



TATIANE CARVALHO ALVARENGA

**MODELO HETEROGÊNEO
AUTORREGRESSIVO: UMA APLICAÇÃO A
DADOS DE MORTALIDADE E A DADOS
CLIMÁTICOS**

LAVRAS - MG

2015

TATIANE CARVALHO ALVARENGA

**MODELO HETEROGÊNEO
AUTORREGRESSIVO: UMA APLICAÇÃO A
DADOS DE MORTALIDADE E A DADOS
CLIMÁTICOS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

Orientadora: Profa. Dra. Thelma Sáfydi

LAVRAS – MG

2015

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Alvarenga, Tatiane Carvalho.

Modelo heterogêneo autorregressivo: : uma aplicação a dados
de mortalidade e a dados climáticos / Tatiane Carvalho Alvarenga.

– Lavras : UFLA, 2015.

85 p. : il.

Dissertação (mestrado acadêmico)—Universidade Federal de
Lavras, 2015.

Orientador(a): Thelma Sáfyadi.

Bibliografia.

1. Óbitos em São Paulo. 2. Poluição em São Paulo. 3. Modelo
HAR(h). 4. Série temporal. I. Universidade Federal de Lavras. II.
Título.

TATIANE CARVALHO ALVARENGA

**MODELO HETEROGÊNEO
AUTORREGRESSIVO: UMA APLICAÇÃO A
DADOS DE MORTALIDADE E A DADOS
CLIMÁTICOS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

APROVADO em 13 de fevereiro de 2015.

Dr. Renato Ribeiro de Lima

UFLA

Dr. Washington Santos Silva

IFMG

Dra. Thelma Sáfadi

Orientadora

LAVRAS – MG

2015

LISTA DE FIGURAS

Figura 1: Modelo aditivo com tendência e sazonalidade	17
Figura 2: Modelo multiplicativo com tendência e sazonalidade.....	18
Figura 3: Modelo aditivo com tendência $Y_t = T_t + \varepsilon_t$	19
Figura 4: Modelo aditivo com sazonalidade $Y_t = S_t + \varepsilon_t$	19
Figura 5: Número diário de óbitos por problemas respiratórios e cardiovasculares (01/01/94 a 31/12/97).....	50
Figura 6: Séries de temperaturas mínimas (temp_min), umidade (humid_med), monóxido de carbono (co) e material particulado (pm10), na cidade de São Paulo.....	51
Figura 7: Série temporal do número de óbitos por problemas respiratórios na cidade de São Paulo, de 1º de janeiro de 1997 a 31 de dezembro de 1997	53
Figura 8: Série temporal do número de óbitos por problemas cardiovasculares na cidade de São Paulo, de 1º de janeiro de 1997 a 31 de dezembro de 1997	54
Figura 9: Série temporal da temperatura mínima da cidade de São Paulo, de 1º de janeiro de 1997 a 31 de dezembro de 1997.....	55
Figura 10: Série temporal de pm10 da cidade de São Paulo, de 1º de janeiro de 1997 a 31 de dezembro de 1997	55
Figura 11: Série temporal de co da cidade de São Paulo, de 1º de janeiro de 1997 a 31 de dezembro de 1997	56
Figura 12: Série temporal da umidade da cidade de São Paulo, de 1º de janeiro de 1997 a 31 de dezembro de 1997.....	57
Figura 13: Séries transformadas com h=1 do número de óbitos.....	58
Figura 14: Séries climáticas transformadas com h=1.....	58
Figura 15: Séries transformadas com h=7 do número de óbitos.....	59

Figura 16: Séries climáticas transformadas com $h=7$	60
Figura 17: Séries transformadas com $h=28$ do número de óbitos.	61
Figura 18: Séries climáticas transformadas com $h=28$	61
Figura 19: Função de autocorrelação e autocorrelação parcial do resíduo do HAR(1).....	65
Figura 20: Função de autocorrelação e autocorrelação parcial do resíduo do HAR (1,7).....	66
Figura 21: Função de autocorrelação e autocorrelação parcial do resíduo do HAR (1,7,28).....	67
Figura 22: Série temporal com uma diferença do número de óbitos por problemas respiratórios na cidade de São Paulo, sendo resp quando $h=1$, sem_resp quando $h=7$ e mes_resp quando $h=28$	72
Figura 23: Função de autocorrelação e autocorrelação parcial da série diferenciada do número de óbitos por problemas respiratórios com $h=1$	72
Figura 24: Função de autocorrelação e autocorrelação parcial da série diferenciada do número de óbitos por problemas respiratórios com $h=7$	73
Figura 25: Função de autocorrelação e autocorrelação parcial da série diferenciada do número de óbitos por problemas respiratórios com $h=28$	73
Figura 26: Série temporal com uma diferença do número de óbitos por problemas cardiovasculares na cidade de São Paulo, sendo card quando $h=1$, sem_card quando $h=7$ e mes_card quando $h=28$	74
Figura 27: Função de autocorrelação e autocorrelação parcial da série diferenciada do número de óbitos por problemas cardiovasculares com $h=1$	74
Figura 28: Função de autocorrelação e autocorrelação parcial da série diferenciada do número de óbitos por problemas cardiovasculares com $h=7$	75
Figura 29: Função de autocorrelação e autocorrelação parcial da série diferenciada do número de óbitos por problemas cardiovasculares com $h=28$. .	75

Figura 30: Série temporal com uma diferença da temperatura mínima da cidade de São Paulo, sendo tempmin quando h=1, sem_temp quando h=7 e mes_temp quando h=28 .	76
Figura 31: Função de autocorrelação e autocorrelação parcial da série diferenciada da temperatura mínima com h=1.	76
Figura 32: Função de autocorrelação e autocorrelação parcial da série diferenciada da temperatura mínima com h=7.	77
Figura 33: Função de autocorrelação e autocorrelação parcial da série diferenciada da temperatura mínima com h=28.	77
Figura 34: Série temporal com uma diferença de pm10 da cidade de São Paulo, sendo pm10 quando h=1, sem_pm quando h=7 e mes_pm quando h=28.	78
Figura 35: Função de autocorrelação e autocorrelação parcial da série diferenciada de pm10 com h=1.	78
Figura 36: Função de autocorrelação e autocorrelação parcial da série diferenciada de pm10 com h=7.	79
Figura 37: Função de autocorrelação e autocorrelação parcial da série diferenciada de pm10 com h=28.	79
Figura 38: Série temporal com uma diferença de co da cidade de São Paulo, sendo co quando h=1, sem_co quando h=7 e mes_co quando h=28.	80
Figura 39: Função de autocorrelação e autocorrelação parcial da série diferenciada de co com h=1.	80
Figura 40: Função de autocorrelação e autocorrelação parcial da série diferenciada de co com h=7.	81
Figura 41: Função de autocorrelação e autocorrelação parcial da série diferenciada de co com h=28.	81
Figura 42: Série temporal com uma diferença da umidade da cidade de São Paulo, sendo humid quando h=1, sem_hum quando h=7 e mes_hum quando h=28.	82

Figura 43: Função de autocorrelação e autocorrelação parcial da série diferenciada de hum com $h=1$	82
Figura 44: Função de autocorrelação e autocorrelação parcial da série diferenciada de hum com $h=7$	83
Figura 45: Função de autocorrelação e autocorrelação parcial da série diferenciada de hum com $h=28$	83

LISTA DE TABELAS

Tabela 1: Exemplo fictício de séries transformadas.....	36
Tabela 2: Análise descritiva das séries de médias do número de óbitos, índices de temperatura, PM10, CO e umidade em São Paulo no ano de 1997.....	62
Tabela 3: Teste da raiz unitária (ADF), sem constante, com uma diferença	63
Tabela 4: Modelo HAR(1) do número de óbitos por problemas respiratórios. ..	84
Tabela 5: Modelo HAR (1,7) do número de óbitos por problemas cardiovasculares.....	84
Tabela 6: Modelo HAR (1, 7, 28) do número de óbitos por problemas respiratórios	85

AGRADECIMENTOS

Primeiramente agradeço a Deus, pelo dom da vida, pela oportunidade de inserção na UFLA, a melhor universidade de Minas Gerais, agradeço pela saúde, ânimo e sabedoria durante a jornada de Mestrado.

Aos meus pais, Vander Mendes de Alvarenga e Sônia Maria Carvalho Alvarenga, pelo apoio e amor incondicional, fazendo de mim uma pessoa cada vez melhor, ou seja, uma pessoa do bem, com os princípios de Deus.

A minhas irmãs, Taís Aparecida Alvarenga e Talita Aparecida Carvalho Alvarenga, pela amizade, carinho e bondade em todos os momentos, desde a infância até os dias de hoje.

A minha orientadora da dissertação, Profa. Dra. Thelma Sáfi, pela compreensão, confiança e orientação no decorrer do curso de Mestrado em Estatística e Experimentação Agropecuária.

Aos membros da banca Prof. Dr. Renato Ribeiro de Lima e Prof. Dr. Washington Santos Silva pela participação e contribuição no trabalho final.

A todos os professores do Departamento de Ciências Exatas da Universidade Federal de Lavras, pelo conhecimento e amizade no decorrer do curso, que me ajudaram muito no crescimento como aluna de pós-graduação.

A todos os colegas e amigos do curso de pós-graduação em Estatística Experimentação Agropecuária pela amizade, solidariedade e companheirismo nos estudos e congressos, em especial, Flaviano Teixeira, Ricardo Mattos, Thaís Ribeiro e Thiago Furtado.

Enfim, obrigada a todos que participaram direta ou indiretamente dessa conquista.

