses parâmetros podem ser modelados com efeitos

aleatórios e/ou termos paramétricos e/ou não paramétricos, em que a média da variável resposta y_t é uma função não necessariamente linear de x_t , como splines cúbicos ou alisamento loess, como apresentado em [Conceição et al. \[2001\]](#) que utilizou o modelo MAG.

O modelo GAMLSS também permite utilizar outras distribuições para y_t além da família exponencial, como, por exemplo, as distribuições da família t .

Enquanto os modelos MLG e GARMA têm somente uma função de ligação monótona $g(\cdot)$, que é utilizada para a média, no modelo GAMLSS é possível ter uma função de ligação monótona $g_k(\cdot)$, com $k = 1, 2, \dots, r$, para cada um dos r parâmetros $\boldsymbol{\vartheta}' = (\vartheta_1, \vartheta_2, \dots, \vartheta_r)$ da função de probabilidade (densidade) da população $f(y|\boldsymbol{\vartheta})$. O modelo assume que as observações y_t são condicionalmente independentes dado $\boldsymbol{\vartheta}_t$, em que $\boldsymbol{\vartheta}'_t = (\vartheta_{1t}, \vartheta_{2t}, \dots, \vartheta_{rt})$ é um vetor de r parâmetros relacionados a variáveis explicativas e efeitos aleatórios. Se os valores das covariáveis são estocásticos ou as observações de y_t dependem de seus valores passados, então $f(y_t|\boldsymbol{\vartheta}_t)$ é entendido como sendo condicional a estes valores. No caso de efeitos aleatórios é usada a verossimilhança penalizada para a estimação dos parâmetros ([Cox \[1975\]](#)). Quando só há efeitos fixos, a função de verossimilhança penalizada é simplificada à função de verossimilhança usual.

O modelo GAMLSS é dado pela equação:

$$g_k(\vartheta_k) = \eta_k = X_k\beta_k + \sum_{j=1}^{J_k} Z_{jk}\gamma_{jk}, \tag{2.30}$$

em que ϑ_k e η_k são vetores de tamanho T , isto é, $\vartheta'_k = (\vartheta_{k1}, \vartheta_{k2}, \dots, \vartheta_{kT})$, $\beta'_k = (\beta_{k1}, \beta_{k2}, \dots, \beta_{km_k})$ é um vetor de parâmetros de tamanho m_k , X_k é uma matriz de planejamento de dimensão $T \times m_k$, Z_{jk} é uma matriz de planejamento fixa conhecida de dimensão $T \times q_{jk}$ e γ_{jk} é um vetor de variáveis aleatórias q_{jk} -dimensional que corresponde aos efeitos aleatórios.

No exemplo do [Capítulo 3](#), foi adotado o modelo paramétrico e ligação logaritmo para permitir fácil interpretação. Considerando $J_k = 0$ para $k = 1, 2, \dots, r$ e incluindo o offset, o modelo (2.30) se reduz ao modelo paramétrico:

$$g_k(\vartheta_k) = \eta_k = \mathbf{X}_k\boldsymbol{\beta}_k + \delta_k \tag{2.31}$$

O modelo (2.31) foi utilizado para a média e variância da distribuição Normal e para a média e o parâmetro de dispersão da Binomial Negativa. Como a distribuição de Poisson só tem um parâmetro, que é o mesmo para a média e a variância, não há necessidade de usar o GAMLSS, porque seu modelo paramétrico é equivalente ao MLG.

Assim como no modelo linear generalizado (ver equação 2.13) é a possível reestruturar a equação 2.31 com ligação logaritmo para que a interpretação seja relacionada à taxa média (ver equação 2.32).

$$\log [E(tx_t)] = \mathbf{X}_k\boldsymbol{\beta}_k \tag{2.32}$$

em que $E(tx_t)$ é a taxa esperada.

O logaritmo da função de verossimilhança para o GAMLSS paramétrico (2.31) é dado pela equação:

$$l(\boldsymbol{\vartheta}; y) = \sum_{t=1}^T \log f(y_t|\boldsymbol{\vartheta}_t) \tag{2.33}$$

Para evitar múltiplos máximos locais, são usados diferentes valores iniciais. A maximização da verossimilhança é feita pelo método de Newton-Raphson/Score de Fisher.

2.6 Resíduos

O resíduo do modelo é a diferença entre o valor observado e o valor esperado: $y_t - \hat{\mu}_t$. Estes resíduos podem não ter variância constante e ter distribuições muito diferentes da Normal, dificultando o diagnóstico do modelo. Portanto, é desejável transformá-los em resíduos com distribuição Normal padrão. Eles precisam também ser independentes.

Nesse trabalho foi utilizado o resíduo quantílico desenvolvido por [Dunn e Smyth \[1996\]](#), que, para distribuições contínuas, é dado por:

$$\begin{aligned} u_t &= F_Y(y_t | \hat{\mu}_t, \hat{\vartheta}_{t2}) \\ r_t &= F_N^{-1}(u_t), \end{aligned} \quad (2.34)$$

em que a média estimada é $\hat{\mu}_t$, $\hat{\vartheta}_{t2}$ é o parâmetro de dispersão estimado, F_Y é a distribuição acumulada da função de probabilidade assumida no modelo (a distribuição Normal, por exemplo) e F_N^{-1} é a função inversa da distribuição acumulada da Normal padrão.

Para distribuições discretas, simula-se a variável u_t segundo uma distribuição uniforme para evitar resíduos discretos, o resíduo r_t é dado por:

$$\begin{aligned} u_t &= U[F_Y(y_t - 1 | \hat{\mu}_t, \hat{\vartheta}_{t2}), F_Y(y_t | \hat{\mu}_t, \hat{\vartheta}_{t2})] \\ r_t &= F_N^{-1}(u_t), \end{aligned} \quad (2.35)$$

$F_Y(y_t | \hat{\mu}_t, \hat{\vartheta}_{t2})$ continua definido como função ajustada da distribuição acumulada com a mesma distribuição do modelo, podendo ser, por exemplo, Poisson ou Binomial Negativa, e u_t é um valor sorteado no intervalo $[F_Y(y_t - 1 | \hat{\mu}_t, \hat{\vartheta}_{t2}), F_Y(y_t | \hat{\mu}_t, \hat{\vartheta}_{t2})]$.

Logo, quanto mais distante de 50% for $F_Y(y_t | \hat{\mu}_t, \hat{\vartheta}_{t2})$, mais distante será o resíduo r_t de 0 e portanto, será maior em valor absoluto. Já que $u_t = 0,5$ implica $r_t = 0$.

Como sugerido por [Dunn e Smyth \[1996\]](#), essa aleatorização foi realizada quatro vezes e para cada uma foram gerados gráficos de resíduos. Qualquer padrão dos resíduos observado em um gráfico diferente dos três demais é ignorado.

Caso a suposição de que $F_Y(\cdot)$ ser a distribuição acumulada do modelo não for rejeitada, r_t terá distribuição Normal padrão. Caso contrário, conclui-se que o modelo não foi bem ajustado.

Além da suposição de normalidade, é necessário examinar se há ausência de correlação serial e se há homocedasticidade, ou seja, variância constante nos resíduos padronizados. Foram utilizados os seguintes métodos para o diagnóstico dos resíduos padronizados:

1. Gráficos de dispersão;
2. Gráfico da função de autocorrelação;
3. Teste de homoscedasticidade da variância;
4. Gráfico do núcleo da densidade;
5. Gráfico quantil-quantil da Normal;
6. Medidas resumo: média, variância, curtose, assimetria dos resíduos padronizados.

Para estudar a independência entre os resíduos padronizados foram utilizados gráficos de dispersão desses resíduos em relação ao tempo t e em relação ao valor ajustado, foi investigado se os resíduos apresentaram algum tipo de tendência, o que indicaria que eles não são independentes.

O gráfico de função de autocorrelação (FAC) foi usado conjuntamente aos gráficos de dispersão. Ele mostra a autocorrelação dos resíduos para k defasagens, ou seja, corresponde à correlação entre r_t e r_{t-k} . Esta autocorrelação assume valores entre -1 e 1, valores próximos de zero indicam baixa correlação.

Esse gráfico também apresenta um intervalo de confiança assintótico em torno de zero com coeficiente de confiança de 95%. Se as autocorrelações estiverem fora do intervalo para alguma defasagem, considera-se que a autocorrelação é significativa, ou seja, o erro no tempo t é correlacionado com o erro no tempo $t - k$.

Se a dispersão dos resíduos não se mostrar constante ao longo do tempo ou em relação ao valor ajustado nos gráficos de dispersão, é um indicio de que a variância não é constante.

Também foi apresentado o gráfico FAC dos resíduos padronizados ao quadrado para verificar a heteroscedasticidade condicional, como no modelo ARCH (Morettin [2008]). A interpretação é análoga à usada para a autocorrelação dos resíduos padronizados: se a autocorrelação for significativa para alguma defasagem dos resíduos padronizados ao quadrado conclui-se que a variância é autocorrelacionada, o que é indicio de heteroscedasticidade condicional.

A variância do resíduo no tempo t pode depender dos resíduos passados, mas também de variáveis explicativas. Os resíduos padronizados ao quadrado também foram usados para checar esta suposição. Usou-se a regressão auxiliar linear simples tendo os resíduos ao quadrado como variável resposta, como nos testes de White e Breusch Pagan (Wooldridge [2012]), e foi testado o efeito de cada variável explicativa, separadamente, nos resíduos ao quadrado. Desta forma, verificou-se se a variância tem relação linear com alguma variável utilizada no modelo.

Para concluir que os resíduos padronizados possuem distribuição Normal e, portanto, a distribuição utilizada é apropriada, primeiro é preciso ter conhecimento dos valores de algumas medidas resumo da distribuição Normal padrão. A distribuição Normal padrão tem média igual a zero, variância igual a um, o coeficiente de assimetria é igual a zero, por se tratar de uma distribuição simétrica, e o coeficiente de curtose é igual a três.

Então para os resíduos terem distribuição Normal padrão, suas medidas de média, variância, assimetria e curtose devem estar próximas das calculadas para essa distribuição. Além dessas medidas, a suposição de normalidade pode ser checada por gráficos, como o gráfico do núcleo da densidade, que estima a função de densidade dos dados (ver Apêndice 6), e o gráfico quantil-quantil da Normal, que compara os quantis dos resíduos com os quantis da Normal padrão.

2.7 Erro quadrático médio e intervalo de confiança

É importante que as suposições do modelo sejam atendidas para que as estimativas sejam calculadas corretamente. Porém, além do modelo ter sido bem ajustado, é desejável que seus valores estimados estejam próximos dos valores observados. Para esse estudo, foram utilizados o erro quadrático médio (EQM) e o intervalo de confiança.

O EQM calcula a média da diferença ao quadrado entre o valor observado e o esperado. Foi calculado o EQM para as contagens para todos os modelos. Para os modelos MLG, GARMA e GAMLSS, ele é calculado pela seguinte fórmula:

$$EQM = \sum_{t=1}^T \frac{(y_t - \hat{\mu}_t)^2}{T}. \quad (2.36)$$

Os modelos regressão linear e ARMA têm como variável resposta $\log(tx_t)$, é preciso fazer uma transformação na fórmula do EQM para que se possa calcular o erro da contagem y_t . Então para esses dois modelos o EQM é dado pela seguinte fórmula:

$$EQM = \sum_{t=1}^T \frac{\{y_t - \exp(\hat{\mu}_t + \delta_t)\}^2}{T}. \quad (2.37)$$

em que $\hat{\mu}_t$ é a média esperada de y no dia t . Quanto menor o valor do EQM mais os valores estimados se aproximaram do observado.

O intervalo de confiança (IC) para o valor predito cria uma região em torno do valor estimado. Neste trabalho, foi usado um intervalo de confiança de 95%, ou seja, se fosse calculado o intervalo para um grande número de amostras de mesma distribuição (e parâmetros), 95% dos intervalos

de confiança englobaria o valor verdadeiro. Os intervalos foram calculados usando as estimativas das médias e dos parâmetros de dispersão e os quantis de 2,5% para o limite inferior e 97,5% para o limite superior por meio das funções `qNO`, `qPO` e `qNBI` da biblioteca `gamlss.dist` do R (Stasinopoulos *et al.* [2010]) para as distribuições Normal, Poisson e Binomial Negativa respectivamente. Para os modelos de regressão linear e ARMA, o intervalo foi feito para os logaritmos da taxa e para os modelos MLG, GARMA e GAMLSS, o intervalo foi calculado para o número de internações. Esta medida é usada para quantificar quanto o modelo acerta, levando em consideração a média e variabilidade da estimativa, diferentemente do EQM que só considera a média.

Também se estudou a diferença entre os desvios padrões dos modelos, esta medida é diretamente proporcional à amplitude do intervalo de confiança.

2.8 Imputação

Em bancos de dados reais é comum não ter informação completa das variáveis, tendo dias em que estas variáveis não são observadas. Em modelos de regressão como regressão linear, MLG e GAMLSS, que não contemplam a presença de correlação serial, a falta de informação já é prejudicial às estimativas, pois para quaisquer y_t é preciso ter todos os x_t para determinado tempo t , caso contrário o y_t não será analisado. Em modelos que utilizam informação passada, como o ARMA e o GARMA, os dados faltantes são ainda mais prejudiciais, porque se, por exemplo, retirar um y_t da análise por não se ter as variáveis explicativas deste mesmo tempo t , impossibilitará o uso das variáveis resposta dos tempo futuros, por precisarem da informação de y_t .

A solução para esse problema está em imputar os dados faltantes. Substitui-se este dado faltante por uma estimativa para que seja possível prosseguir com a análise.

Entre os possíveis métodos de estimação, foi escolhido o algoritmo EM (Dempster *et al.* [1977]) para a distribuição Normal condicional. Esse método é interessante por possibilitar a inclusão da estrutura temporal da série, situação muito comum quando se trabalha com taxa diária de internações ou de óbitos.

O algoritmo EM é uma ferramenta computacional utilizada para o cálculo do estimador de máxima verossimilhança de forma iterativa e é principalmente utilizado em problemas envolvendo dados incompletos. Cada iteração do estimador é dividida em duas etapas: o passo E, que calcula o valor esperado do logaritmo de máxima verossimilhança dado $(\mu^{(k)}, \vartheta_2^{(k)})$, e o passo M, que maximiza este valor, obtendo $(\mu^{(k+1)}, \vartheta_2^{(k+1)})$, em que k é o número da iteração. O processo é finalizado quando se atinge a convergência.

Para imputar os dados supõe-se que eles tenham distribuição Normal multivariada. Tendo as variáveis observadas como explicativas e a variável que contém o dado faltante como resposta, por meio do algoritmo EM, obteve-se as estimativas para os parâmetros. Então, tendo os parâmetros estimados das variáveis explicativas obtêm-se a estimativa para o dado faltante.

Dada a matriz de dados $\mathbf{Z}_{(T,m)}$ em que \mathbf{Z} é a matriz das m variáveis utilizadas para a imputação e T é o número de dias observados, a estatística suficiente consiste em somas das colunas de \mathbf{Z} , totalizando m estatísticas lineares, mais estatísticas quadráticas formadas pelas somas de quadrados e somas de produtos correspondentes a cada coluna e pares de colunas de \mathbf{Z} , totalizando $\frac{1}{2}m(m+1)$ estatísticas quadráticas. É necessário substituir as partes faltantes das somas e das somas dos quadrados e produtos pela esperança condicional às observações e aos parâmetros populacionais ajustados. Então, para cada valor faltante é preciso computar a média, a média ao quadrado e a média dos produtos dos valores faltantes de sua respectiva linha e das observações passadas a qual ele depende.

Este método foi aplicado aos dados do exemplo do Capítulo 3 utilizando a função `mnimput` da biblioteca `mtsdi` do R (Junger e Junger [2012]).

Capítulo 3

Aplicação

3.1 Dados utilizados

Neste trabalho foi analisado o número diário de internações por causas respiratórias em pessoas com 65 anos ou mais, residentes no município de São Paulo, Brasil, de janeiro de 2006 a dezembro de 2011 (2191 dias). Este estudo tem como objetivo comparar os modelos apresentados no Capítulo 2 para avaliar o impacto de variáveis de poluição, climáticas, dias da semana e feriados na taxa diária de internações.

O banco de dados com informações de poluição e variáveis climáticas foi obtido pela Companhia de Tecnologia de Saneamento Ambiental (CETESB), ligada à Secretaria do Meio Ambiente do governo paulista. Ele contém medidas diárias de material particulado, gás carbônico, gás ozônio, temperatura mínima, umidade relativa do ar, ácido nítrico, dióxido de nitrogênio, óxidos de nitrogênio e dióxido de enxofre. O número diário de internações por problemas respiratórios em pessoas com 65 anos ou mais foi obtido pelo Sistema de Informação Hospitalar (SIH) do Programa de Aprimoramento das Informações de Mortalidade (PRO-AIM) da Secretaria de Saúde do Município de São Paulo.

Dos 2191 dias, 293 não tinham medição de dióxido de enxofre, 136 não tinham medição de óxidos de nitrogênio, 85 de gás carbônico, 69 de material particulado, 56 de gás ozônio e 2 não tinham o número de internações.

As variáveis de óxido de nitrogênio e dióxido de enxofre possuem muitos dados faltantes, tendo grandes sequências de dias sem informação. A maior delas foi para o dióxido de enxofre, que teve 73 dias consecutivos sem dados, compreendendo do dia 05/02/2006 ao dia 18/04/2006, além de outras duas sequências maiores que duas semanas. As variáveis de óxido de nitrogênio também tiveram três sequências maiores que duas semanas, sendo a maior delas de 23 dias, entre 18/01/2011 e 04/02/2011. Neste período, também não houve informações de dióxido de enxofre, material particulado e gás carbônico. O gás ozônio não teve nenhuma sequência maior que duas semanas sem medições.

Como foram utilizados modelos para séries temporais, em que a variável resposta depende de observações passadas, todas as variáveis têm que ter informação completa. Como a imputação pode ficar incorreta para variáveis com longos períodos com falta de informação, algumas variáveis explicativas não foram incluídas nos modelos, como óxido de nitrogênio e dióxido de enxofre. As variáveis de gás carbônico, material particulado, gás ozônio e internações foram imputadas utilizando uma modificação do algoritmo EM, em que se possibilita considerar a estrutura de correlação de dados de séries temporais, como apresentado no capítulo anterior.

3.2 Análise descritiva

Nesta primeira parte da análise, foi verificada de forma descritiva a associação entre a taxa de internações e as variáveis explicativas e se avaliou o efeito dos dias da semana e de feriados nas taxas.

Os nomes das variáveis estão abreviados nas tabelas. As terminações min, med e max significam respectivamente a medição diária mínima, média e máxima das variáveis de poluição ou climáticas. A nomenclatura PM10, CO, T e Ur significam respectivamente material particulado, gás carbônico, temperatura e umidade relativa do ar.

Na Tabela 6.1 do Apêndice, estão os coeficientes de correlação linear de Pearson entre as variáveis. A correlação varia de -1 a 1. Coeficientes de Pearson positivos indicam que quanto maior for uma variável, maior será o valor, em média, da outra variável, considerando uma relação linear entre as variáveis. Para valores negativos ocorre o contrário, quanto maior o valor de uma variável, menor será o valor, em média, da outra variável. Variáveis com correlações próximas de zero são pouco correlacionadas. É importante que as variáveis explicativas tenham pouca correlação entre si, pois, caso contrário, tem-se multicolinearidade e os modelos ficam instáveis, com erros padrões elevados, o que infla o intervalo de confiança e reduz a significância da variável para explicar a taxa de internação diária.

Foi verificado que há grande correlação entre as variáveis do mesmo tipo de poluente ou variável climática, como PM10max e PM10med (corr = 0,821) e COmax e COmed (corr = 0,811). Devido a essa alta correlação, foi incluída nos modelos apenas a variável (dentre as variáveis de mesmo tipo) que tem maior correlação em absoluto com a internação diária, ou seja, aquela que tem maior associação com o número de internações. Sendo assim, segue as variáveis que foram incluídas nos modelos e suas respectivas correlações com a internação: PM10med (corr = 0,283), COmax (corr = 0,271), O3max (corr = 0,057), Tmin (corr = -0,278) e Urmin (corr = -0,167).

As variáveis PM10med, COmax e O3max têm uma correlação positiva com o número de internações, enquanto Tmin e Urmin têm correlação negativa. Essas correlações fazem sentido, já que quanto maior a concentração de poluentes e menor umidade e temperatura espera-se maior número médio de internações.

Este resultado também é observado nos gráficos de dupla escala apresentados na Figura 3.1, que mostram o comportamento ao longo do tempo da série das variáveis explicativas selecionadas e do número de internações. Do lado esquerdo está o eixo de internações e do direito o eixo da variável explicativa.

Por esses gráficos, observa-se que as variáveis PM10med e COmax têm comportamento semelhante ao das internações, que vão aumentando até o inverno e então diminuem no verão. A partir de 2009 há uma diminuição desta associação. Já a Tmin parece ter comportamento inverso, o aumento de internações está associado à queda de temperatura. A variável Urmin parece ter um efeito negativo no número de internações. A variável O3max parece não ter associação com o número de internações.

Como a variável internação apresenta um comportamento sazonal, com ciclos anuais, que se assemelham às funções periódicas seno e cosseno, foram usadas estas funções com periodicidade de um ano para explicar a sazonalidade das internações, como proposto em Serfling [1963] e também utilizado em diversos outros trabalhos na literatura. As variáveis PM10med, COmax e Tmin também apresentam comportamento sazonal senoidal. As variáveis O3max e Urmin não tem um comportamento sazonal tão nítido.

A Figura 6.1 apresenta os gráficos de dispersão entre a variável logaritmo da taxa de internações e as variáveis de poluição e climáticas. A partir de seus gráficos parece haver relação não linear entre o logaritmo da taxa e as variáveis de poluição e climáticas. Optou-se por não usar modelos não lineares (como os modelos GAM com lowess); porque queremos comparar os modelos lineares, já que estes possuem melhor interpretação.

Outro aspecto importante a ser considerado é o efeito do dia da semana (variável categórica) e da ocorrência de feriados (variável dicotômica) sobre as internações, pois o número de internações pode depender destas variáveis, como estudado, por exemplo, por Katsouyanni *et al.* [1996].

Como há forte sazonalidade anual no número de internações, a sazonalidade foi removida para avaliar o impacto dos dias da semana e feriados, já que os efeitos dessas variáveis podem estar confundidos com o efeito sazonal. Para isso, foi ajustado um modelo de regressão linear tendo o número de internações como variável resposta e seno e cosseno como explicativas. O resíduo deste

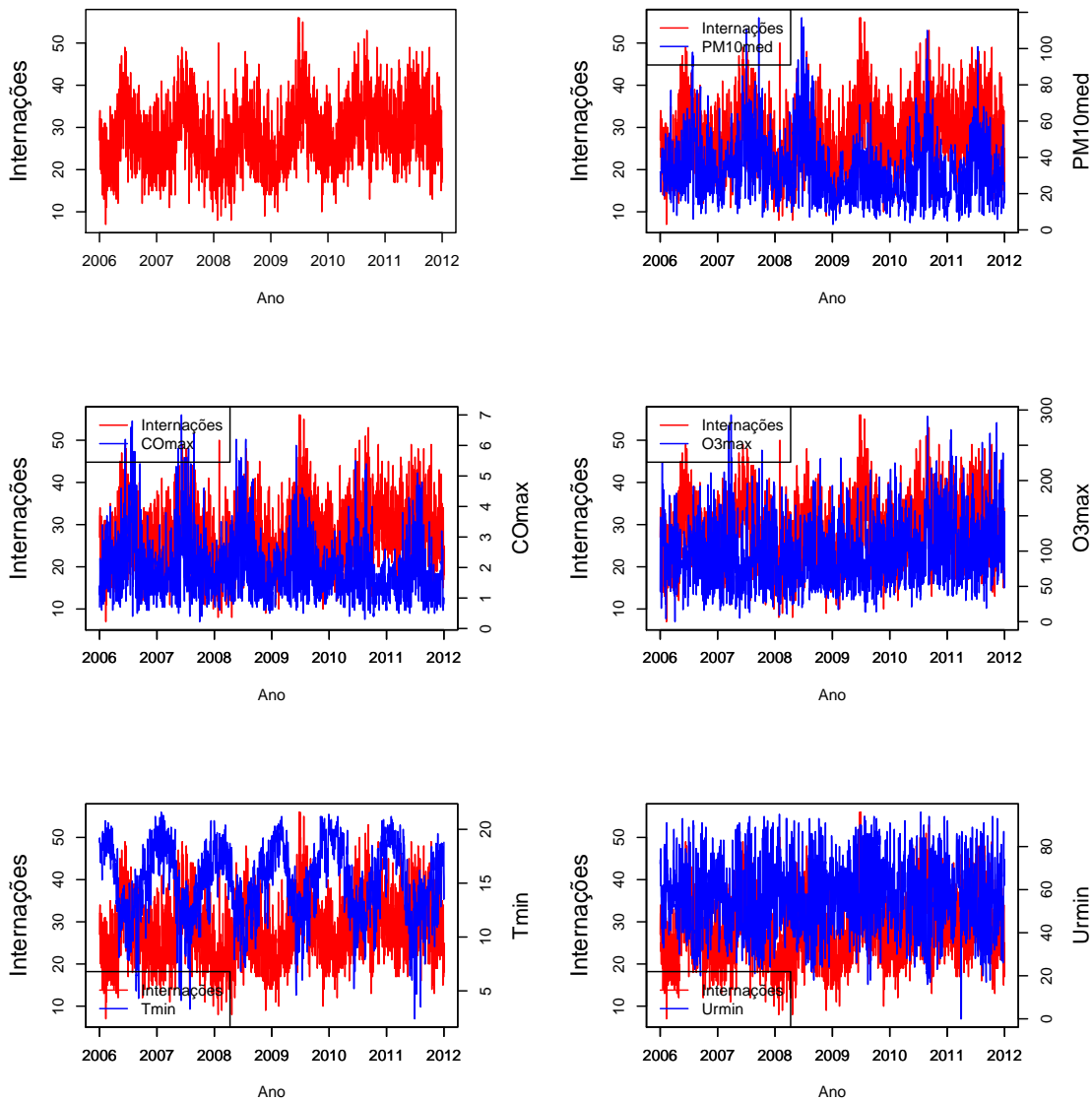


Figura 3.1: *Séries diárias*

modelo mais a média das internações foram chamados de série de internações dessazonalizada. Os boxplots da série dessazonalizada estão apresentados nas Figuras 3.2 e 3.3 e medidas descritivas na Tabela 3.1.

Sábado, domingo e feriados tiveram o menor número médio de internações. Segunda foi o dia com maior número médio de internações, isso é devido ao fato de sábado e domingo terem menores números de internações. Esse resultado também foi observado, por exemplo, por [Pereira *et al.* \[1998\]](#). A distribuição da série dessazonalizada entre terça e sexta-feira é bem semelhante. Então, no modelo para internações, foram incluídas somente as variáveis indicadoras: domingo, segunda e sábado. Assim, as internações de terça a sexta-feira foram utilizadas como referência para analisar os efeitos dos dias da semana.

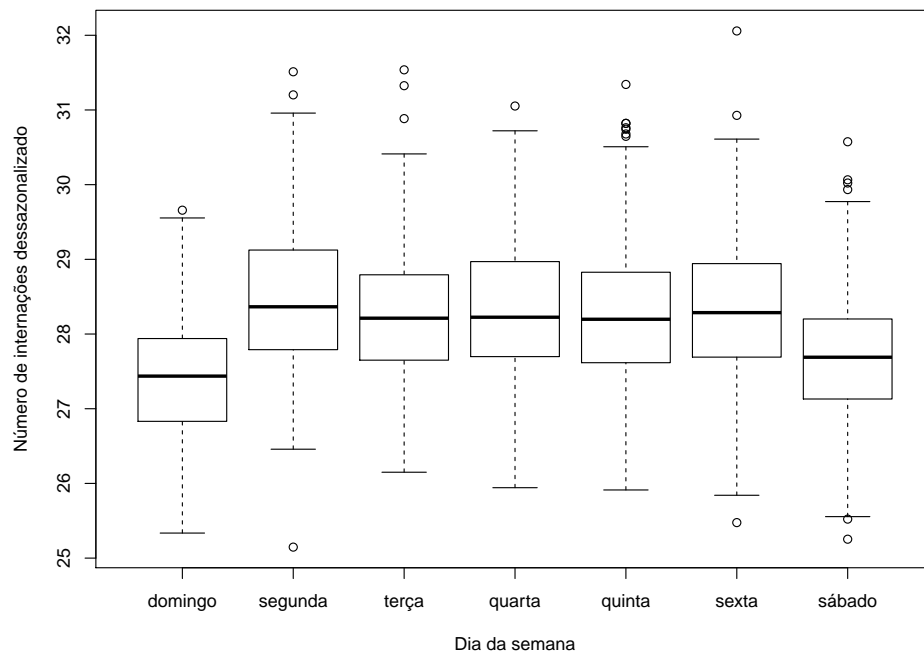


Figura 3.2: *Box-plot da variável interação dessazonalizada segundo dia da semana*

Categorias	n	Média	DP	Q3-Q1	Mínimo	1º quartil	Mediana	3º quartil	Máximo
Domingo	313	27,44	0,82	1,11	25,34	26,83	27,44	27,94	29,66
Segunda	313	28,47	0,99	1,33	25,15	27,79	28,36	29,12	31,51
Terça	313	28,28	0,92	1,14	26,15	27,65	28,21	28,79	31,54
Quarta	313	28,31	0,94	1,27	25,94	27,70	28,23	28,97	31,05
Quinta	313	28,28	0,99	1,21	25,91	27,62	28,20	28,83	31,34
Sexta	313	28,32	0,98	1,25	25,48	27,69	28,29	28,94	32,06
Sábado	313	27,69	0,89	1,07	25,25	27,13	27,69	28,20	30,57
Feriado	84	27,70	0,96	1,34	25,15	27,03	27,57	28,37	29,92

Tabela 3.1: *Medidas resumo do número de interações segundo dias da semana e feriado*

A Tabela 6.2 apresenta as correlações das variáveis seno, cosseno e das variáveis qualitativas domingo, segunda, sábado e feriados com todas as variáveis que serão utilizadas. A maior correlação em absoluto foi entre T_{min} e cosseno com 0,722, como esperado pois a temperatura tem comportamento sazonal, bem como diversas outras variáveis explicativas.

Pelas Tabelas 6.1 e 6.2, observa-se que as variáveis explicativas que tiveram correlação maior que 0,6 em absoluto foram: PM10_{med} e CO_{max} (corr = 0,704) e PM10_{med} e Ur_{min} (corr = -0,621).

