

Roberto Rodrigues Costa

Efeito dos focos de queimadas no nascimento de bebês prematuros: uma aplicação usando a suavização “*spline*”, a partir de Modelos Aditivos Generalizados.

Niterói - RJ, Brasil 28 de março de 2016

Roberto Rodrigues Costa

Efeito dos focos de queimadas no nascimento de bebês prematuros: uma aplicação usando a suavização “*spline*”, a partir de Modelos Aditivos Generalizados.

Trabalho de Conclusão de Curso

Monografia apresentada para obtenção do grau de Bacharel em Estatística pela Universidade Federal Fluminense.

Orientador: Prof. Ludmilla da Silva Viana Jacobson

Roberto Rodrigues Costa

Efeito dos focos de queimadas no nascimento de bebês prematuros: uma aplicação usando a suavização “*spline*”, a partir de Modelos Aditivos Generalizados

Monografia de Projeto Final de Graduação sob o Efeito dos focos de queimadas no nascimento de bebês prematuros: uma aplicação usando a suavização “*spline*”, a partir de Modelos Aditivos Generalizados, defendida por Roberto Rodrigues Costa e aprovada em 28 de março de 2016, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

Prof^ª. Dra. Ludmilla Viana Jacobson
Departamento de Estatística - UFF

Prof^ª. Dra. Márcia Marques de Carvalho
Departamento de Estatística - UFF

Niterói, 28 de março de 2016

M 149 Costa, Roberto Rodrigues

Efeito dos focos de queimadas no nascimento de bebês prematuros: uma aplicação usando a suavização “*spline*”, a partir de Modelos Aditivos Generalizados /Roberto Rodrigues Costa. - Niterói: [s. n.], 2016. 39f.

Trabalho de Conclusão de Curso - (Bacharelado em Estatística) – Universidade Federal Fluminense, 2016.

1. Modelo Aditivo Generalizado. 2. Spline 2.
CDD. 519.536

Resumo

A classe dos Modelos Aditivos Generalizados (MAG) pode ser vista como uma extensão dos Modelos Lineares Generalizados (MLG), e permite que o preditor linear seja substituído por um preditor aditivo, composto pela soma de funções suaves. Neste trabalho, em particular, as funções não-paramétricas utilizadas para suavização são *splines*. Com isso, pode-se inserir no preditor variáveis que não influenciam a resposta média de forma linear, criando-se uma estrutura mais flexível para modelagem dos dados.

O procedimento de inferência seguiu o enfoque clássico para exemplificar o método. Os modelos propostos foram ajustados a um conjunto de dados de contagens de nascimento de bebês prematuros na região do município de Tangará da Serra – Mato Grosso, com observações feitas no período de 01 de janeiro de 2000 a 31 de dezembro de 2010. Pressupõe-se distribuição de Poisson para a contagem de nascimentos prematuros e a média é explicada em função da variável número de focos de queimada mensal da região de Tangará da Serra, ajustado por *spline*.

Todas as séries temporais mencionadas no período acima foram analisadas e divididas por ano e por mês. Para avaliar a relação e o comportamento das principais variáveis nesse trabalho, foi feita uma análise exploratória para Focos de queimadas e Nascimentos prematuros no município de Tangará da Serra. Quanto a autocorrelação foram realizados testes nos pontos temporais das séries para verificar em que pontos os dados da série estariam melhor relacionados. Também foi utilizado o teste Box-Ljung para testar a hipótese de independência e distribuição dos resíduos e o critério de Akaike para a escolha do melhor modelo.

Foi escolhido um modelo pela regressão de Poisson para explicar a prematuridade com 2 parâmetros constituído pelo mês e números de focos (ajustado linearmente pois por *spline* não é significativo), e um efeito do tempo ajustado por *spline* cúbica com 3 nós. Mesmo ajustando linearmente o variável foco através do Risco Relativo, percebe-se um aumento mínimo na prematuridade quando ocorre um aumento do número de focos de queimadas em Tangará da Serra – MS.

Agradecimentos

Primeiramente a Deus, por ter me dado saúde e força para superar dificuldades.

A minha mãe Josélia, pelo amor, incentivo e apoio incondicional.

A minha orientadora Prof^a. Dra. Ludmilla pelo suporte no pouco tempo que lhe coube, pelas suas correções, incentivos e paciência.

Aos amigos que fiz durante minha graduação e nos períodos de estágio.

E a todos que direta ou indiretamente fizeram parte da minha formação, o meu muito obrigado.

Sumário

Lista de Figuras.	09
Lista de Tabelas.	11
1 Introdução.	12
2 Objetivos.	14
2.1 Objetivo Geral.	14
2.2 Objetivos Específicos.	14
3 Materiais e Métodos.	15
3.1 Área de estudo.	15
3.2 Período de estudo.	16
3.3 Focos de queimadas.	16
3.4 Prematuridade.	17
3.5 Series Temporais.	18
3.5.1 Função de autocorrelação (ACF).	22
3.5.2 Função de autocorrelação parcial (PACF)	23
3.6 Modelo Aditivo Generalizado.	24
3.7 Função estimada de suavização: <i>Splines</i>	25
3.8 Critério de decisão – Akaike (AIC)	28
3.9 Diagnóstico do modelo	29
4 Resultados	30
4.1 Análise Exploratória das series temporais	30
4.1.1 Prematuridade	30
4.1.2 Focos de Queimadas	33
4.1.3 Prematuridade X Focos de queimadas	35
4.2 Estimação do Modelo aditivo generalizado	37
4.2.1 Ajuste da tendência.	37
4.2.2 Ajuste da sazonalidade	39
4.2.3 Ajuste do efeito da poluição	39

4.2.4 Análise de Resíduos	43
5 Conclusão	46
Referências.....	47

Lista de Figuras

1: Exemplo de Série Temporal Estacionária	20
2: Exemplo de Série Temporal Não Estacionária.	20
3: Série temporal de passageiros de uma companhia aérea, Componente Sazonal e Série Ajustada	24
4: Efeito de X sobre a variável Y (Exemplo de <i>Spline</i>).	33
5: Distribuição da média mensal do número de prematuros por ano. Tangará da Serra, 2000 a 2010..	30
6: Distribuição mensal do número de focos de queimadas, Tangará da Serra, 2000 a 2010.	31
7: Gráficos das funções de autocorrelação e autocorrelação parcial para a série de prematuridade.	32
8: Gráfico de linha da média mensal do número de focos de queimadas por ano.	33
9: Distribuição mensal do número de focos de queimadas. Tangará da Serra, 2000 a 2010.	34
10: Gráficos das funções de autocorrelação e autocorrelação parcial para a série de focos de queimadas..	35
11: Gráfico de linha comparativo entre as séries mensais de Nascimentos prematuros e Número de queimadas.Tangará da Serra, 2000 a 2010.	36
12: Gráfico de dispersão das variáveis correlacionadas.	36
13: Autocorelação residual do modelo de regressão de Poisson simples.	37
14: Gráfico de ajuste de tendência para a série prematuridade.	38

15: Quadro comparativo dos ajustes do modelo para a variável foco de queimada de acordo com diferentes graus de liberdade	40
16: Análises dos resíduos para o modelo final escolhido.	44
17: Gráfico de autocorrelação modelo final escolhido.	44
18: Gráfico das distancias de Cook.	45

Lista de Tabelas

1: Média das temperaturas mínimas e máximas por mês no estado do Mato Grosso no ano de 2002.	15
2: Percentual de nascidos vivos com duração da gestação inferior a 37 semanas, Brasil e Regiões, 2000 a 2011.	18
3: Autocorrelação e Teste de Box-Ljung para a variável Prematuridade.	32
4: Passo-a-passo do ajuste da Tendência.	38
5: Efeitos estimados para cada mês com Janeiro como mês de referência.	39
6: Estatísticas referentes ao ajuste do efeito do número de focos de queimadas.	42

1 Introdução

Modelos estatísticos são frequentemente usados para o resumo e interpretação dos dados. O Modelo Aditivo Generalizado (MAG) é uma técnica eficaz para a modelagem não linear de séries temporais. É uma extensão do modelo linear generalizado (MLG) que baseia-se em funções não paramétricas, denominadas curvas de alisamento, em que a forma da associação é definida pelos próprios dados (Hastie & Tibshirani, 1990). As funções não paramétricas podem ser qualquer uma dentre aquelas que fizerem parte da família exponencial. Assim, o MAG é estatisticamente mais sofisticado, pois permite ajustar os dados de acordo com a forma da sua distribuição. Isso permite a construção de modelos aditivos para variáveis com distribuição Normal, Poisson ou Binomial (Guisan et al. 2002; Lehmann et al. 2002; Leathwick et al. 2006).

Desde a década de 90, o MAG tem sido muito utilizado na análise de dados de pesquisas médicas, especialmente em modelos de regressão logística binários através do seu potencial como método preditivo eficiente de modelagem. Um estudo em botânica usando espécies de planta do gênero *Inga* mostra a facilidade de execução dessa estratégia de modelagem utilizando a metodologia GRASP (predição espacial por regressão generalizada) para determinar a distribuição das espécies. (De Marco Junior, 2007). Outro exemplo foi o estudo da relação entre estruturas de geologia, como elementos químicos coletados em águas superficiais, com a ocorrência de doenças. Neste contexto, o estudo investigou a associação entre dados geoquímicos de águas superficiais em microbacias do Estado do Paraná, coletados pela MINEROPAR, e o registro de casos de câncer de fígado nas microrregiões do Paraná, obtido do DATASUS no ano de 2004. A análise foi feita através de uma análise exploratória, na qual sugeriu possíveis transformações nas covariáveis e obtém, por exemplo, mapas suavizados, indicando as regiões de maior ou menor incidência de casos da doença (Picharski et al., 2004).

A qualidade do ar atmosférico é comprometida por diversos fatores, dentre eles os eventos climáticos e a poluição ambiental, que pela queima de combustíveis fósseis, desmatamentos e queimadas, podem afetar tanto a saúde humana quanto o meio ambiente. Sendo assim, em especial, as emissões de material particulado e monóxido de carbono originado pelos focos de calor, podem contribuir para a má qualidade do ar, influenciando na dinâmica da incidência dos agravos respiratórios, internações hospitalares e até mesmo na

saúde de bebês ao nascer (Botelho C, Correa AL, Silva AMC, 2003)

Em estudos onde o objetivo é avaliar a relação entre a poluição atmosférica sobre alguma particularidade ligada a saúde, a regressão linear simples não é indicada, pois existem diversos fatores que podem influenciar no ajuste do modelo, como o efeito do tempo nas variáveis observadas. O grande desafio é encontrar um modelo estatístico capaz de considerar todos os fatores envolvidos, pois é necessário levar em consideração que cada pessoa reage de forma diferente a uma determinada concentração de poluente (Tadano et al., 2009).

A variável resposta do modelo aditivo generalizado geralmente é uma série de eventos ligados aos danos a saúde, por exemplo mortalidade, prematuridade, internações, problemas respiratórios ou qualquer outra contagem ou ocorrência de determinado evento. Quanto à prematuridade, que é a variável de desfecho deste trabalho de conclusão de curso, observam-se vários estudos utilizando MAG ligados a saúde materna, para verificar relações que indiquem e possam explicar os nascimentos antes do período previsto. Peso da mãe, histórico de doenças crônicas, uso de tabaco e álcool e estresse são variáveis potencialmente significantes sobre a saúde fetal e também sobre o período de nascimento. Porém, em relação às condições externas, como por exemplo, a poluição atmosférica, ainda existem poucos estudos que visam a avaliar a ligação entre a quantidade de elementos químicos poluentes ao meio ambiente ingeridos pela mãe no período de gravidez e ao longo da vida, com o nascimento prematuro de seus filhos. Assim, uma vez que medir os níveis de poluentes na atmosfera exige aparelhos sofisticados, os quais não estão disponíveis para este trabalho, o objetivo principal deste estudo é utilizar a variável “número de focos de queimadas” como variável *proxy*¹ dos níveis de poluentes na atmosfera, em uma região do Brasil, e avaliar sua associação com o nascimento de bebês prematuros através dos MAGs.

¹ *proxy*: substitui e explica, aproximadamente, o comportamento da outra variável que tem-se interesse em analisar.

2 Objetivos

2.1 Objetivo Geral

Explorar o Modelo Aditivo Generalizado (MAG) utilizando as funções não paramétricas de suavização chamadas *splines*, para verificar a associação entre os níveis de poluição atmosférica, medidos pelo número de focos de queimadas, e o registro de nascimentos de bebês prematuros no município de Tangará da Serra, no estado do Mato Grosso.

2.2 Objetivos Específicos

- Analisar as séries temporais mensais do número de nascimentos de bebês prematuros e o número mensal de focos de queimadas na última década;
- Implementar o MAG (utilizando a técnica de suavização por *splines*) para descrever a associação entre a variável resposta (nascimento de bebês prematuros) com a variável preditora ou independente do modelo (número de focos de queimadas).