RESUMO

Uma série temporal é todo conjunto de observações ordenadas no tempo, na qual existe dependência entre as observações e tem-se como objetivo modelar essa dependência. Comumente, em séries temporais, a modelagem é feita com séries de mesma frequência temporal. A partir dos estudos de Corsi (2009) buscou-se agregar ao modelo diferentes frequência de variações temporais, o que foi denominado de modelo heterogêneo autorregressivo (HAR). Existem regras e restrições na modelagem do HAR, nas diferentes frequências temporais. Assim, buscou-se modelar o número de óbitos por problemas respiratórios e número de óbitos por problemas cardiovasculares, em função de temperatura mínima, material particulado, monóxido de carbono e umidade relativa do ar, sendo todas as séries diárias e da cidade de São Paulo, no período de 1º de janeiro de 1997 a 31 dezembro de 1997. Dessa maneira, foi possível verificar que as séries de temperatura mínima, material particulado e umidade influenciaram o número de mortes por problemas respiratórios, a série de temperatura mínima influenciou o número de mortes por problemas cardiovasculares e a série de monóxido de carbono não teve influência nas séries de mortes.

Palavras-chave: Óbitos em São Paulo. Poluição em São Paulo. Modelo HAR(h). Série temporal.

ABSTRACT

A time series is all the sets of observations sorted in time, in which there is dependency among observations and one has a goal to model that dependency. Commonly, in time series, the modeling is done with series of the same time frequency. From the studies by Corsi (2009), different sources of time variations were intended to aggregate into the model, which was named heterogeneous autoregressive model (HAR). There are both rules and restrictions in the HAR modeling, in the different time frequencies. Thus, it was intended to model the number of deaths by respiratory problems and number of deaths by cardiovascular problems due to the minimal temperature, particulate material, carbon monoxide and relative air humidity, all the series being daily and of the city of São Paulo in the period from 1st January, 1997 to December, 1997. In this way, it was possible to verify that the series of minimum temperature, particulate material and humidity influenced the number of deaths by respiratory problems, the series of minimum temperature influenced the number of deaths by cardiovascular problems and the series of carbon monoxide had no influence upon the series of deaths.

Key words: Deaths in São Paulo. Pollution in São Paulo. HAR(h) Model. Time series.

SUMÁRIO

1. INTRODUÇÃO	14
2. REFERENCIAL TEÓRICO.....	16
2.1. Séries temporais	16
2.2. Modelos em séries temporais	16
2.2.1. Covariância e correlação.....	19
2.2.2. Processos estacionários	21
2.2.2.1. Teste de estacionariedade.....	22
2.2.3. Função de autocovariância (facv) amostral.....	22
2.2.4. Função de autocorrelação (fac).....	23
2.2.5. Tendência em séries temporais	24
2.2.6. Sazonalidade em séries temporais	26
2.3. Metodologia de Box e Jenkins em séries temporais.....	28
2.3.1. Modelagem ARIMA (p, d, q).....	28
2.3.2. Modelagem ARIMA sazonal - SARIMA	29
2.3.3. Função de autocorrelação de um processo autorregressivo.....	30
2.3.4. Função de autocorrelação parcial.....	32
2.4. Modelo heterogêneo autorregressivo	33
2.4.1. O modelo HAR(h)	34
2.4.1.1. Volatilidade e o modelo HAR.....	38
2.5. Adequabilidade do modelo.....	41
2.6. Mortalidade e clima.....	41
3. MATERIAL E MÉTODOS	46
3.1. Dados	46
3.1.1. Período das séries temporais.....	46
3.2. Software estatístico.....	47
3.3. Métodos	47

4. RESULTADOS E DISCUSSÃO	49
4.1 Análises das séries temporais em estudo.....	49
4.1.1 Séries temporais no período janeiro de 1994 a dezembro de 1997	
.....	49
4.1.2 Séries temporais no período janeiro de 1997 a dezembro de 1997	
.....	53
4.1.2 Séries temporais transformadas	57
4.1.2.1. Séries temporais transformadas com $h = 1$	57
4.1.2.2. Séries temporais transformadas com $h = 7$	59
4.1.2.3. Séries temporais transformadas com $h = 28$	60
4.1.3. Séries temporais estacionárias	63
4.2. Modelagem heterogênea autorregressiva das séries temporais em	
estudo	64
5. CONCLUSÕES.....	68
REFERÊNCIAS.....	69
ANEXOS	72

1. INTRODUÇÃO

A aplicação de séries temporais se estende nas diferentes áreas de conhecimento, entre elas Ciências Exatas e da Terra, Ciências Biológicas, Engenharias, Ciências da Saúde, Ciências Agrárias, Ciências Sociais Aplicadas, Ciências Humanas, Linguística, Letras e Artes, entre outras. A ordem dos dados é fundamental, uma vez que a vizinhança desses dados tem dependência e se almeja analisar e modelar esta dependência com o estudo de séries temporais.

Desenvolvido por Corsi (2009), o *Heterogeneous autoregressive model of realized volatility*, ou seja, o modelo HAR de estimação da volatilidade realizada se baseia na modelagem do comportamento da volatilidade de memória longa de maneira simples e parcimoniosa. Este modelo foi inspirado em hipóteses de mercados heterogêneos e pela assimetria da propagação da volatilidade. Corsi (2009) considera apenas como fonte de heterogeneidade os diferentes horizontes de tempos (dias, semanas e meses), de onde se inicia todo o estudo do modelo heterogêneo autorregressivo (HAR).

O modelo HAR(h) é um processo de memória longa que se caracteriza pela dependência significativa entre observações distantes (em tempo). Nesse caso, existe persistência nas autocorrelações observadas, que decaem lentamente para zero. O h no modelo pode assumir os valores de 1, 7 e 28, que será explicado com detalhes no referencial teórico.

O HAR(h) é utilizado para agregar diferentes frequências de tempo (dia, semana e mês) no mesmo modelo. Para isso, existem algumas regras e restrições para se trabalhar com esta modelagem.

Neste trabalho, a abordagem do modelo HAR, desenvolvido por Corsi (2009), será aplicada a séries temporais que não sejam de dados de finanças, já que na literatura o modelo aparece apenas aplicado a esses dados.

Assim, o objetivo geral foi estudar a teoria do modelo HAR(h) e o objetivo específico, aplicar o HAR(h) a séries de mortalidade e de clima. O banco de dados utilizados refere-se ao número diário de óbitos por problemas respiratórios e cardíacos, temperaturas mínimas, material particulado (pm10), umidade relativa do ar e monóxido de carbono da cidade de São Paulo, no período de janeiro a dezembro de 1997.

A dissertação está dividida em sete seções. Na primeira, a introdução, relatam-se os pontos estratégicos e faz-se uma abordagem resumida do trabalho. A segunda seção diz respeito ao referencial teórico do trabalho, abordando os principais pontos de séries temporais e do modelo em estudo, o modelo heterogêneo autorregressivo (HAR). Na sequência, tem-se a seção de metodologia, na qual se apresentam os passos que o banco de dados passou até a modelagem do HAR. Na seção de resultados e discussões relatam-se os resultados inéditos do modelo HAR(h). Por fim, apresentam-se as conclusões, referências bibliográficas e anexos.

2. REFERENCIAL TEÓRICO

2.1. Séries temporais

Segundo Morettin e Toloi (2006), uma série temporal é todo conjunto de observações ordenadas no tempo, na qual existe dependência entre as observações. Uma série pode ser representada da forma $Y = \{y_1, y_2, \dots, y_n\}$, em que $y_1, y_2, \dots, y_n$ representam os valores assumidos pelas observações ordenadas no tempo, em intervalos equidistantes.

As séries podem ser discretas, ou seja, quando o intervalo de tempo (valores diários, mensais, trimestrais, anuais entre outros) está explícito na série ou contínuas, caso contrário.

Muitas séries contínuas podem ser transformadas em discretas. Basta considerar $N = \frac{T}{\Delta t}$, em que T = tempo amostrado $[0, s]$ e Δt = intervalo para a obtenção de N observações.

Entre os objetivos para as séries temporais tem-se: investigar o comportamento da série, descobrir o mecanismo gerador da mesma, encontrar periodicidades na série temporal e fazer previsões para observações futuras.

Em todos os objetivos de séries temporais, podem-se construir modelos probabilísticos ou estocásticos.

2.2. Modelos em séries temporais

A Estatística reúne a metodologia utilizada no planejamento dos experimentos, bem como na análise e interpretação dos dados, de tal forma que seja garantida a validade das conclusões.

De acordo com Salsburg (2009), a estatística revolucionou a ciência, pois fornece modelos úteis e sofisticados que aprimoraram o processo de pesquisa na direção de melhores parâmetros de investigação, assim permitindo melhor orientação na tomada de decisões.

Um bom modelo deve ser econômico, ou seja, parcimonioso (poucos parâmetros) e sua descrição deve ser relativamente simples e flexível para poder se adaptar ao futuro (incerto) e facilitar o aprendizado.

Uma série temporal, dada por $Y = \{y_1, y_2, \dots, y_n\}$, pode ser desagregada em uma soma de três componentes não observáveis, em que essas componentes são independentes. O modelo é da forma aditiva, sendo

$$Y_t = T_t + S_t + \varepsilon_t,$$

em que T_t e S_t representam, respectivamente, tendência e sazonalidade e ε_t é a componente aleatória, de média zero e variância constante.

Na Figura 1, as observações seguem um modelo na forma aditiva, com as componentes de tendência e sazonalidade.

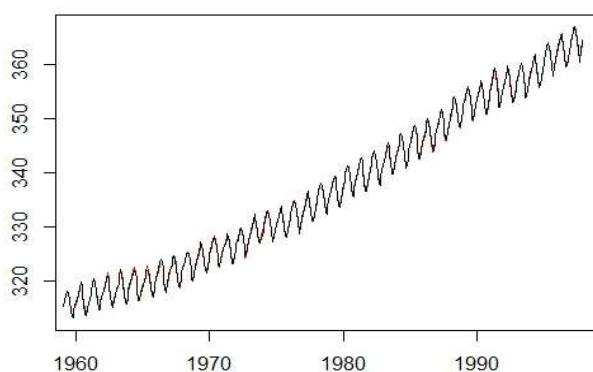


Figura 1: Modelo aditivo com tendência e sazonalidade

No caso de as componentes sazonais variarem de acordo com a tendência, o modelo mais adequado é o multiplicativo, cujo comportamento é ilustrado na Figura 2, dado por $Y_t = T_t S_t \varepsilon_t$.