As variáveis explicativas utilizadas nos modelos são:

1. Função seno - com periodicidade anual;
2. Função cosseno - com periodicidade anual;
3. Dias da semana - variáveis indicadoras de domingo, segunda e sábado, ou seja, cada variável é igual a um para o dia correspondente e zero, caso contrário. A categoria 'terça a sexta' será usada como referência;
4. Feriados - variável igual a 1 se feriado e 0, caso contrário;

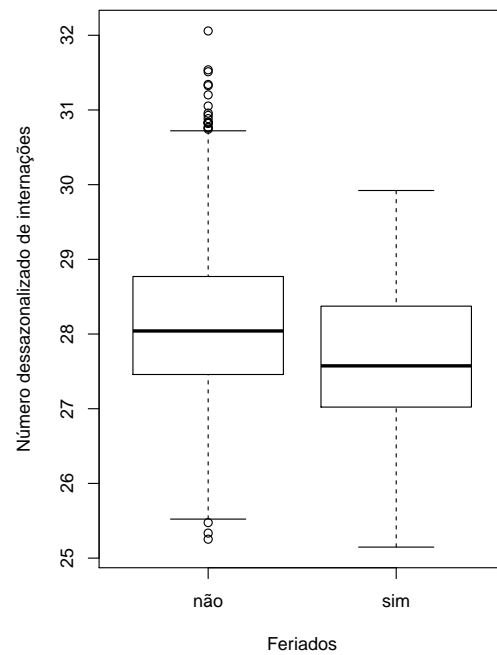


Figura 3.3: *Box-plot da variável internação dessazonalizada segundo feriado*

5. Material particulado média (PM10med) - medida em $\mu g/m^3$;
6. Gás carbônico máximo (COmax) - medida em $\mu g/m^3$;
7. Gás ozônio máximo (O3max) - medida em $\mu g/m^3$;
8. Temperatura mínima (Tmin) - medida em $^{\circ}C$;
9. Umidade relativa do ar mínima (Urmin) - medida em porcentagem.

3.3 Análise Inferencial

Para estudar a associação da poluição e variáveis climáticas com a série do número diário de internações, foram ajustados os seguintes modelos:

1. Regressão linear para o logaritmo da taxa;
2. Modelo autorregressivo e média móvel (ARMA) para o logaritmo da taxa;
3. Modelo linear generalizado (MLG) com as distribuições: Normal, Poisson e Binomial Negativa para a taxa com função de ligação logaritmo;
4. Modelo generalizado autorregressivo e média móvel (GARMA) com as distribuições: Normal, Poisson e Binomial Negativa para a taxa com função de ligação logaritmo;
5. Modelo generalizado aditivo para localização, escala e forma (GAMLSS) com as distribuições: Normal e Binomial Negativa para a taxa com função de ligação logaritmo;

Optou-se por não dessazonalizar as variáveis explicativas, para facilitar a interpretação; como foi feito em Bakonyi *et al.* [2004] e Gouveia *et al.* [2006].

Todas as análises foram realizadas no programa R versão 3.0.1. Para os modelos regressão linear, MLG e GAMLSS, foi usada a função `gamlss` da biblioteca `gamlss` e para os modelos ARMA e GARMA, foi usada a função `garmaFit` da biblioteca `gamlss.util`. Para os testes, utilizou-se um nível de significância de 5%.

O modelo mais simples e mais utilizado na prática para verificar a relação entre as variáveis climáticas e de poluição com a taxa de internações é o de regressão linear múltipla. Devido à natureza dos dados, é razoável que a variação das variáveis explicativas não tenha efeito aditivo na taxa de hospitalizações, mas sim, implique variação percentual da taxa. Portanto, o modelo apresenta o logaritmo da taxa de internações como variável resposta e para verificar o impacto das variáveis explicativas, foi estimada a variação percentual da taxa de internações. No modelo ARMA a variável resposta também foi o logaritmo da taxa de internações.

Os modelos MLG, GARMA e GAMLSS permitem a análise da contagem, que consiste no número diário de internações como variável resposta e admitem a inclusão de uma função de ligação e do termo offset. Para esses modelos foi adotada a função de ligação logaritmo neperiano e incluído como offset o logaritmo da população por cem mil habitantes. Desta forma, é possível analisar o efeito de variáveis explicativas sobre a taxa de internações, apesar de o modelo ter a contagem como resposta, e reestruturar a equação para que a interpretação seja relacionada à taxa média e não, à média dos logaritmos das taxas.

Nos modelos GAMLSS, foi utilizada a ligação logaritmo neperiano, não só para a equação da média μ (ou ϑ_1), mas também para a equação do parâmetro ϑ_2 que na distribuição Binomial Negativa, representa o parâmetro de dispersão e na distribuição Normal, representa o desvio padrão σ , que também é uma medida de dispersão. Para o cálculo da variância na distribuição Binomial Negativa, utiliza-se a média e o parâmetro ϑ_2 (ver equação 2.21). Primeiramente, foi foram incluídas somente as variáveis seno e cosseno na equação do parâmetro ϑ_2 , pois pelo gráfico da série de internações parece que os dias com maior internação tem variabilidade maior. Outra forma de avaliar quais variáveis influenciam a variância é utilizar um método de seleção de variáveis como o backward ou forward na equação do parâmetro ϑ_2 do modelo GAMLSS. Foi utilizado o método forward por ser mais apropriado para o caso em que as variáveis explicativas são multicolineares, já que as variáveis são acrescentadas uma a uma em vez de se iniciar o processo de seleção com todas as variáveis no modelo, inclusive as que apresentam multicolinearidade. Como o modelo GAMLSS foi usado principalmente para explicar também o desvio padrão, não foi utilizada a distribuição Poisson, já que esta distribuição só tem um parâmetro, de modo que a variância é igual à média.

Pela seleção das variáveis utilizando o método forward, foram incluídas no modelo para a variabilidade somente as variáveis cosseno, domingo e COMax para a distribuição Normal e feriados e Urmin para a distribuição Binomial Negativa.

Para simplificar a apresentação dos resultados, os modelos GAMLSS com distribuição Normal foram denominados modelos 1, 2 e 3 e os com distribuição Binomial Negativa de modelos 4, 5 e 6. As equações do parâmetro ϑ_2 dos modelos 1 e 4 contêm somente as variáveis seno e cosseno e as variáveis selecionadas no método forward. Foram utilizadas somente as variáveis seno e cosseno nos modelos 2 e 5 e somente as variáveis selecionadas pelo método forward nos modelos 3 e 6. O objetivo de considerar três modelos para cada distribuição foi avaliar se mudavam as estimativas e os erros padrões entre esses modelos e em relação aos modelos MLG e GARMA.

Os parâmetros de todos os modelos foram estimados usando o método de estimação de máxima verossimilhança (EMV). As estimativas dos coeficientes das variáveis explicativas foram transformadas usando a equação 2.1 para obter as variações percentuais da taxa de internações. O erro padrão das variações percentuais foi calculado utilizando-se o método delta, conforme equação 2.2.

3.3.1 Ajuste dos modelos

As tabelas a seguir apresentam as estimativas e seus respectivos erros padrões das variações percentuais da taxa média. Os resultados dos ajustes dos modelos de regressão linear, ARMA, MLG, GARMA e GAMLSS (Normal e Binomial Negativa) se encontram respectivamente nas Tabelas 3.2, 3.3, 3.4, 3.5 e 3.6. As estimativas das variáveis explicativas para a média foram transformadas em variações percentuais (usando a Equação 2.1) e seus respectivos erros padrões foram obtidos pelo método delta (por meio da Equação 2.2). As estimativas das variáveis explicativas da equação do parâmetro ϑ_2 foram exponencializadas e seus respectivos erros padrões também foram obtidos pelo método delta.

Foram escolhidos os modelos ARMA(1,1) e GARMA(1,1) para as três distribuições - Normal, Poisson e Binomial Negativa - por terem tido o melhor ajuste pela função de autocorrelação.

A versão de 2013 da função `gammaFit` do programa R, que foi utilizada para os modelos ARMA e GARMA, apresentou problemas de singularidade nos modelos GARMA com distribuições Poisson e Binomial Negativa para mais do que 1 parâmetro autorregressivo quando se tem 1 parâmetro de médias móveis e mais do que 2 parâmetros de autorregressivos quando não se tem parâmetros de médias móveis. Na última versão disponível, o problema foi praticamente corrigido havendo poucos casos de singularidade para, pelo menos, 5 parâmetros de médias móveis e 14 parâmetros autorregressivos.

Coeficientes	Regressão linear		ARMA	
	Estimativa (EP)	Valor p	Estimativa (EP)	Valor p
Intercepto	1,051 (0,048)	<0.001	1,038 (0,051)	<0.001
Seno	-4,601 (0,727)	<0.001	-4,782 (2,102)	0,026
Cosseno	-16,977 (0,942)	<0.001	-17,820 (1,975)	<0.001
Domingo	-19,387 (1,202)	<0.001	-20,377 (1,094)	<0.001
Segunda	4,219 (1,514)	0,004	4,239 (1,384)	0,002
Sábado	-14,633 (1,249)	<0.001	-14,859 (1,139)	<0.001
Feriados	-14,029 (2,203)	<0.001	-17,264 (1,960)	<0.001
PM10med	0,060 (0,050)	0,230	0,067 (0,054)	0,211
COMax	0,738 (0,719)	0,303	0,237 (0,721)	0,742
O3max	0,027 (0,013)	0,034	0,016 (0,012)	0,208
Tmin	1,075 (0,228)	<0.001	1,411 (0,248)	<0.001
Urmin	-0,116 (0,040)	0,004	-0,142 (0,039)	<0.001
ϕ			0,962 (0,010)	<0.001
θ			-0,863 (0,017)	<0.001
ϑ_2	0,229 (0,003)	<0.001	0,218 (0,003)	<0.001

Tabela 3.2: Estimativas dos modelos de regressão linear e ARMA.

Os efeitos das variáveis cosseno, domingo, segunda, sábado, feriados, Tmin e Urmin foram significativos em todos os modelos. A variável seno somente não foi significativa no modelo GARMA com distribuição Normal. As variáveis PM10med e COMax não foram significativas em nenhum modelo

Coeficientes	Normal		Poisson		Binomial negativa	
	Estimativa (EP)	Valor p	Estimativa (EP)	Valor p	Estimativa (EP)	Valor p
Intercepto	1,083 (0,044)	<0.001	1,074 (0,039)	<0.001	1,072 (0,045)	<0.001
Seno	-3,724 (0,679)	<0.001	-4,051 (0,599)	<0.001	-4,138 (0,696)	<0.001
Cosseno	-16,052 (0,900)	<0.001	-16,229 (0,783)	<0.001	-16,278 (0,908)	<0.001
Domingo	-19,022 (1,303)	<0.001	-19,034 (1,061)	<0.001	-19,041 (1,210)	<0.001
Segunda	4,897 (1,337)	<0.001	4,563 (1,208)	<0.001	4,471 (1,412)	0,001
Sábado	-13,927 (1,289)	<0.001	-14,039 (1,077)	<0.001	-14,069 (1,236)	<0.001
Feriados	-12,018 (2,438)	<0.001	-12,089 (1,986)	<0.001	-12,104 (2,263)	<0.001
PM10med	0,067 (0,043)	0,123	0,070 (0,040)	0,079	0,071 (0,047)	0,129
COmax	1,105 (0,644)	0,085	0,976 (0,578)	0,09	0,942 (0,674)	0,161
O3max	0,015 (0,011)	0,182	0,019 (0,010)	0,06	0,021 (0,012)	0,092
Tmin	0,945 (0,209)	<0.001	0,988 (0,186)	<0.001	1,001 (0,216)	<0.001
Urmin	-0,095 (0,037)	0,009	-0,096 (0,032)	0,003	-0,096 (0,038)	0,011
ϑ_2	6,141 (0,094)	<0.001			0,012 (0,001)	<0.001

Tabela 3.3: Estimativas dos modelos lineares generalizados.

Coeficientes	Normal		Poisson		Binomial negativa	
	Estimativa (EP)	Valor p	Estimativa (EP)	Valor p	Estimativa (EP)	Valor p
Intercepto	3,249 (0,049)	<0.001	3,228 (0,044)	<0.001	3,228 (0,050)	<0.001
Seno	-3,531 (2,085)	0,096	-4,974 (1,161)	<0.001	-5,107 (1,214)	<0.001
Cosseno	-16,430 (1,929)	<0.001	-17,070 (1,228)	<0.001	-17,109 (1,321)	<0.001
Domingo	-19,607 (1,191)	<0.001	-20,122 (0,997)	<0.001	-20,130 (1,124)	<0.001
Segunda	4,663 (1,220)	<0.001	4,402 (1,129)	<0.001	4,382 (1,296)	<0.001
Sábado	-14,186 (1,173)	<0.001	-14,491 (1,003)	<0.001	-14,534 (1,132)	<0.001
Feriados	-13,430 (2,283)	<0.001	-14,478 (1,902)	<0.001	-14,392 (2,130)	<0.001
PM10med	0,092 (0,048)	0,058	0,043 (0,046)	0,35	0,036 (0,052)	0,494
COmax	0,077 (0,645)	0,905	0,225 (0,612)	0,712	0,221 (0,701)	0,753
O3max	0,008 (0,012)	0,524	0,020 (0,011)	0,064	0,023 (0,013)	0,062
Tmin	1,396 (0,236)	<0.001	1,308 (0,227)	<0.001	1,270 (0,259)	<0.001
Urmin	-0,122 (0,037)	<0.001	-0,108 (0,035)	0,002	-0,105 (0,039)	0,007
ϕ	0,946 (0,011)	<0.001	0,810 (0,033)	<0.001	0,762 (0,045)	<0.001
θ	-0,815 (0,019)	<0.001	-0,606 (0,040)	<0.001	-0,549 (0,050)	<0.001
ϑ_2	5,924 (0,090)	<0.001			0,010 (0,001)	<0.001

Tabela 3.4: Estimativas dos modelos GARMA.

para explicar a taxa média. A variável O3max foi significativa somente no modelo de regressão linear.

As estimativas são mais semelhantes entre os modelos do que os erros padrões. Isso é esperado, porque os estimadores foram obtidos pelo método de verossimilhança, que fornece um estimador assintoticamente não viesado. Porém a variância fica viesada quando houver heteroscedasticidade ou autocorrelação, podendo ser super ou subestimadas. Na seção 3.3.4, foi analisada a suposição de homocedasticidade dos resíduos padronizados dos modelos.

É interessante ressaltar algumas observações sobre os erros padrões:

- As variáveis que apresentaram maior disparidade do erro padrão entre os modelos foram seno e cosseno. Os erros padrões mais altos para essas duas variáveis foram nos modelos ARMA e GARMA com distribuição Normal, em que estiveram em torno de 2. Foram menores nos modelos MLG e GAMLSS, que tiveram erros padrões com valores semelhantes.
- Os erros padrões das estimativas das variáveis indicadoras de dia da semana são semelhantes em relação às mesmas variáveis em modelos diferentes.
- Os erros padrões das estimativas do modelo MLG com distribuição Poisson são menores do

Coeficientes de μ	Modelo 1			Modelo 2		Modelo 3	
	Estimativa (EP)	Valor p		Estimativa (EP)	Valor p	Estimativa (EP)	Valor p
Intercepto	1,080 (0,045)	<0.001		1,081 (0,045)	<0.001	1,080 (0,045)	<0.001
Seno	-3,943 (0,693)	<0.001		-3,853 (0,692)	<0.001	-3,917 (0,688)	<0.001
Cosseno	-16,375 (0,908)	<0.001		-16,384 (0,907)	<0.001	-16,358 (0,905)	<0.001
Domingo	-18,971 (1,176)	<0.001		-18,955 (1,315)	<0.001	-18,967 (1,176)	<0.001
Segunda	4,424 (1,375)	0,001		4,523 (1,349)	<0.001	4,437 (1,374)	<0.001
Sábado	-14,121 (1,317)	<0.001		-14,099 (1,301)	<0.001	-14,107 (1,316)	<0.001
Feriados	-12,531 (2,420)	<0.001		-12,692 (2,433)	<0.001	-12,490 (2,418)	<0.001
PM10med	0,068 (0,046)	0,134		0,063 (0,044)	0,159	0,069 (0,046)	0,132
COmax	0,927 (0,686)	0,175		1,025 (0,658)	0,118	0,923 (0,685)	0,176
O3max	0,017 (0,012)	0,147		0,017 (0,011)	0,146	0,017 (0,012)	0,146
Tmin	1,005 (0,216)	<0.001		0,979 (0,215)	<0.001	1,003 (0,216)	<0.001
Urmin	-0,103 (0,037)	0,006		-0,097 (0,037)	0,009	-0,103 (0,037)	0,006
ϑ_2							
Intercepto	-0,392 (0,038)	<0.001		-0,330 (0,015)	<0.001	-0,392 (0,038)	<0.001
Seno	0,993 (0,022)	0,764		0,998 (0,023)	0,925		
Cosseno	0,954 (0,021)	0,034		0,936 (0,020)	0,001	0,954 (0,021)	0,034
Domingo	0,862 (0,038)	<0.001				0,862 (0,038)	<0.001
COmax	1,041 (0,018)	0,018				1,040 (0,017)	0,018

Tabela 3.5: Estimativas dos modelos GAMLSS com distribuição Normal.

Coeficientes	Modelo 4			Modelo 5		Modelo 6	
	Estimativa (EP)	Valor p		Estimativa (EP)	Valor p	Estimativa (EP)	Valor p
Intercepto	1,079 (0,045)	<0.001		1,076 (0,045)	<0.001	1,076 (0,045)	<0.001
Seno	-3,854 (0,699)	<0.001		-3,999 (0,700)	<0.001	-3,984 (0,695)	<0.001
Cosseno	-16,339 (0,905)	<0.001		-16,347 (0,908)	<0.001	-16,331 (0,906)	<0.001
Domingo	-18,938 (1,206)	<0.001		-19,006 (1,208)	<0.001	-18,961 (1,210)	<0.001
Segunda	4,710 (1,407)	<0.001		4,522 (1,410)	0,001	4,637 (1,410)	<0.001
Sábado	-14,173 (1,223)	<0.001		-14,051 (1,234)	<0.001	-14,207 (1,225)	<0.001
Feriados	-11,853 (2,773)	<0.001		-11,950 (2,280)	<0.001	-12,264 (2,757)	<0.001
PM10med	0,070 (0,045)	0,119		0,069 (0,046)	0,133	0,071 (0,046)	0,119
COmax	0,929 (0,656)	0,155		0,995 (0,665)	0,133	0,880 (0,662)	0,182
O3max	0,018 (0,012)	0,148		0,018 (0,012)	0,134	0,019 (0,012)	0,118
Tmin	1,002 (0,213)	<0.001		0,985 (0,214)	<0.001	1,017 (0,214)	<0.001
Urmin	-0,104 (0,038)	0,006		-0,097 (0,038)	0,01	-0,104 (0,038)	0,006
ϑ_2							
Intercepto	-7,472 (0,495)	<0.001		-6,566 (0,123)	<0.001	-7,467 (0,467)	<0.001
Seno	1,106 (0,183)	0,541		1,074 (0,178)	0,666		
Cosseno	1,302 (0,241)	0,155		1,338 (0,246)	0,115		
Feriados	3,450 (1,195)	<0.001				3,654 (1,279)	<0.001
Urmin	1,015 (0,008)	0,059				1,014 (0,007)	0,047

Tabela 3.6: Estimativas dos modelos GAMLSS com distribuição Binomial Negativa.

que nos modelos MLG com distribuição Normal e Binomial Negativa, o mesmo acontece com os modelos GARMA, com exceção dos erros relativos aos parâmetros ϕ e θ na distribuição Normal. Este resultado é esperado, porque para esse tipo de dado é comum haver superdispersão e as distribuições Normal e Binomial Negativa permitem que a variância seja maior que a média enquanto para a Poisson a variância é igual a média. Na distribuição Normal, apesar da variância ser constante, ela não está atrelada à média, podendo então ser mais adequada para a superdispersão que a Poisson.

Em geral, proporcionalmente há maior diferença entre os erros padrões do modelo MLG e com relação aos do modelo GARMA do que com relação aos erros padrões do modelo GAMLSS. Ou seja,

modelar a heteroscedasticidade utilizando variáveis explicativas em vez da autocorrelação causou menor impacto nos valores dos erros padrões dos modelos MLG que estão calculados errados por não atender às suposições do modelo, como será mais abordado na Seção 3.3.2.

As variações médias entre os erros padrões dos modelos GARMA em relação aos dos modelos MLG foram maiores para a distribuição Normal.

O ideal seria ter um modelo que pudesse utilizar termos de séries temporais para modelar a média e que a variância fosse modelada com variáveis explicativas, uma generalização do modelo GARMA e GAMLSS.

3.3.2 Análise de resíduos

Os resíduos de todos os modelos foram transformados em resíduos de quantis como apresentado no Capítulo 2.6 para que se possa fazer diagnóstico do modelo.

Para verificar se o modelo está bem ajustado, as seguintes suposições para os resíduos quantílicos padronizados foram checadas:

1. Ausência de correlação serial;
2. Homocedasticidade, ou seja, variância constante;
3. Distribuição Normal;

Os 5 primeiros valores ajustados dos modelos que utilizavam termos de séries temporais (ARMA e GARMA) foram valores aberrantes. Como o valor atual depende da observação passada, a estimação pode ter sido prejudicada para os primeiros valores da série. Por isso, estes valores não serão considerados na análise de resíduos.

3.3.3 Ausência de correlação serial dos resíduos

Para verificar se os resíduos apresentaram correlação serial, foram apresentados no Apêndice gráficos de dispersão dos resíduos padronizados em função do valor ajustado e em relação ao tempo (os dois gráficos superiores das Figuras 6.2 a 6.15) e gráficos da função de autocorrelação (FAC) até a defasagem 33 (Figura 6.16 a 6.20).

Pelas Figuras 6.2 para o modelo regressão linear, 6.4, 6.5, 6.6, para os modelos MLG, e 6.10 a 6.15 para os modelos GAMLSS, os resíduos não estão aleatórios em relação ao tempo. Isso indica que eles podem apresentar dependência temporal. Esta dependência pode ser vista mais facilmente nos gráficos da função de autocorrelação de resíduos (Figuras 6.16, 6.17, 6.19 e 6.20). Estes modelos apresentaram autocorrelação significativa atingindo valores de aproximadamente 0,2 e com um decaimento lento. Os picos de autocorrelação foram para as defasagens: 2, 7, 9, 14, 15, 16 e 26. Os gráficos FAC dos modelos GAMLSS com distribuição Normal são parecidos (Figura 6.19), o mesmo se observa nos gráficos dos modelos GAMLSS com distribuição Binomial Negativa (Figura 6.20).

Nos modelos ARMA e GARMA com distribuição Normal, os pontos parecem se distribuir aleatoriamente no gráfico de resíduo padronizado em relação ao tempo, o que indica não haver autocorrelação dos resíduos (Figuras 6.3 e 6.7). Para os modelos GARMA com distribuição Poisson e Binomial Negativa, os pontos parecem estar aleatórios com leve tendência nas últimas 500 observações, o que indica haver leve autocorrelação dos resíduos em relação ao tempo (Figuras 6.8 e 6.9). Ambos os modelos tiveram valores de autocorrelação parecidos (Figura 6.18), prevalecendo baixas autocorrelações em absoluto, com algumas autocorrelações significativas, destacando para a defasagem de um dia com -0.094 para Poisson e -0.101 para a Binomial Negativa. Os modelos ARMA e GARMA com distribuição Normal tiveram melhor ajuste ao considerarmos a FAC (Figuras 6.16 e 6.18).

Em todos os modelos, o gráfico de dispersão dos resíduos padronizados por valor ajustado indica não haver tendência com relação ao valor ajustado.

3.3.4 Homocedasticidade

A suposição de variância constante dos resíduos foi investigada utilizando um modelo de regressão linear simples auxiliar tendo os resíduos ao quadrado como variável resposta e foi testado o efeito de cada variável explicativa, separadamente, como nos testes Breusch Pagan (Wooldridge [2012]). Esta é uma análise inicial para estudar a associação entre a variância dos resíduos e cada variável explicativa. Uma análise mais completa requer o uso do modelo GAMLSS, que permite incluir variáveis para explicar a variância.

A Tabela 3.7 indica quais variáveis explicativas foram significativas na regressão auxiliar tendo resíduo padronizado ao quadrado como variável resposta. As tabelas com os resultados dessas regressões estão nas Tabelas 6.4, 6.5, 6.6, 6.7 e 6.8 do Apêndice.

	Seno	Cosseno	Dom	Seg	Sáb	Feriados	PM10med	COmax	O3max	Tmin	Urmin
Reg. Lin.		x				x			x	x	x
ARMA		x				x				x	x
MLG - No		x	x				x	x			
MLG - Po						x					
MLG - BN						x					x
GARMA - No			x				x	x			
GARMA - Po		x				x					x
GARMA - BN		x				x				x	x
Modelo 1											
Modelo 2			x				x	x			
Modelo 3											
Modelo 4							x	x			
Modelo 5						x					
Modelo 6											

Tabela 3.7: Variáveis significativas na regressão auxiliar do resíduo ao quadrado.

Os modelos 3 e 6 são modelos reduzidos, em que a equação da variância só contém as variáveis selecionadas pelo método forward. Como esperado, o resíduo padronizado ao quadrado desse modelo não está associado a nenhuma variável explicativa. Somente as variâncias desses dois modelos e do modelo GAMLSS 1 não são influenciadas pelas variáveis explicativas. Este resultado é interessante, pois mostra a necessidade de utilizar o modelo GAMLSS para obtermos resíduos padronizados homoscedásticos em relação às variáveis explicativas. Também se conclui que utilizar somente as funções seno e cosseno não foi suficiente para explicar a variância.

No modelo 4, em que a equação do desvio padrão está em função das variáveis selecionadas pelo método forward e mais as funções seno e cosseno, as variáveis PM10med e COmax foram significativas para explicar os resíduos ao quadrado. Portanto, utilizar mais variáveis explicativas do que foram escolhidas pelo método de seleção pode, em vez de melhorar, prejudicar o ajuste do modelo.

Apesar dos efeitos das variáveis PM10med e COmax terem sido significativos na regressão auxiliar para os modelos com distribuição Normal MLG, GARMA e o modelo 2, o PM10max não foi selecionado na equação da variância do modelo 3, somente o COmax. Isso pode ter ocorrido devido à alta correlação entre eles ($\text{corr}=0,704$).

Também foi investigada a possível presença de heterocedasticidade condicional construindo-se as funções de autocorrelação dos resíduos padronizados ao quadrado.

Pelos gráficos de FAC dos resíduos padronizados ao quadrado (Figura 6.21, 6.22, 6.23, 6.24 e 6.25), observa-se que todos os modelos tiveram autocorrelações significativas dos resíduos padronizados ao quadrado para alguma defasagem. Apesar disso, as autocorrelações foram próximas de zero, com variação entre -0,042 a 0,085. Os modelos que não levaram em consideração a dependência temporal das observações atingiram autocorrelações mais altas do que os modelos ARMA e GARMA.

3.3.5 Normalidade

Os gráficos do núcleo da densidade, gráfico quantil-quantil da Normal e as medidas resumo: média, variância, curtose, assimetria dos resíduos foram usados para verificar a suposição de normalidade. Como se supõe que esses resíduos pertençam à distribuição Normal(0,1), espera-se que o coeficiente de assimetria seja igual a 0 e o de curtose igual a 3. Os dois gráficos da parte inferior das figuras de análise de resíduos padronizados apresentadas no Apêndice (6.2 a 6.15) mostram o núcleo da densidade e o gráfico quantil-quantil da Normal. As medidas resumo estão na Tabela 6.9.

Como o método de estimação por máxima verossimilhança é assintoticamente não viesado, todos os modelos tiveram resíduo com média próxima de 0, sendo que o valor mais distante do 0 foi de -0,0114 para o modelo MLG com distribuição Poisson. Todos os modelos tiveram variância próxima de 1, com exceção dos modelos MLG e GARMA ambos com distribuição Poisson tendo variâncias iguais a 1,34 e 1,26 respectivamente.

Os modelos com distribuição Binomial Negativa tiveram os menores coeficientes de assimetria. Deste, o maior valor foi igual a 0,0699 para o modelo 5. O modelo de regressão linear e o ARMA tiveram os maiores coeficientes de assimetria, -0,414 e -0,417 respectivamente, como pode ser observado na Figura 6.3. As demais distribuições tiveram assimetria maior que 0,29. As assimetrias podem ser mais facilmente observadas nos gráficos de núcleo de densidade.

Os modelos ARMA, regressão linear e GARMA com distribuição Normal tiveram o coeficientes de curtose mais distantes de 3 (coef: 3,94; 3,77 e 3,13 respectivamente). Os demais modelos tiveram coeficientes entre 2,93 e 3,07.

Pelas medidas resumo e pelos gráficos de núcleo e quartil-quartil, conclui-se que os modelos com distribuição Binomial Negativa tiveram resíduos padronizados mais próximos da distribuição Normal.

3.3.6 Erro quadrático médio e intervalo de confiança

Os erros quadráticos médios (EQM) para a contagem e os intervalos de confiança foram usados para avaliar a qualidade das estimativas.

A Tabela 6.10 apresenta os erros quadráticos dos modelos. O modelo ARMA teve o menor EQM, igual a 33,86, seguido pelos modelos GARMA (EQM = 34,38; 35,26; 35,54 respectivos às distribuições Normal, Poisson e Binomial Negativa). Quanto menor o valor do EQM, mais os valores preditos ficaram próximos dos valores observados. Os demais modelos tiveram EQM entre 37 e 38.

Os intervalos de confiança de 95% para o valor predito foram calculados usando os quantis de 2,5% para o limite inferior e 97,5% para o limite superior. Para os modelos regressão linear e ARMA, o intervalo foi feito para os logaritmos da taxa e para os modelos MLG, GARMA e GAMLSS, o intervalo foi calculado para o número de internações. A partir desses intervalos foi calculada a proporção de dias em que as observações estavam fora do intervalo de confiança. A amplitude do intervalo foi analisada estudando os desvios padrões ao longo do tempo, medida que é diretamente proporcional à amplitude do intervalo.

As Figuras 6.26, 6.27, 6.28, 6.29 e 6.30 respectivas aos modelos regressão linear e ARMA, MLG, GARMA, GAMLSS com distribuição Normal e GAMLSS com distribuição Binomial Negativa apresentam os gráficos com as observações, os valores ajustados e seus respectivos intervalos de confiança, a proporção de observações fora dos intervalos estão na Tabela 6.12 e os gráficos do desvio padrão ao longo do tempo foram apresentados nas Figuras 6.31, 6.32, 6.33, 6.34, 6.35, respectivas aos modelos regressão linear e ARMA, MLG, GARMA, GAMLSS com distribuição Normal e GAMLSS com distribuição Binomial Negativa, e a Tabela 6.11 contem as medidas resumo da média, mínimo, 1º quartil, mediana, 3º quartil e máximo dos desvios padrões.

Contanto que o modelo tenha poucos pontos fora do intervalo de confiança, é preferível que tenha intervalos pequenos.

Os modelos com distribuição Binomial Negativa têm desvios padrões maiores que o da Poisson, como já é esperado por se tratar de um modelo de superdispersão. Porém a média e mediana dos desvios padrões dos modelos a distribuição Binomial Negativa são menores que as dos modelos com

distribuição Normal.

Observando as porcentagens da Tabela 6.12, se verificou que os modelos que tiveram mais que 5% das observações fora do intervalo de confiança foram: modelo MLG com distribuição Poisson (7,5%), modelo GARMA com distribuição Poisson (6,8%) e o modelo 2 (5,2%). Os modelos com distribuições Binomial Negativa tiveram as menores proporções de erros, destes o menor valor ocorreu para o modelo GARMA (3,7%) e o maior para o modelo MLG (4%). Então, apesar dos modelos de distribuição Normal terem amplitude maior no intervalo de confiança, os modelos com distribuições Binomial Negativa tiveram maior proporção de observações dentro do intervalo e, levando em consideração o fato de que seus resíduos padronizados tiveram melhor aproximação para a Normal, são preferíveis em relação às distribuições Normal e Poisson.