3 Materiais e Métodos

3.1 Área de Estudo

Tangará da Serra é um município brasileiro localizado na Região Centro-Oeste do Brasil, no estado de Mato Grosso. O município é o quinto mais populoso do estado, com população de 102 056 habitantes conforme a estimativa do IBGE, em 2014. Sua área é de 11.323,640 km² e a distância até Cuiabá, capital administrativa estadual, é de 240 quilômetros.

Sua economia baseia-se na prestação de serviços, agroindústria e agricultura, com destaque para a produção de soja e cana-de-açúcar. O comércio é considerado um dos mais estruturados no interior de Mato Grosso. O município é um polo regional, sendo uma das cidades mais ricas do estado, possuindo também diversos atrativos turísticos em seu interior, como cachoeiras, pousadas e parques (<http://www.tangaradaserra.mt.gov.br/>, último acesso em 20/07/2015).

O clima do município é o tropical chuvoso quente e úmido, dividido em dois períodos bem definidos: chuvas entre setembro e abril, e estiagem entre maio e agosto. Com temperaturas médias entre 16 e 36 graus, variando de acordo com a época do ano, é considerada uma cidade quente. No inverno, Tangará da Serra costuma ficar entre as cidades que registram as menores temperaturas no estado (Tabela 1).

Tabela 1: Média das temperaturas mínimas e máximas por mês no estado do Mato Grosso no ano de 2002

	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Ano
Máxima temperatura média °C	32,1	31,7	31,7	33,6	32,1	30,9	31,7	33,4	32,7	32,9	33,7	33,2	32,4
Mínima temperatura média °C	21,3	22	20,9	20,6	18,6	15,9	15,7	17,5	19,3	21,2	21,8	21,3	19,6
Precipitação média mm³	312,4	271,6	231,2	148,7	47,4	18,2	11,1	12,7	59,1	138,6	196,7	278	1448,2

Fonte: <http://www.tangaradaserra.mt.gov.br/>

3.2 Período de Estudo

Devido a disponibilidade dos dados sobre o número mensal de nascimentos de bebês e o número de focos de queimadas, o período de estudo foi de janeiro de 2000 a dezembro de 2010 (132 meses).

3.3 Focos de Queimadas

Segundo o site <http://basilio.fundaj.gov.br> (último acesso em 28/02/2016), a queimada “é um processo de queima de biomassa que pode ocorrer por razões naturais ou ser provocada pelo homem. Sua evolução passa pelos estágios de ignição, chamas, brasas e extinção. A ignição depende do material a ser queimado (biomassa) e de fatores ambientais como temperatura, umidade relativa do ar e vento”.

O monitoramento dos focos de queimadas no Brasil é realizado pelo Instituto Nacional de Pesquisas Espaciais (INPE), utilizando imagens de satélites (neste trabalho foi utilizado o satélite de referência NOAA) que geram um conjunto de imagens ao longo do dia. Estas imagens são a principal ferramenta para apoiar os órgãos ambientais e ao governo na identificação dos focos de queimadas e poluição atmosférica, minimizando assim danos

sociais, econômicos e ambientais. O INPE faz o monitoramento e disponibiliza as informações em sua página de internet (<http://www.dpi.inpe.br/proarco/bdqueimadas/>, último acesso em 16/07/2015)

Neste sentido, a variável, utilizada neste trabalho, que mede a exposição à poluição atmosférica é o número de focos de queimadas, registrados mensalmente pelo satélite NOAA no período de 2000 a 2010.

3.4 Prematuridade

O bebê nascido vivo prematuro é aquele com idade gestacional igual ou inferior a 36 semanas e 6 dias, que, depois da expulsão ou da extração completa do corpo materno, manifesta algum sinal vital (ANS – Agência Nacional de Saúde Suplementar). Como o estudo tem como variável resposta uma contagem, ou seja, o número de nascimento de bebês prematuros em cada mês é possível considerar que esta variável tem uma distribuição de Poisson.

As informações sobre os nascimentos prematuros tanto no Brasil, como na região Centro-Oeste (Tangará da Serra) foram retirados do SINASC – Sistema de Informações sobre Nascidos Vivos, contendo informações do ano 2000 a 2011.

A variável do SINASC que foi usada para determinar o número de prematuros no mês foi “Gestação”. Esta variável originalmente é categórica, onde cada categoria representa um período da gestação, em semanas. Por exemplo: 9: Ignorado; 1: Menos de 22 semanas; 2: 22 a 27 semanas; 3: 28 a 31 semanas; 4: 32 a 36 semanas; 5: 37 a 41 semanas; 6: 42 semanas e mais. Para este trabalho as categorias foram agrupadas da seguinte forma:

- Os dados agregados das categorias 1, 2, 3 e 4 se referem a prematuridade, ou seja, nascimentos com menos de 37 semanas.
- Enquanto que as categorias 5 e 6 se referem aos nascimentos a termo.

Por fim, as informações foram agregadas por mês. Sendo assim, a variável resposta prematuridade é definida como:

$$Y_{it} = \text{Número de bebês prematuros no mês } i, \text{ no ano } t.$$

Observa-se um aumento na prematuridade registrada no SINASC em todas as regiões brasileiras (Tabela 2). Tudo indica que este aumento represente um ajuste na qualidade da informação, pois tradicionalmente as pesquisas nacionais vêm apontando que os dados do SINASC, até 2010, subestimam a prematuridade no país (SINASC, 2011)

Tabela 2: Percentual de nascidos vivos com duração da gestação inferior a 37 semanas, Brasil e Regiões, 2000 a 2011.

R. Ocorrência	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
R. Norte	6,3	5,9	5,4	5,2	5,1	5,1	5,1	4,5	5,1	5,2	5,5	9,9
R. Nordeste	6,2	5,6	5,4	5,2	5,5	5,6	5,6	5,4	5,7	5,8	6,0	10,5
R. Sudeste	7,0	6,8	6,9	7,1	7,4	7,4	7,5	7,7	7,9	8,2	8,2	9,4
R. Sul	7,0	6,8	6,9	7,2	7,3	7,3	7,6	7,7	8,0	7,8	8,1	9,3
R. C-Oeste	6,0	6,7	7,1	6,7	6,5	6,6	6,3	6,5	6,6	7,1	6,8	10,0
Brasil	6,7	6,3	6,3	6,3	6,5	6,6	6,6	6,6	6,9	7,0	7,1	9,8

Com base nas informações apresentadas na Tabela 2, decidiu-se estudar a série temporal dos dados de prematuridade no período de 2000 a 2010.

3.5 Séries Temporais

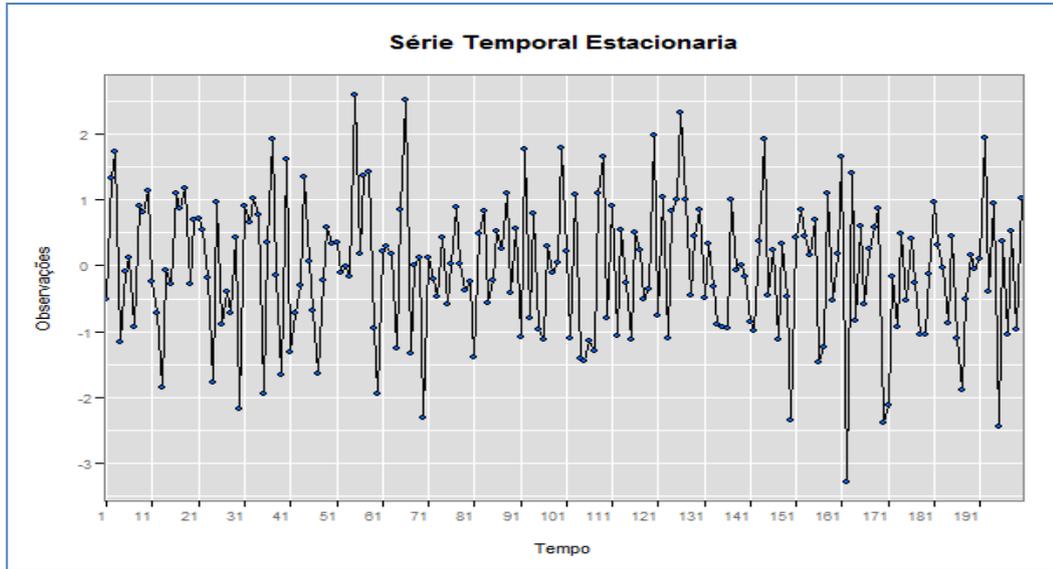
Neste trabalho, os conceitos e técnicas da análise de séries temporais foram usados para descrever o comportamento das séries, a partir de gráficos e estatísticas, com o objetivo de verificar a existência de tendências, ciclos e variações sazonais, assim como periodicidades relevantes nos dados (Morettin et al., 1981).

Uma série temporal é qualquer conjunto de observações ordenadas no tempo. O banco de dados deste trabalho inclui o número mensal de focos de queimadas e o número mensal de nascimentos de bebês prematuros ao longo de 10 anos (2000 a 2010).

Uma das suposições mais frequentes que se faz a respeito de uma série temporal é a de que ela é estacionária, ou seja, ela se desenvolve no tempo aleatoriamente ao redor de uma média constante, refletindo alguma forma de equilíbrio estável. Todavia, as maiores partes das

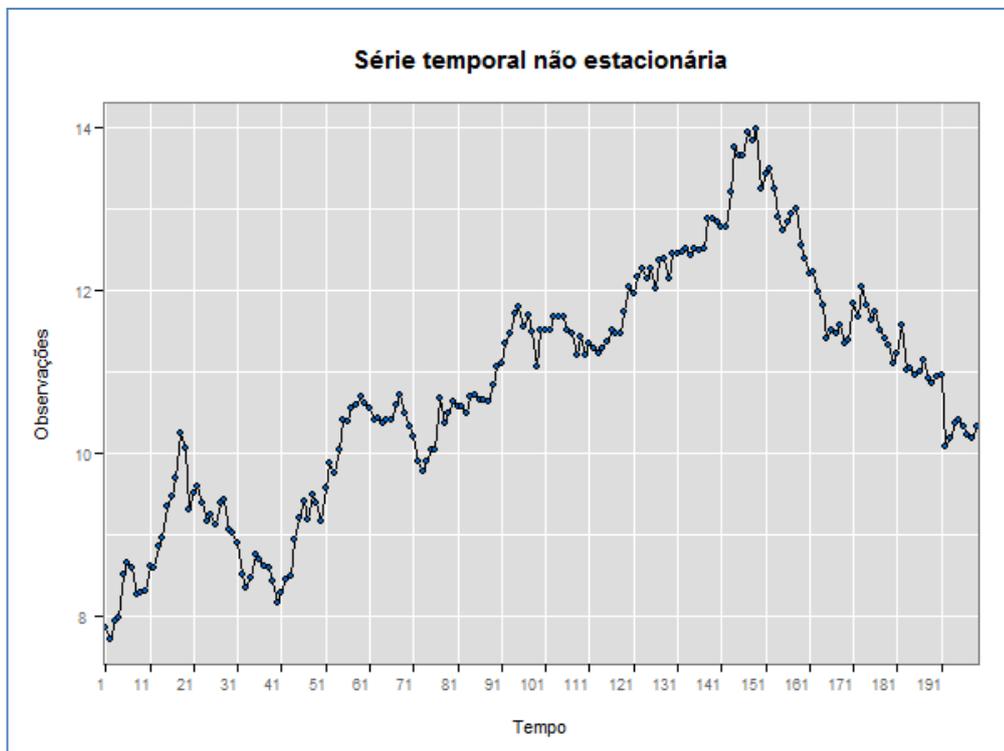
séries que encontramos, na prática apresentam alguma forma de não estacionariedade, por exemplo, tendências.

Uma série pode ser estacionária durante um período muito longo, mas pode ser estacionária apenas em períodos muito curtos, mudando de nível e/ou inclinação (Morettin; Tolo, 2006). Exemplos são apresentados nas Figuras 1 e 2.



Fonte: Morettin; Tolo, 2006

Figura 1: Exemplo de Série Temporal Estacionária



Fonte: Morettin; Tolo, 2006

Figura 2: Exemplo de Série Temporal Não Estacionária

A tendência refere-se ao movimento gradual de longo prazo da série e o cálculo de sua estimativa é realizado pela identificação de uma equação que descreva este movimento. A plotagem dos dados passados permite a identificação desta equação, que pode ser linear ou não (exponencial, parabólica, logarítmica, etc.) (Tubino, 2000). Para tratar previsões de séries que apresentam tendência existem 2 técnicas mais empregadas, sendo uma baseada em equação linear e outra a partir de suavização não paramétrica.

O primeiro passo na análise de uma série temporal é a construção de seu gráfico, que revelará características importantes, como tendência, sazonalidade, variabilidade, “outliers”, etc. Além dessa inspeção gráfica, é possível utilizar testes de hipóteses estatísticos para verificar se existe tendência na série. Isto pode ser feito de duas maneiras:

- a) antes da estimação da Tendência.
- b) depois que se obtém uma estimativa de Tendência.