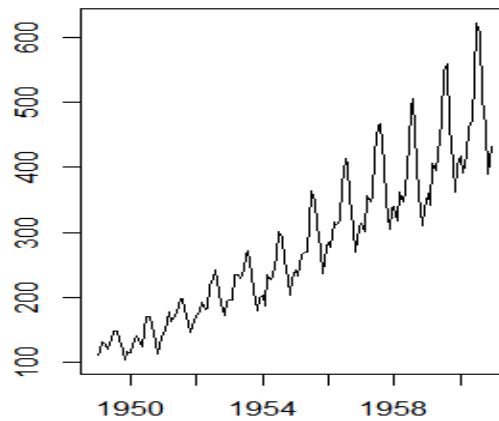


Figura 2: Modelo multiplicativo com tendência e sazonalidade

Geralmente, no caso do modelo multiplicativo ocorre um acréscimo na variância à medida que o tempo passa, conforme visualizado na Figura 2, na qual a amplitude no início da série temporal é menor, comparada com a amplitude no final da série.

No modelo multiplicativo são utilizadas as transformações para estabilizar a variância e também para tornar o efeito sazonal aditivo. Neste caso, a transformação logarítmica é aconselhável. Exemplo: $Y_t = T_t S_t \varepsilon_t$ aplicando o logaritmo: $\ln(Y_t) = \ln(T_t S_t \varepsilon_t) \Rightarrow \ln(Y_t) = \ln(T_t) + \ln(S_t) + \ln(\varepsilon_t)$. Assim, estabiliza-se a variância e torna-se o efeito sazonal aditivo.

Os modelos também podem ter apenas a componente de tendência, ou apenas a sazonalidade, conforme apresentado nas Figuras (3) e (4), respectivamente.

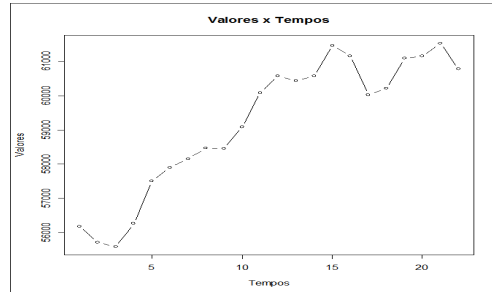


Figura 3: Modelo aditivo com tendência $Y_t = T_t + \varepsilon_t$

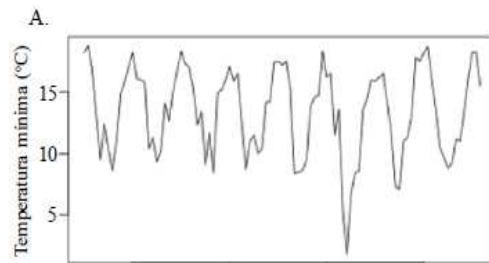


Figura 4: Modelo aditivo com sazonalidade $Y_t = S_t + \varepsilon_t$

2.2.1. Covariância e correlação

Modelos são processos estocásticos, isto é, processos controlados por leis probabilísticas, que descrevem as séries temporais. Dado $Y = \{Y(t), t \in T\}$ a série temporal, para $t \in T$, tem-se uma variável aleatória que tem média $\mu(t)$ e variância $\sigma^2(t)$.

Para medir a dependência entre as variáveis aleatórias (v.a), utilizam-se a covariância e a correlação, já que uma das pressuposições de séries temporais é que as v.a apresentem dependência.

A covariância é conhecida por medir a dependência linear entre as duas variáveis aleatórias.

Considerem-se os pares de valores dados por $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Dessa maneira, a covariância entre as variáveis aleatórias, X e Y , é

$$\text{dada por } \text{cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n}, \text{ ou}$$

seja, a média dos produtos dos valores centrados das variáveis. Considerando que x_i e y_i são os valores da variável no tempo i , $\bar{x}$ e $\bar{y}$ são as médias, n número de variáveis da série, $E(X)$ e $E(Y)$ são as esperanças matemáticas.

A partir da covariância, surge o coeficiente de correlação $(\rho(X, Y))$, que é dado por $\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}$.

Se $\text{cov}(X, Y) = 0$, então, as variáveis aleatórias X e Y são ditas não correlacionadas.

No caso de $E(X) = 0$ e $E(Y) = 0$, então, $\text{cov}(X, Y) = E(XY)$. Se X e Y são independentes, então, $\text{cov}(X, Y) = E(X)E(Y)$.

Calcula-se a correlação entre as variáveis aleatórias, X e Y , por

$$\text{corr}(X, Y) = \frac{\sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma(X)} \right) \left(\frac{y_i - \bar{y}}{\sigma(Y)} \right)}{n}, \text{ ou seja, a média dos produtos dos}$$

valores padronizados das variáveis. Dado que x_i e y_i são os valores da variável no tempo i , $\bar{x}$ e $\bar{y}$ são as médias, $\sigma(X)$ e $\sigma(Y)$ os desvios padrões respectivamente de X e Y , e n é o número de variáveis da série.

A correlação indica a força e a direção do relacionamento linear entre duas variáveis aleatórias. Assim, a correlação está inserida no intervalo $[-1,1]$.

Se $\text{corr}(X,Y)=1$, significa uma correlação positiva perfeita entre as duas variáveis; se x aumenta, então y também aumenta, idem se a variável diminui. Caso $\text{corr}(X,Y)=0$, as variáveis não dependem linearmente uma da outra. E, se $\text{corr}(X,Y)=-1$, significa uma correlação negativa perfeita entre as duas variáveis, isto é, se uma aumenta, a outra sempre diminui, mantendo a proporcionalidade.

2.2.2. Processos estacionários

Uma das suposições que se faz a respeito de uma série temporal é que ela seja estacionária, ou seja, se desenvolve no tempo ao redor de uma média constante.

De acordo com a independência ou não relativa à origem dos tempos, pode-se classificar se a série é estacionária ou não. Um processo Y_t é estacionário se ele se desenvolve no tempo, de modo que a escolha de uma origem de tempo não é importante, ou seja, as características de $Y_{t+\Gamma}$, $\forall \Gamma$ são as mesmas de Y_t .

Um processo estocástico $Y = \{Y_t, t \in T\}$ diz-se fracamente estacionário ou estacionário de segunda ordem, se e somente se satisfaz (i-iii):

- i) $E\{Y(t)\} = \mu(t) = \mu$, constante para todo $t \in T$.
- ii) $E\{Y^2(t)\} < \infty$, para todo $t \in T$.
- iii) $\gamma(t_1, t_2) = \text{Cov}(Y(t_1), Y(t_2))$ é uma função apenas da distância $|t_1 - t_2|$.

Contudo, a maioria das séries apresenta componentes como tendência e sazonalidade. Para torná-las estacionárias, aplicam-se diferenças na série original, buscando-se, assim, retirar essas componentes da série. Dessa forma, de acordo com Morettin & Toloi (2006), a metodologia para ajuste de modelos de Box e Jenkins pode ser aplicada.

2.2.2.1. Teste de estacionariedade

Um dos testes que se destacam para analisar a estacionariedade de uma série temporal é o Teste de *Dickey-Fuller* aumentado ou Teste ADF (do acrônimo em inglês *Augmented Dickey-Fuller*) que é um teste de raiz unitária. Na maioria dos testes, a hipótese nula é de que a série tenha raiz unitária e, portanto, não estacionária e como hipótese alternativa, a série é estacionária.

O Teste ADF introduz um operador de defasagens para resolver o problema de autocorrelação do termo de erro ε_t , tal como,

$$\Delta Y_t = \alpha_0 + \alpha_1 t + \gamma Y_{t-1} + \sum_{i=1}^m \Delta Y_{t-i} + \varepsilon_t, \text{ na qual pode ser estimada também}$$

sem constata e ou tendência.

Para testar a raiz unitária, Dickey e Fuller (1979) tabularam os valores críticos, através de simulações de Monte Carlo, e desenvolveram uma estatística τ (Tau). Assim, se o valor absoluto da estatística τ calculado for maior que o valor absoluto tabulado por ADF, aceita-se a hipótese nula, logo a série é não estacionária, caso contrário a série é estacionária.

2.2.3. Função de autocovariância (*facv*) amostral

Considere um processo estacionário real discreto, de média zero e função de autocovariância (*facv*) dada por $\gamma_\Gamma = E(Y_t Y_{t+\Gamma})$. γ_Γ é *facv* se e somente se satisfaz às propriedades:

- i) $\gamma_0 > 0$
- ii) $\gamma_{-\Gamma} = \gamma_\Gamma$
- iii) $|\gamma_\Gamma| \leq \gamma_0$
- iv) γ_Γ é não negativa, definida no sentido que $\sum_{j=1}^n \sum_{k=1}^n a_j a_k \gamma_{\Gamma_j - \Gamma_k}$ para quaisquer números reais $a_1, \dots, a_n, \Gamma_1, \dots, \Gamma_n$ de $\mathbb{Z}$.

2.2.4. Função de autocorrelação (*fac*)

A função de autocorrelação (*fac*) do processo é definida por $\rho_\Gamma = \frac{\gamma_\Gamma}{\gamma_0}$, $\forall \Gamma \in \mathbb{Z}$ e tem as mesmas propriedades da γ_Γ , exceto na origem, na qual $\rho_0 = 1$.

Considerando as observações da série temporal $Y = \{Y_t, t \in T\}$, dadas por, $y_1, y_2, y_3, \dots, y_n$, a *fac* (ρ_Γ) é estimada por $r_\Gamma = \frac{c_\Gamma}{c_0}$,

$$\Gamma = 0, 1, 2, \dots, n-1, \text{ em que } c_\Gamma = \frac{\sum_{i=1}^{n-\Gamma} (y_i - \bar{y})(y_{i+\Gamma} - \bar{y})}{n}.$$

2.2.5. Tendência em séries temporais

Num primeiro momento, para se analisar uma série temporal, é necessário construir seu gráfico, que revelará características importantes, como presença das componentes, tendência e/ou sazonalidade e também variabilidade e valores atípicos (*outliers*).