Os modelos GAMLSS com distribuição Binomial Negativa tiveram desvios padrões parecidos, porém o modelo 5 - em que se usou somente as variáveis seno e cosseno para modelar o desvio padrão - teve um comportamento mais suave, os outros dois modelos tiveram dias em que o desvio padrão aumentou repentinamente.

Os modelos de regressão linear e ARMA tiveram intervalos de confiança parecidos com a mesma proporção de pontos fora do intervalo de confiança 4,5%. Ambos tiveram mais valores observados abaixo do limite inferior do que acima do limite superior, o que causou assimetria negativa.

Os modelos 1 e 3 (GAMLSS com distribuição Normal) tiveram desvios padrões muito semelhantes tendo uma amplitude maior no verão. O modelo 2 teve um comportamento senoidal.

3.4 Escolha e ajuste do modelo final

Os modelos GAMLSS, ARMA e GARMA tiveram bons ajustes. Os modelos GAMLSS são os únicos que possibilitam modelar a variância com variáveis explicativas. Por outro lado, os modelos ARMA e GARMA são os únicos que incluíram parâmetros para a autocorrelação da série. Estes modelos que consideram a autocorrelação dos resíduos também tiveram os menores EQMs.

Em relação às distribuições, os resíduos padronizados dos modelos com distribuição Binomial Negativa foram os que tiveram melhor aproximação para a Normal. Esta distribuição também teve maior proporção de observações dentro do intervalo de confiança e, além disso, em média, tiveram desvios padrões menores que a Normal.

Portanto, o modelo que aliou bom ajuste e boa capacidade de acerto foi o modelo GARMA com distribuição Binomial Negativa.

Comparado os modelos e escolhido o GARMA com distribuição Binomial Negativa, este foi utilizado para obter o modelo reduzido contendo somente variáveis significativas.

Inicialmente foi feita a seleção de variáveis explicativas pelo método forward no modelo sem termos de séries temporais. As variáveis que ficaram neste modelo por ordem de entrada foram:

- Cosseno;
- Domingo;
- Sábado;
- PM10med (material particulado médio);
- Feriados;
- Seno;
- Tmin (temperatura mínima);
- Segunda;
- Urmin (umidade relativa mínima do ar).

Após essa etapa, se incluiu termos autorregressivos e de médias móveis a fim de retirar a autocorreção dos resíduos. O melhor ajuste foi o GARMA(1,1), como nos modelos com todas as variáveis explicativas.

Neste modelo a variável PM10med deixou de ser significativa e foi retirada do modelo. Foram incluídas no modelo, as variáveis COmax (gás carbônico máximo) e O3max (gás ozônio máximo), separadamente. Das duas variáveis, somente o efeito de O3max foi significativo e, então, a variável foi acrescentada no modelo.

A equação do preditor linear, com função de ligação logaritmo, do modelo reduzido é:

$$\begin{aligned} \mathbf{x}'_t \hat{\boldsymbol{\beta}} &= 3,251 - 0.05359 \times \sin\left(\frac{2\pi t}{365}\right) - 0.1928 \times \cos\left(\frac{2\pi t}{365}\right) - 0.2288 \times \text{Domingo}_t \\ &+ 0.04139 \times \text{Segunda}_t - 0.1589 \times \text{Sabado}_t - 0.1572 \times \text{Feriados}_t + 0.0002478 \times \text{O3max}_t \\ &+ 0.01298 \times \text{Tmin}_t - 0.001283 \times \text{Urmin}_t \\ \eta_t &= \mathbf{x}'_t \hat{\boldsymbol{\beta}} + \log\left(\frac{\text{Pop}_t}{100000}\right) + 0.7570 \times \left\{ \log(y_{t-j}) - \mathbf{x}'_{t-1} \hat{\boldsymbol{\beta}} - \log\left(\frac{\text{Pop}_t}{100000}\right) \right\} \\ &- 0.5440 \times \{ \log(y_{t-1}) - \eta_{t-1} \}. \end{aligned} \quad (3.1)$$

$$(3.2)$$

O valor esperado de internações (μ_t) é obtido exponencializando a função de ligação η_t .

As estimativas do intercepto, do ϕ e do θ e as variações percentuais das estimativas dos parâmetros das variáveis explicativas deste modelo reduzido estão apresentadas na Tabela 3.8.

Coeficientes	Modelo reduzido	
	Estimativa (EP)	Valor p
Intercepto	3,251 (0,045)	<0,001
Seno	-5,218 (1,193)	<0,001
Cosseno	-17,538 (1,242)	<0,001
Domingo	-20,452 (1,078)	<0,001
Segunda	4,226 (1,285)	<0,001
Sábado	-14,692 (1,117)	<0,001
Feriados	-14,548 (2,122)	<0,001
O3max	0,025 (0,012)	0,044
Tmin	1,306 (0,257)	<0,001
Urmin	-0,128 (0,033)	<0,001
ϕ	0,757 (0,045)	<0,001
θ	-0,544 (0,050)	<0,001
ϑ_2	0,010 (0,001)	<0,001

Tabela 3.8: Estimativas do modelo final reduzido.

Em relação à terça a sexta feira, a taxa de internação diminui, em média, 14,692% aos sábados e 20,452% aos domingos, e aumenta, em média, 4,226% às segundas feiras. Nos feriados a taxa se reduz, em média, 14,548%.

A taxa de internação, em média, aumenta 0,025% para o aumento de $1\mu\text{g}/\text{m}^3$ de gás ozônio máximo, 1,306% para o aumento de 1°C da temperatura mínima e diminui 0,128% para o aumento de 1% na umidade mínima relativa do ar.

O efeito positivo da variável temperatura, diferente do esperado pela análise descritiva, ocorre ao controlar a sazonalidade anual da taxa de internações. Essa relação foi mais investigada gerando três modelos MLG com distribuições Binomial Negativa. O primeiro foi para a série de número de internações e com as variáveis explicativas seno, cosseno, temperatura mínima, o segundo modelo foi para a mesma variável resposta e somente temperatura mínima como variável explicativa, por fim, o último modelo foi para a série do número de internações dessazonalizada com a variável temperatura mínima dessazonalizada como explicativa. A sazonalidade foi extraída utilizando seno e cosseno no modelo MLG com distribuição Normal. Dentre esses três modelos, somente o segundo teve o efeito

da temperatura mínima negativo. Ou seja, a correlação negativa da temperatura mínima com o número de internações é devido a sazonalidade anual, e ao controlá-la através das funções seno e cosseno, o efeito passa a ser positivo. O efeito positivo da temperatura no número de internações de pessoas idosas por problemas respiratórios também foi observado por [Michelozzi *et al.* \[2009\]](#), [Baccini *et al.* \[2008\]](#), [Linares e Díaz \[2008\]](#) e [Kovats *et al.* \[2004\]](#).

A combinação das funções de seno e cosseno aumenta em média a taxa de internações no inverno e diminui no verão. No início do ano (verão), o valor de seno está próximo de zero e o do cosseno está próximo de um, então a taxa de internação reduz aproximadamente 17,538%; no início de março, o seno está em torno de um e o cosseno de zero, logo a taxa de internação reduz aproximadamente 5,218%; seguindo este raciocínio, no meio do ano (inverno), taxa aumenta aproximadamente 17,538%; e no início de outubro, a taxa de internação aumenta aproximadamente 5,218%.

Portanto, em média, aos sábados, domingos e feriados há uma redução das taxas de internação e segunda há um aumento. Aumento da umidade mínima relativa do ar está associado à redução da taxa de internação e o aumento da temperatura mínima aumenta a taxa de internação. Esse último efeito não era esperado, pois se espera que quanto menor a temperatura, maior a taxa de internação.

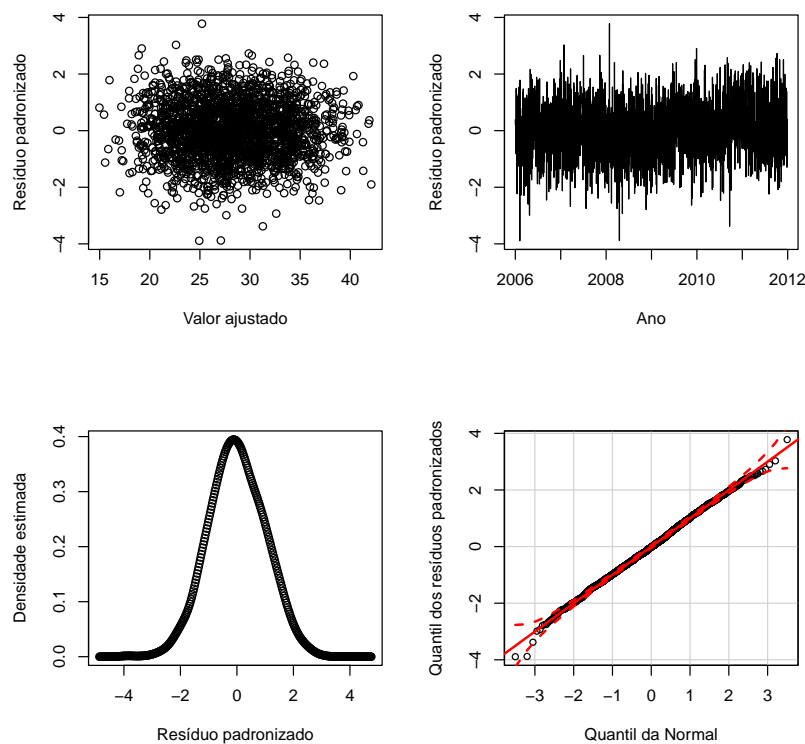


Figura 3.4: Análise dos resíduos padronizados do modelo reduzido

Pela Figura 3.4, os resíduos parecem estarem próximos da distribuição Normal. A Figura 3.5 mostra autocorrelações significativas tanto no gráfico de FAC para o resíduo padronizado quanto para o resíduo padronizado ao quadrado. Apesar disso, as autocorrelações estão entre -0,1 e 0,1, e por serem muito baixas não comprometem o cálculo da variância.

Pelo intervalo de confiança da Figura 3.6, conclui-se que o modelo conseguiu explicar bem as observações. Contendo um único valor muito fora do IC, o do dia 01/02/2008, em que teve 50 internações e o limite superior foi de 37.

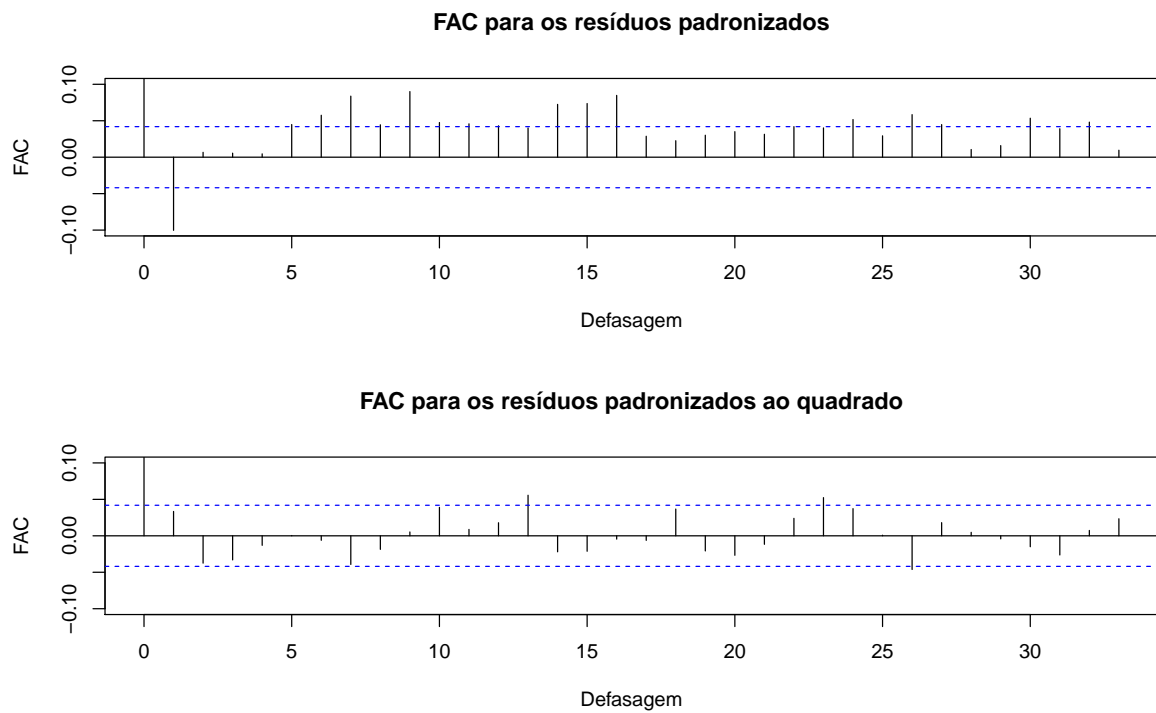


Figura 3.5: *FAC dos resíduos e dos resíduos ao quadrado do modelo reduzido.*

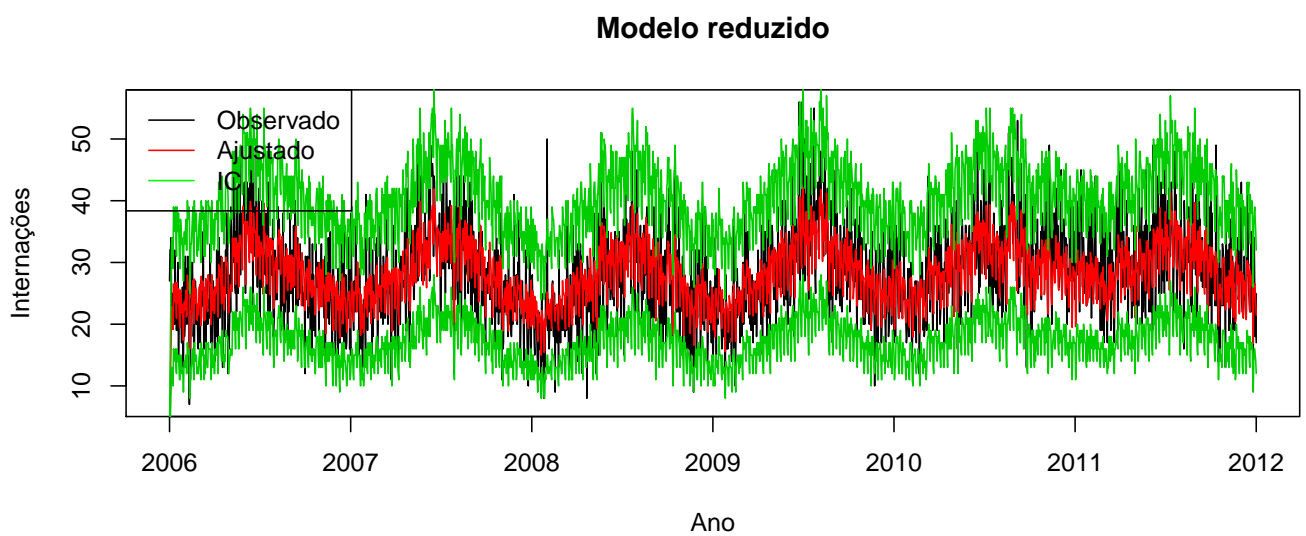


Figura 3.6: *Intervalo de confiança do modelo reduzido.*

Capítulo 4

Simulação

Em geral, para dados de séries temporais, é usual que as observações apresentem correlação serial. Se um modelo for proposto para a série temporal e a suposição de que os erros sejam independentes, ou mesmo não correlacionados, não for válida, os erros padrões apresentados na estimação são incorretos, podendo ser super ou subestimados. Nesse capítulo, foram realizadas simulações com o intuito de verificar a influência da autocorrelação nas estimativas dos parâmetros e nos erros padrões dos estimadores utilizados.

A série de dados foi simulada a partir do modelo GARMA com e sem autocorrelação, por meio da função `garsim` da biblioteca `gsarima` do programa R (Briet *et al.* [2000]). Para todas essas séries foram ajustados modelos que consideraram autocorrelação (modelos GARMA) e que não consideraram a autocorrelação (modelos MLG), utilizando respectivamente as funções `gammaFit` da biblioteca `gamlss.util` e `gamlss` da biblioteca `gamlss` do programa R.

Primeiramente, foi pensado para a simulação, utilizar as estimativas do modelo GARMA(1,1) com distribuição Binomial Negativa da Tabela 3.4 do Capítulo 3, por esse modelo ter tido o melhor ajuste. Essas séries seriam geradas utilizando a distribuição Binomial Negativa, com as variáveis explicativas e o offset da população. Porém, ao ajustar os modelos GARMA(1,0) e GARMA(1,1) com distribuições Normal e Binomial Negativa para essas séries simuladas, foi verificado que, ao incluir o offset, as estimativas estavam muito diferentes do valor verdadeiro tanto para as séries sem, como para as séries com autocorrelação. Era esperado que as estimativas estivessem muito próximas do valor verdadeiro, já que se utilizou o estimador de máxima verossimilhança que é assintoticamente não viesados. Para resolver este problema, optou-se por não incluir o offset nas simulações.

Logo, as séries foram geradas a partir do modelo GARMA(1,0) com distribuição Binomial Negativa e função de ligação logaritmo, usando as variáveis explicativas utilizadas no exemplo do Capítulo 3. A equação do preditor linear do modelo GARMA(1,0) que será usado para a simulação é dada por:

$$\begin{aligned} \eta_t = & 3,228 - 0,05242 \times \sin\left(\frac{2\pi t}{365}\right) - 0,1876 \times \cos\left(\frac{2\pi t}{365}\right) - 0,2248 \times \text{Domingo}_t \\ & + 0,04289 \times \text{Segunda}_t - 0,1571 \times \text{Sabado}_t - 0,1554 \times \text{Feriado}_t + 0,0003568 \times \text{PM10med}_t \\ & + 0,002204 \times \text{COmax}_t + 0,0002342 \times \text{O3max}_t + 0,01262 \times \text{Tmin}_t - 0,001055 \times \text{Urmin}_t \\ & + \log\left(\frac{\text{Pop}_t}{100000}\right) + \phi \times \log(y_{t-1}). \end{aligned} \quad (4.1)$$

e o parâmetro de dispersão é $\vartheta_2 = 0,010068$.

Os valores de ϕ foram fixados iguais a 0, 0,3 e 0,5 e para cada ϕ foram simuladas 1000 séries temporais de tamanho 2191, que é o mesmo tamanho da série do exemplo do Capítulo 3.

Para cada série simulada, foram ajustados os modelos GARMA(1,1), GARMA(1,0) e MLG com distribuições Normal e Binomial Negativa usando as variáveis explicativas do preditor linear

(4.1) e função de ligação logaritmo, totalizando 18000 modelos ajustados. Os efeitos das variáveis explicativas são expressos por meio da variação percentual na taxa média de internações devido ao acréscimo de uma unidade em cada variável explicativa. Essas variações foram calculadas usando a Equação (2.1).

As estimativas do modelo GARMA(1,1) com distribuição Binomial Negativa do exemplo do Capítulo 3 usadas nas simulações, chamadas neste capítulo de valores verdadeiros, e a variação percentual das estimativas das variáveis explicativas são apresentadas na Tabela 4.1.

Parâmetros fixados	Valor verdadeiro	Erro padrão	Var. % do valor verdadeiro	Erro padrão da var. %
Intercepto	32,281	0,0504		
Seno	-0,0524	0,0128	-5,107	1,214
Cosseno	-0,1876	0,0159	-17,109	1,321
Domingo	-0,2248	0,0141	-20,130	1,124
Segunda	0,0429	0,0124	4,382	1,296
Sábado	-0,1571	0,0132	-14,534	1,132
Feriado	-0,1554	0,0249	-14,392	2,130
PM10med	0,0004	0,0005	0,036	0,052
COmax	0,0022	0,0070	0,221	0,701
O3max	0,0002	0,0001	0,023	0,013
Tmin	0,0126	0,0026	1,270	0,259
Urmin	-0,0011	0,0004	-0,105	0,039
ϑ_2	0,0101	0,0014		

Tabela 4.1: Estimativas do modelo GARMA(1,1) com distribuição Binomial Negativa da Tabela 3.4.

A Tabela 4.2 mostra o número de erros de estimação durante as simulações. Foram analisados somente os resultados dos ajustes que não apresentaram problemas.

ϕ	Modelo	Distribuição	Nº de erros de estimação
0	GARMA(1,1)	Normal	6
0	GARMA(1,0)	Normal	9
0	MLG	Normal	1
0	GARMA(1,1)	Binomial Negativa	9
0	GARMA(1,0)	Binomial Negativa	6
0	MLG	Binomial Negativa	1
0,3	GARMA(1,1)	Normal	8
0,3	GARMA(1,0)	Normal	4
0,3	MLG	Normal	0
0,3	GARMA(1,1)	Binomial Negativa	6
0,3	GARMA(1,0)	Binomial Negativa	3
0,3	MLG	Binomial Negativa	0
0,5	GARMA(1,1)	Normal	9
0,5	GARMA(1,0)	Normal	2
0,5	MLG	Normal	0
0,5	GARMA(1,1)	Binomial Negativa	11
0,5	GARMA(1,0)	Binomial Negativa	4
0,5	MLG	Binomial Negativa	0

Tabela 4.2: Número de modelos com erro na estimação.

As medidas resumo: média, mínimo, 1º quartil, mediana, 3º quartil e máximo das estimativas dos modelos com distribuição Normal estão nas Tabelas 4.3, 4.7 e 4.11 e são respectivas à ϕ verdadeiro igual a 0, 0,3 e 0,5. Analogamente as medidas resumo das estimativas dos modelos com distribuição Binomial Negativa estão nas Tabelas 4.5, 4.9 e 4.13. Nas últimas colunas dessas tabelas, foram

incluídas as variações percentuais das médias e das medianas das estimativas em relação aos valores verdadeiros dos parâmetros. Os valores verdadeiros do intercepto, do ϕ , do θ e do ϑ_2 e os valores verdadeiros das variações percentuais correspondentes às variáveis explicativas são apresentados na primeira coluna, abaixo do nome de cada variável.

O desvio padrão estimado utilizando o modelo GARMA(1,1) com distribuição Binomial Negativa, cujas estimativas estão na Tabela 3.4, foi calculado utilizando a expressão (2.22). Depois foi calculada a média dos desvios padrões ao longo do tempo e obteve-se o valor 6,023.

Os parâmetros ϑ_2 dos modelos com distribuição Normal correspondem aos desvios padrões, enquanto que este parâmetro representa o parâmetro de dispersão da Binomial Negativa. Então, nas Tabelas 4.3, 4.7 e 4.11, que apresentam as estimativas dos modelos com distribuição Normal, o valor verdadeiro apresentado para o parâmetro ϑ_2 é a média dos desvios padrões ao longo do tempo.

Portanto, é possível comparar o parâmetro ϑ_2 dos modelos com distribuição Normal com a média do desvio padrão verdadeiro.

As Tabelas 4.4, 4.6, 4.8, 4.10, 4.12 e 4.14 apresentam as mesmas medidas resumo das estimativas para os erros padrões dos estimadores. Sendo as Tabelas 4.4, 4.8 e 4.12 referentes à distribuição Normal e as Tabelas 4.6, 4.10 e 4.14 referentes as distribuições Binomial Negativa. As duas últimas colunas dessas tabelas correspondem às variações percentuais das médias e das medianas dos erros padrões dos estimadores dos modelos GARMA(1,1) e MLG em relação, respectivamente, às médias e medianas dos erros padrões dos estimadores do modelo GARMA(1,0).

Como as estimativas do parâmetro de dispersão ϑ_2 dos modelos com distribuição Binomial Negativa tiveram valores muito baixos, as tabelas apresentam seus valores multiplicados por 100.

Pelas Tabelas 4.3, 4.5, 4.7, 4.9, 4.11 e 4.11, observa-se que, como esperado, as médias e medianas das estimativas do intercepto e das variáveis explicativas tiveram pouca diferença em relação ao valor verdadeiro. Com exceção da variável PM10med e COmax. A variável PM10med teve maiores variações percentuais da média para $\phi = 0,5$, atingindo o valor de 10,14% para o modelo GARMA(1,0) com distribuição Normal. Já a variável COmax teve variações percentuais da média maiores que 9% em todos os modelos, excluindo o GARMA(1,1) com distribuição Normal e ϕ verdadeiro igual a 0,3 e os modelos GARMA(1,1) e GARMA(1,0) com distribuições Normal e Binomial Negativa e ϕ verdadeiro igual a 0,5. Porém, as distribuições das variações percentuais das estimativas dessas duas variáveis estão assimétricas, nesses casos, é preferível analisar a mediana, por ser uma medida mais robusta. Analisando as variações percentuais da mediana, somente a variável COmax do modelo MLG com ϕ verdadeiro igual a 0,5 tiveram variações maiores que 10%, sendo iguais a 18,28% e 21,27% respectivamente às distribuições Normal e Binomial Negativa, note que esses valores são maiores que a variação percentual da média (9,17% e 12,34% respectivamente às distribuições Normal e Binomial Negativa). As demais variáveis, incluindo a variável PM10med, tiveram variações percentuais da mediana menor que 10%. Portanto, a autocorrelação de 0,5 na série afetou principalmente as estimativas do parâmetro da variável COmax nos modelos MLG, que tiveram variação percentual grande em relação ao valor verdadeiro.

As estimativas dos parâmetros ϕ , em média, ficaram próximas do valor verdadeiro. Quando havia autocorrelação nas séries, os valores mínimos das estimativas desse parâmetro ficaram muito próximos de zero. As estimativas máximas de ϕ tanto dos modelos GARMA(1,1) como dos modelos GARMA(1,0) com distribuição Normal ficaram mais afastadas do valor verdadeiro do que os da distribuição Binomial Negativa. As médias das estimativas de θ ficaram próximas de zero tanto para a distribuição Normal como para a distribuição Binomial Negativa, para os três ϕ fixados.

Para $\phi = 0$, houve pequena variação percentual na estimativa do parâmetro ϑ_2 . Porém, para $\phi = 0,3$, o modelo MLG com distribuição Binomial Negativa teve variação da média das estimativas de 43% em relação ao valor verdadeiro do parâmetro de dispersão, e para $\phi = 0,5$, a média das estimativas do parâmetro ϑ_2 para esse modelo teve variação de 155,86%.

O modelo GARMA(1,0) também com distribuição Binomial Negativa teve variação de 10,06% e 31,63% da média das estimativas do parâmetro ϑ_2 , respectivamente aos ϕ verdadeiros iguais a 0,3 e 0,5. Vale observar que, para essas duas autocorrelações, as distribuições das estimativas do

parâmetro ϑ_2 dos modelos GARMA(1,0) com distribuição Binomial Negativa estão assimétricas devido a altos valores que afetaram a média. Por isso, é preferível analisar a mediana. As variações percentuais das medianas das estimativas do parâmetro de dispersão dos modelos GARMA(1,0) com distribuição Binomial Negativa foram 3,5% e 2,25% respectivamente a $\phi = 0, 3$ e $\phi = 0, 5$.

As médias das estimativas de ϑ_2 dos modelos com distribuição Normal estiveram próximas do valor verdadeiro.

Logo, o modelo MLG foi o único modelo que não obteve estimativas próximas do valor real do parâmetro ϑ_2 quando houve autocorrelação nas séries.

A maioria das estimativas dos parâmetros das variáveis explicativas e do intercepto dos modelos com distribuição Normal tiveram os valores mínimo, 1º quartil, 3º quartil e máximo mais afastados do valor verdadeiro dos parâmetros do que os da distribuição Binomial Negativa.

Pelas Tabelas 4.4, 4.6, 4.8, 4.10, 4.12 e 4.14, verifica-se que, em geral, as distribuições dos erros padrões das estimativas das variáveis explicativas estão simétricas. As maiores assimetrias estão nas variáveis seno e cosseno do modelo GARMA(1,0) com distribuição Normal para ϕ verdadeiro igual a 0, fazendo com que nesse caso, as variações percentuais da média nos modelos GARMA(1,1) e MLG em relação ao modelo GARMA(1,0) sejam bastante diferentes da variação percentual da mediana. Para estes casos mencionados, a maior variação da média do erro padrão foi -11,82% para a variável seno do modelo MLG em relação ao modelo GARMA(1,0) com distribuição Normal, entretanto a variação percentual da mediana foi apenas -1,73%.

Todas as variações percentuais da mediana ficaram próximas de zero para $\phi = 0$. A maior variação foi -5,20% na variável O3max do modelo MLG em relação ao modelo GARMA(1,0) com distribuição Normal.

Como esperado, os erros padrões das estimativas foram afetados quando não se considera a autocorrelação serial. As variações percentuais média do erro padrão entre os modelos GARMA(1,1) e GARMA(1,0) não foram grandes (com exceção das variáveis seno e cosseno devido à assimetria).

As variáveis seno e cosseno foram as que tiveram maior variação percentual da média do erro padrão entre o modelo MLG e GARMA(1,0) quando havia autocorrelação nas séries, sendo maior para $\phi = 0, 5$, atingindo a variação de -37,43% para a variável seno e -27,10% para a variável cosseno, ambos para a distribuição Normal. Os resultados foram semelhantes para a distribuição Binomial Negativa.

As variáveis dicotômicas domingo, segunda, sábado e feriado tiveram grandes variações percentuais do erro padrão entre o modelo MLG e GARMA(1,0) para ambas distribuições quando ϕ era igual a 0,5. Sendo as variações maiores na variável feriado (24,29% e 19,53% respectivamente às distribuições Normal e Binomial Negativa) e menores na variável domingo (10,54% e 7,10% respectivamente às distribuições Normal e Binomial Negativa).

A variável Tmin também teve grandes variações percentuais que ficaram entre -14% e -12% para as duas distribuições quando havia autocorrelação nas séries, porém a variação não aumentou com o aumento da autocorrelação.

As distribuições dos erros padrões das estimativas dos parâmetros ϑ_2 para $\phi = 0, 5$ dos modelos GARMA(1,0) com distribuições Normal e Binomial Negativa estão assimétricas. As variações das medianas para esse parâmetro entre os modelos GARMA(1,1) e GARMA(1,0) são menores que 1%, diferentemente das variações percentuais da média, que foram iguais a -6,73% e -11,49%, respectivamente às distribuições Normal e Binomial Negativa. Porém, a variação percentual da mediana para o parâmetro ϑ_2 entre os modelos MLG e GARMA(1,0) foram de 16,15% e 35,31%, respectivamente às distribuições Normal e Binomial Negativa, valores maiores do que os da variação percentual da média, que foram iguais a 3,72% e 13,25%, respectivamente às distribuições Normal e Binomial Negativa.

Então, pode-se concluir que a omissão do parâmetro que considera autocorrelação dos dados fez com que o erro padrão não fosse estimado corretamente podendo ser super ou subestimado.