No segundo caso é possível efetuar testes formais somente no caso do ajuste polinomial, pois aqui se tem uma teoria desenvolvida para os estimadores de mínimos quadrados. Desta maneira, podemos obter intervalos de confiança para os parâmetros do polinômio, bem como testar hipóteses a respeito destes parâmetros.

Assim, podemos testar a hipótese que $\beta_1 = 0$, ou seja, que não existe tendência, contra a alternativa que $\beta_1 > 0$, ou seja, existe uma tendência crescente.

Existem alguns testes não paramétricos que são úteis para se verificar se há tendência em um conjunto de observações. Contudo, em geral, se baseiam em hipóteses que podem não estar verificadas para o caso de uma série temporal. Uma suposição comum é que as observações constituem uma amostra de uma população e assim elas serão independentes. Deste modo, estes testes devem ser utilizados com cautela e, em geral, são pouco poderosos para detectar alternativas de interesse.

Neste trabalho, verificaremos indicações de tendência a partir dos gráficos da série original e usaremos as *splines* para ajustar a tendência nos modelos de regressão.

Dizemos que uma série temporal possui sazonalidade estocástica quando a componente sazonal da série varia com o tempo. Este procedimento pode ser utilizado e normalmente é utilizado quando temos um padrão sazonal constante (Morettin et al., 1981).

Dado a série temporal Z_t , seja T_t um estimador para a tendência calculado previamente, consideramos a seguinte série temporal:

$$Y_t = Z_t - T_t \tag{1}$$

Assim, considerando o caso em que temos um padrão sazonal constante, utilizaremos Y_t para estimar S_t . Inicialmente, considerando os dados fornecidos anualmente, tomamos a média dos meses:

$$Y_j = \frac{1}{n_j} \sum_{i=1}^{n_i} Y_{ij} , \quad j = 1, \dots, 12 ; \quad (2)$$

em geral, a soma dos Y_j não é zero e, portanto, tomamos como estimativas das constantes sazonais

$$S_j = Y_j - Y , \quad \text{onde} \quad Y = \frac{1}{12} \sum_{j=1}^{12} Y_j \quad (3)$$

Assim, o modelo da série original Z_t pode ser escrito na forma aditiva por

$$Z_t = T_t + S_j + a_t \quad (4)$$

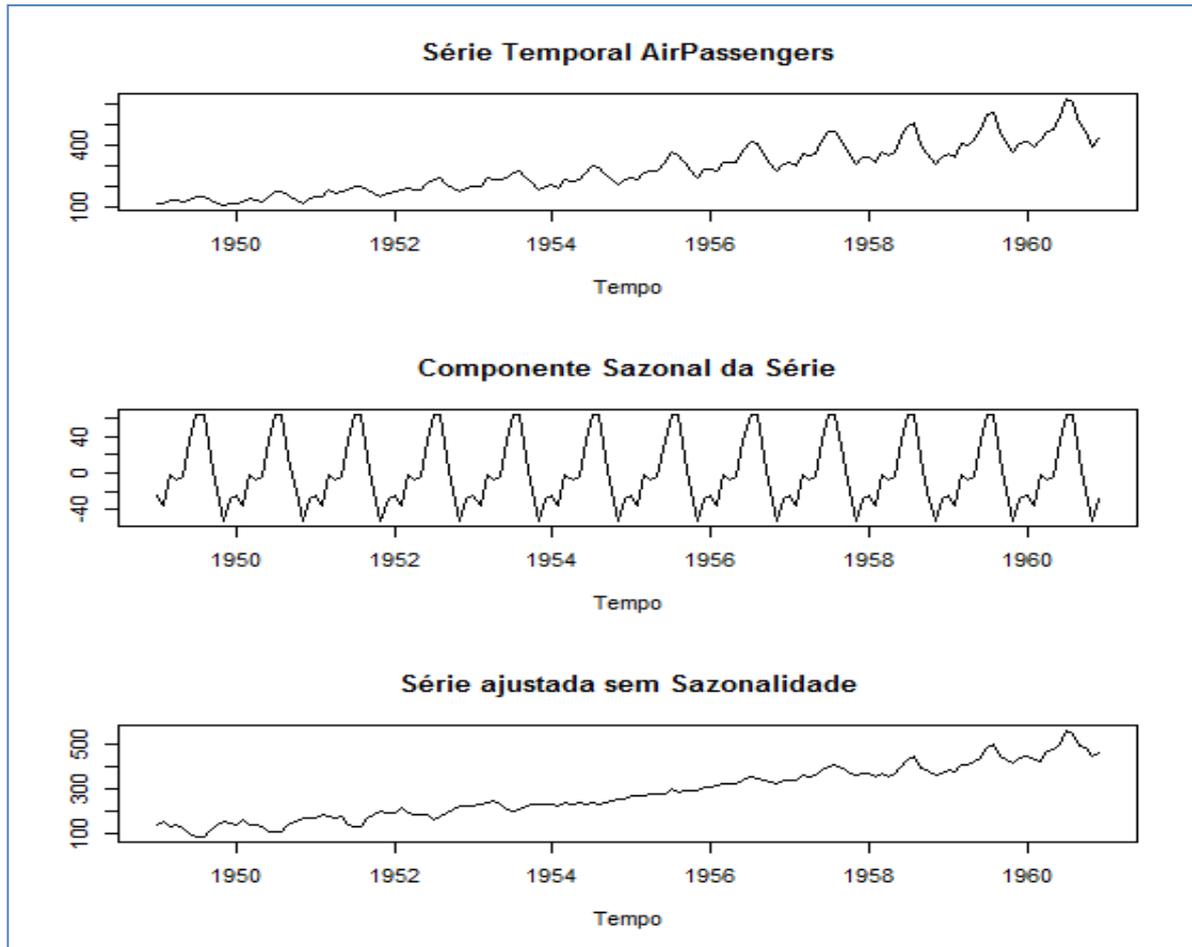
com $t = 12i + j$, $i = 0, 1, \dots, p - 1$, $j = 1, \dots, 12$, onde p é o número de anos.

Então, a série livre de sazonalidade pode ser escrita como:

$$Z_t^* = Z_t - S_t \quad (5)$$

Exemplos são apresentados na Figura 3.

Neste trabalho, verificaremos indicações de sazonalidades a partir dos gráficos da série original e usaremos a variável “Mês” como “*dummies*” nos modelos de regressão para ajustar a possível sazonalidade.



Fonte: Morettin; Toloí, 2006

Figura 3: Série temporal de passageiros de uma companhia aérea, Componente Sazonal e Série Ajustada

3.5.1 Função de autocorrelação (ACF)

A função de autocorrelação é dada por:

$$\rho_k = \frac{\gamma_k}{\gamma_0} ; k = 0,1,2, \dots \quad (6)$$

Onde:

$$\gamma_0 = var(Z_t)$$

Logo, ρ_k será uma medida padrão de dependência com $|\rho_k| \leq 1$, $k = 1,2, \dots$

O Gráfico da ACF, chamado de correlograma, é utilizado para explorar o padrão de tendência e sazonalidade dos dados.

3.5.2 Função de autocorrelação parcial (PACF)

Box & Jenkins (1970) propuseram um segundo método de análise: A Função de Auto-Correlação Parcial (PACF): ϕ_{kk}

As PACF's podem ser calculadas a partir das equações de Yule-Walke:

$$\begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & & & \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} \quad (7)$$

Esta análise é realizada em conjunto com a função de Autocorrelação para identificar os componentes de um modelo ARIMA.

PACF para modelos AR, MA e ARMA.

- Um processo AR(p) tem pacf $\phi_{kk} \neq 0$ para k menor ou igual a p, e $\phi_{kk} = 0$ para k maior que p;
- Um processo MA(q) tem pacf que se comportam de maneira similar às pacf de um processo AR(p): são dominadas por exponenciais e/ou senoides amortecidas;
- Um processo ARMA (p,q) tem pacf que se comportam como a pacf de um processo MA puro.

3.6 Modelo Aditivo Generalizado

O modelo aditivo generalizado é uma extensão do modelo linear generalizado proposto por Hastie & Tibshirani (1990), em que o termo $x_i^T \beta = \sum_{j=0}^p \beta_j x_{ij}$ é substituído por $\sum f_j(x_{ij})$, com $f_j(x_{ij})$ denotando uma função não paramétrica (por exemplo, cuja forma não é especificada) estimada através de curvas de alisamento. Com essa substituição, não é necessário assumir uma relação linear entre $g(\mu_i)$ e as variáveis explicativas, como no MLG (Dobson, 2002). De fato, não é necessário nem mesmo conhecer previamente a forma dessa relação, mas é possível estimá-la a partir de um conjunto de dados. Essa função estimada $\hat{f}_j(x_{ij})$, também chamada de curva alisada, em muitas situações, nada mais é do que algum tipo de média dos valores Y_i na vizinhança de um dado valor x_i (Conceição et al., 2001).

A curva alisada permite então descrever a forma, e também revelar possíveis não linearidades nas relações estudadas, uma vez que não apresenta a estrutura rígida de uma função paramétrica.

Os procedimentos de estimação para os MAG's são semelhantes àqueles adotados na estimação sob os MLG's, valendo-se basicamente de uma combinação do algoritmo "Scoring" de Fisher com métodos de alisamento (Conceição et al., 2001)

$$g(\mu_i) = \sum f_j(x_{ij}) \quad (8)$$

$f(x_{ij})$ é uma função não paramétrica (cuja forma não é especificada)

$f_j(x_{ij})$ é estimada através de curvas de alisamento (ex: médias móveis, *splines*)

Neste trabalho, em particular, as funções não paramétricas utilizadas para a suavização serão *splines*

3.7 Função estimada de suavização: *Splines*

A utilização do alisamento *spline* polinomial nos modelos não paramétricos generalizados foi investigado em vários contextos. Stone (1986) verificou pela primeira vez a taxa de convergência das *splines* polinomiais estimadas para o modelo aditivo generalizado. Stone (1994) e Huang (1998) propuseram uma estimação voltada para a estimativa de *spline* polinomial utilizando o modelo da ANOVA generalizada, enquanto Huang et al. (2000) e Huang e Liu (2006) consideraram o modelo da ANOVA generalizada e o modelo de índice único. A suavização polinomial *spline* é um método de alisamento global, que aproxima as funções desconhecidas via *splines* polinomiais caracterizadas por apenas um número finito de coeficientes.

Splines é uma classe de funções flexíveis que são adequadas para modelar as medições longitudinais. Um número de pontos chamados *nós* são definidos no intervalo onde a função *spline* é definida; entre cada par de nós adjacentes a função *spline* é um polinômio de grau p , enquanto estes polinômios ficam unidos uns aos outros de uma forma suave, e que as suas funções genéricas de ordem $p-1$, serão contínuas também nos *nós*. Para $p = 1$ isto leva de novo a funções lineares definidas em partes (piecewise linear functions), mas para $p > 1$ aparecem funções que são lisas, mas que não têm a aparência de funções lineares por partes.

Um exemplo de uma *spline* quadrática (de ordem $p = 2$) foi dado no Capítulo 8 do livro Multilevel Analysis (Snijders e Bosker ,1999) pelas equações:

$$\begin{cases} 2.233x + 0.0246x^2 & \text{se } x < 0 \\ 2.223x - 0.306x^2 & \text{se } x \geq 0 \end{cases} \quad (9)$$

e a figura 4. Esta equação e figura representam uma função de x , que é quadrática para $x < 0$ e também para $x > 0$, mas que tem diferentes coeficientes, tais que a função e a sua derivada são contínuas também no *nó*, como pode ser visto a partir da figura 4. *Splines* cúbicas ($p = 3$) também são frequentemente utilizadas.