A tendência (T_t) de uma série pode ser entendida como uma mudança gradual (aumento ou diminuição) nos valores das observações ao longo do tempo. Para estimar T_t são utilizados métodos como ajuste polinomial e ou suavização.

Para testar a hipótese de tendência na série temporal, destacam-se Teste de Wald Wolfowitz, Teste de Cox-Stuart e Teste de Mann-Kendall (MORETTIN; TOLOI, 2006).

Para o Teste de Cox-Stuart é considerado o conjunto de observações $y_1, y_2, \dots, y_n$. Agrupam-se as observações em pares $(y_1, y_{1+c}), (y_2, y_{2+c}), \dots,$

(y_{N-c}, y_N) , em que $c = \frac{N}{2}$, no caso de N par e $c = \frac{N+1}{2}$, caso N seja ímpar.

A cada par de observações associa-se o sinal "+", se $y_i < y_{i+c}$ e o sinal "-", se $y_i > y_{i+c}$. Caso $y_i = y_{i+c}$, desconsidera-se o par.

Considere N o número de pares em que $y_i < y_{i+c}$. Assim, deseja-se testar:

$$H_0 : P(y_i < y_{i+c}) = P(y_i > y_{i+c}), \forall_i, \text{ não existe tendência.}$$

$$H_1 : P(y_i < y_{i+c}) \neq P(y_i > y_{i+c}), \forall_i, \text{ existe tendência.}$$

Utiliza-se a distribuição binomial para avaliar o teste, com os parâmetros $p = 0,5$ e $n = N$. Trata-se de um teste bilateral, em que a estatística do teste é dada por $T = \text{número de pares com sinal "+"}$. Assim, rejeita-se H_0 quando $T > n - t$, na qual t é encontrado na tabela da distribuição binomial com $p = 0,5$ e n , para um dado nível de significância α . Caso contrário, se aceita H_0 .

No caso de ser diagnosticada tendência na série temporal, para retirar essa componente basta utilizar a diferença, tal como $\Delta^d Y_t = \Delta(\Delta^{d-1} Y_t)$, em que $\Delta Y_t : Y_t - Y_{t-1}$, Δ : operador diferença, Y_t : série em estudo e d : número de diferenças necessárias para transformar a série em série estacionária.

Os casos mais simples são os de tendência linear, em que a série flutua ao redor de uma reta com inclinação positiva ou negativa. Por exemplo, $T_t = \beta_0 + \beta_1 t$, em que é possível escrever $Y_t = \beta_0 + \beta_1 t + \varepsilon_t$, sendo

Y_t : modelo com tendência linear

β_0 : coeficiente linear

β_1 : coeficiente angular, não nulo por hipótese

t : instante de tempo e

ε_t : erro aleatório.

Observa-se que, ao fazer uma diferença na série com tendência linear, obtém-se uma série estacionária, ou seja,

$$\Delta Y_t = \beta_0 + \beta_1 t + \varepsilon_t - (\beta_0 + \beta_1(t-1) + \varepsilon_{t-1})$$

$\Delta Y_t = \beta_1 + \varepsilon_t - \varepsilon_{t-1}$, que é uma série sem tendência.

2.2.6. Sazonalidade em séries temporais

A sazonalidade (S_t) pode ser interpretada como sendo as flutuações ocorridas ao longo de um tempo determinado, ou seja, os valores assumidos pela série temporal que se repetem a cada k elementos, devido a fatores exógenos. Pode ocorrer em períodos anuais, interanuais, seja trimestralmente, bimestralmente, mensalmente, diariamente, entre outros.

Para estimar a componente (S_t) são utilizados métodos como o de regressão ou de médias móveis, para a sazonalidade determinística e a sazonalidade estocástica, respectivamente (MORETTIN e TOLOI, 2006).

A sazonalidade determinística é pressuposta quando a sazonalidade é regular e estável no tempo e, dessa forma, o comportamento sazonal pode ser previsto perfeitamente a partir de dados anteriores. Quando a componente sazonal da série varia com o tempo é dito que a sazonalidade é estocástica.

Para testar a hipótese de sazonalidade na série temporal, podem-se utilizar os testes de Kruskal-Wallis, de Friedman ou de Fisher.

Considere a série temporal, $Y = \{Y_t, t \in T\}$, interpretada na forma do modelo aditivo. Neste caso, escrevem-se as observações da série temporal sem a componente de tendência no seguinte modelo:

$$Y_t = S_j + \varepsilon_{ij} \quad , i = 1, \dots, n_j, j = 1, \dots, k.$$

Serão testadas as seguintes hipóteses no teste de Fisher:

H_0 : não existe sazonalidade

H_1 : existe sazonalidade

Considere a estatística dada por

$$g = \frac{\max_{j=1}^{N/2} I_j}{\sum_{j=1}^{N/2} I_j}$$

em que

I_j : valor do periodograma no período j ;

N : número de observações da série.

A estatística do teste é dada por

$$z_\alpha = 1 - \left(\frac{\alpha}{n} \right)^{\frac{1}{n-1}}$$

Na qual α é o nível de significância do teste e $n = \frac{N}{2}$.

Se $g > z_\alpha$, rejeita-se H_0 , isto é, a série apresenta sazonalidade no período j . Caso contrário, se aceita H_0 .

No caso de ser diagnosticada e estimada a sazonalidade ($\hat{S}_t$), para retirar essa componente basta realizar a diferença que é apresentada em (1) quando o modelo é aditivo e (2) quando o modelo é multiplicativo.

$$Y_t^{\text{SA}} = Y_t - \hat{S}_t \quad (1)$$

$$Y_t^{\text{SA}} = Y_t / \hat{S}_t \quad (2)$$

2.3. Metodologia de Box e Jenkins em séries temporais

Em Morettin & Toloi (2006) é apresentada uma importante metodologia adotada na análise de modelos paramétricos, denominada abordagem de Box e Jenkins. Tal metodologia consiste em ajustar modelos autorregressivos e de médias móveis a uma série temporal estacionária, em que p é a ordem do modelo autorregressivo $AR(p)$ e q é a ordem do modelo de médias móveis $MA(q)$. Sua forma geral é dada por $\phi(B)Y_t = \theta(B)\varepsilon_t$, em que B é o operador de translação para o passado, tal que $BY_t = Y_{t-1}$, $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ e $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ são, respectivamente, o polinômio autorregressivo de ordem p e o polinômio de médias móveis de ordem q e ε_t é o ruído branco, ou seja, uma sequência de erros aleatórios, independentes e identicamente distribuídos, com média zero e variância constante.

2.3.1 Modelagem ARIMA (p, d, q)

ARIMA é o nome dado a um modelo muito utilizado na modelagem e nas previsões de séries temporais. O termo deriva do inglês *autoregressive integrated moving average*, que significa modelo autorregressivo integrado de média móvel. O modelo foi sistematizado em 1976, pelos estatísticos George Box e Gwilym Jenkins, o que o tornou conhecido também por Modelo de Box-Jenkins (MORETTIN; TOLOI, 2006).

O modelo ARIMA é uma generalização do modelo autorregressivo de média móvel (ARMA). A representação ARIMA (p, d, q) refere-se, respectivamente, às ordens de autorregressão (p), de integração (d) e de média móvel (q).

Chama a atenção, na literatura, que a ordem d , comumente nomeada de integração, se refere ao número de diferenças necessárias tomadas na série até torna-lá estacionária.

No caso de $d = 0$, advém o modelo ARMA (p, q) , modelo autorregressivo de médias móveis, dado por

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p},$$

em que, no caso do modelo ARIMA, no qual $d = 0$ e $q = 0$, tem-se o modelo AR(p), modelo autorregressivo de ordem p , dado por:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t.$$

Já no caso de $d = 0$ e $p=0$ no modelo ARIMA (p, d, q) , aparece como resultado o modelo MA(q), média móvel q , dado por

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}.$$

2.3.2. Modelagem ARIMA sazonal - SARIMA

De acordo com Morettin e Toloi (2006), um dos modelos mais utilizados que consideram a sazonalidade de uma determinada série temporal é o chamado modelo ARIMA sazonal ou SARIMA. Esses modelos são importantes, pois levam em consideração a sazonalidade estocástica dos dados. Quando o

período $s=12$, o modelo denominado SARIMA de ordem $(p; d; q) \times (P; D; Q)_{12}$ é dado por

$$\phi(B)\Phi(B^{12})\Delta^d\Delta_{12}^D Y_t = \theta(B)\Theta(B^{12})\varepsilon_t,$$

em que $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ é o operador autorregressivo (AR) de ordem p ; $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ é o operador médias móveis (MA) de ordem q ; $\Phi(B^{12}) = 1 - \Phi_1 B^{12} - \Phi_2 B^{24} - \dots - \phi_p B^{12P}$ é o operador AR-sazonal de ordem P ; $\Theta(B^{12}) = 1 - \Theta_1 B^{12} - \Theta_2 B^{24} - \dots - \Theta_p B^{12Q}$ é o operador MA-sazonal de ordem Q ; $\Delta^d = (1 - B)^d$ é o operador diferença; $\Delta_{12}^D = (1 - B^{12})^D$ é o operador diferença sazonal e ε_t é um ruído branco.