Variável/ Parâmetro	Modelo	Média	Mínimo	1º quartil	Mediana	3º quartil	Máximo	Var. % média	Var. % mediana
Intercepto 3,228	GARMA(1,1)	3,230	3,106	3,202	3,232	3,259	3,382	0,06	0,12
	GARMA(1,0)	3,230	3,107	3,202	3,232	3,259	3,383	0,06	0,12
	MLG	3,230	3,107	3,202	3,232	3,259	3,383	0,06	0,12
Seno -5,107	GARMA(1,1)	-5,094	-7,088	-5,566	-5,105	-4,636	-2,924	-0,25	-0,04
	GARMA(1,0)	-5,093	-7,088	-5,572	-5,106	-4,632	-2,926	-0,27	-0,02
	MLG	-5,092	-7,094	-5,569	-5,100	-4,637	-2,926	-0,29	-0,14
Cosseno -17,109	GARMA(1,1)	-17,040	-19,600	-17,670	-17,050	-16,420	-14,490	-0,40	-0,34
	GARMA(1,0)	-17,040	-19,610	-17,650	-17,040	-16,420	-14,470	-0,40	-0,40
	MLG	-17,040	-19,610	-17,650	-17,040	-16,430	-14,480	-0,40	-0,40
Domingo -20,130	GARMA(1,1)	-20,150	-25,150	-20,940	-20,190	-19,370	-16,500	0,10	0,30
	GARMA(1,0)	-20,150	-25,150	-20,940	-20,170	-19,380	-16,500	0,10	0,20
	MLG	-20,150	-25,190	-20,940	-20,160	-19,380	-16,510	0,10	0,15
Segunda 4,382	GARMA(1,1)	4,393	-0,266	3,551	4,457	5,250	8,398	0,24	1,71
	GARMA(1,0)	4,405	-0,243	3,556	4,474	5,260	8,393	0,52	2,10
	MLG	4,401	-0,250	3,559	4,472	5,251	8,381	0,43	2,06
Sábado -14,534	GARMA(1,1)	-14,520	-19,050	-15,310	-14,500	-13,760	-10,970	-0,10	-0,24
	GARMA(1,0)	-14,520	-19,040	-15,320	-14,510	-13,750	-10,740	-0,10	-0,17
	MLG	-14,520	-19,010	-15,320	-14,510	-13,760	-10,740	-0,10	-0,17
Feriado -14,392	GARMA(1,1)	-14,454	-20,718	-15,884	-14,460	-12,923	-7,694	0,43	0,47
	GARMA(1,0)	-14,451	-20,562	-15,886	-14,462	-12,922	-7,635	0,41	0,49
	MLG	-14,452	-20,562	-15,881	-14,458	-12,940	-7,635	0,42	0,46
PM10med 0,036	GARMA(1,1)	0,035	-0,101	0,005	0,034	0,064	0,205	-2,07	-4,30
	GARMA(1,0)	0,035	-0,100	0,005	0,035	0,065	0,204	-1,82	-2,12
	MLG	0,035	-0,103	0,005	0,035	0,065	0,202	-1,93	-2,35
COmax 0,221	GARMA(1,1)	0,194	-2,073	-0,272	0,205	0,656	2,456	-12,00	-7,01
	GARMA(1,0)	0,194	-2,083	-0,278	0,206	0,649	2,453	-12,22	-6,51
	MLG	0,196	-2,083	-0,278	0,201	0,656	2,452	-11,27	-8,78
O3max 0,023	GARMA(1,1)	0,023	-0,180	0,015	0,023	0,031	0,058	-2,58	-3,05
	GARMA(1,0)	0,023	-0,012	0,015	0,023	0,031	0,059	-1,25	-1,94
	MLG	0,023	-0,013	0,016	0,023	0,031	0,058	-0,19	-1,77
Tmin 1,270	GARMA(1,1)	1,262	0,587	1,110	1,257	1,411	1,969	-0,65	-1,00
	GARMA(1,0)	1,261	0,588	1,111	1,256	1,411	1,969	-0,73	-1,13
	MLG	1,261	0,588	1,112	1,255	1,412	1,969	-0,68	-1,17
Urmin -0,105	GARMA(1,1)	-0,106	-0,218	-0,131	-0,106	-0,081	0,007	0,32	0,50
	GARMA(1,0)	-0,106	-0,220	-0,131	-0,106	-0,081	0,006	0,22	0,62
	MLG	-0,106	-0,219	-0,131	-0,106	-0,081	0,007	0,11	0,73
ϕ 0,000	GARMA(1,1)	0,017	-0,233	-0,049	0,002	0,068	0,811		
	GARMA(1,0)	0,092	-0,069	-0,009	0,013	0,160	0,601		
θ 0,000	GARMA(1,1)	0,009	-0,437	-0,046	0,010	0,062	0,988		
ϑ_2 6,023	GARMA(1,1)	6,071	5,790	6,010	6,072	6,129	6,378	0,80	0,81
	GARMA(1,0)	6,072	5,791	6,011	6,073	6,131	6,378	0,82	0,83
	MLG	6,073	5,791	6,010	6,073	6,131	6,378	0,83	0,83

Tabela 4.3: Medidas resumo das estimativas das simulações com distribuição Normal e $\phi = 0,0$.

Variável	Modelo	Média	Mínimo	1º quartil	Mediana	3º quartil	Máximo	Var. % média	Var. % mediana
Intercepto	GARMA(1,1)	0,044	0,002	0,043	0,043	0,044	0,055	-3,52	-1,19
	GARMA(1,0)	0,045	0,040	0,043	0,044	0,047	0,056		
	MLG	0,043	0,041	0,043	0,043	0,043	0,045	-4,84	-1,62
Seno	GARMA(1,1)	0,675	0,282	0,643	0,655	0,668	1,305	-9,18	-1,71
	GARMA(1,0)	0,743	0,597	0,648	0,666	0,773	1,553		
	MLG	0,655	0,621	0,648	0,655	0,662	0,690	-11,82	-1,73
Cosseno	GARMA(1,1)	0,887	0,212	0,852	0,868	0,885	1,459	-6,90	-1,75
	GARMA(1,0)	0,952	0,786	0,859	0,883	0,992	1,629		
	MLG	0,867	0,811	0,857	0,867	0,878	0,915	-8,92	-1,83
Domingo	GARMA(1,1)	1,260	0,027	1,243	1,256	1,272	1,435	-0,57	-0,71
	GARMA(1,0)	1,267	1,191	1,245	1,265	1,286	1,363		
	MLG	1,254	1,192	1,241	1,254	1,267	1,317	-1,03	-0,87
Segunda	GARMA(1,1)	1,297	0,022	1,283	1,297	1,313	1,415	0,09	0,14
	GARMA(1,0)	1,296	1,228	1,282	1,295	1,312	1,371		
	MLG	1,295	1,231	1,281	1,294	1,310	1,370	-0,08	-0,08
Sábado	GARMA(1,1)	1,242	0,019	1,230	1,243	1,258	1,309	0,34	0,36
	GARMA(1,0)	1,238	1,128	1,224	1,239	1,254	1,309		
	MLG	1,244	1,177	1,231	1,244	1,258	1,309	0,48	0,40
Feriado	GARMA(1,1)	2,339	0,010	2,323	2,349	2,375	2,464	0,76	0,44
	GARMA(1,0)	2,321	1,900	2,300	2,339	2,369	2,464		
	MLG	2,346	2,240	2,322	2,346	2,370	2,452	1,08	0,30
PM10med	GARMA(1,1)	0,043	0,000	0,042	0,043	0,044	0,051	-2,80	-1,09
	GARMA(1,0)	0,045	0,039	0,043	0,044	0,046	0,055		
	MLG	0,043	0,041	0,042	0,043	0,043	0,045	-4,31	-2,13
COmax	GARMA(1,1)	0,623	0,004	0,615	0,623	0,632	0,725	-1,58	-1,01
	GARMA(1,0)	0,633	0,298	0,619	0,629	0,649	0,805		
	MLG	0,622	0,592	0,616	0,622	0,629	0,654	-1,74	-1,13
O3max	GARMA(1,1)	0,012	0,000	0,011	0,012	0,012	0,029	-0,99	-0,82
	GARMA(1,0)	0,012	0,010	0,012	0,012	0,012	0,013		
	MLG	0,011	0,011	0,011	0,011	0,011	0,012	-5,45	-5,20
Tmin	GARMA(1,1)	0,209	0,002	0,203	0,206	0,210	0,286	-4,85	-1,36
	GARMA(1,0)	0,220	0,191	0,205	0,209	0,231	0,292		
	MLG	0,206	0,197	0,204	0,206	0,208	0,218	-6,36	-1,53
Urmin	GARMA(1,1)	0,036	0,000	0,036	0,036	0,036	0,043	-1,67	-0,97
	GARMA(1,0)	0,037	0,034	0,036	0,036	0,038	0,040		
	MLG	0,036	0,034	0,035	0,036	0,036	0,037	-2,92	-2,06
ϕ	GARMA(1,1)	0,082	0,000	0,068	0,078	0,092	0,205		
	GARMA(1,0)	0,022	0,017	0,021	0,021	0,022	0,033		
θ	GARMA(1,1)	0,078	0,000	0,065	0,076	0,091	0,210		
	GARMA(1,0)	0,092	0,001	0,091	0,092	0,093	0,107	-0,83	-0,42
ϑ_2	GARMA(1,1)	0,092	0,001	0,091	0,092	0,093	0,107	-0,83	-0,42
	GARMA(1,0)	0,093	0,088	0,091	0,092	0,094	0,108		
	MLG	0,093	0,089	0,092	0,093	0,094	0,097	0,11	0,66

Tabela 4.4: Medidas resumo dos erros padrões das simulações com distribuição Normal e $\phi = 0,0$.

Variável/ Parâmetro	Modelo	Média	Mínimo	1º quartil	Mediana	3º quartil	Máximo	Var. % média	Var. % mediana
Intercepto 3,228	GARMA(1,1)	3,230	3,112	3,202	3,231	3,259	3,386	0,06	0,09
	GARMA(1,0)	3,230	3,112	3,203	3,231	3,259	3,386	0,06	0,09
	MLG	3,230	3,112	3,203	3,231	3,259	3,387	0,06	0,09
Seno -5,107	GARMA(1,1)	-5,091	-7,041	-5,556	-5,104	-4,622	-2,864	-0,31	-0,06
	GARMA(1,0)	-5,091	-7,039	-5,548	-5,106	-4,624	-2,867	-0,31	-0,02
	MLG	-5,090	-7,031	-5,548	-5,103	-4,624	-2,867	-0,33	-0,08
Cosseno -17,109	GARMA(1,1)	-17,040	-19,540	-17,630	-17,030	-16,430	-14,700	-0,40	-0,46
	GARMA(1,0)	-17,040	-19,540	-17,640	-17,020	-16,420	-14,670	-0,40	-0,52
	MLG	-17,040	-19,540	-17,630	-17,020	-16,410	-14,690	-0,40	-0,52
Domingo -20,130	GARMA(1,1)	-20,150	-25,110	-20,930	-20,180	-19,360	-16,460	0,10	0,25
	GARMA(1,0)	-20,150	-25,110	-20,930	-20,180	-19,370	-16,450	0,10	0,25
	MLG	-20,150	-25,180	-20,920	-20,190	-19,370	-16,460	0,10	0,30
Segunda 4,382	GARMA(1,1)	4,390	-0,400	3,557	4,446	5,263	8,625	0,18	1,47
	GARMA(1,0)	4,398	-0,370	3,582	4,452	5,262	8,520	0,37	1,60
	MLG	4,397	-0,375	3,578	4,459	5,272	8,546	0,35	1,76
Sábado -14,534	GARMA(1,1)	-14,520	-19,050	-15,360	-14,470	-13,760	-10,830	-0,10	-0,44
	GARMA(1,0)	-14,520	-19,020	-15,340	-14,460	-13,740	-10,830	-0,10	-0,51
	MLG	-14,520	-19,020	-15,340	-14,460	-13,750	-10,820	-0,10	-0,51
Feriado -14,392	GARMA(1,1)	-14,452	-20,506	-15,851	-14,446	-13,035	-7,915	0,42	0,38
	GARMA(1,0)	-14,447	-20,502	-15,838	-14,433	-13,029	-7,887	0,38	0,28
	MLG	-14,454	-20,259	-15,870	-14,456	-13,023	-7,887	0,43	0,44
PM10med 0,036	GARMA(1,1)	0,035	-0,099	0,006	0,035	0,065	0,203	-1,16	-2,88
	GARMA(1,0)	0,035	-0,098	0,005	0,034	0,066	0,202	-0,84	-3,44
	MLG	0,035	-0,093	0,005	0,034	0,065	0,200	-1,94	-3,91
COmax 0,221	GARMA(1,1)	0,195	-1,965	-0,266	0,206	0,655	2,266	-11,77	-6,47
	GARMA(1,0)	0,193	-1,990	-0,272	0,207	0,636	2,262	-12,68	-6,38
	MLG	0,196	-1,991	-0,263	0,204	0,638	2,262	-11,36	-7,37
O3max 0,023	GARMA(1,1)	0,024	-0,013	0,016	0,023	0,032	0,061	0,71	-1,64
	GARMA(1,0)	0,024	-0,013	0,017	0,024	0,032	0,060	3,23	1,09
	MLG	0,023	-0,013	0,016	0,023	0,031	0,057	-0,02	-2,07
Tmin 1,270	GARMA(1,1)	1,261	0,601	1,108	1,258	1,414	1,896	-0,73	-0,93
	GARMA(1,0)	1,260	0,601	1,108	1,258	1,414	1,896	-0,76	-0,96
	MLG	1,260	0,601	1,109	1,257	1,413	1,895	-0,76	-1,06
Urmin -0,105	GARMA(1,1)	-0,106	-0,221	-0,131	-0,105	-0,081	-0,005	0,07	-0,63
	GARMA(1,0)	-0,105	-0,222	-0,130	-0,104	-0,081	-0,006	-0,23	-1,39
	MLG	-0,106	-0,222	-0,131	-0,104	-0,081	-0,005	0,26	-1,15
ϕ 0,000	GARMA(1,1)	0,006	-0,228	-0,052	-0,002	0,059	0,327		
	GARMA(1,0)	0,039	-0,072	-0,010	0,006	0,039	0,417		
θ 0,000	GARMA(1,1)	0,000	-0,277	-0,050	0,005	0,056	0,302		
$\vartheta_2 \times 100$ 1,007	GARMA(1,1)	0,977	0,593	0,889	0,977	1,062	1,403	-2,98	-3,03
	GARMA(1,0)	0,980	0,595	0,890	0,980	1,065	1,406	-2,70	-2,69
	MLG	0,981	0,595	0,893	0,980	1,066	1,406	-2,56	-2,70

Tabela 4.5: Medidas resumo das estimativas das simulações com distribuição Binomial Negativa e $\phi = 0,0$.

Variável	Modelo	Média	Mínimo	1º quartil	Mediana	3º quartil	Máximo	Var. % média	Var. % mediana
Intercepto	GARMA(1,1)	0,044	0,041	0,043	0,044	0,044	0,052	-1,68	-0,63
	GARMA(1,0)	0,045	0,041	0,043	0,044	0,045	0,053		
	MLG	0,044	0,042	0,044	0,044	0,044	0,046	-1,77	-0,41
Seno	GARMA(1,1)	0,670	0,609	0,652	0,665	0,677	0,966	-3,91	-0,98
	GARMA(1,0)	0,697	0,608	0,657	0,671	0,693	1,072		
	MLG	0,666	0,631	0,658	0,665	0,673	0,700	-4,57	-0,91
Cosseno	GARMA(1,1)	0,878	0,795	0,857	0,871	0,888	1,174	-3,02	-1,06
	GARMA(1,0)	0,905	0,792	0,862	0,880	0,906	1,261		
	MLG	0,872	0,818	0,862	0,872	0,883	0,920	-3,58	-0,91
Domingo	GARMA(1,1)	1,164	1,101	1,149	1,163	1,177	1,271	-0,43	-0,34
	GARMA(1,0)	1,169	1,102	1,151	1,167	1,185	1,264		
	MLG	1,161	1,100	1,147	1,161	1,174	1,226	-0,68	-0,51
Segunda	GARMA(1,1)	1,365	1,289	1,348	1,365	1,384	1,453	0,22	0,22
	GARMA(1,0)	1,362	1,282	1,346	1,362	1,380	1,448		
	MLG	1,364	1,288	1,347	1,364	1,382	1,450	0,15	0,15
Sábado	GARMA(1,1)	1,192	1,112	1,179	1,192	1,207	1,254	0,17	0,17
	GARMA(1,0)	1,190	1,112	1,176	1,190	1,205	1,255		
	MLG	1,193	1,112	1,179	1,193	1,208	1,254	0,25	0,25
Feriado	GARMA(1,1)	2,163	2,007	2,132	2,163	2,193	2,292	0,19	0,09
	GARMA(1,0)	2,159	1,944	2,129	2,161	2,191	2,293		
	MLG	2,156	2,004	2,124	2,156	2,184	2,297	-0,14	-0,23
PM10med	GARMA(1,1)	0,045	0,039	0,044	0,045	0,046	0,060	-1,29	-0,51
	GARMA(1,0)	0,046	0,034	0,045	0,045	0,046	0,059		
	MLG	0,045	0,043	0,045	0,045	0,046	0,047	-1,62	-0,60
COmax	GARMA(1,1)	0,649	0,083	0,641	0,649	0,657	0,734	-0,81	-0,49
	GARMA(1,0)	0,655	0,132	0,643	0,652	0,664	0,730		
	MLG	0,649	0,619	0,642	0,649	0,656	0,683	-0,87	-0,57
O3max	GARMA(1,1)	0,012	0,011	0,012	0,012	0,012	0,012	-0,68	-0,43
	GARMA(1,0)	0,012	0,011	0,012	0,012	0,012	0,016		
	MLG	0,012	0,011	0,012	0,012	0,012	0,012	-0,17	0,00
Tmin	GARMA(1,1)	0,211	0,195	0,207	0,210	0,213	0,265	-2,40	-0,85
	GARMA(1,0)	0,216	0,195	0,208	0,212	0,217	0,274		
	MLG	0,210	0,201	0,208	0,210	0,213	0,222	-2,77	-0,71
Urmin	GARMA(1,1)	0,037	0,035	0,036	0,037	0,037	0,040	-0,86	-0,41
	GARMA(1,0)	0,037	0,035	0,036	0,037	0,037	0,041		
	MLG	0,037	0,035	0,036	0,037	0,037	0,039	-0,76	-0,24
ϕ	GARMA(1,1)	0,078	0,024	0,072	0,077	0,084	0,132		
	GARMA(1,0)	0,021	0,008	0,021	0,021	0,021	0,024		
θ	GARMA(1,1)	0,075	0,011	0,069	0,075	0,082	0,131		
$\vartheta_2 \times 100$	GARMA(1,1)	0,135	0,123	0,132	0,135	0,138	0,149	-0,07	-0,07
	GARMA(1,0)	0,135	0,123	0,132	0,135	0,138	0,149		
	MLG	0,136	0,124	0,133	0,136	0,139	0,150	0,67	0,74

Tabela 4.6: Medidas resumo dos erros padrões das simulações com distribuição Binomial Negativa e $\phi = 0,0$.

Variável/ Parâmetro	Modelo	Média	Mínimo	1º quartil	Mediana	3º quartil	Máximo	Var. % média	Var. % mediana
Intercepto 3,228	GARMA(1,1)	3,230	3,059	3,195	3,231	3,264	3,390	0,06	0,09
	GARMA(1,0)	3,228	3,029	3,192	3,229	3,263	3,385	0,00	0,03
	MLG	3,222	3,029	3,187	3,223	3,261	3,384	-0,19	-0,16
Seno -5,107	GARMA(1,1)	-5,062	-8,221	-5,689	-5,080	-4,395	-1,269	-0,88	-0,53
	GARMA(1,0)	-5,062	-8,091	-5,708	-5,068	-4,392	-1,474	-0,88	-0,76
	MLG	-5,086	-8,091	-5,750	-5,106	-4,404	-1,474	-0,41	-0,02
Cosseno -17,109	GARMA(1,1)	-17,060	-21,080	-17,850	-17,040	-16,290	-12,930	-0,29	-0,40
	GARMA(1,0)	-17,080	-21,070	-17,880	-17,070	-16,310	-12,950	-0,17	-0,23
	MLG	-17,150	-20,950	-17,930	-17,150	-16,370	-13,140	0,24	0,24
Domingo -20,130	GARMA(1,1)	-20,080	-24,640	-20,920	-20,110	-19,210	-16,560	-0,25	-0,10
	GARMA(1,0)	-20,090	-24,410	-20,930	-20,100	-19,220	-16,410	-0,20	-0,15
	MLG	-20,110	-24,410	-20,980	-20,110	-19,270	-16,410	-0,10	-0,10
Segunda 4,382	GARMA(1,1)	4,391	-1,026	3,429	4,419	5,342	9,372	0,20	0,84
	GARMA(1,0)	4,379	-1,026	3,393	4,408	5,328	9,376	-0,07	0,59
	MLG	4,291	-1,026	3,313	4,311	5,261	9,257	-2,08	-1,62
Sábado -14,534	GARMA(1,1)	-14,510	-18,660	-15,310	-14,430	-13,770	-10,270	-0,17	-0,72
	GARMA(1,0)	-14,510	-18,660	-15,350	-14,450	-13,750	-10,280	-0,17	-0,58
	MLG	-14,500	-18,690	-15,310	-14,480	-13,670	-10,450	-0,24	-0,37
Feriado -14,392	GARMA(1,1)	-14,390	-20,900	-15,772	-14,382	-13,059	-7,263	-0,01	-0,07
	GARMA(1,0)	-14,394	-20,912	-15,837	-14,417	-13,033	-7,372	0,01	0,17
	MLG	-14,404	-22,270	-16,019	-14,458	-12,943	-6,843	0,08	0,46
PM10med 0,036	GARMA(1,1)	0,036	-0,121	0,002	0,036	0,072	0,189	1,50	0,14
	GARMA(1,0)	0,037	-0,124	0,001	0,038	0,074	0,188	4,99	6,34
	MLG	0,038	-0,140	0,001	0,036	0,077	0,216	5,92	0,03
COmax 0,221	GARMA(1,1)	0,206	-1,736	-0,283	0,235	0,680	2,595	-6,56	6,58
	GARMA(1,0)	0,191	-2,403	-0,302	0,199	0,680	2,632	-13,31	-9,91
	MLG	0,190	-2,403	-0,324	0,210	0,705	2,632	-13,81	-4,93
O3max 0,023	GARMA(1,1)	0,024	-0,013	0,016	0,024	0,033	0,062	2,50	4,04
	GARMA(1,0)	0,025	-0,012	0,016	0,025	0,034	0,065	6,98	5,79
	MLG	0,024	-0,022	0,015	0,024	0,033	0,073	1,86	3,01
Tmin 1,270	GARMA(1,1)	1,254	0,388	1,084	1,254	1,436	2,260	-1,28	-1,29
	GARMA(1,0)	1,253	0,379	1,074	1,253	1,438	2,262	-1,30	-1,35
	MLG	1,259	0,379	1,071	1,255	1,447	2,190	-0,84	-1,15
Urmin -0,105	GARMA(1,1)	-0,107	-0,225	-0,135	-0,108	-0,079	0,021	1,57	2,52
	GARMA(1,0)	-0,106	-0,230	-0,135	-0,107	-0,079	0,021	0,81	1,79
	MLG	-0,107	-0,231	-0,137	-0,108	-0,077	0,032	1,52	1,98
ϕ 0,300	GARMA(1,1)	0,289	0,002	0,258	0,294	0,329	0,564		
	GARMA(1,0)	0,278	0,000	0,267	0,292	0,314	0,694		
θ 0,000	GARMA(1,1)	0,013	-0,258	-0,029	0,005	0,044	0,443		
ϑ_2 6,023	GARMA(1,1)	6,070	5,760	5,982	6,052	6,129	6,723	0,78	0,48
	GARMA(1,0)	6,143	5,760	6,018	6,106	6,260	6,705	1,99	1,38
	MLG	6,347	5,998	6,268	6,347	6,424	6,728	5,38	5,38

Tabela 4.7: Medidas resumo das estimativas das simulações com distribuição Normal e $\phi = 0,3$.

Variável	Modelo	Média	Mínimo	1º quartil	Mediana	3º quartil	Máximo	Var. % média	Var. % mediana
Intercepto	GARMA(1,1)	0,050	0,046	0,049	0,050	0,051	0,060	-0,10	0,16
	GARMA(1,0)	0,050	0,044	0,049	0,050	0,051	0,059		
	MLG	0,045	0,043	0,045	0,045	0,046	0,048	-9,68	-9,14
Seno	GARMA(1,1)	0,922	0,696	0,891	0,914	0,939	1,535	-0,75	0,66
	GARMA(1,0)	0,929	0,660	0,872	0,908	0,938	2,036		
	MLG	0,690	0,648	0,680	0,690	0,699	0,740	-25,69	-23,99
Cosseno	GARMA(1,1)	1,133	0,926	1,102	1,124	1,153	1,652	-0,39	0,46
	GARMA(1,0)	1,137	0,870	1,085	1,119	1,154	2,044		
	MLG	0,912	0,852	0,898	0,913	0,926	0,986	-19,75	-18,42
Domingo	GARMA(1,1)	1,306	1,218	1,283	1,301	1,320	1,550	-0,53	-0,61
	GARMA(1,0)	1,313	1,232	1,290	1,309	1,332	1,445		
	MLG	1,321	1,247	1,303	1,322	1,338	1,419	0,61	0,99
Segunda	GARMA(1,1)	1,308	1,226	1,286	1,303	1,323	1,461	-0,91	-0,76
	GARMA(1,0)	1,320	1,234	1,293	1,313	1,345	1,456		
	MLG	1,363	1,289	1,346	1,363	1,381	1,461	3,26	3,81
Sábado	GARMA(1,1)	1,229	1,154	1,211	1,226	1,243	1,351	-1,21	-0,65
	GARMA(1,0)	1,244	1,136	1,216	1,234	1,267	1,380		
	MLG	1,311	1,236	1,294	1,311	1,327	1,403	5,39	6,24
Feriado	GARMA(1,1)	2,261	2,055	2,227	2,258	2,292	2,554	-1,48	-0,62
	GARMA(1,0)	2,295	1,905	2,233	2,272	2,336	2,596		
	MLG	2,472	2,343	2,441	2,472	2,503	2,639	7,71	8,80
PM10med	GARMA(1,1)	0,048	0,013	0,047	0,048	0,049	0,064	-0,41	-0,02
	GARMA(1,0)	0,048	0,032	0,047	0,048	0,049	0,082		
	MLG	0,045	0,042	0,044	0,045	0,045	0,048	-7,22	-6,68
COmax	GARMA(1,1)	0,663	0,195	0,654	0,663	0,671	0,739	-0,78	-0,45
	GARMA(1,0)	0,668	0,509	0,656	0,666	0,676	1,175		
	MLG	0,655	0,621	0,647	0,655	0,663	0,694	-1,95	-1,64
O3max	GARMA(1,1)	0,012	0,010	0,012	0,012	0,012	0,013	-0,75	-0,66
	GARMA(1,0)	0,012	0,007	0,012	0,012	0,012	0,015		
	MLG	0,012	0,011	0,012	0,012	0,012	0,012	-4,35	-4,10
Tmin	GARMA(1,1)	0,253	0,217	0,248	0,252	0,256	0,314	0,32	0,28
	GARMA(1,0)	0,252	0,209	0,245	0,251	0,257	0,312		
	MLG	0,217	0,205	0,214	0,217	0,220	0,230	-13,89	-13,54
Urmin	GARMA(1,1)	0,039	0,032	0,038	0,039	0,039	0,042	-0,51	-0,44
	GARMA(1,0)	0,039	0,036	0,038	0,039	0,039	0,043		
	MLG	0,038	0,036	0,037	0,038	0,038	0,040	-3,45	-3,20
ϕ	GARMA(1,1)	0,051	0,021	0,047	0,051	0,054	0,116		
	GARMA(1,0)	0,021	0,020	0,020	0,020	0,021	0,026		
θ	GARMA(1,1)	0,051	0,006	0,048	0,051	0,055	0,102		
	GARMA(1,0)	0,092	0,087	0,090	0,091	0,093	0,113	-3,41	-0,90
ϑ_2	GARMA(1,1)	0,092	0,087	0,090	0,091	0,093	0,113	-3,41	-0,90
	GARMA(1,0)	0,096	0,087	0,091	0,092	0,102	0,114		
	MLG	0,097	0,092	0,096	0,097	0,098	0,103	1,32	5,04

Tabela 4.8: Medidas resumo dos erros padrões das simulações com distribuição Normal e $\phi = 0,3$.

Variável/ Parâmetro	Modelo	Média	Mínimo	1º quartil	Mediana	3º quartil	Máximo	Var. % média	Var. % mediana
Intercepto 3,228	GARMA(1,1)	3,229	3,047	3,195	3,229	3,264	3,389	0,03	0,03
	GARMA(1,0)	3,228	3,047	3,191	3,228	3,262	3,392	0,00	0,00
	MLG	3,222	3,016	3,188	3,221	3,261	3,390	-0,19	-0,22
Seno -5,107	GARMA(1,1)	-5,060	-8,084	-5,662	-5,082	-4,394	-1,397	-0,92	-0,49
	GARMA(1,0)	-5,060	-8,089	-5,674	-5,075	-4,378	-1,384	-0,92	-0,63
	MLG	-5,082	-8,001	-5,728	-5,102	-4,409	-1,496	-0,49	-0,10
Cosseno -17,109	GARMA(1,1)	-17,070	-20,640	-17,810	-17,060	-16,350	-12,900	-0,23	-0,29
	GARMA(1,0)	-17,090	-20,920	-17,820	-17,070	-16,330	-12,920	-0,11	-0,23
	MLG	-17,150	-20,920	-17,950	-17,120	-16,370	-13,260	0,24	0,07
Domingo -20,130	GARMA(1,1)	-20,070	-24,370	-20,900	-20,100	-19,210	-16,400	-0,30	-0,15
	GARMA(1,0)	-20,090	-24,370	-20,950	-20,110	-19,240	-16,400	-0,20	-0,10
	MLG	-20,120	-24,300	-21,000	-20,120	-19,280	-16,330	-0,05	-0,05
Segunda 4,382	GARMA(1,1)	4,401	-1,139	3,467	4,432	5,316	8,911	0,43	1,14
	GARMA(1,0)	4,381	-1,139	3,475	4,415	5,300	8,901	-0,03	0,75
	MLG	4,284	-1,139	3,397	4,282	5,241	8,901	-2,24	-2,28
Sábado -14,534	GARMA(1,1)	-14,490	-18,670	-15,280	-14,460	-13,720	-10,220	-0,30	-0,51
	GARMA(1,0)	-14,490	-18,670	-15,300	-14,470	-13,700	-10,240	-0,30	-0,44
	MLG	-14,500	-18,620	-15,310	-14,470	-13,690	-10,370	-0,24	-0,44
Feriado -14,392	GARMA(1,1)	-14,406	-21,907	-15,785	-14,454	-13,079	-7,233	0,10	0,43
	GARMA(1,0)	-14,435	-21,927	-15,791	-14,517	-13,039	-6,882	0,30	0,87
	MLG	-14,417	-22,751	-15,964	-14,442	-12,930	-6,882	0,17	0,35
PM10med 0,036	GARMA(1,1)	0,037	-0,111	0,002	0,037	0,070	0,187	4,28	4,98
	GARMA(1,0)	0,037	-0,107	0,004	0,037	0,072	0,191	4,76	3,40
	MLG	0,038	-0,126	0,003	0,037	0,075	0,206	6,65	2,94
COmax 0,221	GARMA(1,1)	0,199	-2,117	-0,240	0,218	0,629	2,352	-9,87	-1,39
	GARMA(1,0)	0,198	-2,118	-0,260	0,222	0,639	2,396	-10,36	0,78
	MLG	0,188	-2,118	-0,327	0,216	0,695	2,638	-15,03	-2,34
O3max 0,023	GARMA(1,1)	0,024	-0,013	0,016	0,024	0,032	0,068	3,66	3,53
	GARMA(1,0)	0,025	-0,014	0,017	0,025	0,033	0,071	7,03	7,84
	MLG	0,024	-0,021	0,015	0,024	0,033	0,069	1,56	3,27
Tmin 1,270	GARMA(1,1)	1,256	0,460	1,083	1,257	1,434	2,163	-1,11	-0,99
	GARMA(1,0)	1,256	0,459	1,079	1,254	1,437	2,119	-1,13	-1,28
	MLG	1,261	0,406	1,086	1,257	1,440	2,119	-0,72	-1,06
Urmin -0,105	GARMA(1,1)	-0,106	-0,228	-0,135	-0,107	-0,080	0,018	0,75	1,51
	GARMA(1,0)	-0,106	-0,227	-0,135	-0,107	-0,078	0,036	0,50	1,53
	MLG	-0,107	-0,230	-0,136	-0,109	-0,078	0,041	1,36	3,57
ϕ 0,300	GARMA(1,1)	0,283	0,002	0,256	0,292	0,325	0,500		
	GARMA(1,0)	0,267	0,003	0,263	0,293	0,311	0,508		
θ 0,000	GARMA(1,1)	0,021	-0,222	-0,025	0,007	0,044	0,532		
$\vartheta_2 \times 100$ 1,007	GARMA(1,1)	1,023	0,603	0,893	0,993	1,112	1,848	1,61	-1,41
	GARMA(1,0)	1,108	0,612	0,928	1,042	1,262	1,882	10,06	3,50
	MLG	1,441	0,971	1,327	1,439	1,557	2,015	43,15	42,92

Tabela 4.9: Medidas resumo das estimativas das simulações com distribuição Binomial Negativa e $\phi = 0,3$.