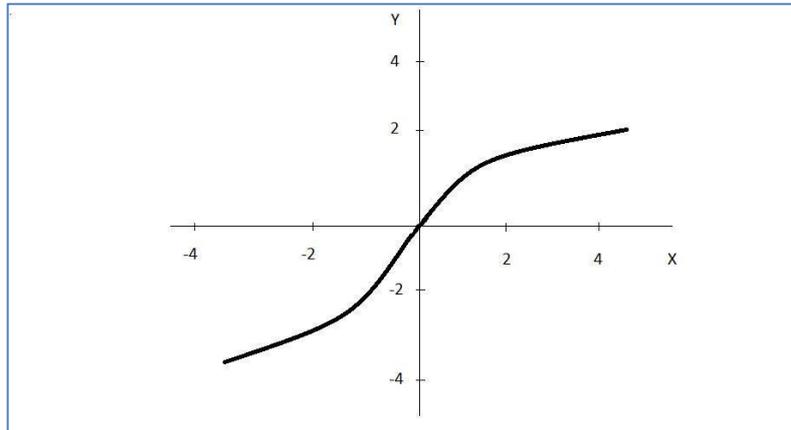


Figura 4: Efeito de X sobre a variável Y

Suponha que alguém esteja investigando o desenvolvimento de algumas características sobre a idade de 12 a 17 anos. Dentro de cada um dos intervalos de 12 a 15 anos e 15 a 17 anos, as curvas de desenvolvimento podem ser aproximadamente quadráticas (isso pode ser verificado por uma análise de tendência polinomial para os dados relativos a estes intervalos separadamente). No entanto, se considerar todo o intervalo de 12 a 17 anos as curvas suaves sobre o todo intervalo não serão quadráticas. Em tal caso, valeria a pena testar uma *spline* quadrática ($p = 2$) com um nó, em 15 anos. Definindo $t_1 = 15$, as funções básicas podem ser tomadas como:

$$f_1(t) = t \quad (\text{função linear}) \quad (10)$$

$$f_2(t) = \begin{cases} (t - t_1)^2, & (t \leq t_1) \\ 0 & , (t > t_1) \end{cases} \quad (\text{função quadrática à esquerda de } t_1) \quad (11)$$

$$f_3(t) = \begin{cases} 0 & , (t \leq t_1) \\ (t - t_1)^2 & , (t > t_1) \end{cases} \quad (\text{função quadrática à direita de } t_1) \quad (12)$$

As funções f_1 e f_2 são funções quadráticas à esquerda de t_1 e à direita de t_2 , e são contínuas com derivadas contínuas. Ambas também são funções contínuas e têm derivadas contínuas quando o nó $t = t_1$.

As funções de desenvolvimento individuais são modeladas como: se $\beta_{2i} = \beta_{3i}$, a

curva para o indivíduo é exatamente quadrática. A liberdade de ter estes dois coeficientes que diferem uns dos outros permite representar funções que são muito diferentes das funções quadráticas, por exemplo, se estes coeficientes têm sinais opostos, a função será côncava de um lado de t_1 e convexa no outro lado. A equação (9) e a Figura 4 fornecem exatamente um exemplo desta função.

A Fórmula $F_i(t) = \beta_{0i} + \beta_{1i}f_1(t) + \beta_{2i}f_2(t) + \beta_{3i}f_3(t)$ indica que uma *spline* quadrática com um nó tenha mais de um parâmetro do que uma função quadrática (4 em vez de 3). Cada nó adicionado aumentará o número de parâmetros da função por um. Há uma considerável liberdade de escolha ao definir as funções básicas, que são sujeitas à restrições de que elas são quadráticas em cada intervalo entre os nós adjacentes, e são contínuas com derivadas contínuas também nos nós. Para dois nós, t_1 e t_2 , uma possível escolha é uma representação que emprega um valor de referência t_0 sendo um valor arbitrário (conveniente ou significativa). É aconselhável utilizar um t_0 dentro do intervalo de tempo de observação dos dados. As funções básicas são:

$$f_1(t) = t - t_0 \quad (\text{função linear}) \quad (13)$$

$$f_2(t) = (t - t_0)^2 \quad (\text{função linear}) \quad (14)$$

$$f_3(t) = \begin{cases} (t - t_1)^2, & (t \leq t_1) \\ 0, & (t > t_1) \end{cases} \quad (\text{função quadrática à esquerda de } t_1) \quad (15)$$

$$f_4(t) = \begin{cases} 0, & (t \leq t_2) \\ (t - t_2)^2, & (t > t_2) \end{cases} \quad (\text{função quadrática à direita de } t_2) \quad (16)$$

Estas quatro funções são utilizadas na representação a seguir:

$$F_i(t) = \beta_{0i} + \beta_{1i}f_1(t) + \beta_{2i}f_2(t) + \beta_{3i}f_3(t) + \beta_{4i}f_4(t) \quad (17)$$

onde o coeficiente β_{2i} é o coeficiente quadrático no intervalo entre t_1 e t_2 , enquanto β_{3i} e β_{4i} são as alterações no coeficiente quadrático que ocorrem quando o tempo t passa com os nós t_1 e t_2 respectivamente. O coeficiente quadrático para $t < t_1$ é $\beta_{2i} + \beta_{3i}$, e para $t > t_2$ é $\beta_{2i} + \beta_{4i}$.

Um exemplo para um ajuste das funções básicas, para *spline* cúbica ($p=3$) com um nó, tendo t_0 como de referência (ou nó), é:

$$f_1(t) = t - t_0 \quad (\text{função linear}) \quad (18)$$

$$f_2(t) = (t - t_0)^2 \quad (\text{função linear}) \quad (19)$$

$$f_3(t) = \begin{cases} (t - t_1)^3, & (t \leq t_0) \\ 0, & (t > t_0) \end{cases} \quad (\text{função quadrática à esquerda de } t_0) \quad (20)$$

$$f_4(t) = \begin{cases} 0, & (t \leq t_0) \\ (t - t_1)^3, & (t > t_0) \end{cases} \quad (\text{função quadrática à direita de } t_0) \quad (21)$$

Para mais de dois nós, e com uma ordem arbitrária p dos polinômios, as funções básicas de *spline* podem ser escolhidos da seguinte forma (nós denotados de t_1 para t_m):

$$f_k(t) = (t - t_0)^k \quad (k = 1, \dots, p) \quad (22)$$

$$f_{p+k}(t) = \begin{cases} 0, & (t \leq t_k) \\ (t - t_k)^p, & (t > t_k) \end{cases} \quad (k = 1, \dots, M) \quad (23)$$

A escolha dos nós é importante para se obter uma boa aproximação de encaixe. Uma vez que as funções *spline* são funções não lineares dos nós, a otimização formal da colocação do nó é mais complicada do que o encaixe com funções *spline* com os nós definidos pelo pesquisador por exemplo.

3.8 Critério de decisão – Akaike (AIC)

Akaike (1974) definiu seu critério de informação como:

$$AIC_p = -2 \log(L_p) + 2[(p + 1) + 1], \quad (24)$$

em que L_p é a função de máxima verossimilhança do modelo e p é o número de variáveis explicativas consideradas no modelo.

De acordo com Souza e Camargo (2004) o AIC aumenta conforme a Soma de Quadrado dos Resíduos (SQE) aumenta. Além disso, este critério penaliza o modelo com muitas variáveis sendo que valores menores de AIC são preferíveis.

O AIC foi utilizado para decidir o melhor modelo, de acordo com o ajuste das variáveis explicativas.

3.9 Diagnóstico do Modelo

Para verificar a qualidade do ajuste do modelo foram utilizados os gráficos da função de autocorrelação, resíduos padronizados e Distância de Cook. O teste Box & Ljung foi utilizado para verificar a independência entre os resíduos. (Ljung & Box, 1978).

4 Resultados

Para avaliar a relação e o comportamento das principais variáveis nesse projeto, foram calculadas algumas estatísticas descritivas para Focos de queimadas e Nascimentos prematuros no município de Tangará da Serra, no estado do Mato Grosso.

4.1 Análise Exploratória das series Temporais

4.1.1 Prematuridade

Foram observados os números de nascimentos prematuros em Tangará da Serra de 2000 a 2010, por ano e por mês. Os gráficos de linha (Figuras 5 e 6) abaixo ilustram o comportamento da variável prematuridade ao longo do tempo.

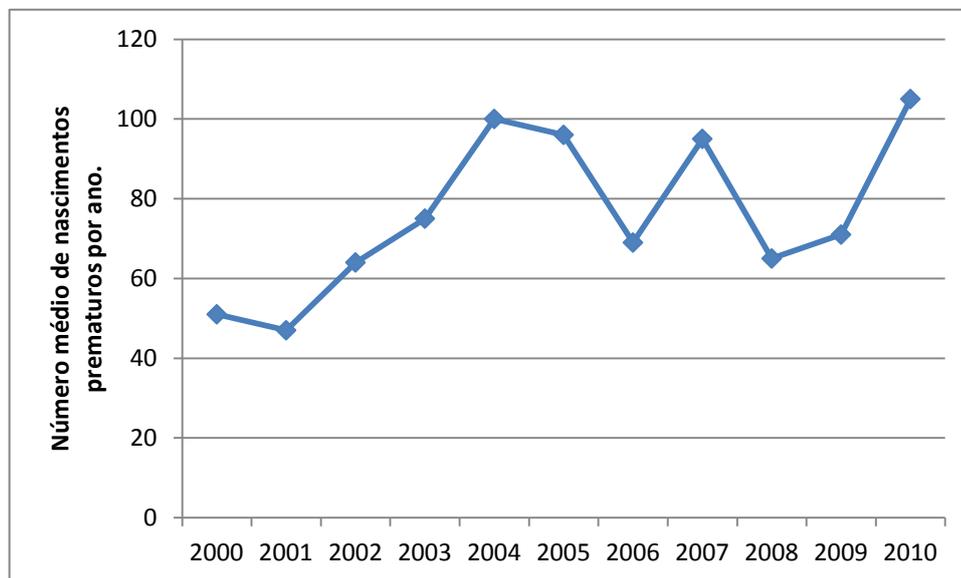


Figura 5: Distribuição da média mensal do número de prematuros por ano. Tangará da Serra, 2000 a 2010.

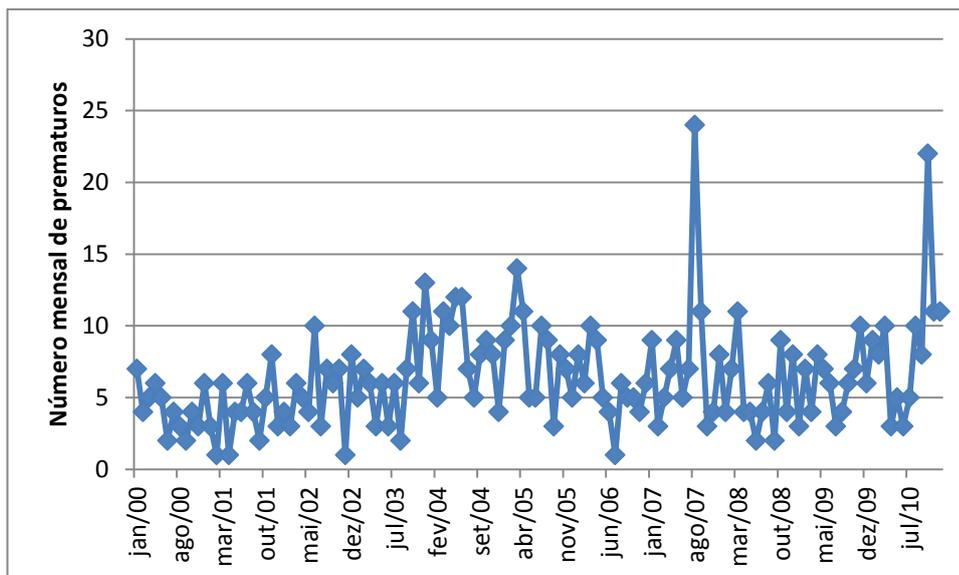


Figura 6: Distribuição mensal do número de prematuros. Tangará da Serra, 2000 a 2010.

Com base nos gráficos de linha, observa-se em geral ausência de estacionariedade e indícios de tendências nos anos de 2001 a 2004, onde os nascimentos prematuros crescem gradativamente. Em 2005 a 2010 a média do número de nascimentos prematuros apresenta um padrão de aumento e “queda” nesses anos. Em relação a variável de prematuridade mensal, observamos estacionariedade ao longo de meses que se estende por anos. Porém em agosto de 2007 e outubro de 2010, a série apresenta um salto considerável, mostrando um aumento dos nascimentos prematuros nesses meses.

Em relação a figura 7, o coeficiente de correlação começa com um valor acima do limite superior na defasagem 1, tendo a maioria dos valores dentro do intervalo, sugerindo uma série temporal estacionária.

Quanto a autocorrelação foram realizados testes nos pontos temporais da série de Prematuridade para verificar em que pontos os dados da série estariam melhor relacionados. Também foi utilizado o teste Box-Ljung para testar a hipótese de independência e distribuição dos resíduos. Ambos estão representados na tabela 3.