2.3.3. Função de autocorrelação de um processo autorregressivo

A função de autocorrelação (*fac*) de um processo autorregressivo é construída de uma mistura de polinômios, exponenciais e senoides amortecidas. Para encontrar a função de autocorrelação (ρ_j), necessita-se, primeiramente, da

função de autocovariância (γ_Γ), uma vez que $\rho_\Gamma = \frac{\gamma_\Gamma}{\gamma_0}$, tal que

$$\begin{aligned}\gamma_\Gamma &= E(Y_t Y_{t-\Gamma}) = E[(\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t) Y_{t-\Gamma}] \\ \gamma_\Gamma &= \phi_1 E(Y_{t-1} Y_{t-\Gamma}) + \phi_2 E(Y_{t-2} Y_{t-\Gamma}) + \dots + \phi_p E(Y_{t-p} Y_{t-\Gamma}) + \underbrace{E(\varepsilon_t Y_{t-\Gamma})}_0\end{aligned}$$

$$\gamma_\Gamma = \phi_1 \gamma_{\Gamma-1} + \phi_2 \gamma_{\Gamma-2} + \dots + \phi_p \gamma_{\Gamma-p}.$$

Logo, $\rho_\Gamma = \frac{\gamma_\Gamma}{\gamma_0} = \phi_1 \gamma_{\Gamma-1} + \phi_2 \gamma_{\Gamma-2} + \dots + \phi_p \gamma_{\Gamma-p}$. Observe que

$$\gamma_0 = \text{var}(Y_t) = \phi_1 \gamma_{-1} + \phi_2 \gamma_{-2} + \dots + \phi_p \gamma_{-p} + \sigma_e^2 \quad \text{e,} \quad \text{como} \quad \gamma_j = \gamma_{-j},$$

dividindo-se por γ_0 , tem-se $1 = \phi_1 \rho_1 + \phi_2 \rho_2 + \dots + \phi_p \rho_p + \frac{\sigma_e^2}{\gamma_0}$. Logo,

$$\text{var}(Y_t) = \frac{\sigma_e^2}{1 - \phi_1 \rho_1 - \phi_2 \rho_2 - \dots - \phi_p \rho_p}, \text{ se } \Gamma = 1, \dots, p \text{ resulta nas equações}$$

de Yule-Walker:

$$\begin{cases} \rho_1 = \phi_1 + \phi_2 \rho_1 + \phi_3 \rho_2 + \dots + \phi_p \rho_{p-1} \\ \rho_2 = \phi_1 \rho_1 + \phi_2 + \phi_3 \rho_2 + \dots + \phi_p \rho_{p-2} \\ \cdot \\ \cdot \\ \cdot \\ \rho_p = \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \phi_3 \rho_{p-3} + \dots + \phi_p \cdot \end{cases}$$

Na forma matricial,

$$\begin{bmatrix} 1 & \rho_1 & \cdot & \cdot & \cdot & \rho_{p-1} \\ \rho_1 & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_{p-1} & \rho_{p-2} & \cdot & \cdot & \cdot & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \cdot \\ \cdot \\ \cdot \\ \phi_p \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \cdot \\ \cdot \\ \cdot \\ \rho_p \end{bmatrix}$$

As equações de Yule-Walker são utilizadas para estimar os coeficientes $\phi_1, \phi_2, \phi_3, \dots, \phi_p$ do modelo $AR(p)$ sendo ρ_Γ substituído por r_Γ .

2.3.4. Função de autocorrelação parcial

A função de autocorrelação parcial (*facp*) facilita o processo de determinação da ordem dos modelos. Denota-se ϕ_{jk} o j -ésimo coeficiente de um modelo $AR(k)$, de modo que ϕ_{kk} seja o último coeficiente, isto é,

$$Y_t = \phi_{1k} Y_{t-1} + \phi_{2k} Y_{t-2} + \dots + \phi_{kk} Y_{t-k} + \varepsilon_t .$$

Sabe-se que $\rho_\Gamma = \phi_{1k} \gamma_{\Gamma-1} + \phi_{2k} \gamma_{\Gamma-2} + \dots + \phi_{kk} \gamma_{\Gamma-p}$, $j=1, \dots, k$, a partir das quais se obtêm as equações de Yule-Walker.

$$\begin{bmatrix} 1 & \rho_1 & \cdot & \cdot & \cdot & \rho_{k-1} \\ \rho_1 & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_{k-1} & \rho_{k-2} & \cdot & \cdot & \cdot & 1 \end{bmatrix} \begin{bmatrix} \phi_{1k} \\ \phi_{2k} \\ \cdot \\ \cdot \\ \cdot \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \cdot \\ \cdot \\ \cdot \\ \rho_k \end{bmatrix}$$

Resolvendo estas equações sucessivamente para $k=1,2,\dots$, resulta em

$$\phi_{11} = \rho_1, \phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}, \phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} \text{ e } \phi_{kk} = \underbrace{\frac{\begin{vmatrix} P_k^* \\ P_k \end{vmatrix}}{\begin{vmatrix} P_k \end{vmatrix}}}_{\text{Generalizando}},$$

em que P_k é a matriz de autocorrelação e P_k^* é a matriz P_k com a última coluna substituída pelo vetor de autocorrelação.

A quantidade ϕ_{kk} , encarada como função de k , é chamada função de autocorrelação parcial (*facp*).

As equações de Yule-Walker são utilizadas para estimar os coeficientes $\phi_1, \phi_2, \phi_3, \dots, \phi_p$ do modelo $AR(p)$, sendo ρ_Γ substituído por r_Γ .

2.4. Modelo heterogêneo autorregressivo

Na literatura não se dispõe de uma vasta publicação sobre o modelo heterogêneo autorregressivo (HAR). Assim, será apresentada a ideia do modelo e suas extensões encontradas na literatura.

Comumente, em séries temporais, o modelo tem apenas uma frequência de variação temporal. Assim, a fonte é diária, ou semanal, ou trimestral, ou anual, entre outras, na qual se tem como objetivo analisar e modelar a dependência temporal. Exemplo: modelo autorregressivo de ordem um (AR(1))

- $Y_t = \phi_1 Y_{t-1} + \varepsilon_t$ (3)
- Com ϕ_1 constante.

O significado prático de (3) é explicado pela dependência apenas de um retardo, ou seja, depende apenas da variação temporal anterior, isto é, se os dados são diários, então, teve dependência do dia anterior; ou, se os dados são anuais, teve dependência do ano anterior, o que é explicado pelo índice $(t-1)$. Caso estivesse dependendo de dois retardos anteriores no índice, representaria $(t-2)$. Outros exemplos são

$$AR(2): Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$

$$AR(3): Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \varepsilon_t$$

A partir de estudos do professor Dr. Fulsio Corsi, da Universidade de Veneza, foi desenvolvido o modelo heterogêneo autorregressivo (HAR). O

sentido de heterogêneo no HAR está explicado pela adição no mesmo modelo de diferentes frequências de variações temporais, sendo elas dias, semanas e meses. O sentido de autorregressivo no HAR está explicado pela dependência temporal das observações anteriores.

2.4.1. O modelo HAR(h)

O modelo HAR(h) é um processo de memória longa que se caracteriza pela dependência significativa entre observações distantes (em tempo). Nesse caso, existe persistência nas autocorrelações observadas, que decaem lentamente para zero.

O HAR(h) serve para agregar diferentes frequências de tempo (dia, semana e mês) no mesmo modelo. Para ajustar o HAR (h) são necessárias séries temporais diárias que, a princípio, passam pelo processo

$$z_{t,h} = \frac{z_t + z_{t-1} + z_{t-2} + \dots + z_{t-h+1}}{h}, \quad (4)$$

em que $z_{t,h}$ é a série de médias das últimas h observações.

Os valores típicos de h na equação (4) são 1 para dados diários, 7 para dados semanais e 28 para dados mensais.

Considere a série diária de n observações. Assim,

- Substituindo $h = 1$ em (4), tem-se

$$z_{t,1} = z_t,$$

ou seja, a mesma série em estudo, com n observações.

- Substituindo $h = 7$ em (4), tem-se

$$z_{t,7} = \frac{z_t + z_{t-1} + z_{t-2} + \dots + z_{t-7+1}}{7},$$

assim, obtém-se a média dos últimos 7 dias e se origina uma nova série com $(n-6)$ dias.

- Substituindo $h = 28$ em (4)

$$z_{t,28} = \frac{z_t + z_{t-1} + z_{t-2} + \dots + z_{t-28+1}}{28}$$

obtém-se a média dos últimos 28 dias e se origina uma nova série com $(n-27)$ dias.

A seguir apresenta-se um exemplo fictício para ilustração.

Dada a série com os números naturais $(1, 2, 3, \dots, 365)$, deseja-se obter as séries transformadas, por meio da Equação 4, para $h = 1$, $h = 7$ e $h = 28$. Assim, têm-se as séries apresentadas na Tabela 1.

Observa-se, na Tabela 1, que, a partir da série diária, se obtêm duas séries de médias com características do processo de memória longa semanal e mensal.

A série $h = 1$ será a mesma série diária, com as 365 observações.

A série $h = 7$ tem $(365-6)=359$ observações, ou seja, trabalha-se com a média das sete observações anteriores. O primeiro valor da série $h = 7$, que é igual a 4, é obtido da média das sete primeiras observações, ou seja,

$$z_{t,7} = \frac{7+6+5+4+3+2+1}{7} = 4$$

O segundo valor da série $h = 7$ é obtido por

$$z_{t,7} = \frac{8+7+6+5+4+3+2}{7} = 5$$

Tabela 1: Exemplo fictício de séries transformadas

Série	$h = 1$	$h = 7$	$h = 28$
1	1		
2	2		
3	3		
4	4		
5	5		
6	6		
7	7	4	
8	8	5	
9	9	6	
10	10	7	
.	.	.	
.	.	.	
.	.	.	
27	27	24	
28	28	25	14,5
29	29	26	15,5
30	30	27	16,5
.	.	.	.
.	.	.	.
.	.	.	.
365	365	362	351,5

Assim, a ideia se repete até a última observação da série original. Seguindo o mesmo raciocínio da série $h = 7$, obtém-se a série $h = 28$. A única diferença é que se têm a média das 28 últimas observações e um total de $(365 - 27) = 337$ observações na série.

Uma observação fundamental para a validação do HAR(h) é que, após as transformações em $h = 1$, $h = 7$ e $h = 28$, todas as séries têm que ser estacionárias para o ajuste do modelo.