Variável	Modelo	Média	Mínimo	1º quartil	Mediana	3º quartil	Máximo	Var. % média	Var. % mediana
Intercepto	GARMA(1,1)	0,051	0,046	0,050	0,051	0,052	0,058	0,51	0,16
	GARMA(1,0)	0,051	0,045	0,050	0,051	0,051	0,058		
	MLG	0,046	0,044	0,046	0,046	0,047	0,049	-8,64	-8,66
Seno	GARMA(1,1)	0,936	0,718	0,906	0,930	0,954	1,390	2,41	0,94
	GARMA(1,0)	0,914	0,687	0,888	0,921	0,949	1,341		
	MLG	0,702	0,658	0,692	0,701	0,711	0,744	-23,22	-23,84
Cosseno	GARMA(1,1)	1,139	0,921	1,108	1,133	1,159	1,521	1,79	0,67
	GARMA(1,0)	1,119	0,903	1,091	1,125	1,154	1,482		
	MLG	0,918	0,854	0,904	0,918	0,932	0,984	-17,99	-18,42
Domingo	GARMA(1,1)	1,229	1,139	1,203	1,220	1,242	1,549	0,08	-0,33
	GARMA(1,0)	1,228	1,142	1,207	1,224	1,245	1,331		
	MLG	1,218	1,138	1,201	1,218	1,233	1,308	-0,81	-0,49
Segunda	GARMA(1,1)	1,360	1,273	1,336	1,355	1,377	1,528	-0,95	-0,66
	GARMA(1,0)	1,373	1,280	1,341	1,364	1,399	1,520		
	MLG	1,440	1,358	1,419	1,440	1,460	1,552	4,88	5,57
Sábado	GARMA(1,1)	1,180	1,108	1,161	1,177	1,195	1,293	-1,01	-0,76
	GARMA(1,0)	1,192	1,111	1,167	1,186	1,213	1,311		
	MLG	1,254	1,174	1,237	1,254	1,270	1,342	5,20	5,73
Feriado	GARMA(1,1)	2,103	1,909	2,065	2,102	2,134	2,353	-1,54	-0,66
	GARMA(1,0)	2,136	1,957	2,077	2,116	2,176	2,399		
	MLG	2,260	2,123	2,222	2,258	2,295	2,477	5,81	6,71
PM10med	GARMA(1,1)	0,051	0,006	0,050	0,051	0,051	0,061	-0,22	0,03
	GARMA(1,0)	0,051	0,045	0,050	0,051	0,051	0,061		
	MLG	0,048	0,045	0,047	0,048	0,048	0,050	-6,12	-5,89
COmax	GARMA(1,1)	0,697	0,565	0,688	0,696	0,706	0,756	-0,53	-0,36
	GARMA(1,0)	0,700	0,574	0,688	0,699	0,710	0,806		
	MLG	0,685	0,644	0,676	0,685	0,694	0,725	-2,26	-1,95
O3max	GARMA(1,1)	0,012	0,012	0,012	0,012	0,012	0,013	-0,81	-0,57
	GARMA(1,0)	0,012	0,012	0,012	0,012	0,013	0,013		
	MLG	0,012	0,012	0,012	0,012	0,013	0,013	0,41	0,81
Tmin	GARMA(1,1)	0,259	0,224	0,253	0,257	0,262	0,306	1,17	0,35
	GARMA(1,0)	0,256	0,216	0,251	0,257	0,261	0,301		
	MLG	0,222	0,211	0,219	0,222	0,225	0,234	-13,22	-13,53
Urmin	GARMA(1,1)	0,039	0,036	0,039	0,039	0,040	0,043	-0,48	-0,28
	GARMA(1,0)	0,040	0,038	0,039	0,039	0,040	0,043		
	MLG	0,039	0,037	0,038	0,039	0,039	0,041	-2,22	-1,85
ϕ	GARMA(1,1)	0,051	0,037	0,048	0,050	0,053	0,115		
	GARMA(1,0)	0,020	0,019	0,020	0,020	0,021	0,023		
θ	GARMA(1,1)	0,051	0,023	0,048	0,051	0,054	0,100		
$\vartheta_2 \times 100$	GARMA(1,1)	0,139	0,124	0,133	0,136	0,140	0,181	-4,01	-1,23
	GARMA(1,0)	0,145	0,124	0,134	0,138	0,157	0,191		
	MLG	0,151	0,136	0,147	0,151	0,155	0,168	4,50	9,41

Tabela 4.10: Medidas resumo dos erros padrões das simulações com distribuição Binomial Negativa e $\phi = 0,3$.

Variável/ Parâmetro	Modelo	Média	Mínimo	1º quartil	Mediana	3º quartil	Máximo	Var. % média	Var. % mediana
Intercepto 3,228	GARMA(1,1)	3,224	3,015	3,187	3,225	3,261	3,418	-0,13	-0,10
	GARMA(1,0)	3,220	2,947	3,179	3,220	3,261	3,418	-0,25	-0,25
	MLG	3,208	2,947	3,161	3,208	3,256	3,427	-0,62	-0,62
Seno -5,107	GARMA(1,1)	-5,128	-9,733	-6,045	-5,151	-4,222	0,635	0,40	0,86
	GARMA(1,0)	-5,118	-9,719	-6,039	-5,153	-4,226	0,640	0,21	0,90
	MLG	-5,166	-10,388	-6,128	-5,164	-4,219	0,725	1,16	1,11
Cosseno -17,109	GARMA(1,1)	-17,100	-21,850	-18,170	-17,120	-16,100	-12,450	-0,05	0,07
	GARMA(1,0)	-17,110	-21,840	-18,180	-17,090	-16,110	-12,450	0,01	-0,11
	MLG	-17,220	-22,000	-18,320	-17,200	-16,150	-11,830	0,65	0,53
Domingo -20,130	GARMA(1,1)	-20,180	-23,690	-21,020	-20,220	-19,390	-16,540	0,25	0,45
	GARMA(1,0)	-20,150	-23,700	-21,000	-20,170	-19,350	-16,520	0,10	0,20
	MLG	-20,170	-24,800	-21,070	-20,160	-19,340	-16,790	0,20	0,15
Segunda 4,382	GARMA(1,1)	4,303	0,128	3,380	4,293	5,224	8,387	-1,80	-2,03
	GARMA(1,0)	4,306	-0,147	3,367	4,297	5,244	8,436	-1,75	-1,94
	MLG	4,207	-0,182	3,270	4,182	5,132	8,754	-3,99	-4,57
Sábado -14,534	GARMA(1,1)	-14,580	-17,810	-15,440	-14,600	-13,760	-11,010	0,32	0,45
	GARMA(1,0)	-14,570	-18,630	-15,400	-14,570	-13,760	-10,980	0,25	0,25
	MLG	-14,550	-18,020	-15,450	-14,560	-13,690	-10,420	0,11	0,18
Feriado -14,392	GARMA(1,1)	-14,350	-21,040	-15,770	-14,360	-13,010	-8,110	-0,29	-0,22
	GARMA(1,0)	-14,359	-21,365	-15,784	-14,389	-12,984	-8,166	-0,23	-0,02
	MLG	-14,318	-23,193	-15,909	-14,368	-12,711	-6,073	-0,51	-0,17
PM10med 0,036	GARMA(1,1)	0,038	-0,113	0,002	0,037	0,076	0,223	6,02	3,03
	GARMA(1,0)	0,039	-0,131	0,000	0,038	0,079	0,305	10,14	6,46
	MLG	0,039	-0,175	-0,009	0,037	0,087	0,305	9,18	3,28
COmax 0,221	GARMA(1,1)	0,217	-1,995	-0,264	0,216	0,725	2,435	-1,48	-2,16
	GARMA(1,0)	0,230	-1,996	-0,274	0,217	0,746	3,356	4,14	-1,57
	MLG	0,241	-2,568	-0,449	0,261	0,901	3,356	9,17	18,28
O3max 0,023	GARMA(1,1)	0,024	-0,019	0,015	0,024	0,032	0,064	1,52	2,89
	GARMA(1,0)	0,025	-0,019	0,016	0,025	0,034	0,069	7,54	6,73
	MLG	0,024	-0,035	0,012	0,024	0,035	0,089	1,65	4,17
Tmin 1,270	GARMA(1,1)	1,272	0,488	1,065	1,271	1,471	2,205	0,13	0,07
	GARMA(1,0)	1,269	0,380	1,059	1,269	1,479	2,248	-0,08	-0,08
	MLG	1,270	0,379	1,027	1,262	1,511	2,247	-0,02	-0,61
Urmin -0,105	GARMA(1,1)	-0,104	-0,246	-0,132	-0,104	-0,074	0,035	-1,81	-1,79
	GARMA(1,0)	-0,102	-0,246	-0,131	-0,102	-0,072	0,091	-3,62	-3,27
	MLG	-0,103	-0,310	-0,138	-0,103	-0,067	0,087	-2,52	-2,11
ϕ 0,500	GARMA(1,1)	0,472	0,006	0,469	0,494	0,519	0,627		
	GARMA(1,0)	0,439	0,002	0,470	0,495	0,510	0,884		
θ 0,000	GARMA(1,1)	0,021	-0,098	-0,020	0,007	0,034	0,536		
ϑ_2 6,023	GARMA(1,1)	6,112	5,706	5,971	6,039	6,121	7,341	1,48	0,27
	GARMA(1,0)	6,254	5,709	5,992	6,073	6,249	7,324	3,84	0,83
	MLG	6,995	6,515	6,892	6,995	7,098	7,452	16,14	16,14

Tabela 4.11: Medidas resumo das estimativas das simulações com distribuição Normal e $\phi = 0,5$.

Variável	Modelo	Média	Mínimo	1º quartil	Mediana	3º quartil	Máximo	Var. % média	Var. % mediana
Intercepto	GARMA(1,1)	0,054	0,050	0,053	0,053	0,054	0,066	-0,26	-0,13
	GARMA(1,0)	0,054	0,047	0,053	0,054	0,054	0,075		
	MLG	0,050	0,046	0,050	0,050	0,051	0,054	-6,88	-6,01
Seno	GARMA(1,1)	1,250	0,781	1,207	1,248	1,291	2,277	2,02	0,36
	GARMA(1,0)	1,225	0,741	1,181	1,244	1,287	6,088		
	MLG	0,766	0,710	0,752	0,766	0,780	0,837	-37,43	-38,41
Cosseno	GARMA(1,1)	1,407	1,057	1,366	1,402	1,443	2,285	1,19	0,59
	GARMA(1,0)	1,390	0,975	1,338	1,394	1,439	5,534		
	MLG	1,014	0,933	0,993	1,013	1,033	1,117	-27,10	-27,35
Domingo	GARMA(1,1)	1,308	1,187	1,265	1,286	1,310	1,795	-1,51	-0,39
	GARMA(1,0)	1,328	1,208	1,272	1,291	1,337	1,566		
	MLG	1,468	1,351	1,444	1,467	1,490	1,597	10,54	13,63
Segunda	GARMA(1,1)	1,307	1,187	1,267	1,288	1,308	1,695	-2,02	-0,46
	GARMA(1,0)	1,334	1,212	1,274	1,294	1,342	1,583		
	MLG	1,515	1,388	1,492	1,515	1,537	1,634	13,57	17,08
Sábado	GARMA(1,1)	1,204	1,107	1,170	1,186	1,204	1,517	-2,75	-0,50
	GARMA(1,0)	1,238	1,117	1,174	1,192	1,232	1,523		
	MLG	1,456	1,338	1,435	1,456	1,477	1,563	17,61	22,15
Feriado	GARMA(1,1)	2,118	1,917	2,054	2,089	2,128	2,908	-4,21	-0,67
	GARMA(1,0)	2,211	1,939	2,063	2,103	2,187	2,928		
	MLG	2,748	2,543	2,708	2,748	2,789	2,943	24,29	30,67
PM10med	GARMA(1,1)	0,050	0,002	0,049	0,050	0,051	0,069	-1,03	-0,45
	GARMA(1,0)	0,051	0,023	0,049	0,050	0,051	0,060		
	MLG	0,050	0,046	0,049	0,050	0,051	0,054	-1,60	-0,44
COmax	GARMA(1,1)	0,662	0,162	0,646	0,656	0,665	0,810	-2,09	-0,47
	GARMA(1,0)	0,676	0,582	0,649	0,659	0,682	0,790		
	MLG	0,728	0,673	0,715	0,728	0,739	0,787	7,69	10,49
O3max	GARMA(1,1)	0,012	0,011	0,012	0,012	0,012	0,015	-2,44	-0,59
	GARMA(1,0)	0,012	0,011	0,012	0,012	0,012	0,014		
	MLG	0,013	0,012	0,013	0,013	0,013	0,014	5,54	8,81
Tmin	GARMA(1,1)	0,281	0,240	0,275	0,279	0,283	0,349	0,72	0,18
	GARMA(1,0)	0,279	0,227	0,273	0,278	0,283	0,354		
	MLG	0,241	0,221	0,237	0,241	0,245	0,261	-13,39	-13,26
Urmin	GARMA(1,1)	0,039	0,035	0,038	0,038	0,039	0,051	-1,87	-0,54
	GARMA(1,0)	0,040	0,036	0,038	0,039	0,040	0,046		
	MLG	0,042	0,038	0,041	0,042	0,042	0,045	5,37	7,96
ϕ	GARMA(1,1)	0,035	0,027	0,032	0,033	0,034	0,312		
	GARMA(1,0)	0,020	0,018	0,018	0,019	0,019	0,031		
θ	GARMA(1,1)	0,038	0,018	0,035	0,037	0,038	0,242		
	GARMA(1,0)	0,096	0,086	0,090	0,091	0,092	0,309	-6,73	-0,58
ϑ_2	GARMA(1,1)	0,096	0,086	0,090	0,091	0,092	0,309	-6,73	-0,58
	GARMA(1,0)	0,103	0,086	0,091	0,092	0,094	0,171		
	MLG	0,107	0,099	0,105	0,107	0,108	0,113	3,72	16,15

Tabela 4.12: Medidas resumo dos erros padrões das simulações com distribuição Normal e $\phi = 0,5$.

Variável/ Parâmetro	Modelo	Média	Mínimo	1º quartil	Mediana	3º quartil	Máximo	Var. % média	Var. % mediana
Intercepto 3,228	GARMA(1,1)	3,225	3,045	3,188	3,225	3,262	3,425	-0,10	-0,10
	GARMA(1,0)	3,222	2,967	3,182	3,221	3,261	3,443	-0,19	-0,22
	MLG	3,208	2,967	3,161	3,206	3,255	3,443	-0,62	-0,68
Seno -5,107	GARMA(1,1)	-5,113	-9,629	-6,043	-5,082	-4,230	0,650	0,12	-0,48
	GARMA(1,0)	-5,132	-9,616	-6,056	-5,094	-4,237	0,646	0,49	-0,25
	MLG	-5,164	-10,246	-6,093	-5,152	-4,232	0,580	1,11	0,88
Cosseno -17,109	GARMA(1,1)	-17,080	-21,680	-18,040	-17,070	-16,120	-12,600	-0,17	-0,23
	GARMA(1,0)	-17,120	-21,840	-18,160	-17,090	-16,120	-12,550	0,07	-0,11
	MLG	-17,210	-22,320	-18,350	-17,200	-16,110	-11,590	0,59	0,53
Domingo -20,130	GARMA(1,1)	-20,170	-23,780	-20,970	-20,200	-19,380	-16,910	0,20	0,35
	GARMA(1,0)	-20,170	-23,780	-20,980	-20,210	-19,350	-16,900	0,20	0,40
	MLG	-20,170	-24,340	-21,060	-20,160	-19,350	-16,780	0,20	0,15
Segunda 4,382	GARMA(1,1)	4,334	0,790	3,445	4,318	5,230	8,093	-1,09	-1,47
	GARMA(1,0)	4,311	0,510	3,421	4,287	5,212	8,097	-1,62	-2,16
	MLG	4,224	-0,311	3,298	4,213	5,127	8,237	-3,60	-3,86
Sábado -14,534	GARMA(1,1)	-14,590	-18,440	-15,360	-14,610	-13,760	-10,960	0,38	0,52
	GARMA(1,0)	-14,580	-18,460	-15,350	-14,650	-13,740	-10,840	0,32	0,80
	MLG	-14,540	-18,090	-15,400	-14,590	-13,720	-10,550	0,04	0,38
Feriado -14,392	GARMA(1,1)	-14,355	-21,523	-15,715	-14,359	-13,101	-8,376	-0,26	-0,23
	GARMA(1,0)	-14,314	-22,678	-15,722	-14,364	-12,954	-5,242	-0,54	-0,19
	MLG	-14,304	-22,678	-15,843	-14,370	-12,700	-5,242	-0,61	-0,15
PM10med 0,036	GARMA(1,1)	0,037	-0,123	0,000	0,038	0,076	0,218	5,01	6,32
	GARMA(1,0)	0,038	-0,148	-0,002	0,037	0,078	0,226	5,11	4,24
	MLG	0,039	-0,156	-0,006	0,037	0,084	0,292	9,21	3,09
COmax 0,221	GARMA(1,1)	0,223	-1,953	-0,236	0,212	0,721	2,378	1,15	-3,79
	GARMA(1,0)	0,220	-2,648	-0,258	0,223	0,727	2,351	-0,40	1,05
	MLG	0,248	-2,648	-0,404	0,268	0,888	3,655	12,34	21,27
O3max 0,023	GARMA(1,1)	0,024	-0,021	0,015	0,024	0,031	0,073	2,03	2,16
	GARMA(1,0)	0,024	-0,020	0,015	0,024	0,033	0,085	4,25	4,25
	MLG	0,024	-0,034	0,012	0,024	0,035	0,084	1,14	3,06
Tmin 1,270	GARMA(1,1)	1,268	0,416	1,071	1,265	1,462	2,217	-0,13	-0,41
	GARMA(1,0)	1,273	0,518	1,066	1,269	1,482	2,225	0,24	-0,07
	MLG	1,268	0,312	1,026	1,268	1,514	2,213	-0,15	-0,17
Urmin -0,105	GARMA(1,1)	-0,103	-0,234	-0,130	-0,105	-0,075	0,044	-1,93	-0,24
	GARMA(1,0)	-0,103	-0,246	-0,132	-0,103	-0,075	0,057	-1,91	-1,85
	MLG	-0,102	-0,291	-0,137	-0,104	-0,066	0,084	-2,90	-1,80
ϕ 0,500	GARMA(1,1)	0,481	0,001	0,473	0,496	0,519	0,581		
	GARMA(1,0)	0,415	0,001	0,465	0,493	0,508	0,554		
θ 0,000	GARMA(1,1)	0,014	-0,099	-0,021	0,004	0,029	0,687		
$\vartheta_2 \times 100$ 1,007	GARMA(1,1)	1,048	0,565	0,900	0,990	1,087	3,020	4,11	-1,68
	GARMA(1,0)	1,325	0,569	0,925	1,030	1,242	3,455	31,63	2,25
	MLG	2,576	1,689	2,395	2,576	2,740	3,550	155,86	155,81

Tabela 4.13: Medidas resumo das estimativas das simulações com distribuição Binomial Negativa e $\phi = 0,5$.

Variável	Modelo	Média	Mínimo	1º quartil	Mediana	3º quartil	Máximo	Var. % média	Var. % mediana
Intercepto	GARMA(1,1)	0,055	0,051	0,054	0,054	0,055	0,067	0,29	0,11
	GARMA(1,0)	0,054	0,050	0,054	0,054	0,055	0,060		
	MLG	0,051	0,048	0,051	0,051	0,052	0,056	-5,37	-5,30
Seno	GARMA(1,1)	1,268	0,775	1,229	1,273	1,314	2,047	6,74	1,42
	GARMA(1,0)	1,188	0,739	1,181	1,255	1,299	1,470		
	MLG	0,780	0,722	0,766	0,780	0,793	0,842	-34,35	-37,88
Cosseno	GARMA(1,1)	1,413	1,012	1,377	1,413	1,453	2,074	4,58	1,15
	GARMA(1,0)	1,351	0,970	1,330	1,397	1,441	1,569		
	MLG	1,020	0,932	1,000	1,019	1,038	1,127	-24,53	-27,07
Domingo	GARMA(1,1)	1,234	1,138	1,205	1,223	1,244	1,769	-1,52	-0,57
	GARMA(1,0)	1,253	1,147	1,211	1,230	1,262	1,461		
	MLG	1,342	1,238	1,321	1,341	1,362	1,461	7,10	9,02
Segunda	GARMA(1,1)	1,334	1,224	1,303	1,322	1,342	2,114	-3,40	-0,60
	GARMA(1,0)	1,381	1,236	1,309	1,330	1,368	1,719		
	MLG	1,607	1,453	1,582	1,607	1,633	1,738	16,36	20,83
Sábado	GARMA(1,1)	1,149	1,071	1,125	1,140	1,156	1,513	-3,53	-0,52
	GARMA(1,0)	1,191	1,071	1,129	1,146	1,176	1,466		
	MLG	1,386	1,272	1,365	1,386	1,407	1,482	16,37	20,94
Feriado	GARMA(1,1)	1,984	1,733	1,942	1,972	2,004	3,101	-4,80	-0,70
	GARMA(1,0)	2,084	1,831	1,949	1,986	2,049	2,828		
	MLG	2,491	2,272	2,441	2,491	2,538	2,779	19,53	25,43
PM10med	GARMA(1,1)	0,053	0,013	0,052	0,052	0,053	0,090	-1,15	-0,40
	GARMA(1,0)	0,053	0,045	0,052	0,053	0,054	0,075		
	MLG	0,053	0,049	0,052	0,053	0,054	0,057	-0,02	0,84
COmax	GARMA(1,1)	0,695	0,307	0,684	0,693	0,702	0,847	-2,52	-0,50
	GARMA(1,0)	0,713	0,012	0,686	0,696	0,717	1,552		
	MLG	0,764	0,708	0,751	0,764	0,776	0,826	7,19	9,70
O3max	GARMA(1,1)	0,012	0,011	0,012	0,012	0,012	0,027	-3,06	-0,50
	GARMA(1,0)	0,012	0,011	0,012	0,012	0,012	0,015		
	MLG	0,014	0,013	0,014	0,014	0,014	0,015	10,87	14,45
Tmin	GARMA(1,1)	0,286	0,249	0,282	0,286	0,289	0,342	1,67	0,49
	GARMA(1,0)	0,281	0,239	0,279	0,284	0,288	0,306		
	MLG	0,247	0,227	0,243	0,247	0,251	0,267	-12,26	-13,09
Urmin	GARMA(1,1)	0,039	0,037	0,039	0,039	0,040	0,072	-2,28	-0,46
	GARMA(1,0)	0,040	0,037	0,039	0,039	0,040	0,047		
	MLG	0,043	0,040	0,042	0,043	0,044	0,046	6,83	9,24
ϕ	GARMA(1,1)	0,038	0,022	0,032	0,033	0,034	1,153		
	GARMA(1,0)	0,020	0,017	0,018	0,018	0,019	0,032		
θ	GARMA(1,1)	0,040	0,014	0,035	0,037	0,038	0,890		
$\vartheta_2 \times 100$	GARMA(1,1)	0,146	0,123	0,134	0,137	0,140	1,877	-11,49	-0,87
	GARMA(1,0)	0,165	0,123	0,134	0,138	0,145	0,374		
	MLG	0,187	0,158	0,181	0,187	0,192	0,219	13,25	35,31

Tabela 4.14: Medidas resumo dos erros padrões das simulações com distribuição Binomial Negativa e $\phi = 0,5$.

Capítulo 5

Conclusões e perspectivas

A análise de séries temporais de contagens pode ser aplicada em várias áreas, e em especial para a análise de dados de morbi-mortalidade. A motivação para esse trabalho foi a análise da série do número diário de internações por doenças respiratórias em pessoas com 65 anos ou mais residentes no município de São Paulo de 2006 a 2011. Considerando a população nessa faixa etária no município, a análise pode considerar a taxa de internações.

Diversos aspectos devem ser considerados na análise dessa série temporal de contagem, como as distribuições de probabilidade, possível heterocedasticidade e, principalmente, autocorrelação por tratar-se de uma série temporal.

O primeiro e mais simples modelo a ser considerado é o modelo de regressão linear para a taxa de internações. Esse modelo apresenta estimadores não viesados para os efeitos das variáveis explicativas se a forma linear proposta for apropriada. Entretanto, várias suposições como a de homocedasticidade e a de erros independentes são violadas, o que torna os erros padrões obtidos viesados.

O modelo mais utilizado em vários artigos da área é o modelo linear generalizado usando as distribuições Poisson e Binomial Negativa. Nesses modelos a variância já depende da média. Em geral, encontra-se superdispersão, o que faz a distribuição Binomial Negativa mais apropriada. Entretanto, ao se ajustar esse modelo, a autocorrelação dos resíduos padronizados quantílicos de primeira ordem (defasagem 1) está em torno de 0,2 e não cai rapidamente para zero ao longo das defasagens. Assim, os erros padrões devem ter sido estimados incorretamente, pois necessitam de observações independentes.

Outra proposta interessante de modelo é o modelo GAMLSS que permite a inclusão de uma equação para modelar o desvio padrão em função de variáveis explicativas. Os resíduos padronizados desse modelo são homocedásticos, mas ainda resta a autocorrelação, que não é incluída nesse modelo.

Por fim, podem-se utilizar também modelos de séries temporais como os modelos ARMA e GARMA, que incluem uma estrutura de autocorrelação. O modelo GARMA tem a vantagem em relação ao modelo ARMA por poder utilizar várias distribuições, como Normal, Poisson e Binomial Negativa, e podem ser usadas outras funções de ligação além da identidade.

Considerando a análise das distribuições dos resíduos padronizados quantílicos, os modelos com distribuição Binomial Negativa tiveram a melhor aproximação para a Normal e maior proporção de observações dentro dos intervalos de confiança para os valores preditos.

É usual que o pesquisador da área aplicada utilize o modelo linear generalizado que é mais citado na literatura dessa área. Entretanto, o modelo linear generalizado não estima os erros padrões corretamente quando as suposições de homocedasticidade e independência entre os resíduos padronizados não é atendida, e como apresentaram erros padrões mais distantes dos erros padrões dos modelos GARMA do que dos modelos GAMLSS, se conclui que não considerar a autocorrelação afetou mais as estimativas dos erros padrões do que utilizar as variáveis explicativas para modelar a heterocedasticidade.

Os modelos GARMA e ARMA também tiveram a vantagem de ter erro quadrático médio menor que os outros modelos.

Portanto, o modelo que melhor se adequou à nossa aplicação foi o GARMA com distribuição Binomial Negativa.

A variação no erro padrão dos modelos de séries temporais em relação aos modelos que não utilizam termos de autocorrelação foi corroborada pelas simulações. O aumento da autocorrelação dos dados tende a aumentar a discrepância do erro padrão. Esta variação foi maior na distribuição Normal.

Apesar dos modelos GARMA terem tido um desempenho superior, o GAMLSS tem a vantagem de permitir modelar a variância com variáveis explicativas. O ideal seria ter um modelo que pudesse utilizar termos de séries temporais para modelar a média, como o GARMA e tivesse a flexibilidade do GAMLSS, ou seja, uma generalização do modelo GARMA e GAMLSS.

Capítulo 6

Apêndice

6.1 Gráficos do núcleo da densidade

Este método foi desenvolvido para estimar a função de densidade dos dados. Nas nossas análises ele é utilizado para verificar se a função de densidade dos resíduos está próxima da Normal.

Sendo K uma função núcleo, assume valores não negativos e satisfaz a seguinte condição:

$$\int_{-\infty}^{\infty} K(x)dx = 1 \quad (6.1)$$

Logo, K é uma função de densidade de probabilidade.

O estimador de núcleo é definido por:

$$\hat{f}(x) = \frac{1}{hT} \sum_{i=1}^T K\left(\frac{x - X_i}{h}\right) \quad (6.2)$$

em que, h : largura da janela ou parâmetro de suavização;

O estimador pode ser considerado como uma soma de 'caixas' centradas nas observações, o estimador é uma soma de saliências provocadas pelas observações. A função núcleo determina a forma das saliências, enquanto que a largura da janela h determina sua largura. Quanto maior o h maior é a suavização, o excesso de suavização causa perda de detalhes (espúrias ou não) e o contrário pode deixar visíveis estruturas espúrias.

Como K é uma função de probabilidade, implica que \hat{f} também será e \hat{f} tem todas as propriedades de continuidade e diferenciabilidade do núcleo K .

Mais detalhes podem ser encontrados em [Silverman , 1986].

6.2 Figuras

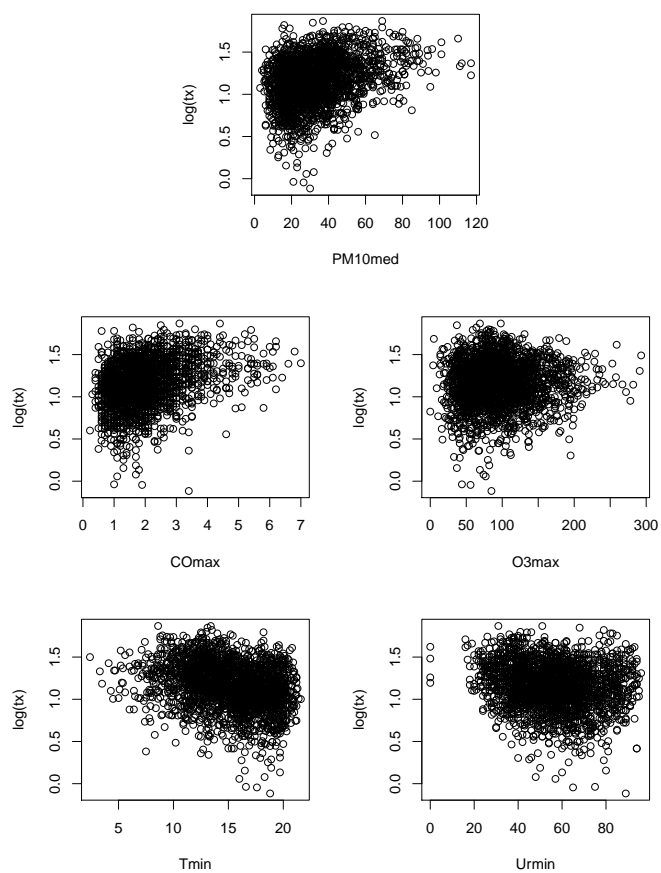


Figura 6.1: Gráficos de dispersão entre a variável logaritmo da taxa de internações e as variáveis de poluição e climáticas.