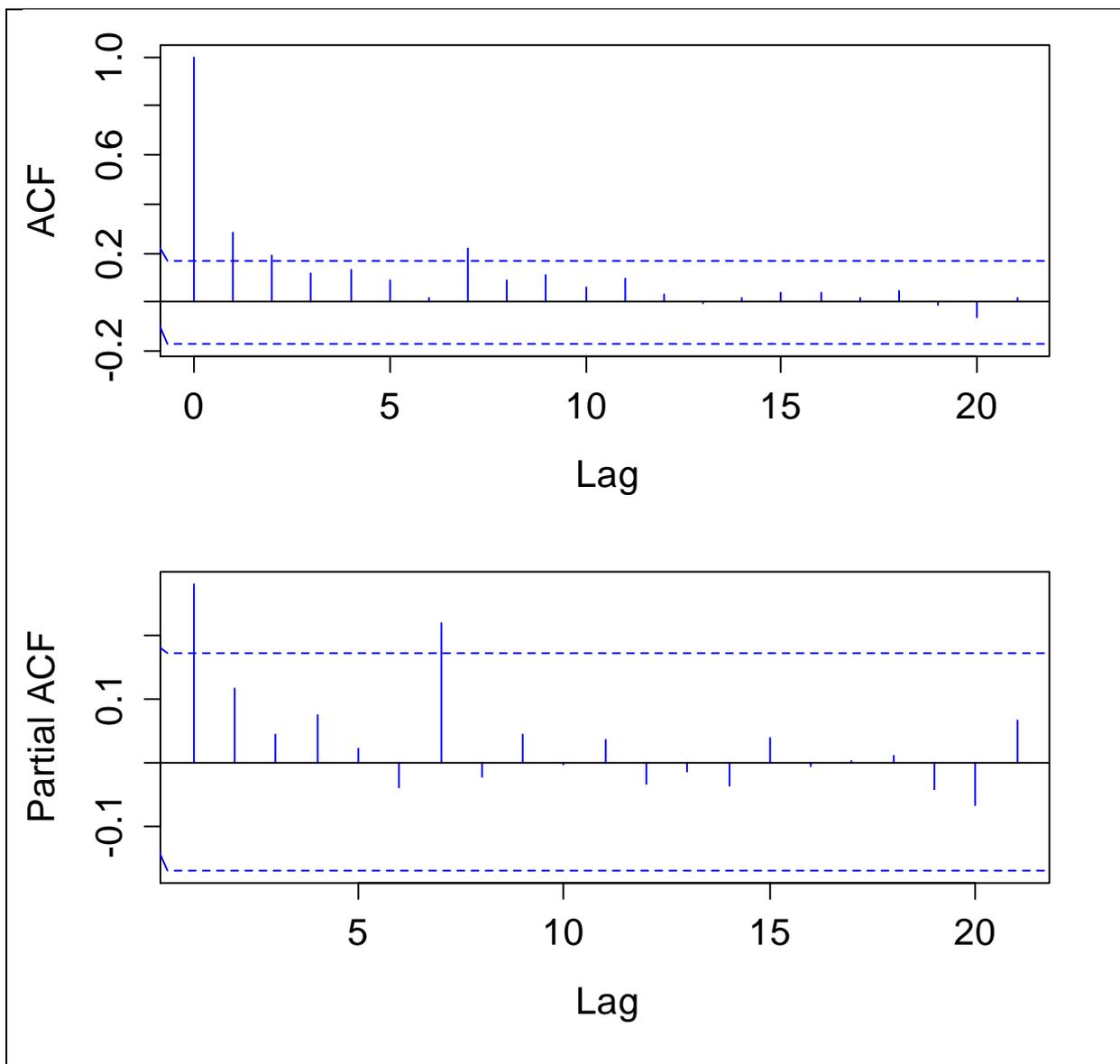


Figura 7. Gráficos das funções de autocorrelação e autocorrelação parcial para a série de prematuridade.

Tabela 3: Autocorrelação e Teste de Box-Ljung para a variável Prematuridade.

Lag	Autocorrelação	Desvio Padrão	Estatística Box-Ljung		
			Valor	gl	p-valor
1	0,281	0,086	10,681	1	0,001
2	0,187	0,086	15,441	2	0,000
3	0,119	0,085	17,380	3	0,001
4	0,130	0,085	19,710	4	0,001
5	0,088	0,085	20,795	5	0,001
6	0,020	0,084	20,852	6	0,002

Lag	Autocorrelação	Desvio Padrão	Estatística Box-Ljung		
			Valor	gl	p-valor
7	0,217	0,084	27,535	7	0,000
8	0,092	0,084	28,754	8	0,000
9	0,114	0,083	30,637	9	0,000
10	0,058	0,083	31,131	10	0,001
11	0,095	0,083	32,441	11	0,001
12	0,035	0,082	32,623	12	0,001
13	-0,002	0,082	32,623	13	0,002
14	0,015	0,082	32,656	14	0,003
15	0,039	0,081	32,882	15	0,005
16	0,039	0,081	33,110	16	0,007

4.1.2 Focos de Queimadas

Foram observados o número de focos de queimadas no estado do Mato Grosso de 2000 a 2010 e mensalmente, também dentro do período anual citado. Os gráficos de linha abaixo (Figuras 8 e 9) ilustram o comportamento da variável focos de queimadas ao longo do tempo.

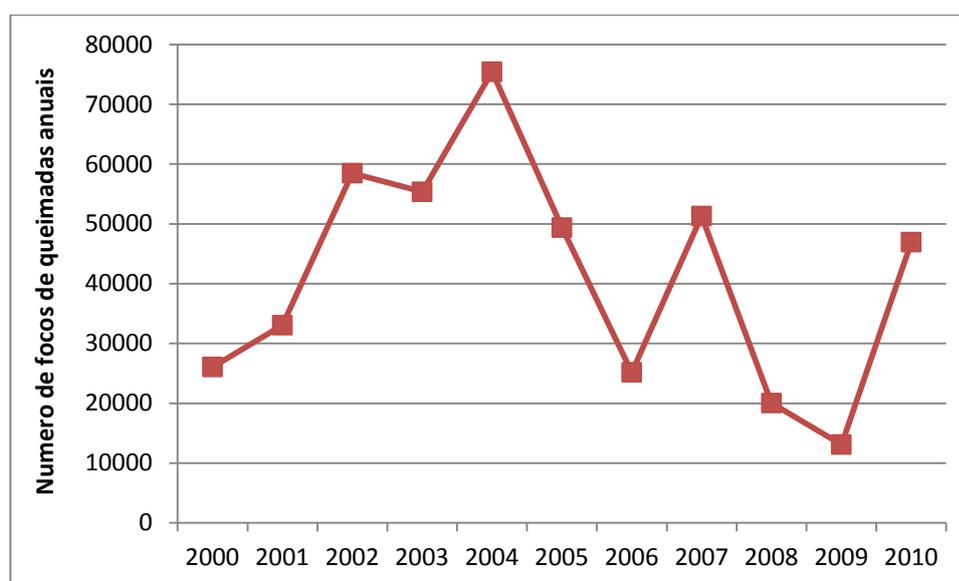


Figura 8: Gráfico de linha da média mensal do número de focos de queimadas por ano

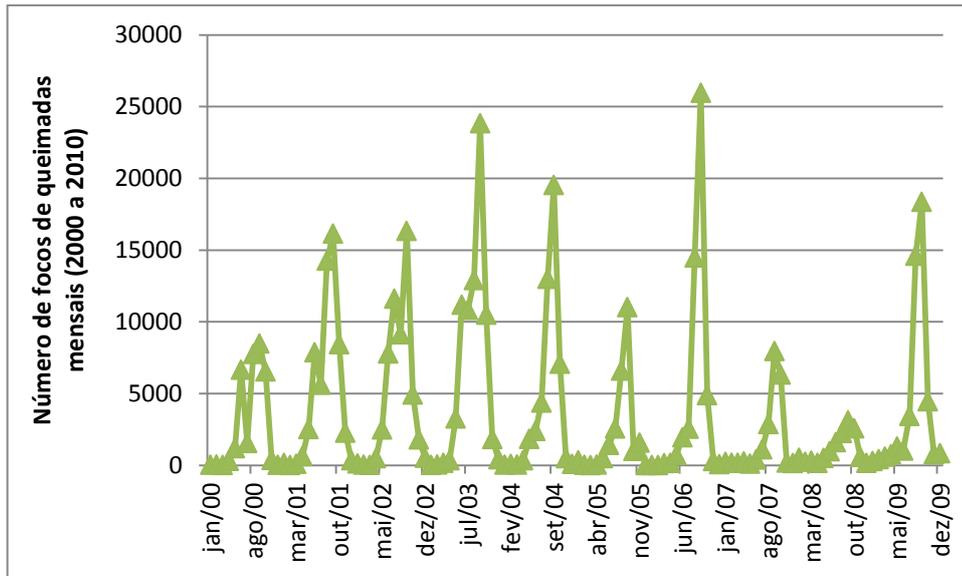


Figura 9: Distribuição mensal do número de focos de queimadas. Tangará da Serra, 2000 a 2010

Com base nos gráficos de linha, observa-se em geral uma tendência nos anos de 2000 a 2005, onde a média do número de focos de queimadas apresenta um padrão de aumento nesses anos. O pico desses focos ocorre no ano de 2004.

Em relação a mesma variável, porém observando-a mensalmente, tem-se sazonalidade ao longo de todos os meses que estende-se por todos os anos. Em outubro de 2004 e outubro de 2007, a série apresenta os maiores números de focos de queimadas, enquanto de janeiro a abril de 2000 foi o período em que o número de focos foi muito baixo. (Figura 10)

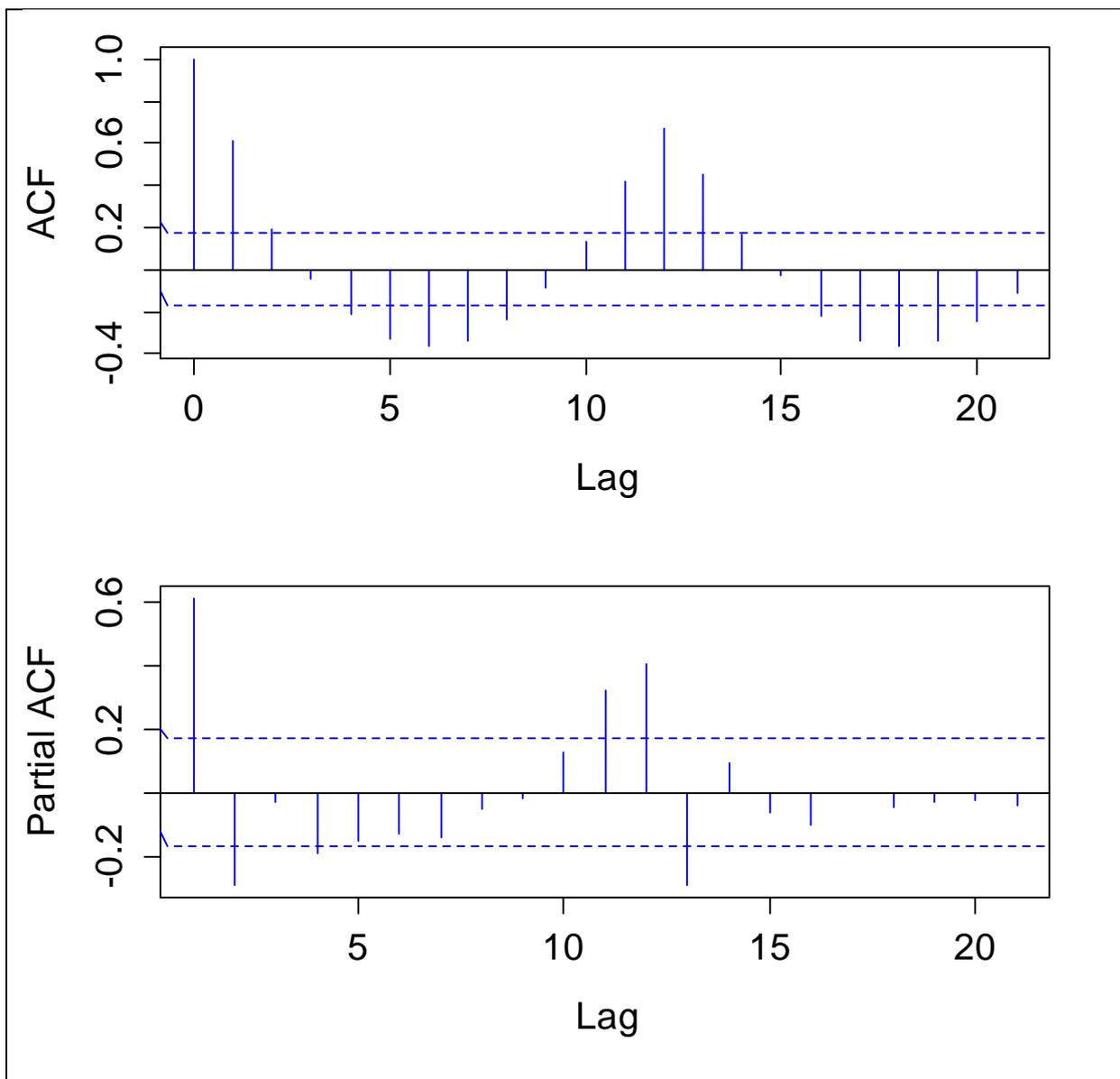


Figura 10: Gráficos das funções de autocorrelação e autocorrelação parcial para a série de focos de queimadas.

4.1.3 Prematuridade X Focos de queimadas

A figura 11 apresenta as séries do número de focos de queimadas e número de bebês prematuros. O gráfico compara o comportamento mensal das variáveis. Verifica-se que principalmente nos meses de seca (junho a Outubro) ocorrem picos nas duas variáveis, sugerindo comportamentos semelhantes.