O modelo HAR(h) é útil para explicar o fenômeno temporal, ou seja, a dependência de uma variável endógena (dependente) em função da exógena (independente), em diferentes intervalos de tempos (dias, meses e anos) no mesmo modelo.

Quando se faz a modelagem apenas para $h = 1$, nomeia-se de HAR(1). Quando se faz a modelagem com dados diários e semanais denomina-se de HAR(1,7) e, quando, no modelo, apresentam-se dados diários, semanais e mensais, denomina-se de HAR(1,7, 28).

Na aplicação empírica do modelo HAR(h), tem-se, conforme exposto em Chang e McAleer (2014), considerando z_t e x_t , respectivamente série dependente e série independente:

$$z_t = \phi_1 + \phi_{21}z_{t-1} + \phi_{22}x_{t-1} + \varepsilon_t$$

(5)

$$z_t = \phi_1 + \phi_{21}z_{t-1} + \phi_{22}x_{t-1} + \phi_{31}z_{t-1,7} + \phi_{32}x_{t-1,7} + \varepsilon_t \quad (6)$$

$$z_t = \phi_1 + \phi_{21}z_{t-1} + \phi_{22}x_{t-1} + \phi_{31}z_{t-1,7} + \phi_{32}x_{t-1,7} + \phi_{41}z_{t-1,28} + \phi_{42}x_{t-1,28} + \varepsilon_t \quad (7)$$

Os modelos apresentados nas Equações (5), (6) e (7) referem-se, respectivamente, aos modelos HAR (1), HAR (1,7) e HAR (1,7,28). Os índices $t-1$; $t-1,7$; $t-1,28$ significam, respectivamente, retardos de um dia, uma semana e um mês, como também os índices $t-n$; $t-n,7$ e $t-n,28$ significam, respectivamente, retardos de n dias, n semanas e n meses. Retardo em séries temporais significa dependência temporal.

2.4.1.1. Volatilidade e o modelo HAR

Existem várias extensões do modelo HAR, tal como HAR-RV (modelo heterogêneo autorregressivo de volatilidade realizada), LHAR-CJ (modelo heterogêneos autorregressivo com volatilidade contínua e saltos), tree-HAR e HAR(h).

No trabalho de Andersen, Bollerslev e Diebold (2006) foram apresentados os resultados das volatilidades realizadas de três mercados. Resultados empíricos em ABDL (2003 apud ANDERSEN; BOLLERSLEV; DIEBOLD, 2006) sugerem fortemente que os modelos de volatilidade realizada superaram os modelos *General Autoregressive Conditional Heterocedasticity*, ou GARCH. Os dados utilizados foram: mercado à vista de câmbio (DM/\$), mercado futuro de ações do S&P500 e mercado futuro de taxas de juros de 30 anos de rendimento do Tesouro dos EUA (*T-bond*).

Corsi (2009) propõe o modelo heterogêneo autorregressivo de componentes de volatilidade, definido sobre diferentes períodos de tempo, denominado de Modelo Heterogêneo Autorregressivo de Volatilidade Realizada (HAR-RV). O conjunto de dados utilizados foram USD/CHF, S&P500 e T-bond. A série USD/CHF abrange 14 anos, no período de dezembro de 1989 a dezembro de 2003. Já para a série S&P500, a fim de evitar a modelagem do

comportamento sazonal da atividade de comércio induzido pelo fim de semana, foi excluída toda a volatilidade realizada, ocorrendo de sexta-feira 21:00 GMT a domingo 22:00 GMT. O período considerado dessa série foi de janeiro de 1990 a julho de 2007. O período da série *T-bond* é de janeiro de 1990 a outubro de 2003. As séries apresentaram uma simples estrutura e ausência de propriedades da memória longa. Os resultados das simulações mostraram que o modelo HAR-RV atinge com sucesso o propósito de reproduzir as principais características empíricas de retornos financeiros (longa memória, caudas pesadas e autossimilaridade) de forma muito simples e parcimoniosa. Além disso, os resultados empíricos mostraram um bom desempenho de previsão.

Corsi e Reno (2009) propõem uma versão estendida do modelo HAR, que considera as respostas assimétricas da volatilidade realizada, não só diariamente anterior aos retornos negativos, mas também para a sua agregação semanal e mensal. Esses autores mostram que a estrutura heterogênea se aplica ao efeito de alavanca, pois é conhecido que a volatilidade, que é uma medida de dispersão dos retornos de um título ou índice de mercado, tende a aumentar mais depois de um choque negativo do que após um choque positivo da mesma magnitude (BOLLERSLEV et al., 2006 apud CORSI; RENO, 2009). O objetivo foi analisar empiricamente o desempenho do modelo *Leverage Heterogeneous Autoregressive with Continuous volatility and Jumps* (LHAR-CJ), ambos dentro da amostra e fora da amostra. O conjunto de dados foi o S&P500. Assim, foi constatado que a presença de saltos (acréscimo ou decréscimo rápido do um título ou índice de mercado) é importante, já que tem um impacto direto sobre a dinâmica da volatilidade e que pode ser explicada pela presença de saltos no preço, o que permite um melhor ajuste da volatilidade realizada, como sugerido por Andersen et al. (2006).

Em Audrino e Corsi (2010) foi testada a precisão do modelo tree-HAR na série de correlações diárias entre S&P500 e US Bond, no período de 1990 a

2008. O modelo tree-HAR é capaz de levar em conta dois fatos importantes de correlações realizadas: forte dependência temporal (que é de memória longa) e quebras estruturais. Além disso, o modelo proposto nesse estudo é utilizado para fazer previsões e foi percebido que correlações condicionais podem ser utilizadas como um modelo de partida para o desenvolvimento de novos melhoramentos.

Wink e Pereira (2011) estudaram os modelos de estimação e previsão de volatilidade realizada, ou seja, o modelo *Heterogeneous Autoregressive Model of Realized Volatility* (HAR-RV), desenvolvido por Corsi (2009) e o *Mixed Data Sampling* (MIDAS-RV), desenvolvido por Ghysels et al. (2004 apud WINK JUNIOR; PEREIRA, 2011). O material utilizado para comparação dos modelos foram os dados intradiários dos ativos mais negociados do BOVESPA, ou seja, dados dos 5 ativos mais líquidos, na Bolsa de Valores de São Paulo (BOVESPA). Foram eles: Bradesco (BBDC4), Petrobrás (PETR4), Vale do Rio Doce (VALE5), Telemar (TNLP4) e Usiminas (USIM5). A amostra abrangeu dados intradiários em três janelas diferentes: intervalos de 5, 15 e 30 minutos. O período analisado foi de 1º de novembro de 2007 até 30 de abril de 2010. Por meio de medidas de comparação de previsão dentro e fora da amostra, constataram-se resultados superiores do modelo MIDAS-RV apenas para previsões dentro da amostra. Para previsões fora da amostra, no entanto, não houve diferença estatisticamente significativa entre os modelos. Também se encontraram evidências de que a utilização da volatilidade realizada induz distribuições dos retornos padronizados mais próximas da normal.

Chang e McAleer (2014) analisaram o quanto a ida de turistas do Japão e dos EUA a Taiwan influenciam na moeda nacional, por meio do efeito preço. Foram utilizados dados diários de 1º de janeiro de 1990 a 31 de dezembro de 2008, para modelar o preço mundial, taxas de câmbio e as chegadas de turistas e sua volatilidade associada. O Modelo Heterogêneo Autorregressivo, $HAR(h)$, foi

utilizado para aproximar as correlações lentamente em decomposição, associadas com as propriedades de memória longa em diárias e taxas de câmbio semanais e mensais.

2.5. Adequabilidade do modelo

Conforme Morettin e Tolo (2006), após o ajuste do modelo para uma série, é necessário verificar se este representa bem o conjunto de dados. Esta verificação é realizada analisando-se os resíduos. Caso o modelo seja adequado, o resíduo é ruído branco, o que significa que o resíduo é não correlacionado.

Dentre vários testes de adequação de modelo, destaca-se o teste de Box-Pierce, que considera as autocorrelações dos resíduos, conforme mostrado em (8).

$$Q(K) = n(n+2) \frac{\sum_{j=1}^K \hat{r}_j^2}{(n-j)}, \quad (8)$$

em que n , $\hat{r}_j^2$ e K representam, respectivamente, o número de observações, a função de autocorrelação e o número de defasagens tomadas.

Se o modelo é adequado, a estatística apresentada em (8) terá, aproximadamente, uma distribuição χ^2 , com $k - p - q$ graus de liberdade. A hipótese de ruído branco é rejeitada para valores grandes de $Q(K)$, ou seja, se $Q(K) < \chi_{k-p-q}^2$, então, o resíduo é ruído branco.

2.6. Mortalidade e clima

A relação entre fatores climáticos e saúde humana na cidade de São Paulo, Brasil, tem sido do interesse público e objeto de vários estudos. Entre eles, destacam-se, na literatura, Saldiva et al. (1994), Saldiva et al. (1995), Miraglia et al. (1997), Pereira et al. (1998), Braga et al. (1999), Lin et al. (1999), Conceição et al. (2001), Gouveia et al. (2003) e Sáfiadi e Peña (2008).

Em Saldiva et al. (1994) foram estudadas séries temporais diárias de poluição atmosférica (umidade relativa, temperatura, monóxido de carbono (CO), dióxido de enxofre(SO₂), material particulado (PM₁₀), ozônio (O₃) e óxidos de nitrogênio (NO_x)) e mortalidade, por problemas respiratórios, em crianças menores de cinco anos de idade, no período de maio de 1990 a abril de 1991. Utilizando métodos de regressão, foi demonstrada uma associação significativa entre os níveis de óxidos de nitrogênio e mortalidade, devido a doenças respiratórias.