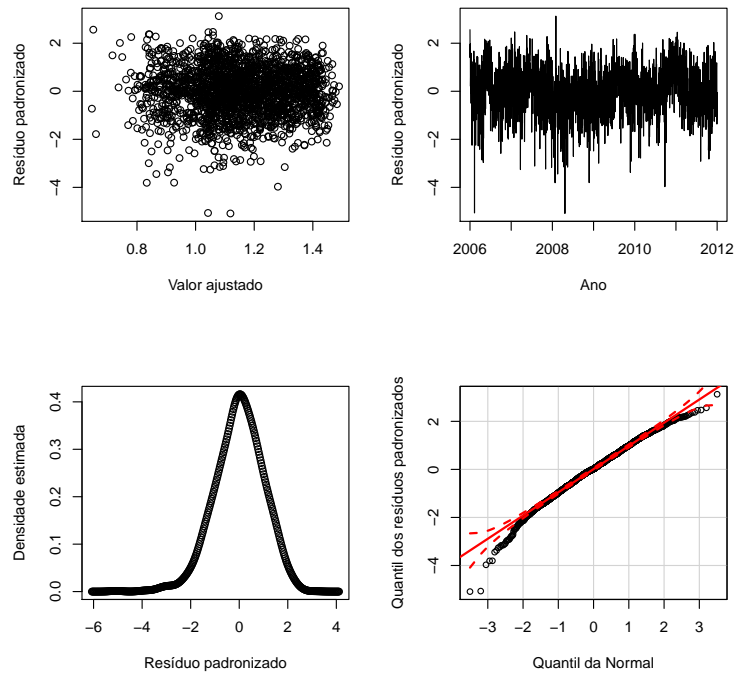


Figura 6.2: *Análise dos resíduos padronizados do modelo regressão linear.*

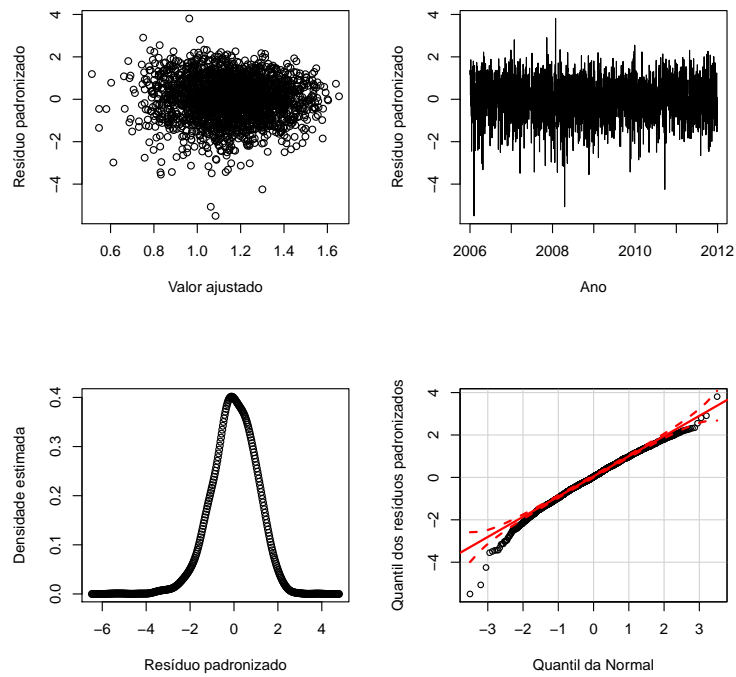


Figura 6.3: *Análise dos resíduos padronizados do modelo ARMA.*

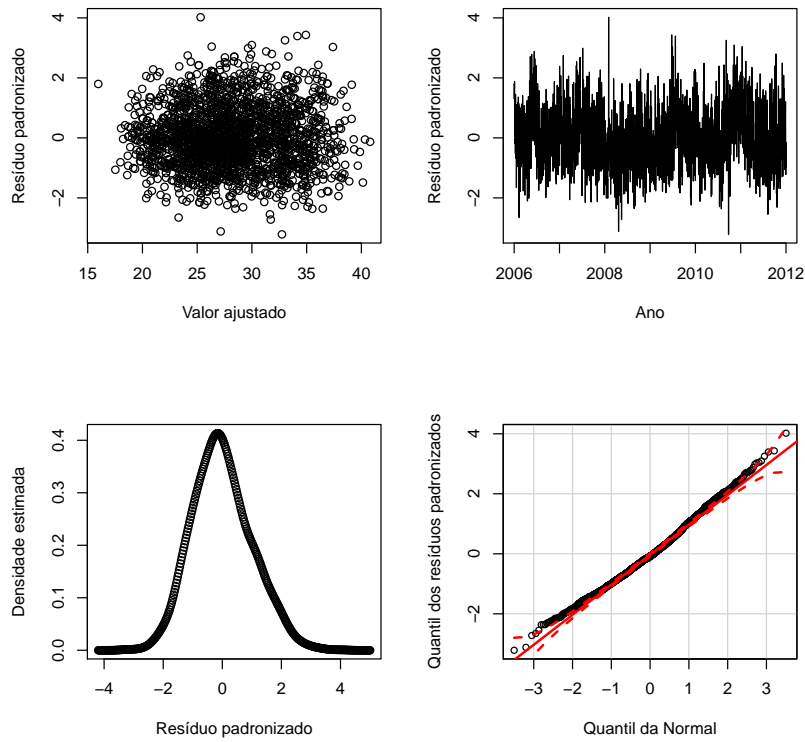


Figura 6.4: Análise dos resíduos padronizados do modelo MLG Normal.

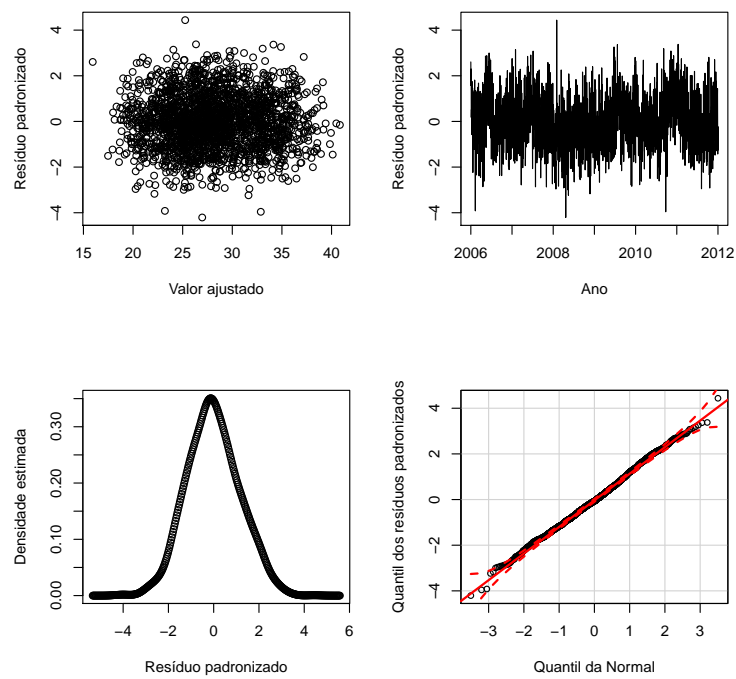


Figura 6.5: Análise dos resíduos padronizados do modelo MLG Poisson.

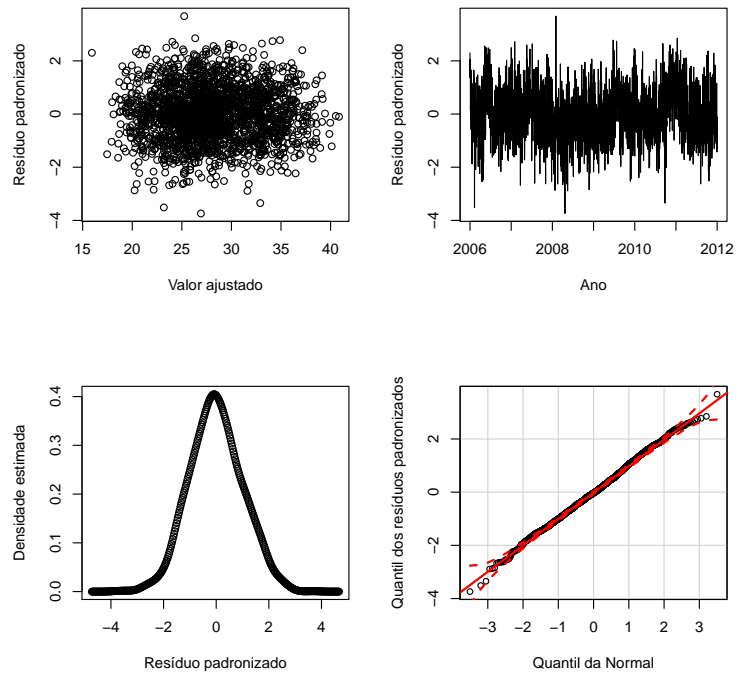


Figura 6.6: Análise dos resíduos padronizados do modelo MLG Binomial Negativa.

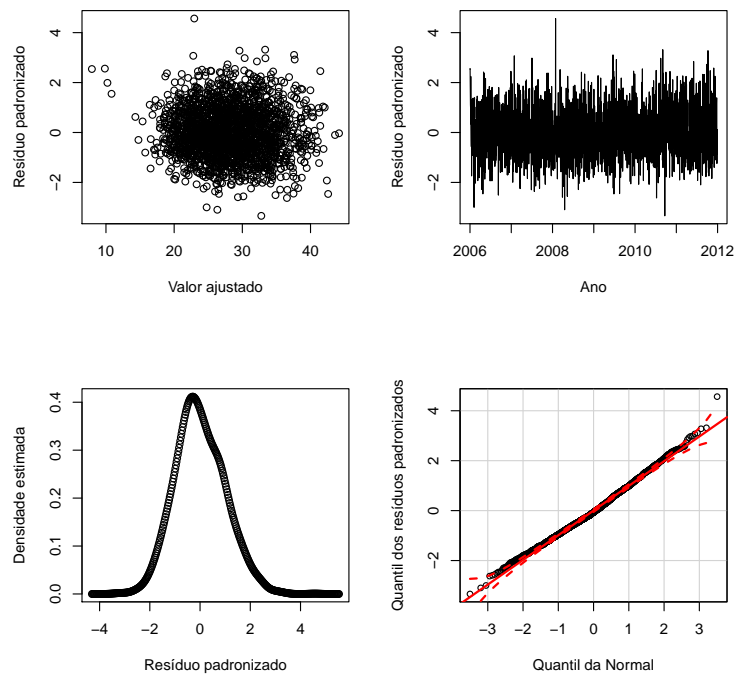


Figura 6.7: Análise dos resíduos padronizados do modelo GARMA Normal.

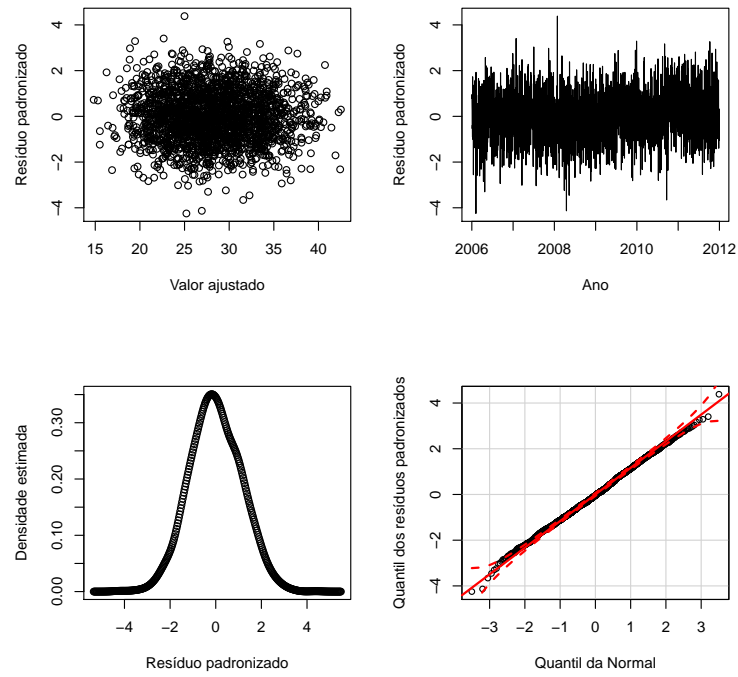


Figura 6.8: Análise dos resíduos padronizados do modelo GARMA Poisson.

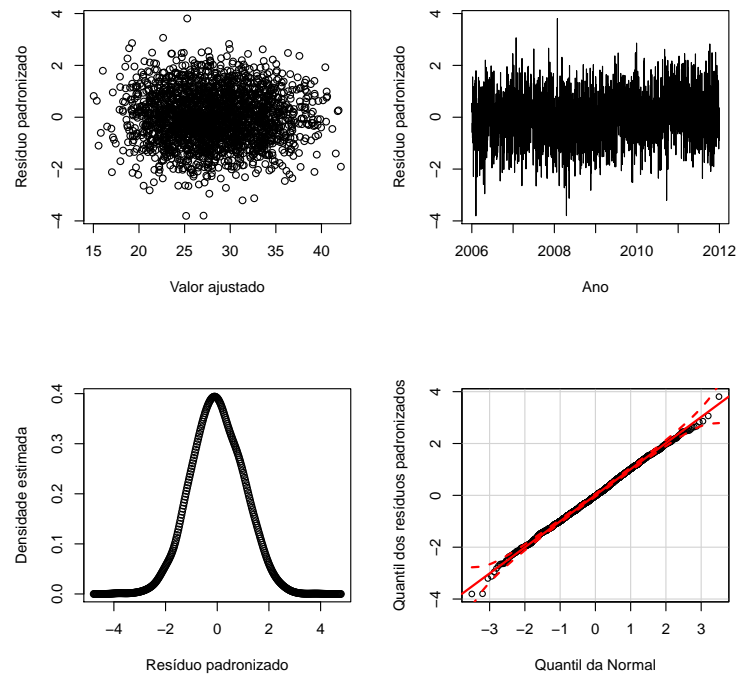


Figura 6.9: Análise dos resíduos padronizados do modelo GARMA Binomial Negativa.

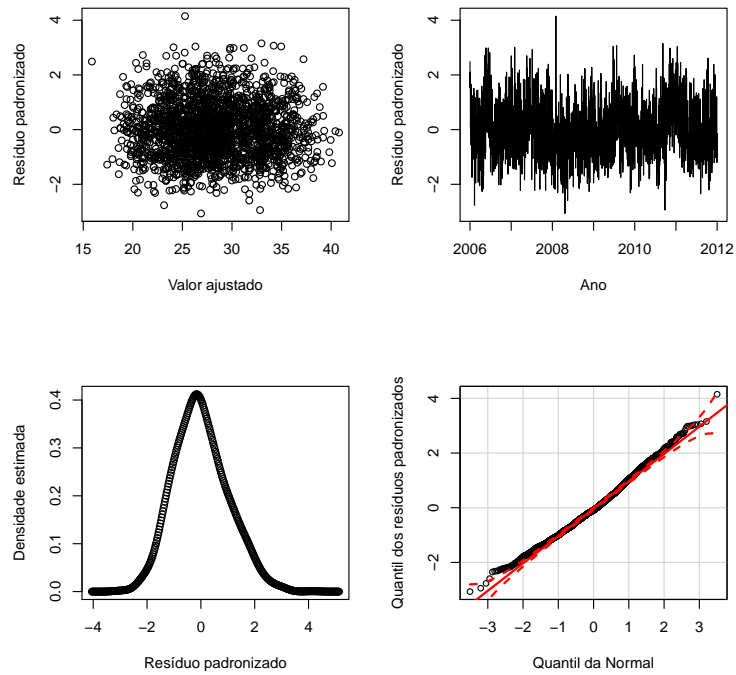


Figura 6.10: Análise dos resíduos padronizados do modelo 1 - GAMLSS Normal.

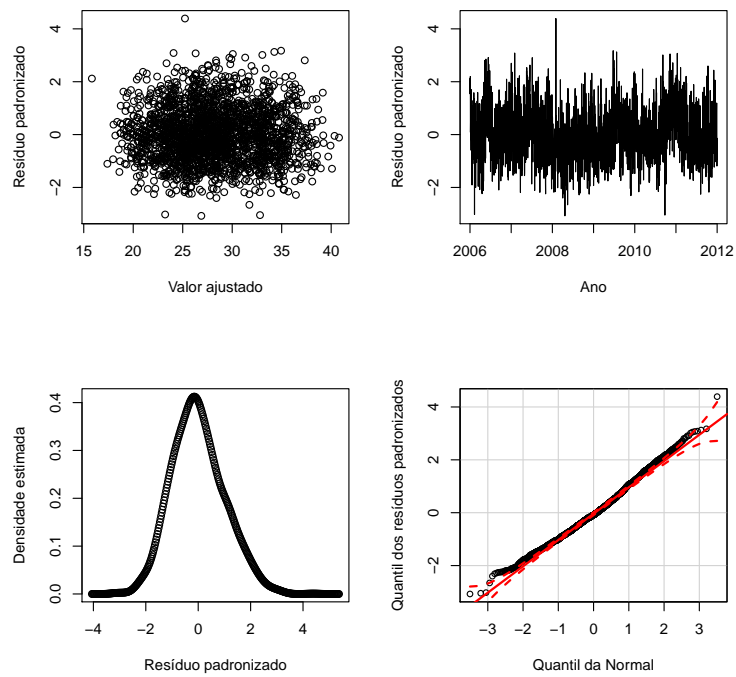


Figura 6.11: Análise dos resíduos padronizados do modelo 2 - GAMLSS Normal.

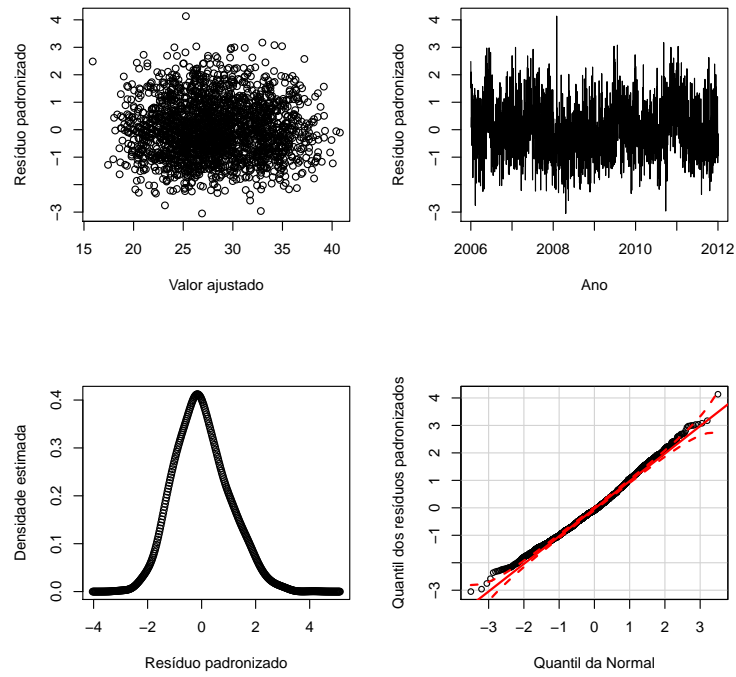


Figura 6.12: Análise dos resíduos padronizados do modelo 3 - GAMLSS Normal.

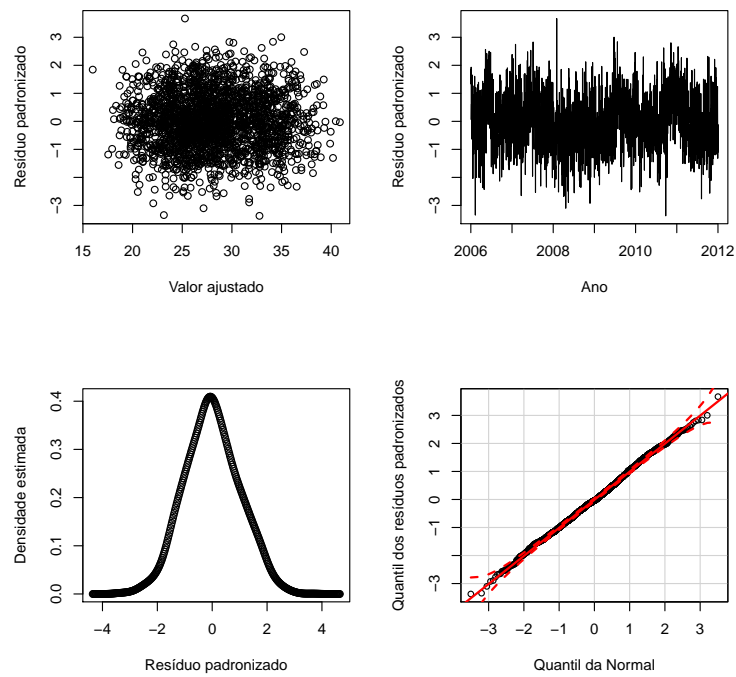


Figura 6.13: Análise dos resíduos padronizados do modelo 4 - GAMLSS Binomial Negativa.

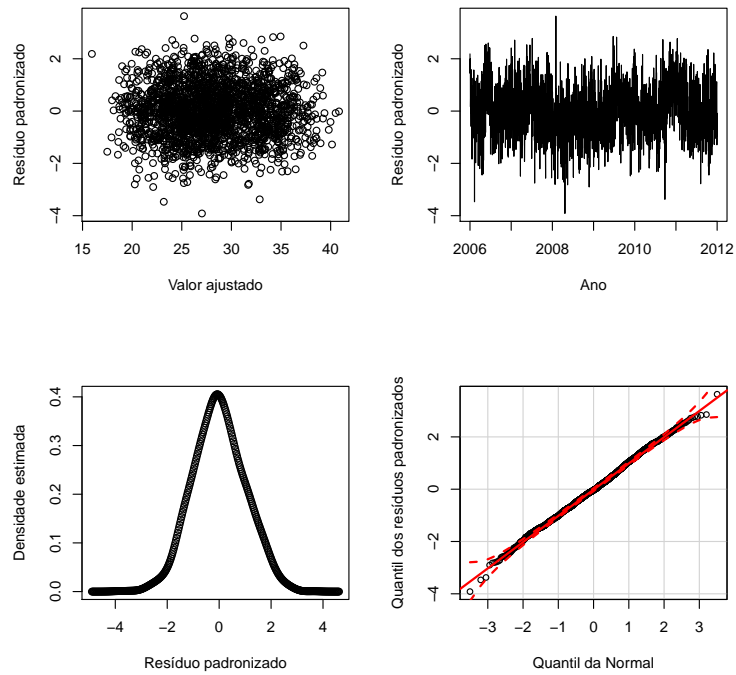


Figura 6.14: Análise dos resíduos padronizados do modelo 5 - GAMLSS Binomial Negativa.

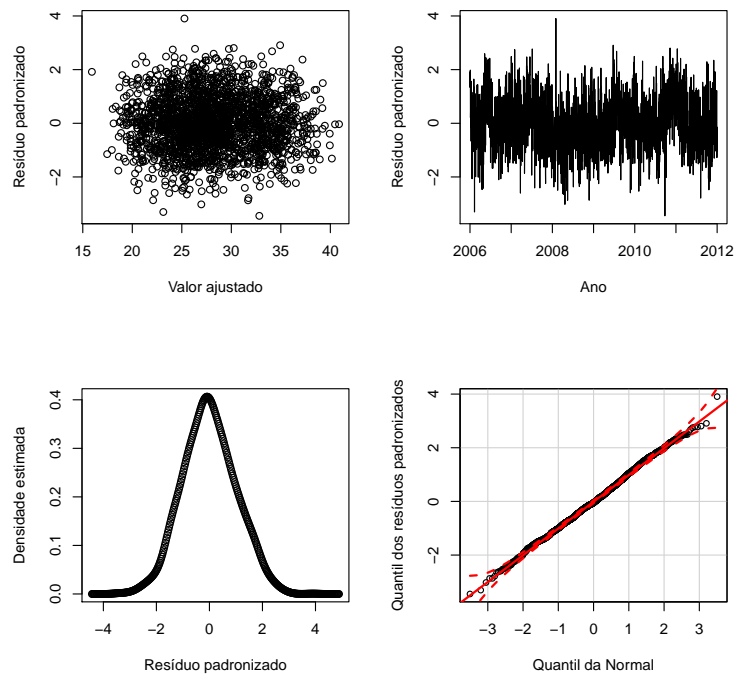


Figura 6.15: Análise dos resíduos padronizados do modelo 6 - GAMLSS Binomial Negativa.

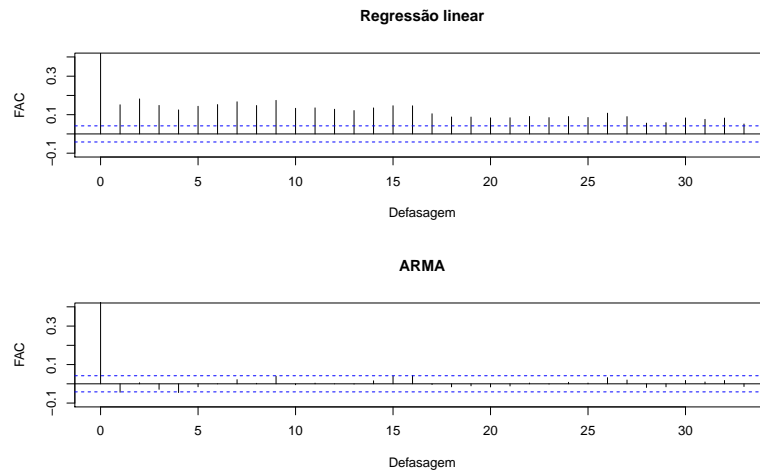


Figura 6.16: *FAC dos resíduos padronizados dos modelos regressão linear e ARMA.*

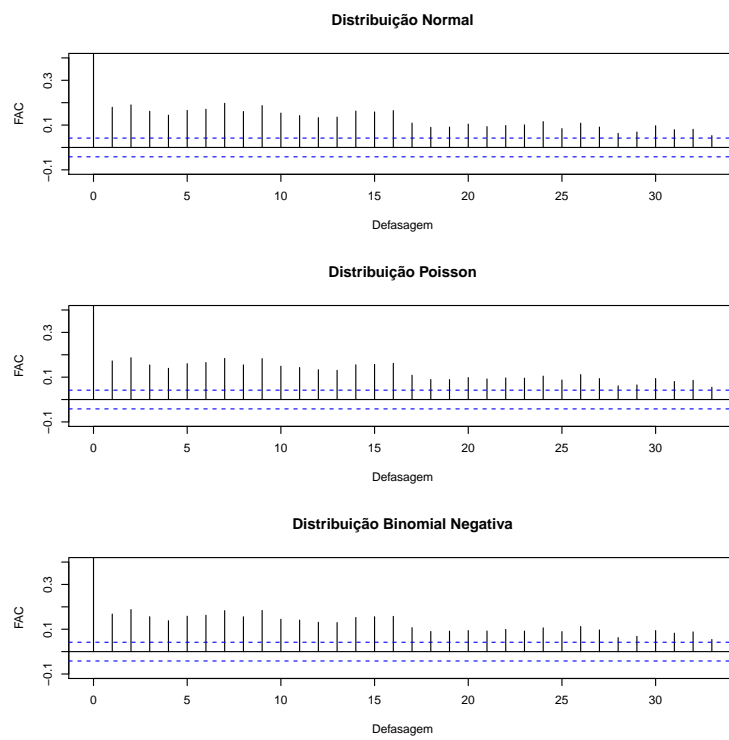


Figura 6.17: *FAC dos resíduos padronizados dos modelos MLG.*

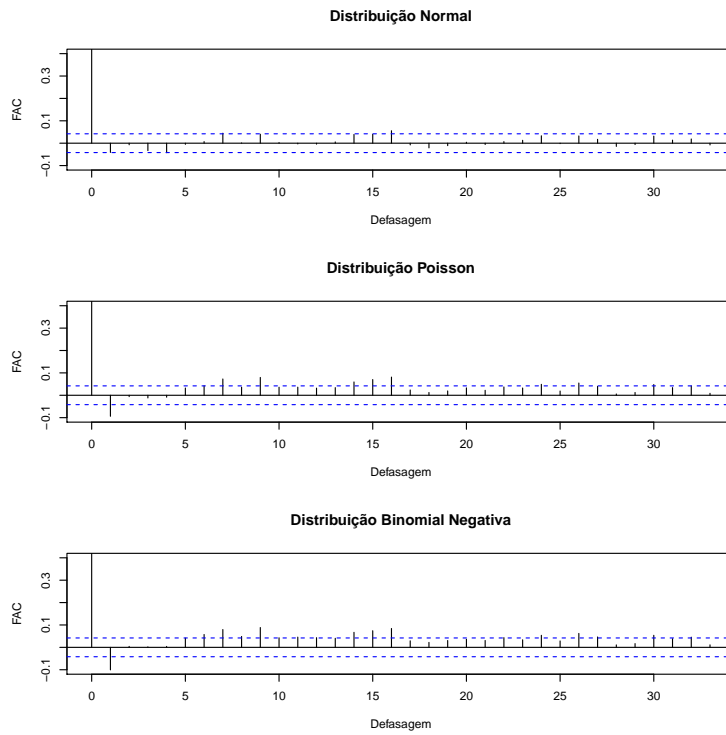


Figura 6.18: *FAC dos resíduos padronizados dos modelos GARMA.*

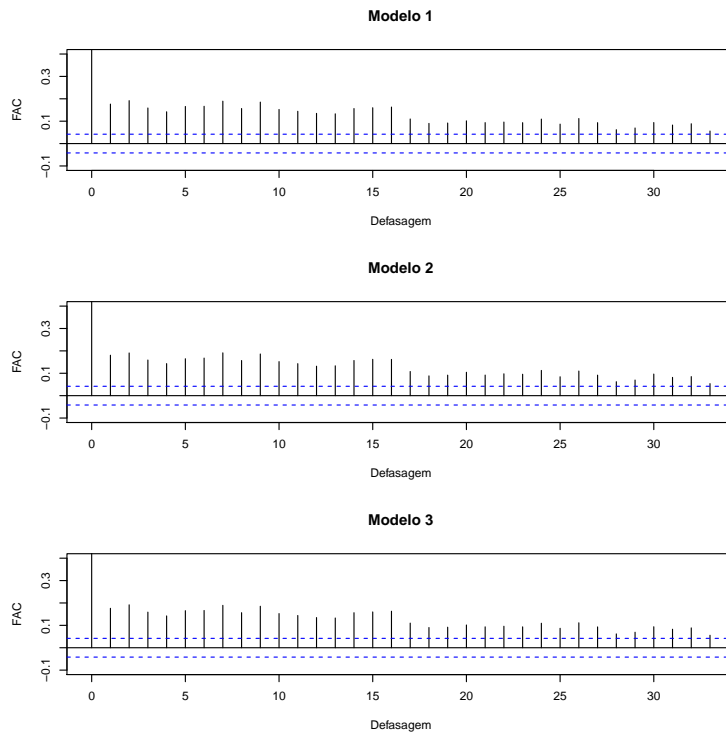


Figura 6.19: *FAC dos resíduos padronizados dos modelos GAMLSS com distribuição Normal.*