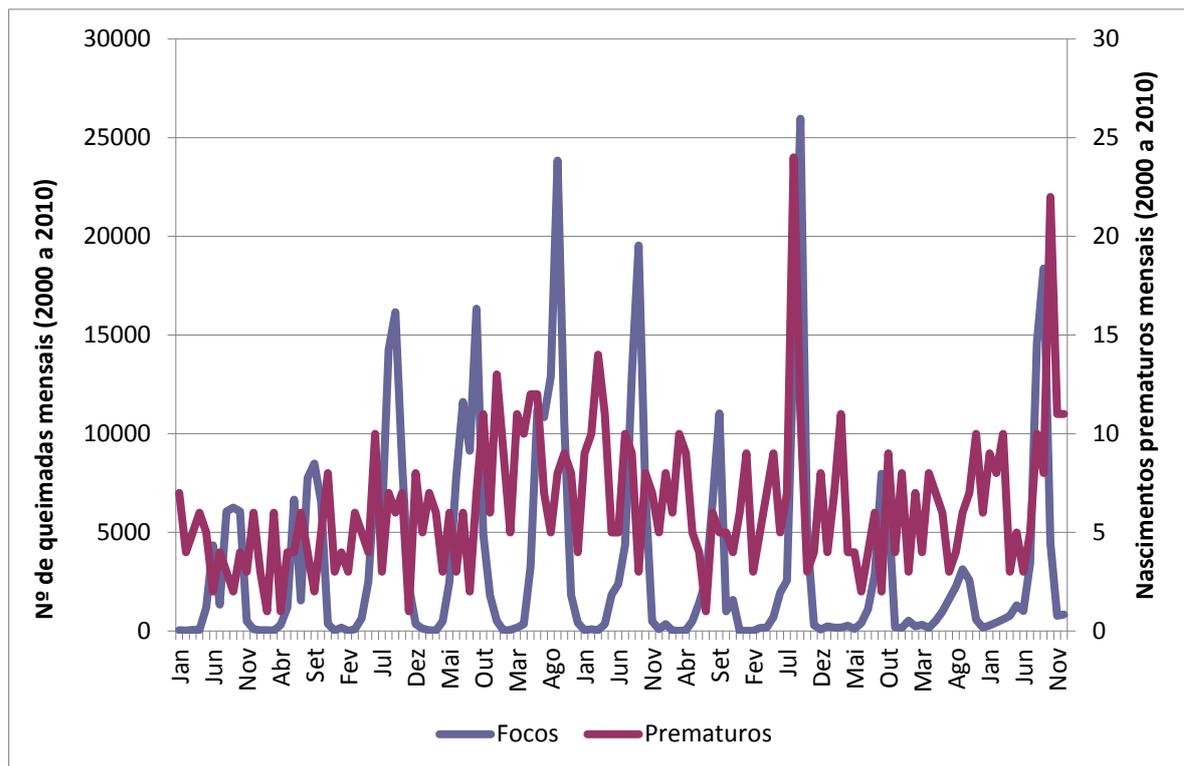


Figura 11: Gráfico de linha comparativo entre as séries mensais de Nascimentos prematuros e Número de queimadas. Tangará da Serra, 2000 a 2010.

O interesse é medir o nível de relação entre as variáveis Prematuridade e Focos de queimadas. Para isso foi utilizado o gráfico de dispersão (Figura 12). É usado para se verificar uma possível relação de causa e efeito, porém não prova que uma variável afeta a outra, mas torna claro se a relação existe e em que intensidade.

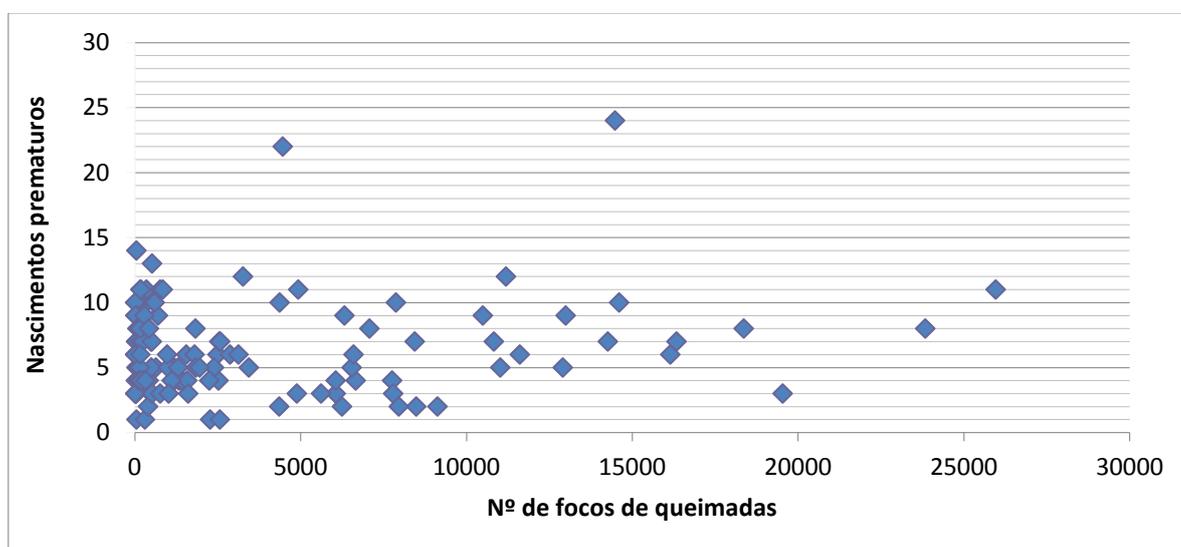


Figura 12: Gráfico de dispersão das variáveis correlacionadas

Considerando um modelo de regressão de Poisson simples, sem ajuste de tendência e sazonalidade, observa-se efeito significativo do número mensal de focos de queimadas sobre o número mensal de prematuros ($\hat{\beta} = 1.496e - 05$; p-valor = 0,0147; AIC = 700,08). No entanto, a análise dos resíduos indicou dependência entre as observações sugerindo que o ajuste do modelo não está adequado (Figura 13). Este resultado é confirmado pelo teste Box-Ljung ($\chi^2=9,69$; gl=1; p-valor=0,002). Gráficos dos resíduos são úteis em algumas situações porque fornecem uma visão do jeito como a dependência da série cai de acordo com a sua distribuição no tempo, ou seja, pelo seu 'lag' ou separação k entre pontos da série.

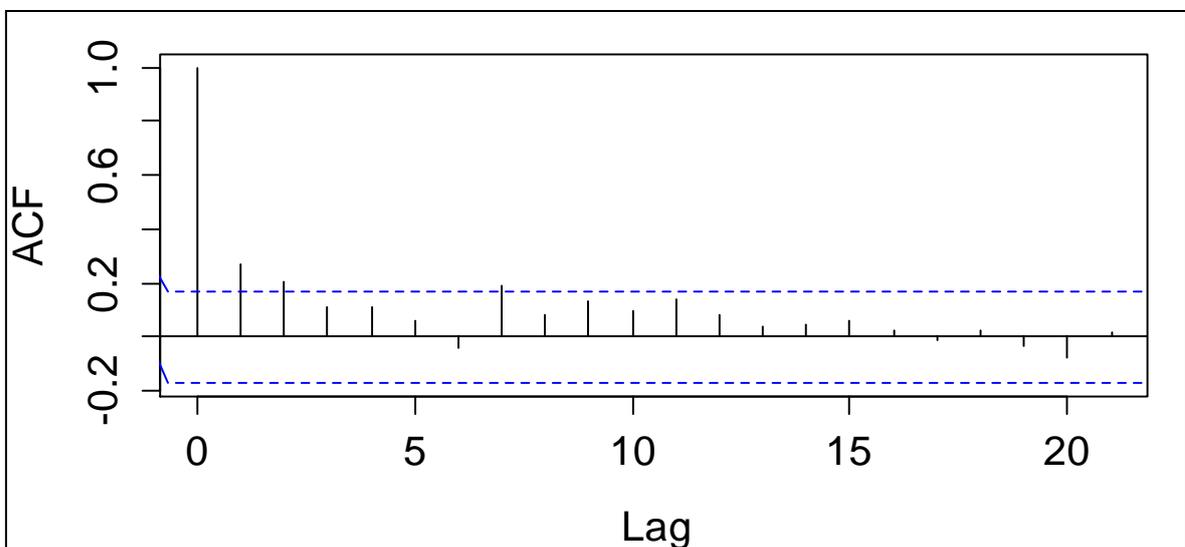


Figura 13: Autocorrelação residual do modelo de regressão de Poisson simples.

4.2 Estimação do Modelo aditivo generalizado

4.2.1 Ajuste da tendência

Para o ajuste da tendência usamos como referência os artigos de Carmo et al. (2013) e Ignotti et al. (2013), nos quais fazem uma análise de séries temporais via os MAGs.

O gráfico de ajuste de tendência da série de prematuridade (Figura 12) não sugere tendência linear. Neste caso, a quantidade de nós estipulados para o ajuste da *spline* será a partir de 1.

A tabela 4 apresenta o passo-a-passo do ajuste da tendência por *spline* cúbica. Verifica-se pelo AIC que quanto maior o número de nós melhor é o ajuste do modelo e, também, o teste Box-Ljung indicou independência entre as observações a partir de 3 nós.

Tabela 4: Passo-a-passo para o ajuste da Tendência.

AIC	GL (Spline Cúbica)	Teste Box-Ljung p-valor
683,43	2	0,03521
672,85	3	0,09048
664,96	4	0,3409
662,89	5	0,4941

Sabe-se que a escolha do modelo também deve-se considerar a parcimônia. Assim, este trabalho considerou o ajuste da tendência por uma *spline* cúbica com 3 nós, representada abaixo na figura 14.

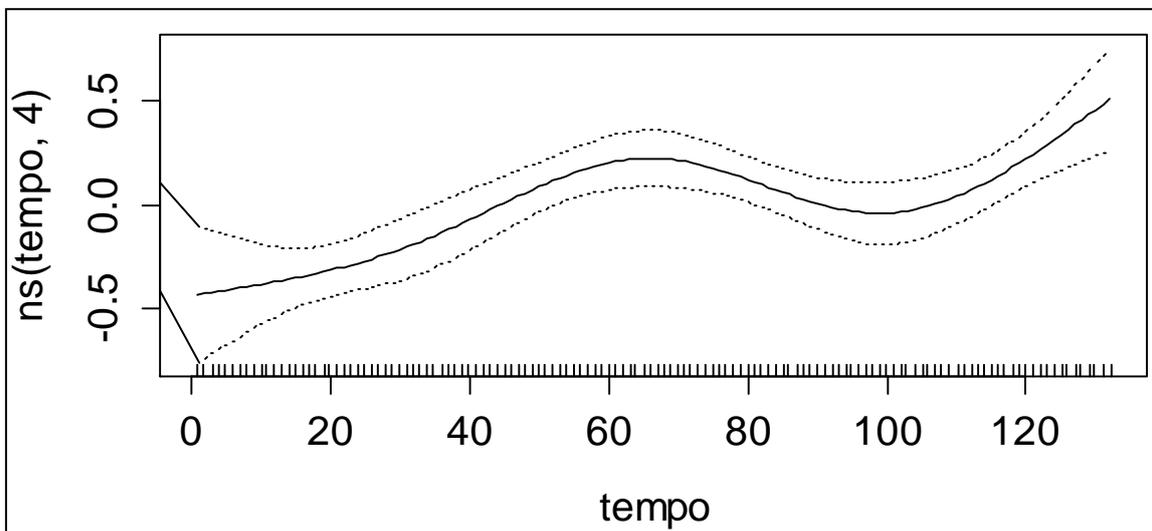


Figura 14: Gráfico de ajuste de tendência para a série prematuridade

4.2.2 Ajuste da sazonalidade

A sazonalidade foi ajustada incluindo no modelo aditivo generalizado a variável indicadora do mês. Os coeficientes foram estimados para cada mês, uma vez que a variável foi transformada em fator, sendo janeiro o mês de referência. Os efeitos estimados de cada mês são apresentados na tabela 5. Verificaram-se efeitos significativos para os meses de junho a setembro.