Nos estudos de Saldiva et al. (1995) trabalhou-se com séries temporais diárias de poluição atmosférica (dióxido de enxofre (SO₂), material particulado (PM₁₀), ozônio (O₃) e óxidos de nitrogênio (NO_x)) e mortalidade em idosos maiores de 65 anos de idade, no período de maio de 1990 a abril de 1991. Utilizando métodos de regressão, foi demonstrada uma associação significativa entre os níveis de material particulado (PM₁₀) e a mortalidade em idosos.

Miraglia et al. (1997) abordaram o impacto do tipo de combustível (etanol e combustíveis fósseis) da poluição atmosférica na saúde humana. Foram consideradas séries de mortalidade por causa respiratória e cardiovascular em idosos em São Paulo, no período de janeiro de 1991 a dezembro de 1994. Por meio da regressão linear, os resultados mostraram uma associação significativa entre a poluição do ar e mortalidade, devido a doenças respiratórias.

Pereira et al. (1998) utilizaram técnicas de regressão de Poisson para associações entre contagens diárias das concentrações de poluente intrauterino (NO₂, SO₂, CO, O₃) e mortalidade. As séries foram investigadas para o período

de janeiro de 1991 até dezembro de 1992, na cidade de São Paulo, Brasil. A associação entre mortalidade intrauterina e poluição atmosférica foi forte para NO₂, mas menor para SO₂ e CO. Assim, os resultados sugerem que a poluição do ar em São Paulo pode promover efeitos adversos para a saúde em fetos.

Braga et al. (1999) objetivaram verificar a relação entre séries de poluição atmosférica (níveis diários de PM₁₀, O₃, SO₂, CO e NO₂) e mortalidade de crianças (menores de 13 anos de idade), em São Paulo, Brasil. Analisaram-se as séries temporais de registros diários de admissões, de crianças, do hospital em São Paulo e níveis de poluição, de outubro de 1992 a outubro de 1993. Utilizando modelos de regressão, foi averiguado que PM₁₀ e O₃ foram os poluentes que exibiram a associação mais robusta com os problemas respiratórios.

Lin et al. (1999) analisaram internatos pediátricos por emergência na cidade de São Paulo, obtidos do Instituto da criança da Universidade de São Paulo e registros diários de CO, SO₂, PM₁₀, O₃ e NO₂, no período de maio de 1991 a abril de 1993. A relação entre as visitas de emergência por problemas respiratórios e a poluição do ar foi avaliada por meio da estatística comparativa simples, análise de correlação simples e, assim, estimando-se uma variedade de modelos de regressão. Observaram-se as associações mais robustas com PM₁₀ e, em menor medida, com O₃. Estas associações eram estáveis por meio de diversas variáveis de controle e especificações de modelo diferente. Assim, foi constatado que a poluição do ar é uma preocupação de saúde pediátrica substancial em São Paulo.

Conceição et al. (2001) avaliaram a associação entre poluição atmosférica, entre eles, PM₁₀, SO₂, CO, O₃, temperatura e umidade relativa do ar, além de marcadores de morbimortalidade na cidade de São Paulo, com séries diárias no período de 1994 a 1997. Utilizaram-se os Modelos Lineares Generalizados (MLG) e os Modelos Aditivos Generalizados (MAG).

Observaram-se associações entre mortalidade e os níveis de CO, SO₂ e, em menor escala, PM₁₀. As relações observadas foram dose-dependente e evidentes após um curto período de exposição.

Gouveia et al. (2003) utilizaram séries temporais diárias sobre mortalidade, internações hospitalares, níveis atmosféricos dos principais poluentes do ar e de variáveis meteorológicas das cidades brasileiras São Paulo e Rio de Janeiro. Estes dados foram analisados utilizando-se técnicas de análise de séries temporais em modelos lineares por meio de Equações de Estimção Generalizada e/ ou por meio de modelos não paramétricos, com a utilização de Modelos Aditivos Generalizados. Assim, foram encontradas relações significantes entre aumentos nos níveis de poluentes atmosféricos e aumentos na mortalidade e nas hospitalizações, por causas respiratórias e cardiovasculares, em crianças e em idosos.

Sáfadi e Peña (2008), utilizando o modelo fatorial dinâmico nas séries de temperatura mínima, umidade relativa do ar, PM₁₀ e CO, mortalidade por problemas respiratórios e cardiovasculares, encontraram forte relação entre poluentes atmosféricos (PM₁₀), umidade relativa do ar e mortalidade por problemas respiratórios na cidade de São Paulo, no período de 1994 a 1997.

Entre os fatores climáticos, podem-se destacar temperaturas mínimas, material particulado, monóxido de carbono e umidade relativa do ar, utilizados nesta dissertação.

Segundo dados do Instituto Nacional de Meteorologia, a cidade de São Paulo tem a terceira menor temperatura, comparada com as outras capitais dos estados do Brasil.

O material particulado (pm₁₀) é um tipo de partícula inalável e constitui um elemento de poluição atmosférica que pode penetrar no aparelho respiratório, provocando inúmeras doenças respiratórias.

O monóxido de carbono (co) é um gás levemente inflamável, incolor, inodoro e muito perigoso, devido à sua grande toxicidade. É produzido pela queima em condições de pouco oxigênio (combustão incompleta) e/ou alta temperatura de carvão ou outros materiais ricos em carbono, como derivados de petróleo. Forma, com a hemoglobina do sangue, um composto mais estável do que ela e o oxigênio, podendo levar à morte por asfixia.

A umidade relativa do ar significa, em termos simplificados, o quanto de água na forma de vapor existe na atmosfera no momento, em relação ao total máximo que poderia existir, na temperatura observada. A umidade do ar é mais baixa, principalmente no final do inverno e no início da primavera, no período da tarde, entre 12 e 16 horas. Entre os problemas decorrentes da baixa umidade do ar estão complicações alérgicas e respiratórias, devido ao ressecamento de mucosas.

Os padrões de qualidade do ar para o estado de São Paulo (Decreto Estadual nº. 8468 de 08/09/76) são:

- PM10..... 240 $\mu\text{g}/\text{m}^3$.
- CO.....40.000 $\mu\text{g}/\text{m}^3$ ou 35 ppm.

3. MATERIAL E MÉTODOS

3.1. Dados

Foram analisadas seis séries temporais diárias obtidas na cidade de São Paulo, SP, Brasil, que são

- número de óbitos por problemas respiratórios;
- número de óbitos por problemas cardiovasculares;
- temperaturas mínimas;
- material particulado (pm10);
- monóxido de carbono e
- umidade do ar.

As séries temporais foram adquiridas do projeto Poluição Atmosférica e Saúde Social (PASSO) na cidade de São Paulo, disponíveis na *home page* do professor Dr. Julio da Motta Singer, da Universidade de São Paulo.

3.1.1. Período das séries temporais

Num primeiro instante foram consideradas as séries temporais no período de 1º de janeiro de 1994 a 31 de dezembro de 1997, totalizando 1.460 observações de cada uma das seis séries temporais.

Para modelar o $HAR(h)$ foi considerado apenas o período de 1º de janeiro de 1997 a 31 de dezembro de 1997, totalizando 365 observações de cada uma das seis séries temporais.

3.2. Software estatístico

O software livre denominado de *Gretl* (acrônimo de *Gnu Regression, Econometrics and Time-series Library*) foi utilizado nas análises, bem como para a produção dos gráficos das séries temporais estudadas.

3.3. Métodos

Em Chang e McAleer (2014) é encontrado o referencial em que se baseia o trabalho desta dissertação. Considerando as séries temporais z_t e x_t , nomeiam-se as séries diárias para modelar o HAR(h), como

z_{1t} : número de óbitos por problemas respiratórios;

z_{2t} : número de óbitos por problemas cardiovasculares;

x_{1t} : temperaturas mínimas;

x_{2t} : material particulado;

x_{3t} : monóxido de carbono e

x_{4t} : umidade relativa do ar.

Assim, cada uma das séries estudadas passou pelos passos de 1 a 6, que foram:

1. visualização e discussão dos gráficos das séries no período de 1º de janeiro de 1994 a 31 de dezembro de 1997;

2. visualização e discussão dos gráficos das séries no período de 1º de janeiro de 1997 a 31 de dezembro de 1997;
3. as séries analisadas anteriormente passaram pela transformação definida na Equação (4), considerando $h = 1$, $h = 7$ e $h = 28$. Assim, foram definidas seis séries com 365 observações, seis séries com $(365-6)=359$ observações e seis séries com $(365-27)=338$ observações, totalizando 18 séries temporais no estudo;
4. as séries definidas no item anterior passaram pelo processo de diferenciação;
5. foi realizada a modelagem HAR(h), mais especificamente com o modelo de séries temporais VAR (autorregressão vetorial) encontrado no software *Gretl*;
6. efetuou-se a análise do resíduo para a verificação de ruído branco.

4. RESULTADOS E DISCUSSÃO

4.1 Análises das séries temporais em estudo

No que segue será feita uma análise descritiva das séries utilizadas no trabalho no período de 1994 a 1997. Dado o caráter sazonal das séries e a ausência de tendência entre os anos optou-se por utilizar apenas o ano de 1997 no ajuste do modelo HAR.

4.1.1 Séries temporais no período janeiro de 1994 a dezembro de 1997

Num primeiro momento, analisaram-se as séries do número de óbitos por problemas respiratórios e cardiovasculares, no período de 1º de janeiro de 1994 a 31 de dezembro de 1997, as quais são apresentadas na Figura 5.

Pode-se observar, por meio da Figura 5, que o número de óbitos, tanto por problemas respiratórios quanto por problemas cardiovasculares, não se desenvolve ao redor de uma média constante, ou seja, o processo não é estacionário, possuindo sazonalidade de 365 dias.

O maior número de óbitos na cidade de São Paulo, no período analisado, ocorreu, principalmente, no mês de junho. O número máximo de óbitos por problemas respiratórios e por problemas cardiovasculares está compreendido, respectivamente, nos intervalos [50, 55) e [90, 100), e como número mínimo de óbitos, entre os intervalos [5, 10) e [20, 30), respectivamente por problemas respiratórios e cardiovasculares.