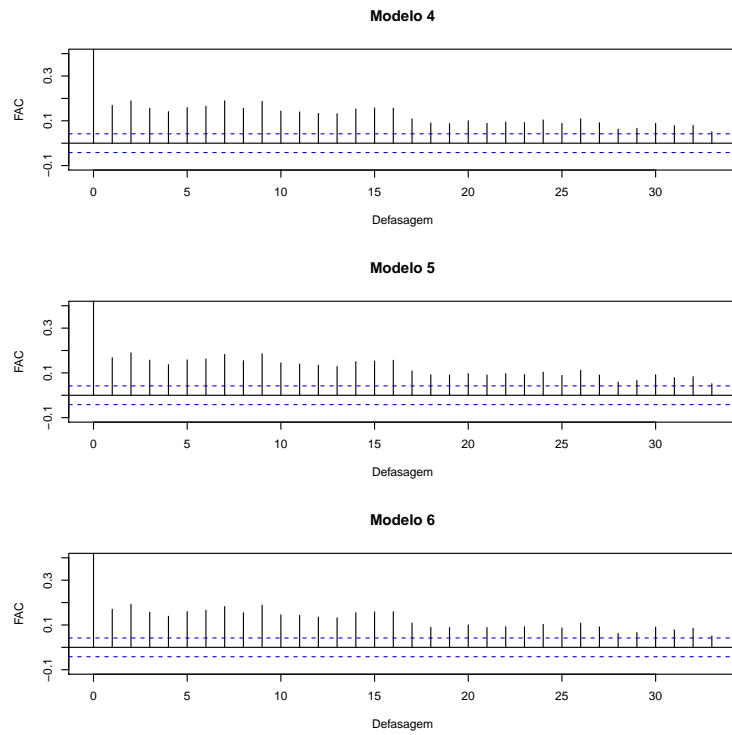


Figura 6.20: *FAC dos resíduos padronizados dos modelos GAMLSS com distribuição Binomial Negativa.*

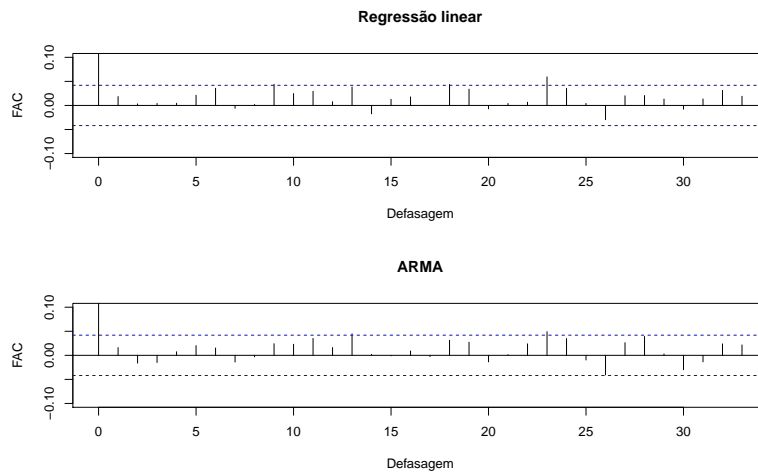


Figura 6.21: *FAC dos resíduos ao quadrado dos modelos regressão linear e ARMA.*

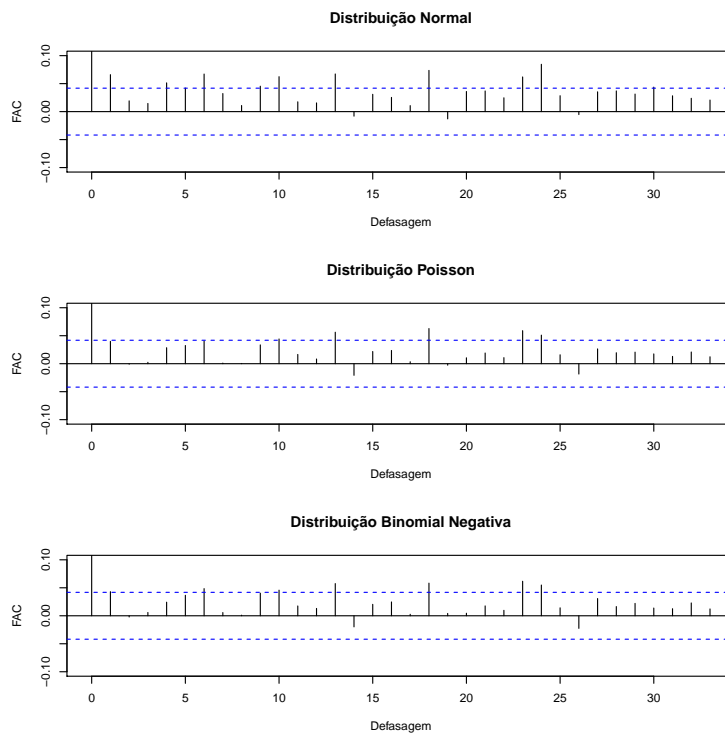


Figura 6.22: *FAC dos resíduos ao quadrado dos modelos MLG.*

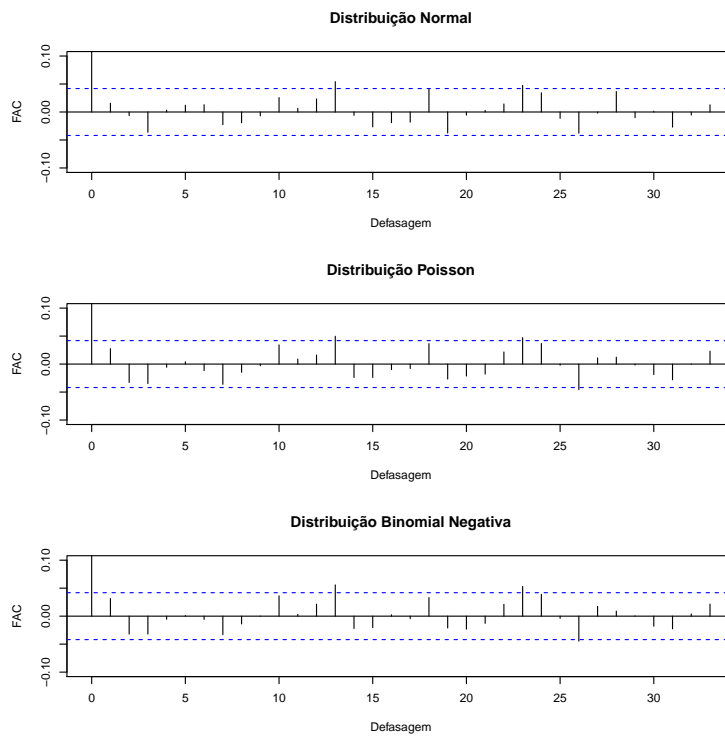


Figura 6.23: *FAC dos resíduos ao quadrado dos modelos GARMA.*

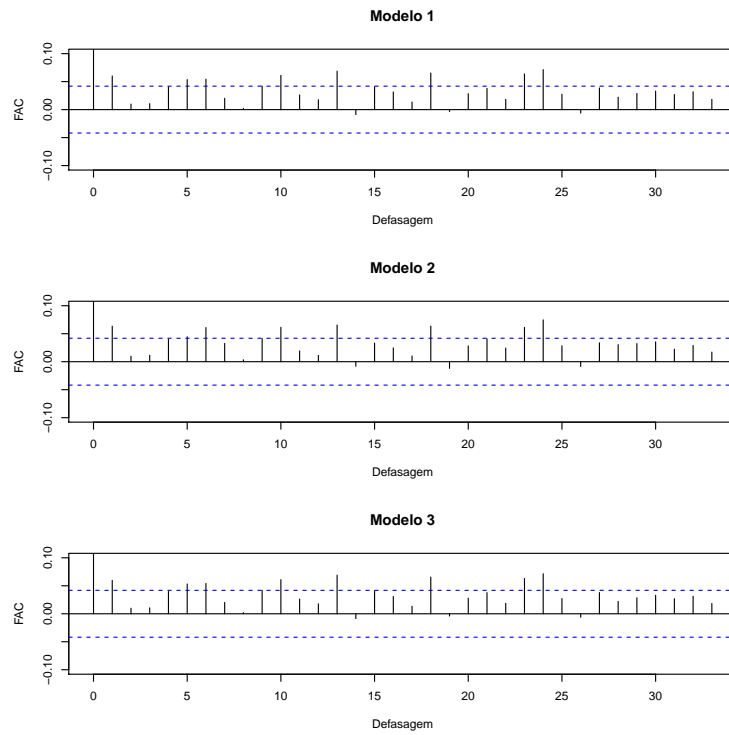


Figura 6.24: *FAC dos resíduos ao quadrado dos modelos GAMLSS com distribuição Normal.*

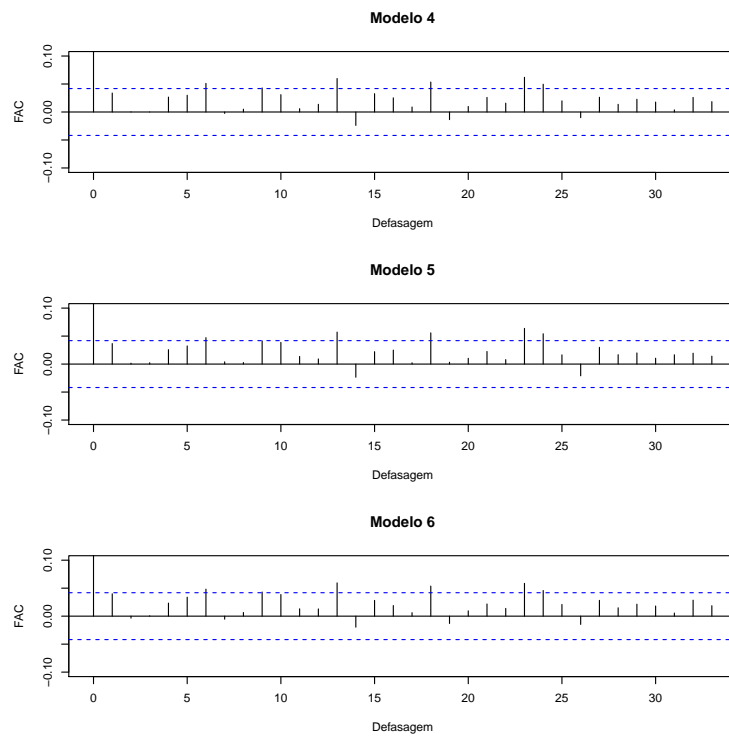


Figura 6.25: *FAC dos resíduos ao quadrado dos modelos GAMLSS com distribuição Binomial Negativa.*

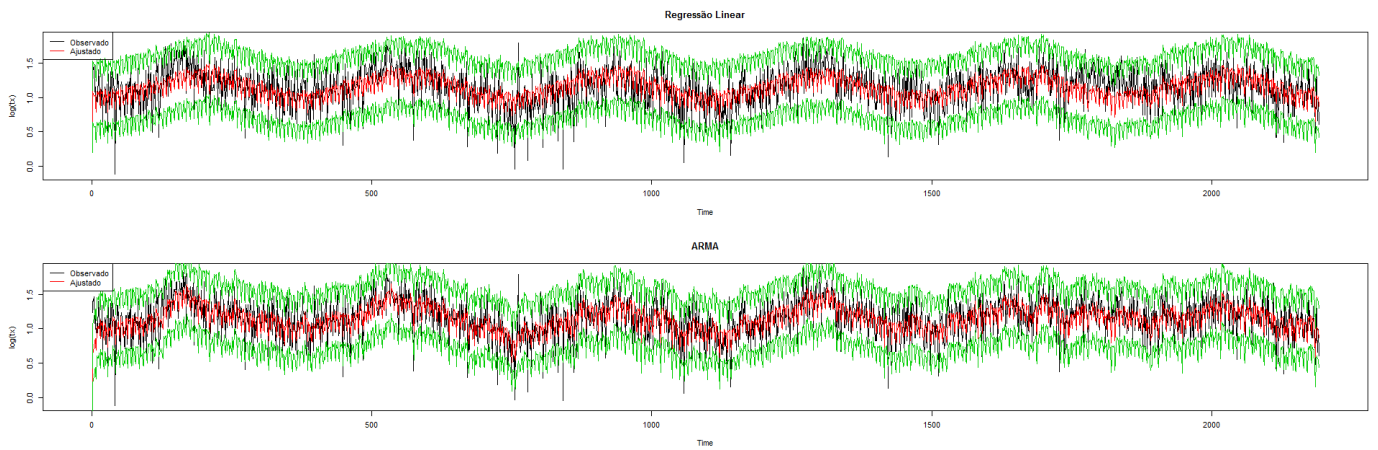


Figura 6.26: Intervalo de confiança dos modelos regressão linear e ARMA.

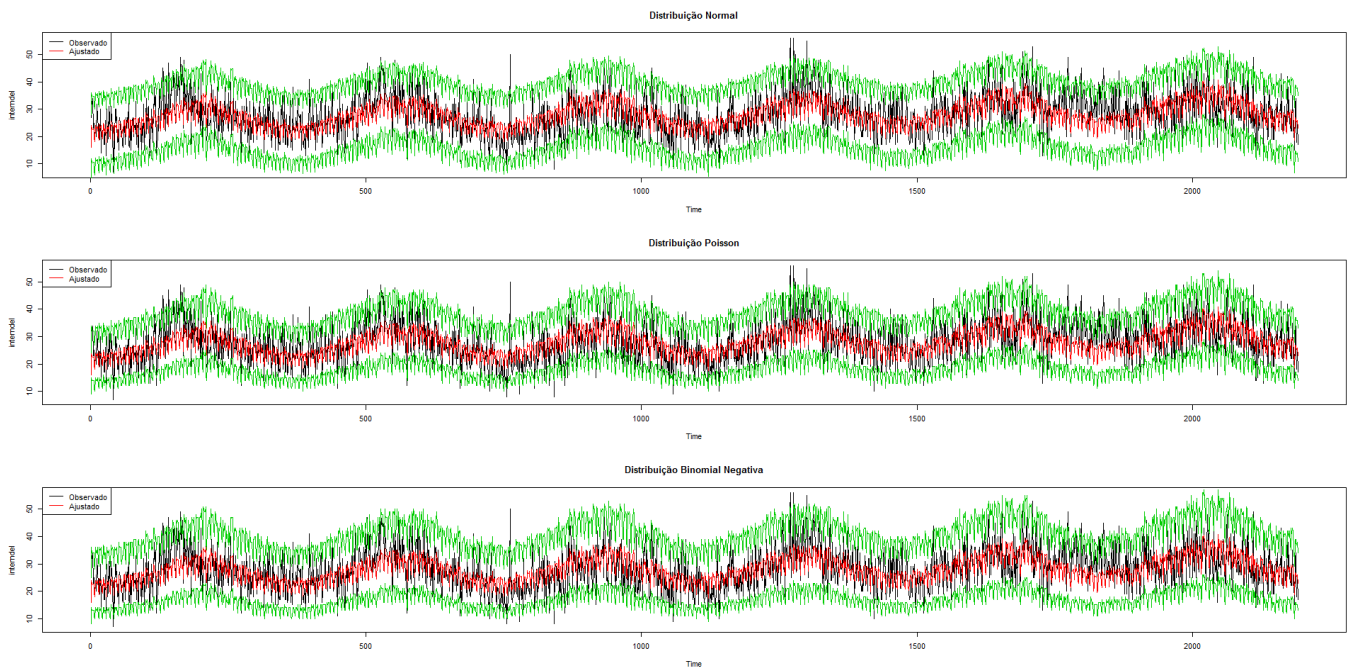


Figura 6.27: Intervalo de confiança dos modelos MLG.

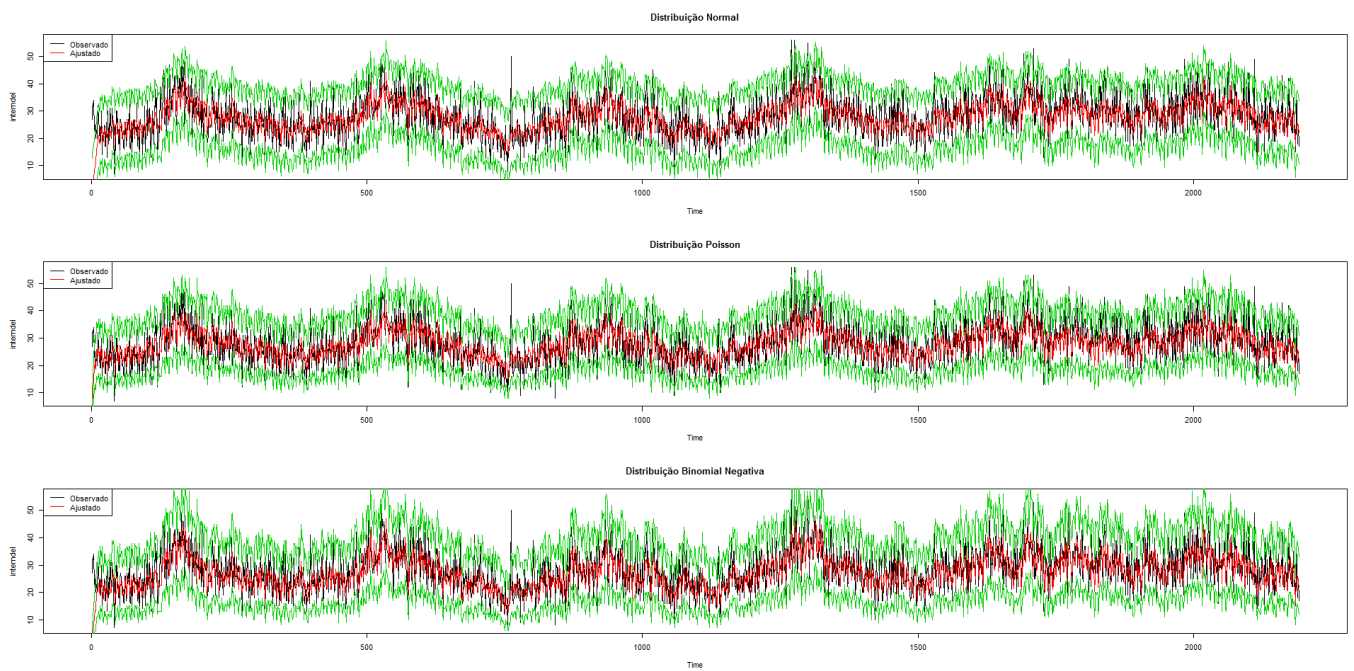


Figura 6.28: Intervalo de confiança dos modelos GARMA.

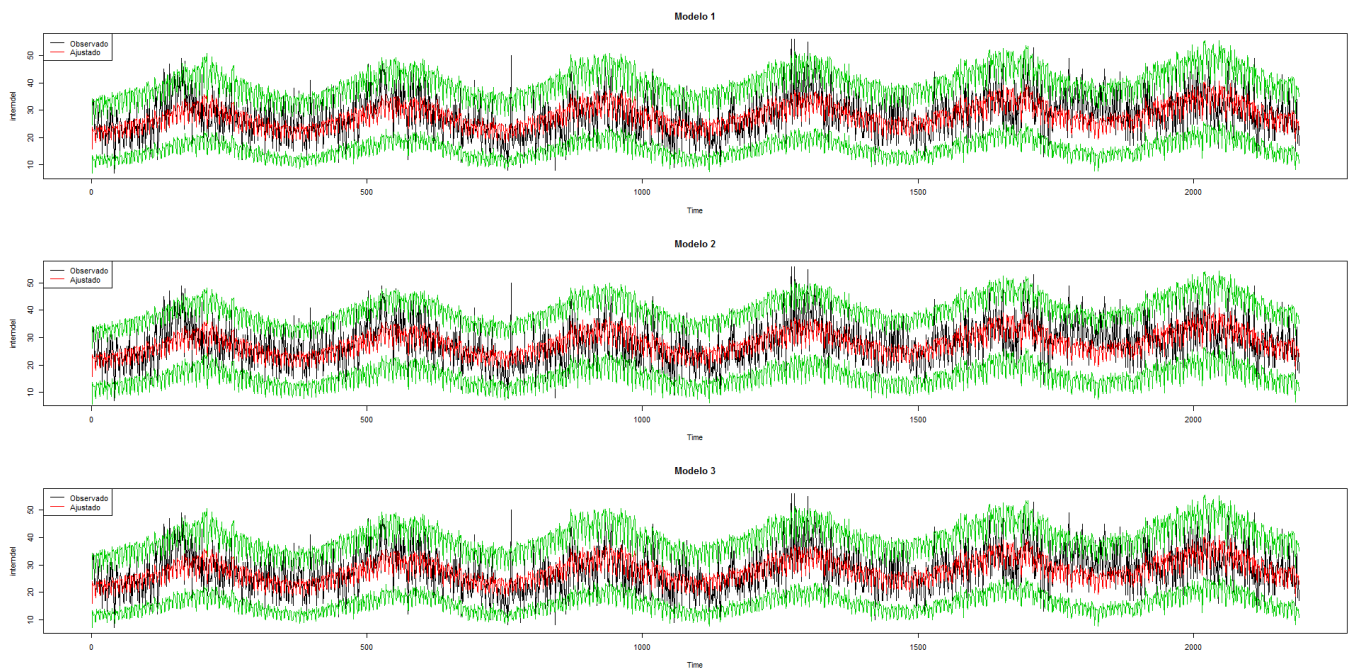


Figura 6.29: Intervalo de confiança dos modelos GAMLSS com distribuição Normal.

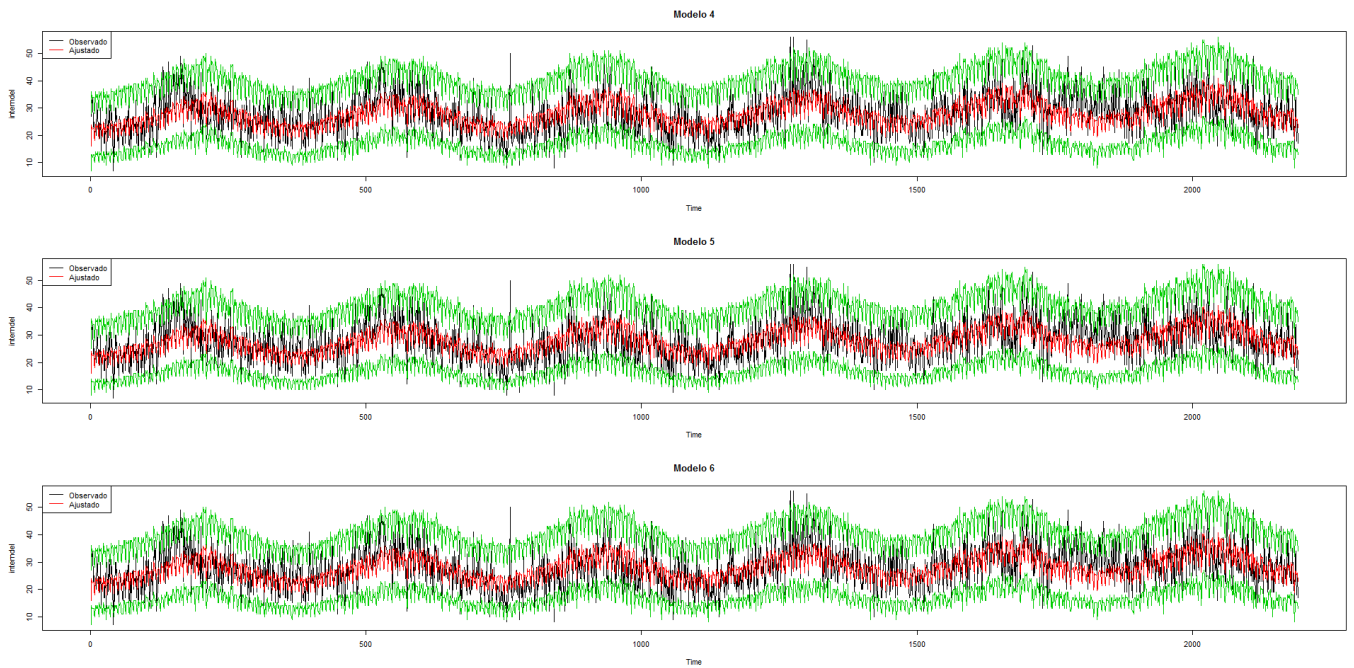


Figura 6.30: Intervalo de confiança dos modelos GAMLSS com distribuição Binomial Negativa.

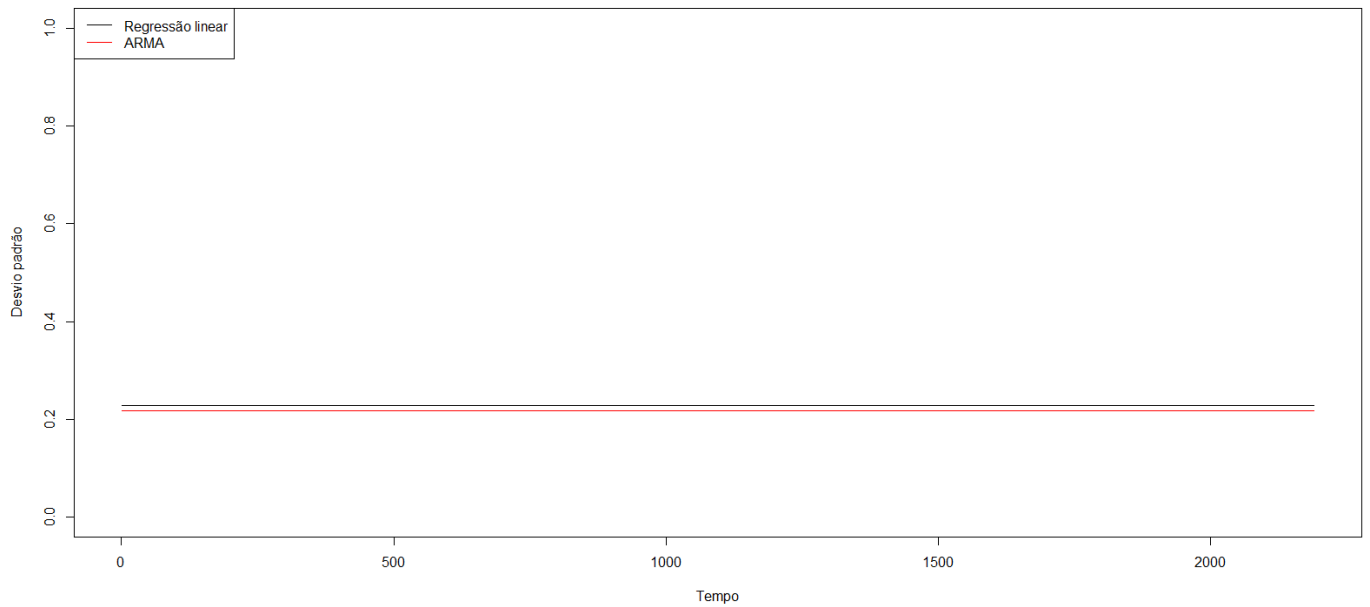


Figura 6.31: Desvio padrão dos modelos regressão linear e ARMA.