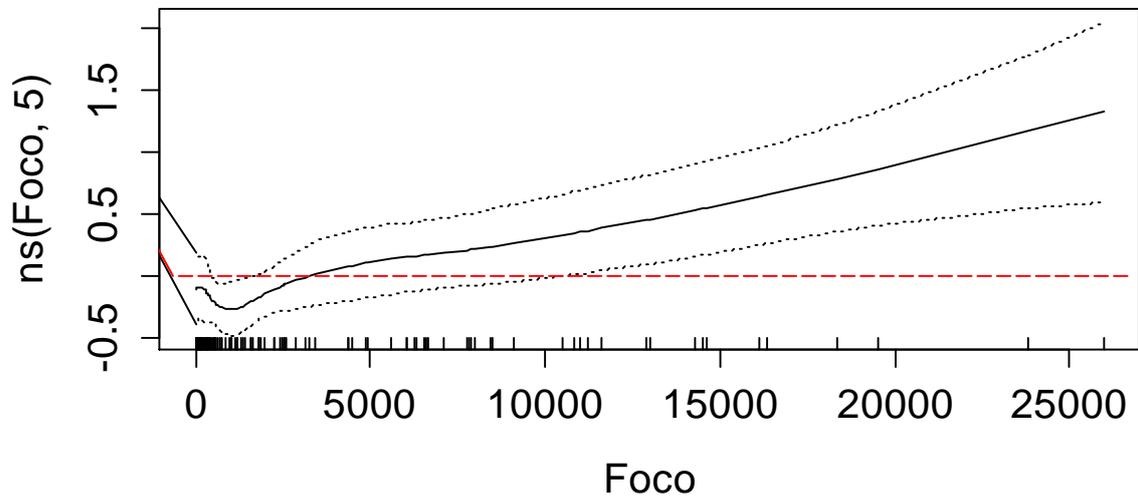
Tabela 5: Efeitos estimados para cada mês com Janeiro como mês de referência

Mês	Estimativa do parâmetro	p-valor
Fevereiro	-0,14024311	0,423
Março	0,21714247	0,175
Abril	-0,07139661	0,676
Maió	-0,15516765	0,368
Junho	-0,48644559	0,010
Julho	-0,51872284	0,006
Agosto	-0,48127757	0,025
Setembro	-1,10954967	0,000
Outubro	-0,12593264	0,476
Novembro	-0,17413390	0,313
Dezembro	0,02413580	0,884

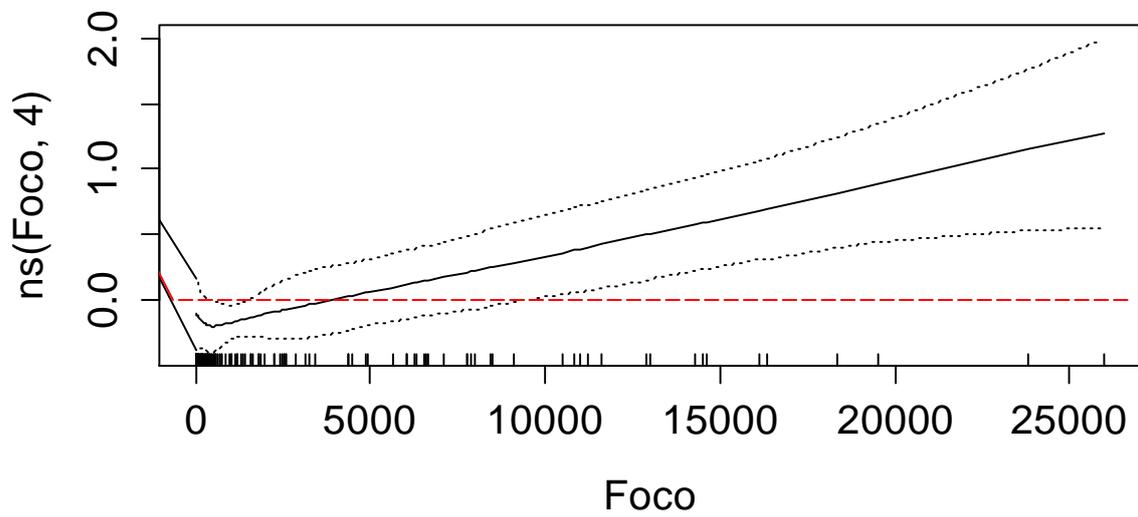
4.2.3 Ajuste do efeito da poluição

A figura 15 compara os ajustes a variável foco de queimada de acordo com diferentes graus de liberdade nas splines cúbicas.

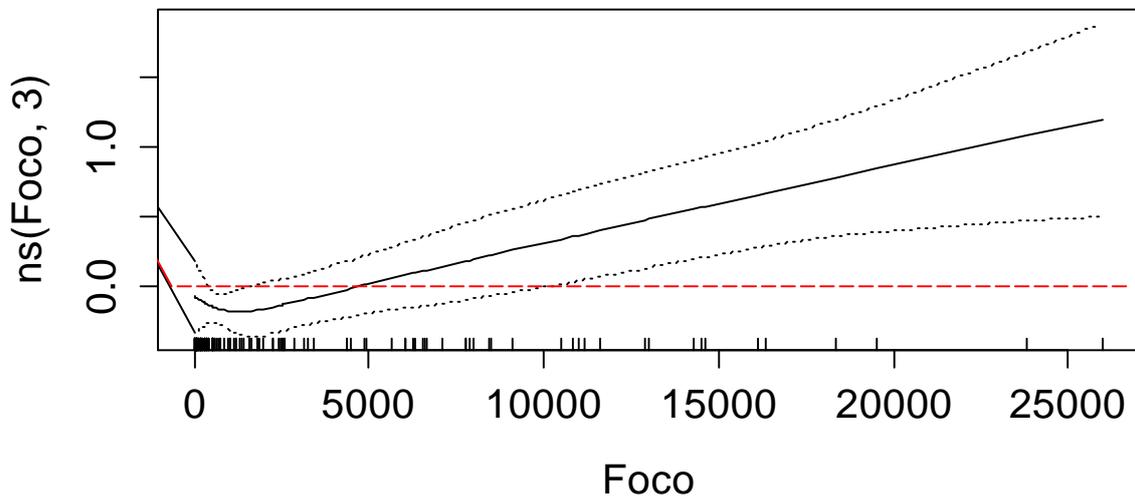
(a) Modelo 1: p-valor = 0,0152



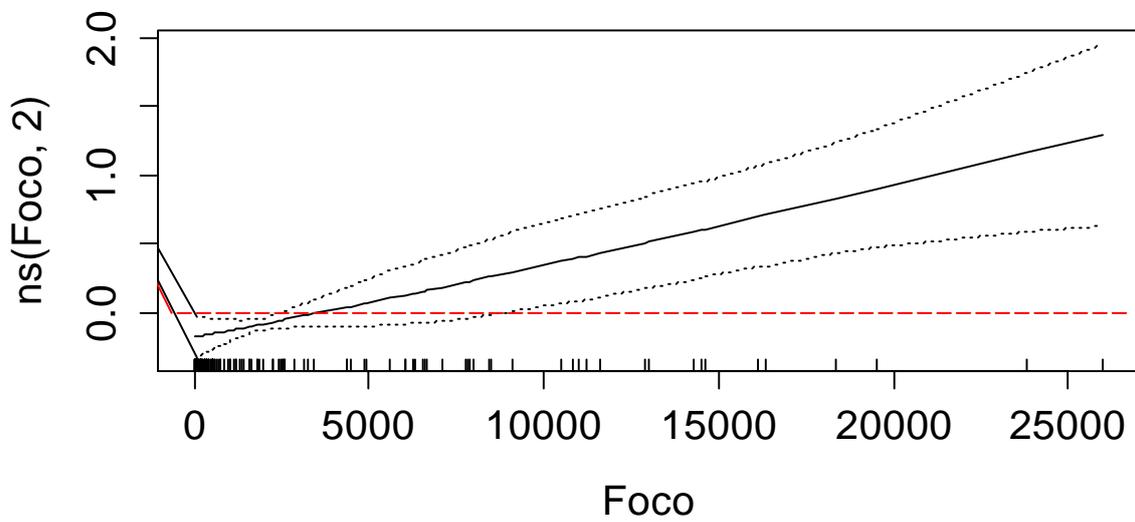
(b) Modelo 2: p-valor = 0,0106



(c) Modelo 3: p-valor = 0,0040



(d) Modelo 4: p-valor = 0,0016



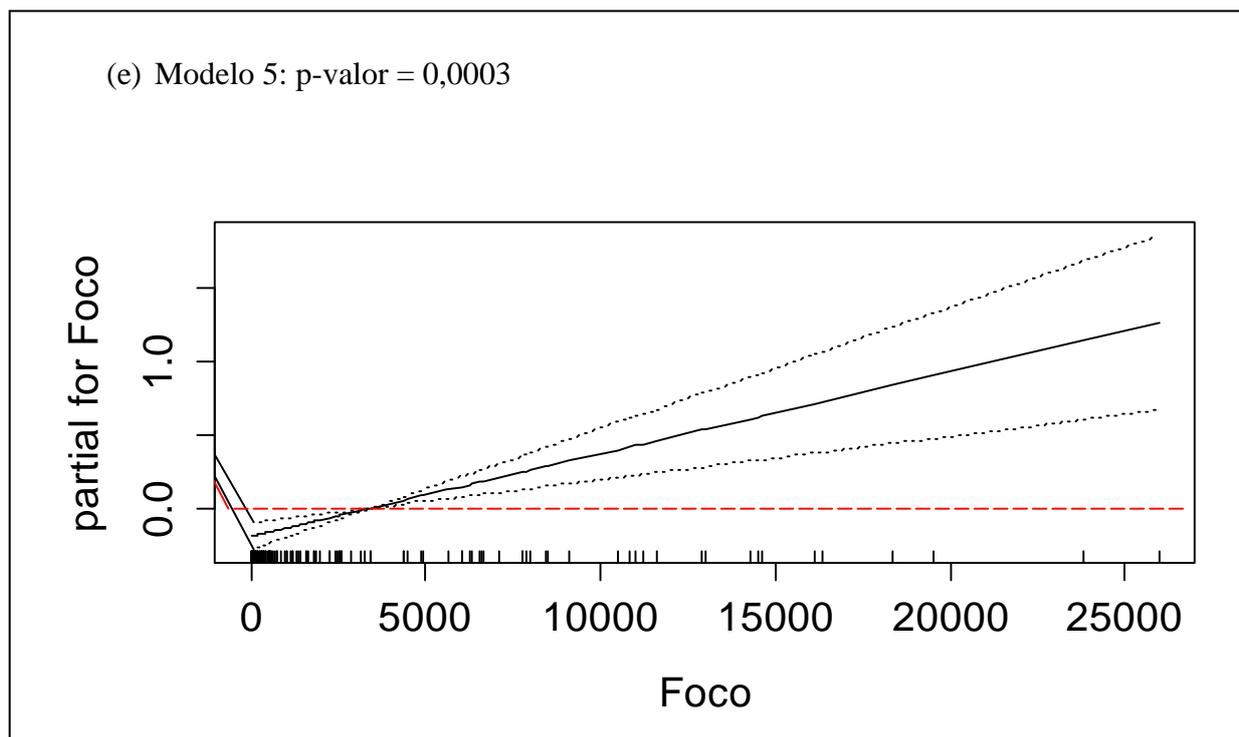


Figura 15: Quadro comparativo dos ajustes do modelo para a variável foco de queimada de acordo com diferentes graus de liberdade

Tabela 6. Estatísticas referentes ao ajuste do efeito do número de focos de queimadas

Modelo	AIC	<i>Spline</i>	ANOVA (Teste F)
		Cúbica	p-valor (Efeito paramétrico para a variável “Focos de queimadas”)
		gl	
1	656,69	5	0,015
2	656,03	4	0,011
3	653,78	3	0,004
4	652,58	2	0,002
5	650,62	1	0,000

A Figura 15 e a Tabela 6 apresentam informações sobre o efeito do número de focos no mês corrente (ou seja, no mesmo mês das ocorrências de prematuridade) sobre o número mensal de bebês prematuros. Verifica-se pelo AIC, que o ajuste do modelo melhora a medida que o número de graus de liberdade diminui. A figura não mostra ganhos importantes no ajuste do efeito por *splines*. Desta forma, podemos concluir que não é necessário o ajuste do efeito dos focos de queimadas via *splines*.

Assim, procedeu-se ao ajuste linear do efeito no mês corrente. O efeito estimado é o risco relativo (RR) correspondente a um aumento de 10 unidades no número de focos de queimadas. O risco relativo é a exponencial da estimativa do parâmetro. Para facilitar a interpretação, o RR foi transformado para aumento percentual. Portanto, o RR percentual estimado foi de 0,06% (Intervalo de confiança de 95%: 0,03% – 0,08%). Ou seja, para cada aumento de 10 unidades no número de focos, o risco de prematuridade aumenta em 0,06%.

4.2.4 Análise de Resíduos

Para verificar se o ajuste dos modelos foi adequado utilizamos diagnósticos habituais para análise de regressão de dados de séries temporais como função de autocorrelação, plotagem dos quantis dos resíduos padronizados e distância de Cook.