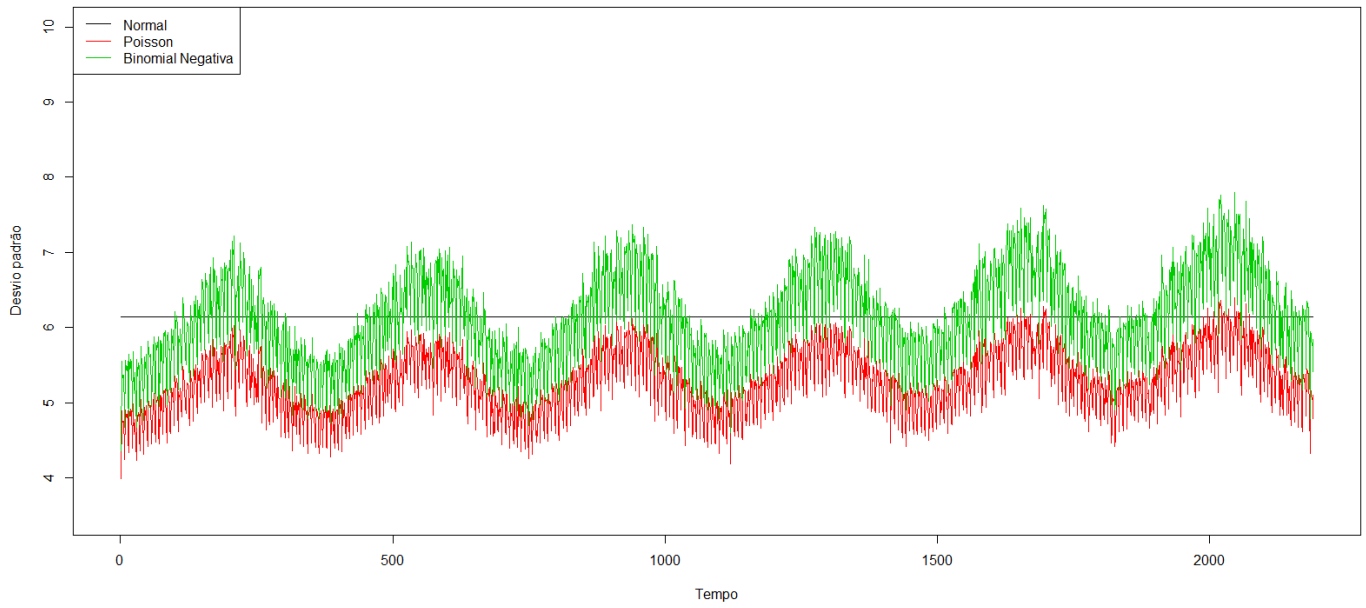


Figura 6.32: *Desvio padrão dos modelos MLG.*

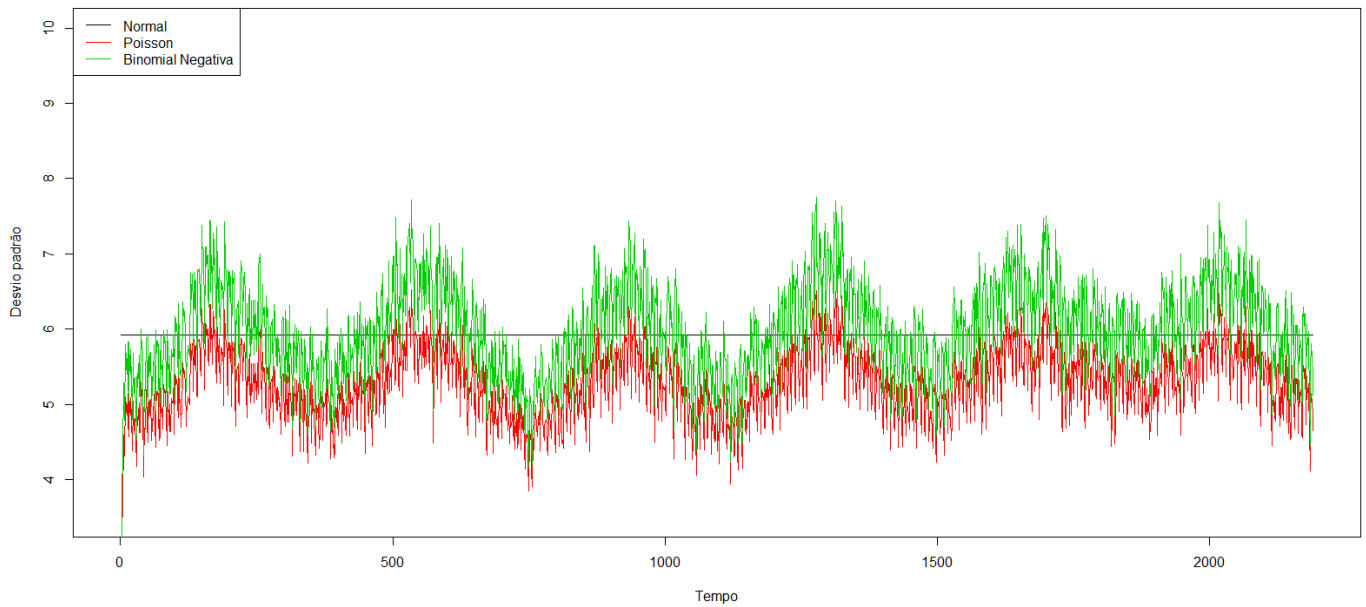


Figura 6.33: *Desvio padrão dos modelos GARMA.*

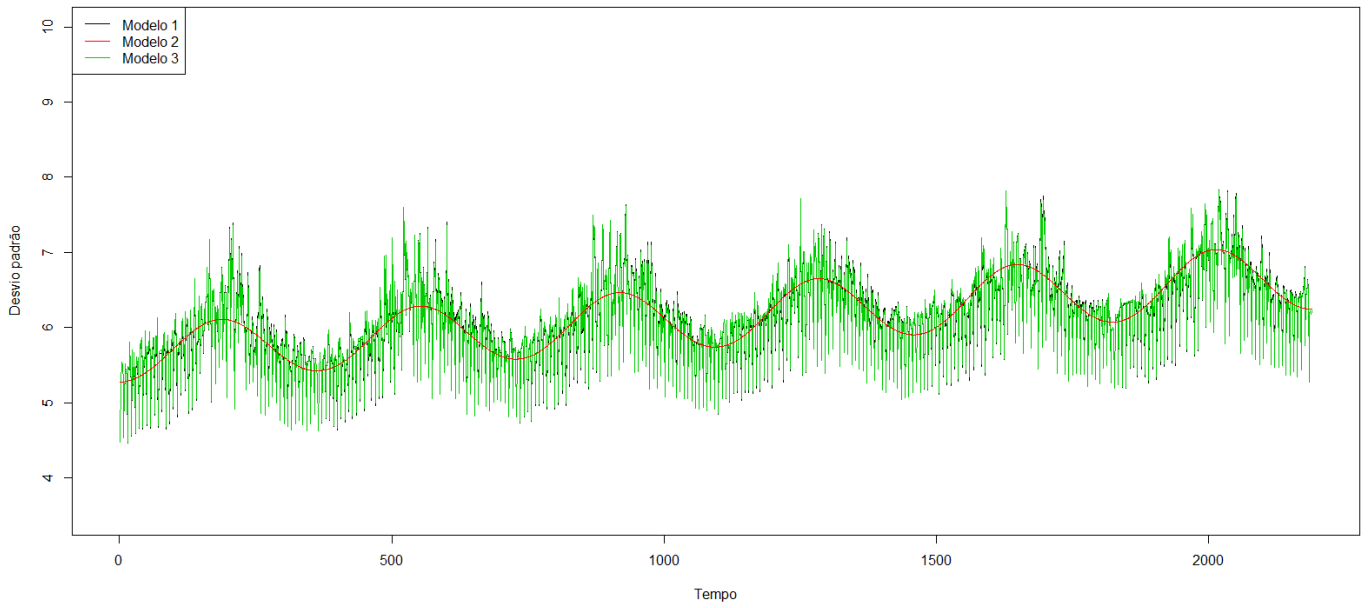


Figura 6.34: Desvio padrão dos modelos GAMLSS com distribuição Normal.

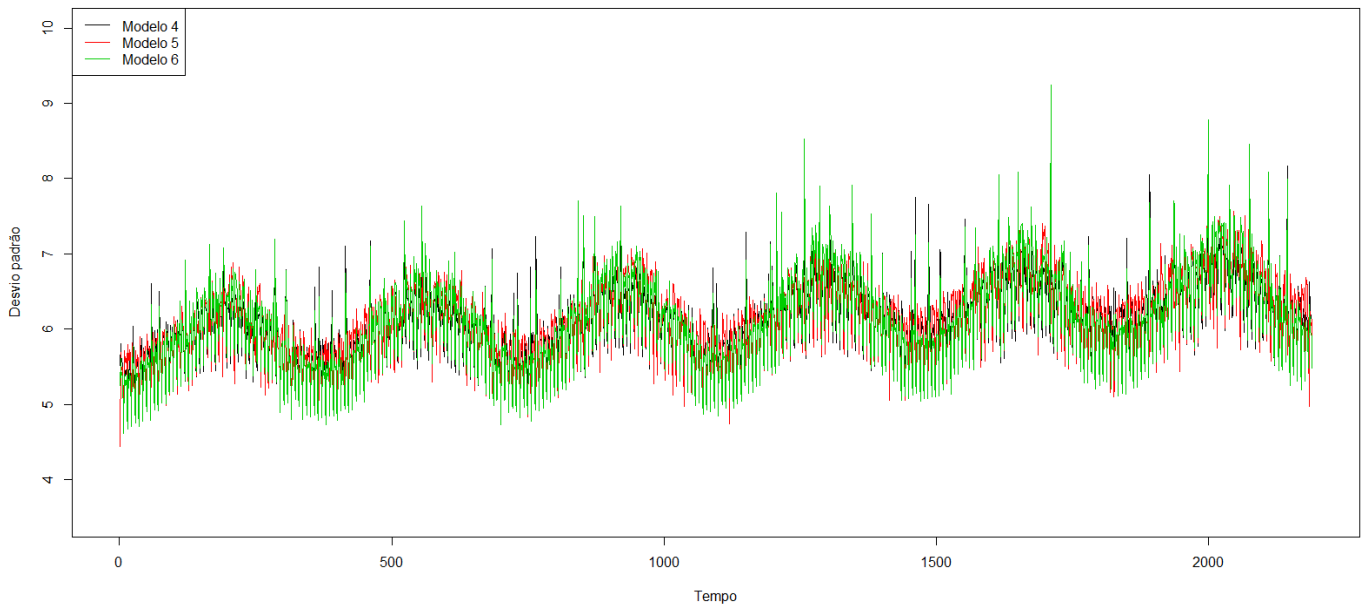


Figura 6.35: Desvio padrão dos modelos GAMLSS com distribuição Binomial Negativa.

6.3 Tabelas

Variáveis	PM10max	PM10med	COmax	COmed	O3max	O3med
log(tx)	0,229	0,306	0,294	0,271	0,039	-0,029
PM10max	1,000	0,821	0,604	0,522	0,193	0,005
PM10med		1,000	0,704	0,676	0,270	0,024
COmax			1,000	0,811	0,074	-0,165
COmed				1,000	0,058	-0,169
O3max					1,000	0,810
O3med						1,000
	Tmax	Tmed	Tmin	Umax	Umed	Umin
log(tx)	-0,062	-0,192	-0,274	-0,063	-0,164	-0,172
PM10max	0,276	0,053	-0,193	-0,058	-0,397	-0,465
PM10med	0,362	0,070	-0,258	-0,085	-0,515	-0,621
COmax	0,260	0,056	-0,179	0,003	-0,336	-0,437
COmed	0,239	0,072	-0,136	0,009	-0,285	-0,373
O3max	0,422	0,260	0,072	-0,034	-0,289	-0,409
O3med	0,284	0,203	0,095	-0,067	-0,213	-0,257
Tmax	1,000	0,878	0,569	-0,113	-0,528	-0,670
Tmed		1,000	0,864	-0,094	-0,284	-0,309
Tmin			1,000	-0,012	0,110	0,126
Umax				1,000	0,626	0,319
Umed					1,000	0,862
Umin						1,000

Tabela 6.1: *Correlação de Pearson para as variáveis de poluição, climáticas e logaritmo da taxa de interações.*

Variáveis	log(tx)	Seno	Cosseno	Domingo	Segunda	Sábado	Feriados
log(tx)	1,000						
Seno	-0,095	1,000					
Cosseno	-0,411	0,000	1,000				
Domingo	-0,252	0,000	0,000	1,000			
Segunda	0,126	0,000	0,000	-0,167	1,000		
Sábado	-0,157	0,000	0,000	-0,167	-0,167	1,000	
Feriados	-0,113	0,029	0,049	-0,020	0,027	-0,041	1,000
PM10max	0,229	-0,110	-0,275	-0,102	-0,037	-0,002	-0,019
PM10med	0,306	-0,111	-0,359	-0,148	-0,043	-0,009	-0,055
COmax	0,294	0,030	-0,362	-0,143	0,005	-0,079	-0,075
Comed	0,271	0,053	-0,330	-0,185	-0,011	-0,094	-0,089
O3max	0,039	-0,054	0,103	0,012	0,000	0,042	0,003
O3med	-0,029	-0,115	0,163	0,100	-0,013	0,063	0,037
Tmax	-0,062	0,133	0,427	0,007	-0,017	0,011	0,039
Tmed	-0,192	0,236	0,650	0,012	-0,009	0,018	0,054
Tmin	-0,274	0,267	0,722	0,014	0,007	0,016	0,049
Urmax	-0,063	0,009	0,025	0,000	-0,005	0,003	0,019
Urmed	-0,164	0,070	0,112	0,018	0,016	-0,006	0,005
Urmin	-0,172	0,031	0,133	0,006	0,018	-0,005	-0,001

Tabela 6.2: Correlação de Pearson para as variáveis sazonais e Feriados.

	Média	DP	Q3-Q1	Mínimo	1º quantil	Mediana	3º quantil	Máximo
Internação	28,112	7,589	10	7	23	27	33	56
PM10med	33,299	16,903	20	3	21	30,5	41	117
COmax	1,991	1,019	1,1	0,2	1,3	1,8	2,4	7
O3max	94,680	43,894	57,5	0	61,5	89	119	293
Tmin	15,109	3,448	5,3	2,4	12,7	15,4	18	21,6
Urmin	55,765	16,850	24	0	43	55	67	96

Tabela 6.3: Tabela descritiva de internação e variáveis de poluição e climáticas.

Coeficientes	Regressão Linear		ARMA	
	Estimativa	Valor p	Estimativa	Valor p
Seno	1,074 (0,054)	0,156	1,076 (0,060)	0,192
Cosseno	1,187 (0,060)	<0,001	1,258 (0,070)	<0,001
Domingo	1,058 (0,108)	0,579	1,120 (0,127)	0,319
Segunda	0,933 (0,095)	0,492	0,967 (0,110)	0,770
Sábado	1,206 (0,122)	0,066	1,179 (0,134)	0,147
Feriados	2,624 (0,483)	<0,001	3,449 (0,708)	<0,001
PM10med	0,997 (0,002)	0,120	0,996 (0,002)	0,110
COmax	0,965 (0,034)	0,304	0,932 (0,036)	0,071
O3max	0,998 (0,001)	0,031	0,999 (0,001)	0,101
Tmin	1,027 (0,011)	0,009	1,036 (0,012)	0,002
Urmin	1,007 (0,002)	0,002	1,009 (0,002)	<0,001

Tabela 6.4: Estimativa pontual, erro padrão e valor p dos efeitos das variáveis do modelos de regressão linear simples para os resíduos ao quadrado dos modelos de regressão linear e ARMA.

Coeficientes	Normal		Poisson		Binomial negativa	
	Estimativa	Valor p	Estimativa	Valor p	Estimativa	Valor p
Seno	0,972 (0,042)	0,505	1,015 (0,057)	0,790	1,027 (0,044)	0,534
Cosseno	0,882 (0,038)	0,004	1,020 (0,058)	0,732	1,056 (0,045)	0,202
Domingo	0,752 (0,066)	0,001	0,851 (0,097)	0,157	0,921 (0,080)	0,343
Segunda	1,124 (0,098)	0,180	1,022 (0,117)	0,846	1,008 (0,087)	0,925
Sábado	0,918 (0,080)	0,331	1,041 (0,119)	0,723	1,047 (0,091)	0,593
Feriados	1,232 (0,196)	0,189	1,888 (0,392)	0,002	1,774 (0,279)	<0,001
PM10med	1,005 (0,002)	0,003	1,002 (0,002)	0,481	1,000 (0,002)	0,992
COmax	1,097 (0,033)	0,002	1,041 (0,041)	0,306	1,012 (0,030)	0,680
O3max	0,999 (0,001)	0,406	0,999 (0,001)	0,13	0,999 (0,001)	0,110
Tmin	0,984 (0,009)	0,070	1,005 (0,012)	0,693	1,009 (0,009)	0,302
Urmin	1,000 (0,002)	0,959	1,004 (0,002)	0,097	1,004 (0,002)	0,043

Tabela 6.5: Estimativa pontual, erro padrão e valor p dos efeitos das variáveis do modelos de regressão linear simples para os resíduos ao quadrado dos modelos MLG.

Coeficientes	Normal		Poisson		Binomial negativa	
	Estimativa	Valor p	Estimativa	Valor p	Estimativa	Valor p
Seno	0,966 (0,049)	0,493	0,999 (0,070)	0,989	1,008 (0,060)	0,888
Cosseno	0,997 (0,051)	0,945	1,190 (0,084)	0,013	1,192 (0,071)	0,003
Domingo	0,758 (0,078)	0,007	0,949 (0,135)	0,712	1,011 (0,122)	0,930
Segunda	1,221 (0,125)	0,052	1,179 (0,168)	0,247	1,154 (0,140)	0,237
Sábado	0,891 (0,091)	0,259	0,994 (0,141)	0,966	1,015 (0,123)	0,904
Feriados	1,433 (0,268)	0,054	2,953 (0,763)	<0,001	2,739 (0,601)	<0,001
PM10med	1,004 (0,002)	0,058	1,000 (0,003)	0,974	0,999 (0,003)	0,762
COmax	1,044 (0,037)	0,224	0,962 (0,047)	0,433	0,949 (0,039)	0,212
O3max	1,000 (0,001)	0,896	1,000 (0,001)	0,856	1,000 (0,001)	0,685
Tmin	0,998 (0,010)	0,857	1,024 (0,015)	0,098	1,024 (0,013)	0,049
Urmin	1,003 (0,002)	0,235	1,007 (0,003)	0,022	1,006 (0,003)	0,010

Tabela 6.6: Estimativa pontual, erro padrão e valor p dos efeitos das variáveis do modelos de regressão linear simples para os resíduos ao quadrado dos modelos GARMA.

Distribuição No Coeficientes	Modelo 1		Modelo 2		Modelo 3	
	Estimativa	Valor p	Estimativa	Valor p	Estimativa	Valor p
Seno	1,000 (0,043)	1,000	1,000 (0,043)	1,000	0,988 (0,042)	0,777
Cosseno	1,000 (0,043)	1,000	1,000 (0,043)	1,000	1,000 (0,043)	1,000
Domingo	1,000 (0,087)	1,000	0,751 (0,066)	0,001	1,000 (0,087)	1,000
Segunda	1,069 (0,093)	0,442	1,118 (0,098)	0,205	1,069 (0,093)	0,439
Sábado	0,893 (0,077)	0,192	0,902 (0,079)	0,240	0,893 (0,077)	0,191
Feriados	1,343 (0,212)	0,061	1,307 (0,209)	0,094	1,341 (0,211)	0,063
PM10med	1,001 (0,002)	0,632	1,004 (0,002)	0,021	1,001 (0,002)	0,608
COmax	1,000 (0,030)	1,000	1,078 (0,032)	0,012	1,000 (0,030)	1,000
O3max	0,999 (0,001)	0,230	0,999 (0,001)	0,322	0,999 (0,001)	0,240
Tmin	1,003 (0,009)	0,698	1,004 (0,009)	0,652	1,003 (0,009)	0,749
Urmin	1,003 (0,002)	0,116	1,001 (0,002)	0,652	1,003 (0,002)	0,120

Tabela 6.7: Estimativa pontual, erro padrão e valor p dos efeitos das variáveis do modelos de regressão linear simples para os resíduos ao quadrado dos modelos GAMLSS com distribuição Normal.

Distribuição BN Coeficientes	Modelo 4		Modelo 5		Modelo 6	
	Estimativa	Valor p	Estimativa	Valor p	Estimativa	Valor p
Seno	1,004 (0,042)	0,919	1,016 (0,043)	0,704	1,027 (0,043)	0,533
Cosseno	0,964 (0,041)	0,379	0,978 (0,042)	0,608	1,026 (0,043)	0,535
Domingo	0,939 (0,080)	0,463	0,926 (0,080)	0,369	0,924 (0,079)	0,355
Segunda	0,975 (0,083)	0,769	0,989 (0,085)	0,900	0,978 (0,083)	0,794
Sábado	1,083 (0,092)	0,350	1,083 (0,093)	0,355	1,089 (0,093)	0,320
Feriados	0,999 (0,155)	0,994	1,700 (0,266)	<0,001	1,023 (0,159)	0,883
PM10med	1,004 (0,002)	0,030	1,001 (0,002)	0,421	1,003 (0,002)	0,089
COmax	1,060 (0,031)	0,047	1,031 (0,030)	0,304	1,050 (0,031)	0,094
O3max	0,999 (0,001)	0,396	0,999 (0,001)	0,054	1,000 (0,001)	0,546
Tmin	0,995 (0,009)	0,544	0,997 (0,009)	0,751	1,004 (0,009)	0,615
Urmin	0,999 (0,002)	0,724	1,003 (0,002)	0,081	1,000 (0,002)	0,779

Tabela 6.8: Estimativa pontual, erro padrão e valor p dos efeitos das variáveis do modelos de regressão linear simples para os resíduos ao quadrado dos modelos GAMLSS com distribuição Binomial Negativa.

	Média	Variância	Coef. de assimetria	Coef. de curtose
Reg linear	2,36E-16	1,00E+00	-4,14E-01	3,77E+00
ARMA	6,61E-05	9,98E-01	-4,3E-01	3,94E+00
MLG - Normal	-1,78E-04	1,00E+00	3,22E-01	3,05E+00
MLG - Poisson	-1,14E-02	1,34E+00	1,06E-01	2,96E+00
MLG - Binomial Negativa	4,77E-05	1,00E+00	5,00E-02	2,99E+00
GARMA - Normal	5,62E-03	9,70E-01	2,53E-01	3,13E+00
GARMA - Poisson	-1,01E-02	1,25E+00	0,03E-02	3,00E+00
GARMA - Binomial Negativa	-1,26E-03	9,89E-01	-3,87E-03	3,02E+00
Modelo 1	-2,05E-05	1,00E+00	3,26E-01	3,01E+00
Modelo 2	-1,67E-05	1,00E+00	3,28E-01	3,06E+00
Modelo 3	-5,53E-05	1,00E+00	3,26E-01	3,01E+00
Modelo 4	1,49E-03	1,00E+00	5,94E-02	2,93E+00
Modelo 5	-1,34E-03	1,01E+00	3,58E-02	2,99E+00
Modelo 6	1,61E-03	1,00E+00	6,99E-02	2,94E+00

Tabela 6.9: Medidas de diagnóstico.

Modelos	EQM
Regressão Linear	37,903
ARMA	33,860
MLG - Normal	37,709
MLG - Poisson	37,718
MLG - Binomial Negativa	37,718
GARMA - Normal	34,381
GARMA - Poisson	35,258
GARMA - Binomial Negativa	35,541
Modelo 1	37,718
Modelo 2	37,715
Modelo 3	37,717
Modelo 4	37,715
Modelo 5	37,717
Modelo 6	37,720

Tabela 6.10: Erro quadrático médio.

	Média	Mínimo	1º quantil	Mediana	3º quantil	Máximo
Regressão Linear	5,497	5,497	5,497	5,497	5,497	5,497
ARMA	5,699	5,699	5,699	5,699	5,699	5,699
MLG - Normal	6,141	6,141	6,141	6,141	6,141	6,141
MLG - Poisson	5,285	3,992	4,990	5,265	5,601	6,391
MLG - Binomial Negativa	5,317	4,015	5,018	5,296	5,635	6,430
GARMA - Normal	5,924	5,924	5,924	5,924	5,924	5,924
GARMA - Poisson	5,279	1,000	4,961	5,273	5,614	6,520
GARMA - Binomial Negativa	5,306	1,005	4,990	5,304	5,637	6,523
Modelo 1	6,134	4,459	5,765	6,153	6,501	7,831
Modelo 2	6,147	5,273	5,841	6,119	6,444	7,036
Modelo 3	6,133	4,470	5,776	6,156	6,493	7,822
Modelo 4	5,318	4,110	5,023	5,297	5,629	6,415
Modelo 5	5,317	4,022	5,022	5,297	5,631	6,422
Modelo 6	5,317	4,081	5,019	5,297	5,634	6,423

Tabela 6.11: Tabela descritiva para o desvio padrão.

	Proporção (em %)
Regressão Linear	4,5
ARMA	4,5
MLG - Normal	4,8
MLG - Poisson	7,5
MLG - Binomial Negativa	4,0
GARMA - Normal	4,6
GARMA - Poisson	6,8
GARMA - Binomial Negativa	3,7
Modelo 1	4,7
Modelo 2	5,2
Modelo 3	4,7
Modelo 4	3,8
Modelo 5	4,0
Modelo 6	3,9

Tabela 6.12: Proporção de pontos fora do intervalo de confiança.

Referências Bibliográficas

- Baccini et al. (2008)** Michela Baccini, Annibale Biggeri, Gabriele Accetta, Tom Kosatsky, Klea Katsouyanni, Antonis Analitis, H Ross Anderson, Luigi Bisanti, Daniela D'Ippoliti, Jana Danova et al. Heat effects on mortality in 15 european cities. *Epidemiology*, 19(5):711–719. Citado na pág. [31](#)
- Bakonyi et al. (2004)** Sonia Maria Cipriano Bakonyi, Inês Moresco Danni-Oliveira, Lourdes Conceição Martins e Alfésio Luís Ferreira Braga. Poluição atmosférica e doenças respiratórias em crianças na cidade de curitiba, pr. *Rev Saúde Pública*, 38(5):695–700. Citado na pág. [22](#)
- Benjamin et al. (2003)** Michael A Benjamin, Robert A Rigby e D Mikis Stasinopoulos. Generalized autoregressive moving average models. *Journal of the American Statistical Association*, 98(461):214–223. Citado na pág. [2](#), [8](#), [11](#)
- Box e Jenkins (1976)** George EP Box e Gwilym M Jenkins. Time series analysis, control, and forecasting, 1976. Citado na pág. [2](#), [5](#), [8](#)
- Box et al. (2013)** George EP Box, Gwilym M Jenkins e Gregory C Reinsel. *Time series analysis: forecasting and control*. Wiley. com. Citado na pág. [9](#)
- Briet et al. (2000)** Olivier Briet, Maintainer Olivier Briet e MASS Imports. Package gsrima. *Australian & New Zealand Journal of Statistics*, 42:479–495. Citado na pág. [33](#)
- Bussab e Morettin (2010)** Wilton de O Bussab e Pedro A Morettin. *Estatística básica; Basic statistics*. Saraiva. Citado na pág. [7](#)
- Conceição et al. (2001)** Gleice Margarete de Souza Conceição, Paulo Hilário Nascimento Saldiva e Julio da Motta Singer. Glm and gam model for analyzing the association between atmospheric pollution and morbidity-mortality markers: an introduction based on data from the city of são paulo. *Revista Brasileira de Epidemiologia*, 4(3):206–219. Citado na pág. [1](#), [13](#)
- Cox (1975)** David R Cox. Partial likelihood. *Biometrika*, 62(2):269–276. Citado na pág. [13](#)
- Dempster et al. (1977)** Arthur P Dempster, Nan M Laird, Donald B Rubin et al. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal statistical Society*, 39(1):1–38. Citado na pág. [16](#)
- Dunn e Smyth (1996)** Peter K Dunn e Gordon K Smyth. Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5(3):236–244. Citado na pág. [14](#)
- Gouveia et al. (2006)** Nelson Gouveia, Clarice Umbelino de Freitas, Lourdes Conceição Martins e Izabel Oliva Marcilio. Hospitalizações por causas respiratórias e cardiovasculares associadas à contaminação atmosférica no município de são paulo, brasil. *Cad Saúde Pública*, 22(12):2669–77. Citado na pág. [22](#)
- Green (1984)** Peter J Green. Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *Journal of the Royal Statistical Society B*, 46(2):149–192. Citado na pág. [7](#)

- Hardin e Hilbe (2007)** James William Hardin e Joseph Hilbe. *Generalized linear models and extensions*. Stata Corp. Citado na pág. 9
- Hastie e Tibshirani (1990)** Trevor J Hastie e Robert J Tibshirani. *Generalized additive models*, volume 43. CRC Press. Citado na pág. 1, 12
- Junger e Junger (2012)** Washington Junger e Maintainer Washington Junger. Package mtsdi, 2012. Citado na pág. 16
- Katsouyanni et al. (1995)** Klea Katsouyanni, D Zmirou, C Spix, J Sunyer, JP Schouten, A Ponka, HR Anderson, Y Le Moullec, B Wojtyniak, MA Vigotti et al. Short-term effects of air pollution on health: a european approach using epidemiological time-series data. the aphea project: background, objectives, design. *European Respiratory Journal*, 8(6):1030–1038. Citado na pág. 1
- Katsouyanni et al. (1996)** Klea Katsouyanni, Joel Schwartz, C Spix, Giota Touloumi, D Zmirou, A Zanobetti, B Wojtyniak, JM Vonk, A Tobias, A Pönkä et al. Short term effects of air pollution on health: a european approach using epidemiologic time series data: the aphea protocol. *Journal of epidemiology and community health*, 50(Suppl 1):S12–S18. Citado na pág. 18
- Kovats et al. (2004)** R Sari Kovats, Shakoor Hajat e Paul Wilkinson. Contrasting patterns of mortality and hospital admissions during hot weather and heat waves in greater london, uk. *Occupational and environmental medicine*, 61(11):893–898. Citado na pág. 31
- Linares e Díaz (2008)** C Linares e J Díaz. Impact of high temperatures on hospital admissions: comparative analysis with previous studies about mortality (madrid). *The European Journal of Public Health*, 18(3):317–322. Citado na pág. 2, 31
- Michelozzi et al. (2009)** Paola Michelozzi, Gabriele Accetta, Manuela De Sario, Daniela D’Ippoliti, Claudia Marino, Michela Baccini, Annibale Biggeri, H Ross Anderson, Klea Katsouyanni, Ferran Ballester et al. High temperature and hospitalizations for cardiovascular and respiratory causes in 12 european cities. *American journal of respiratory and critical care medicine*, 179(5):383–389. Citado na pág. 31
- Mood (1950)** Alexander McFarlane Mood. *Introduction to the Theory of Statistics*. McGraw-hill. Citado na pág. 6
- Morettin e Toloí (2006)** Pedro A Morettin e Clélia Toloí. *Análise de séries temporais*. Blucher. Citado na pág. 8
- Morettin (2008)** Pedro Alberto Morettin. *Econometria financeira: um curso em séries temporais financeiras*. Edgard Blücher. Citado na pág. 15
- Nelder e Wedderburn (1972)** John A Nelder e Robert WM Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, páginas 370–384. Citado na pág. 9
- Paula (2013)** Gilberto A Paula. Modelos de regressão com apoio computacional, 2013. Citado na pág. 1, 9, 10
- Pereira et al. (1998)** LA Pereira, Dana Loomis, GM Conceicao, AL Braga, Rosângela M Arcas, Humberto S Kishi, Júlio M Singer, György M Böhm e PH Saldiva. Association between air pollution and intrauterine mortality in são paulo, brazil. *Environmental Health Perspectives*, 106(6):325. Citado na pág. 19
- Rigby e Stasinopoulos (1996)** RA Rigby e DM Stasinopoulos. A semi-parametric additive model for variance heterogeneity. *Statistics and Computing*, 6(1):57–65. Citado na pág. 6

- Rigby e Stasinopoulos (2005)** RA Rigby e DM Stasinopoulos. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3):507–554. Citado na pág. [2](#), [6](#), [7](#), [12](#)
- Sen et al. (2010)** Pranab K Sen, Julio M Singer e Antonio C Pedroso de Lima. *From finite sample to asymptotic methods in statistics*. Cambridge University Press. Citado na pág. [7](#)
- Serfling (1963)** Robert E Serfling. Methods for current statistical analysis of excess pneumonia-influenza deaths. *Public health reports*, 78(6):494. Citado na pág. [2](#), [18](#)
- Shumway e Stoffer (2000)** Robert H Shumway e David S Stoffer. *Time series analysis and its applications*, volume 3. Springer New York. Citado na pág. [2](#), [8](#)
- Silverman (1986)** Bernard W Silverman. *Density estimation for statistics and data analysis*, volume 26. CRC press. Citado na pág. [51](#)
- Souza et al. (2013)** Maria de Fatima Marinho de Souza, Marc-Alain Widdowson, Airlane P Alencar, Vilma P Gawryszewski, Eduardo Aziz-Baumgartner, Rakhee Palekar, Joseph Breese, Po-Yung Cheng, Jarbas Barbosa, Ana M Cabrera et al. Trends in mortality from respiratory disease in latin america since 1998 and the impact of the 2009 influenza pandemic. *Bulletin of the World Health Organization*, 91(7):525–532. Citado na pág. [2](#)
- Stasinopoulos e Rigby (2007)** D Mikis Stasinopoulos e Robert A Rigby. Generalized additive models for location scale and shape (gamlss) in r. *Journal of Statistical Software*, 23(7):1–46. Citado na pág. [7](#)
- Stasinopoulos et al. (2010)** DM Stasinopoulos, BA Rigby, C Akantziliotou, G Heller, R Ospina e N Motpan. gamlss. dist: Distributions to be used for gamlss modelling. *R package version*, páginas 4–0. Citado na pág. [16](#)
- Stasinopoulos et al. (2013)** Mikis Stasinopoulos, Bob Rigby, Paul Eilers e Maintainer Mikis Stasinopoulos. Package gamlss.util, 2013. Citado na pág. [7](#), [9](#), [11](#), [12](#)
- Van den Bergh et al. (2008)** F Van den Bergh, JP Holloway, M Pienaar, R Koen, CD Elphinstone e S Woodborne. A comparison of various modelling approaches applied to cholera case data. *ORiON: The Journal of ORSSA*, 24(1):17–36. Citado na pág. [1](#)
- Wooldridge (2012)** Jeffrey M Wooldridge. *Introductory econometrics: a modern approach*. Cengage Learning. Citado na pág. [2](#), [5](#), [15](#), [27](#)