A análise dos resíduos considerou o seguinte modelo:

$$\text{Prematuridade} \sim \text{Poisson}(\mu_t)$$

$$\log(\mu_t) = \gamma_0 + \sum_{i=1}^4 \gamma_i f_i + \sum_{i=2}^{12} \beta_{1i}^{\text{Mês}} + \beta_2 X_{2t}^{\text{Foco de queimada}}$$

Onde:

- γ_0 é o intercepto do modelo;
- $\sum_{i=1}^4 \gamma_i f_i$ é o efeito do tempo ajustado por uma *spline* cúbica com 3 nós, $ns(\text{tempo}, df = 4)$;
- β_{1i} é o efeito do i -ésimo nível do fator mês ($i = 2, \dots, 12$) – mês de Janeiro ($i = 1$) como fator de referência;
- β_2 é o efeito do número de focos de queimadas;
- X_{2t} é o número de focos no mês t

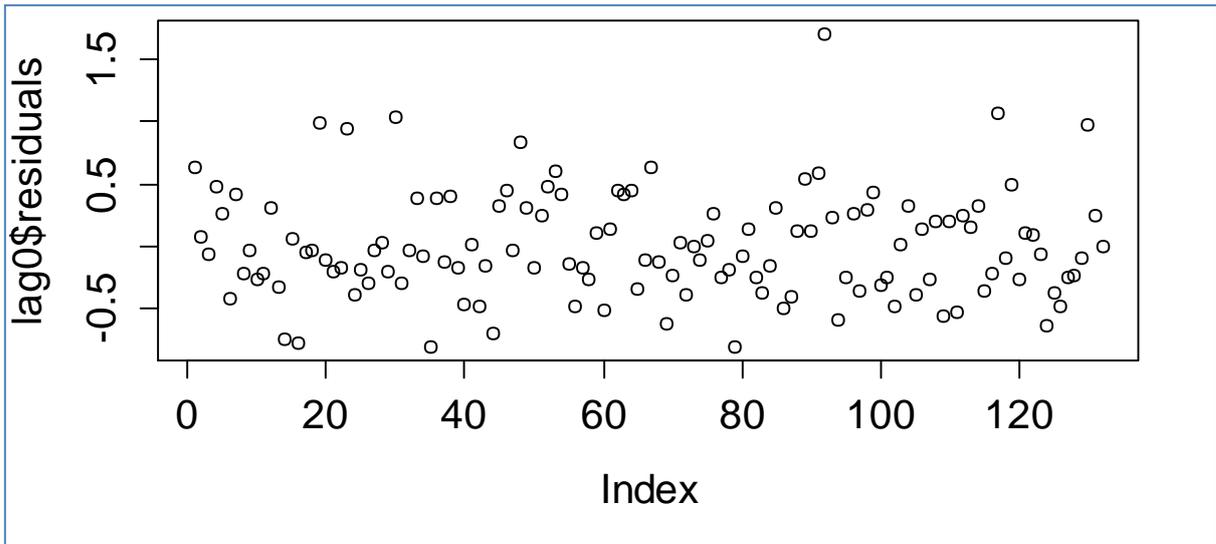


Figura 16: Análises dos resíduos para o modelo final escolhido

Com o gráfico apresentado na figura 16 observamos o comportamento da variância dos resíduos com relação aos valores ajustados. Observa-se uma distribuição constante localizada no intervalo $(-1,1)$, ou seja, as observações estão homogeneamente distribuídas apresentando apenas 2 pontos para fora do intervalo. Então, temos indícios de que nenhuma variável influenciou nos resultados do ajuste do modelo.

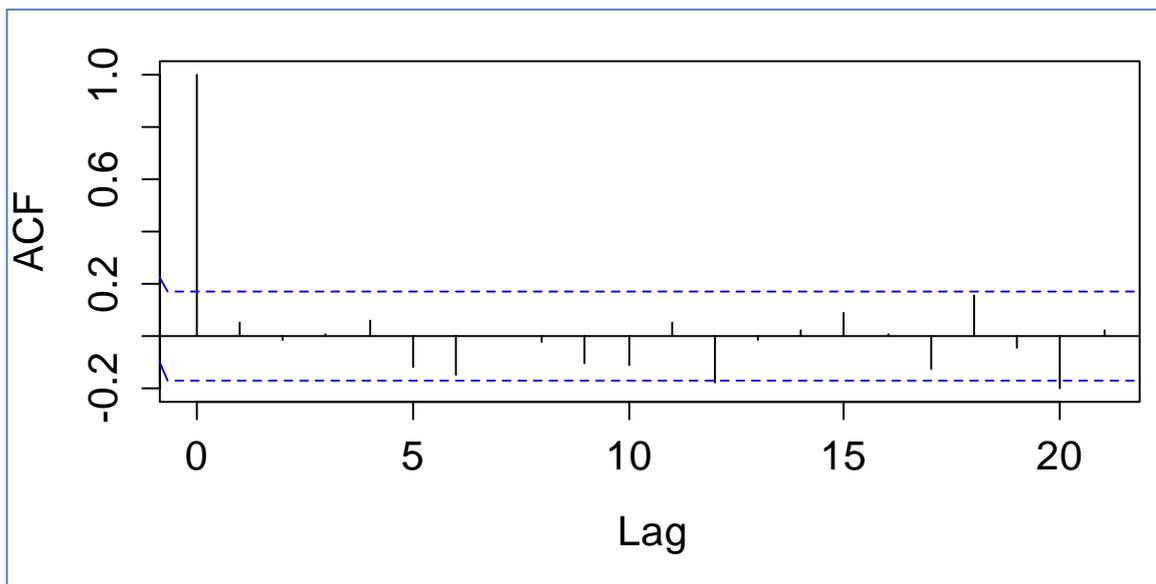


Figura 17: Gráfico de autocorrelação modelo final escolhido

O gráfico apresenta na figura 17 mostra poucas estacas (barras passando do intervalo estipulado) e a maioria das autocorrelações estão próximas de zero, indicando que o modelo

sazonal é apropriado. Note que os valores das autocorrelações estão dentro do intervalo de confiança (existem alguns poucos fora do intervalo, mas “escapam” do IC por pouco), indicando um processo puramente aleatório (ruído branco) com baixas autocorrelações em todos os lags.

Pelo teste de Box-Ljung para os resíduos ($\chi^2=0,313$, $gl=1$, $p\text{-valor}=0,576$), não rejeitamos H_0 , ou seja, os resíduos são ruídos brancos e estão distribuídos em torno 0 garantindo a hipótese de homocedasticidade (resíduos possuem variância constante).

Através do gráfico da distância de cook (Figura 18) para o modelo final escolhido, pela definição, temos que nenhum valor das observações, em módulo, é maior que 1. Logo, pode-se afirmar que nenhuma das observações são pontos influentes, ou seja, não estão influenciando no ajuste do modelo.

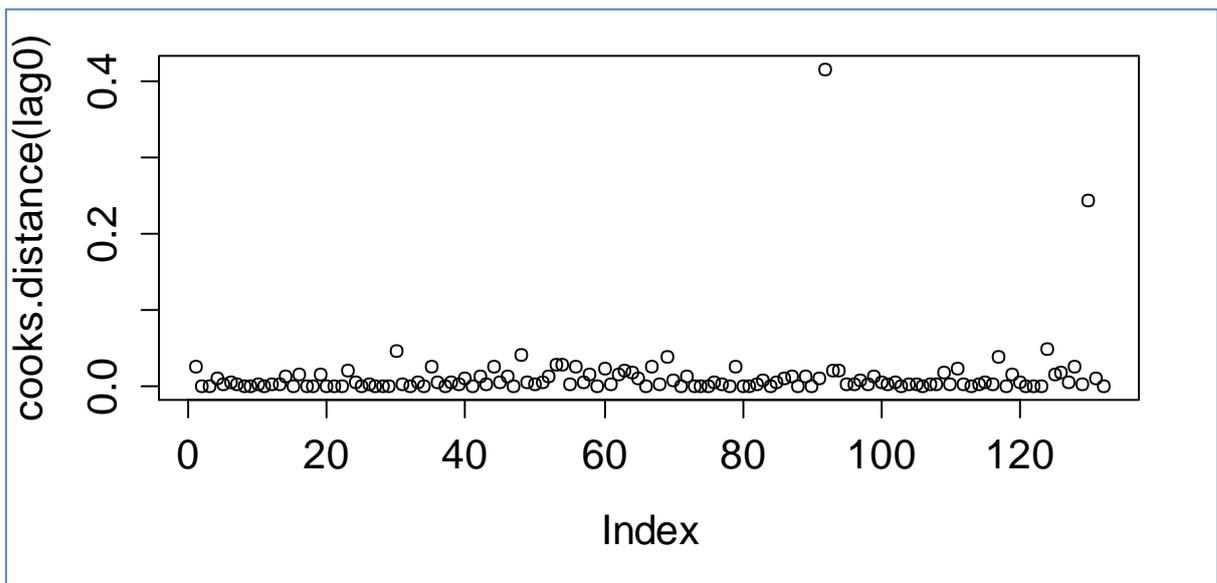


Figura 18: Gráfico das distancias de Cook.

5 Conclusão

Neste trabalho, as contagens mensais de bebês prematuros foram modeladas por regressão de Poisson. Através do gráfico da série temporal de prematuridade, têm-se um aumento do nascimento de bebês prematuros nos meses de Julho e Agosto onde é considerada uma época de seca em Tangará da Serra – MS, o que poderia ser a consequência dos causadores de poluição atmosférica: os focos de queimadas.

Para estimar a associação entre as variações mensais no número de focos de queimadas e os totais mensais de prematuros, foi utilizado um modelo aditivos generalizados. Por este modelo foi possível ajustar efeitos não lineares do tempo e do número de focos de queimadas através de *splines*. O ajuste por *splines* é interessante, pois permite identificar curvas (mudanças de comportamento da função polinomial através dos nós) que seriam difíceis de serem visualizadas a “olho nu”, funcionando como uma espécie de “lupa” para a identificação de mudanças.

Como se observa nos resultados, o ajuste polinomial via *splines* da variável focos de queimadas não é estatisticamente significativo para o ajuste do modelo e, mesmo ajustando linearmente o foco através da medida de Risco Relativo, percebe-se que o efeito do aumento do número de focos de queimadas aumenta minimamente a prematuridade no município de Tangará da Serra – MS.

Referências

- [1] Guisan et al. 2002; Lehmann et al. 2002; Leathwick et al. 2006.
Generalized additive models
- [2] Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman & Hall/CRC. ISBN_HYPERLINK "<https://en.wikipedia.org/wiki/Special:BookSources/978-0-412-34390-2>" 978-0-412-34390-2.
- [3] Gledson Luiz Picharski; Edicleia Furlanetto; Paulo Justiniano Ribeiro Jr.; Silvia Emiko Shimakura (LEG - Laboratório de Estatística e Geoinformação), 2004. *Uma Abordagem de Modelos Estatísticos Aplicados a Geomedicina - Casos de Câncer de Fígado no Estado do Paraná*.
- [4] Yara de Souza Tadano; Cássia Maria Lie Ugaya; Admilson Teixeira Franco (2009). *Método de regressão de Poisson: metodologia para avaliação do impacto da poluição atmosférica na saúde populacional*.
- [5] Snijders, Tom A.B., and Bosker, Roel J. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling, second edition*. London etc.: Sage Publishers, 2012 ISBN 9781849202008 (hardcover), ISBN 9781849202015 (pbk). xii + 368 p.
- [6] Tubino, Dálvio Ferrari. (2000) – *Manual de Planejamento e Controle da Produção*. 2ª ed. São Paulo: Atlas.
- [7] Pedro Alberto Morettin; Clélia Maria de Castro Toloí (1981). *Modelos para previsão de Séries Temporais*.
- [8] Tápia, Milena. *Redes Neurais Artificiais: Uma Aplicação na Previsão de Preços de Ovos*. Dissertação de mestrado, Universidade Federal de Santa Catarina - UFSC, 2000.
- [9] Huang JZ, Liu L. *Polynomial spline estimation and inference of proportional hazards regression models with flexible relative risk form*. *Biometrics*. 2006;62:793–802.
- [10] Huang JZ, Kooperberg C, Stone CJ, Truong YK. *Functional Anova modeling for proportional hazards regression*. *Ann. Statist.* 2000; 28:961–999.
- [11] Edition A.J. Dobson : *An Introduction to Generalized Linear Models*, Second Edition.
- [12] Souza, R.C.; Camargo, M.E. *Análise e previsão de séries temporais: os modelos ARIMA*. Ijuí, RS: Sedigraf, 1996. 242.
- [13] G. M. Ljung; G. E. P. Box (1978). *On a Measure of a Lack of Fit in Time Series Models*. *Biometrika* 65 (2): 297–303. doi:10.1093/biomet/65.2.297.
- [14] Eliane Ignotti, Ageo Mário Cândido da Silva, Inês Echenique Mattos, Sandra de Souza Hacon. *Material particulado originário de queimadas e doenças respiratórias*. *Rev. Saúde Pública* vol.47 no.2 São Paulo Apr. 2013.

[15] Cleber Nascimento do Carmo, Sandra de Souza Hacon. Estudos de séries temporais de poluição atmosférica por queimadas e saúde humana. Ciênc. saúde coletiva vol.18 no.11 Rio de Janeiro Nov. 2013.

[16] Botelho C, Correia AL, Silva AMC, Macedo AG, Silva COS. *Fatores ambientais e hospitalizações em crianças menores de 5 anos com infecção respiratória aguda*. Cad. Saúde Pública, Rio de Janeiro. 2003;19:1771-80.