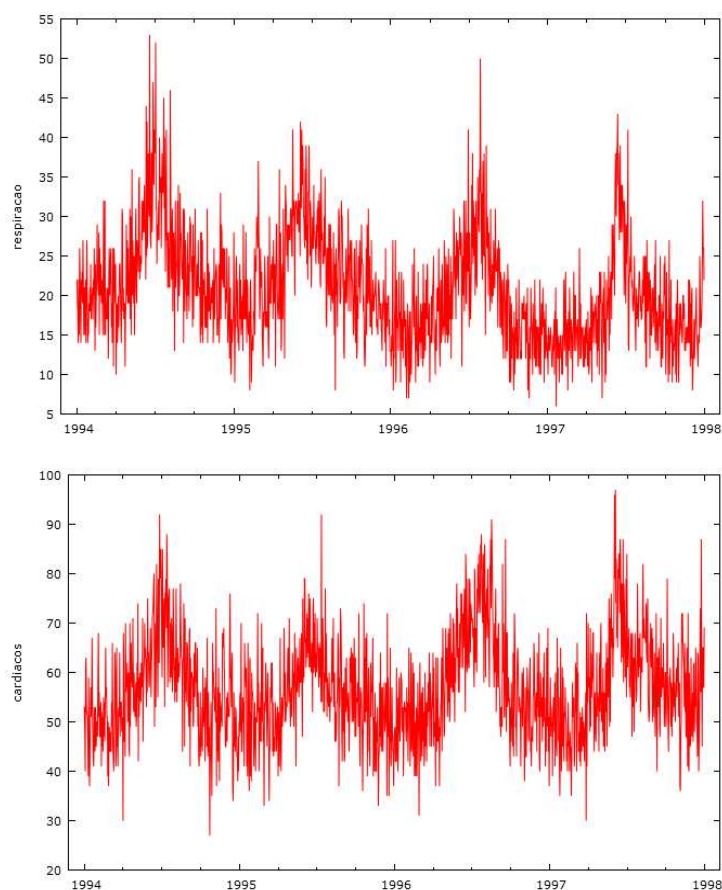


Figura 5: Número diário de óbitos por problemas respiratórios e cardiovasculares (01/01/94 a 31/12/97).

Pode-se observar, na Figura 6, que as séries de temperaturas mínimas (temp_min), umidade (humid_med), monóxido de carbono (co) e material particulado (pm10), analisadas no período de 1º de janeiro de 1994 a 31 de dezembro de 1997, também não se desenvolvem ao redor de uma média constante, ou seja, o processo não é estacionário e pode ser diagnosticada sazonalidade de 365 dias.

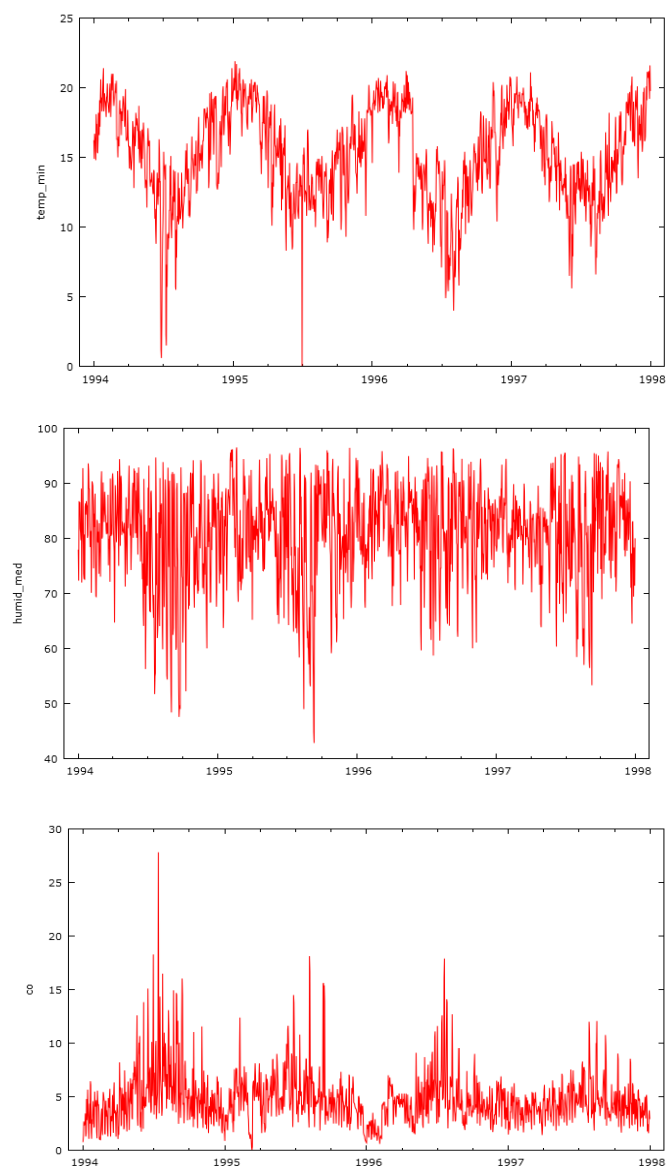
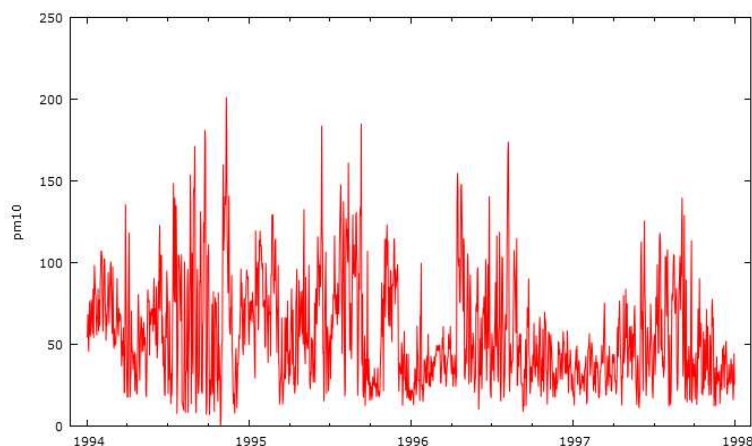


Figura 6: Séries de temperaturas mínimas (temp_min), umidade (humid_med), monóxido de carbono (co) e material particulado (pm10), na cidade de São Paulo.



Continuação Figura 6: Séries de temperaturas mínimas (temp_min), umidade (humid_med), monóxido de carbono (co) e material particulado (pm10), na cidade de São Paulo.

De acordo com a Figura 6, o mês considerado mais crítico para o período analisado é junho, pois, neste mês, houve um aumento ou um decréscimo considerável, comparado aos outros meses dos anos de 1994 a 1997.

A série de temperaturas mínimas tem os intervalos [0° C, 5° C) e [25 ° C, 30 ° C) como valores mínimos e máximos, respectivamente. Pode-se observar que o ano de 1997 teve a temperatura mínima no mês de junho maior, comparada à dos anos 1994, 1995 e 1996.

Já para a série monóxido de carbono (co), os intervalos [0 ppm, 5 ppm) e [20 ppm, 25 ppm) foram seus valores mínimos e máximos, respectivamente. Pode-se observar que o mês de junho 1997 teve menor quantidade do poluente, comparado aos outros anos analisados.

Já a série umidade relativa do ar (hum) apresentou os intervalos [40%, 50%) e [90%, 100%), respectivamente, como valores mínimos e máximos. Pode-se observar que o mês de junho de 1995 teve menor taxa de umidade,

comparado aos outros anos analisados e o mês de junho 1996 teve a maior taxa de umidade.

Para a série de material particulado (pm10) os intervalos $[0 \mu\text{g}/\text{m}^3, 50\mu\text{g}/\text{m}^3)$ e $[200 \mu\text{g}/\text{m}^3, 250 \mu\text{g}/\text{m}^3)$ foram seus valores mínimos e máximos, respectivamente. Pode-se observar que o ano de 1994 teve maior índice de pm10 e o ano de 1997, o menor índice de pm10 nos meses de junho e julho.

Pode-se observar que, enquanto as séries de números de óbitos por problemas respiratórios e cardiovasculares aumentam, as séries de temperaturas mínimas e umidade relativa do ar diminuem e os índices de monóxido de carbono e material particulado crescem.

4.1.2 Séries temporais no período janeiro de 1997 a dezembro de 1997

Pode-se observar, por meio da série temporal (Figura 7), que o maior número de óbitos na cidade de São Paulo ocorreu no mês de junho, ou seja, entre 40 e 45 óbitos, e que a série tende a crescer até o início de junho e a decrescer lentamente no restante dos meses do ano de 1997. Assim, a série não é dita estacionária.

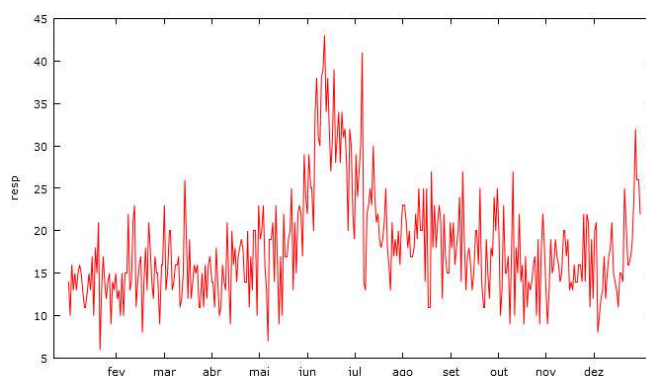


Figura 7: Número de óbitos por problemas respiratórios na cidade de São Paulo, de 1º de janeiro de 1997 a 31 de dezembro de 1997

Num segundo momento, analisou-se a série original do número de óbitos por problemas cardiovasculares.

Na Figura 8 é possível verificar que os maiores números de óbitos por problemas cardiovasculares se encontram no início de junho e os menores índices, no final de março. A série não é estacionária, já que os números de óbitos no decorrer do ano não são constantes ao redor de uma média.

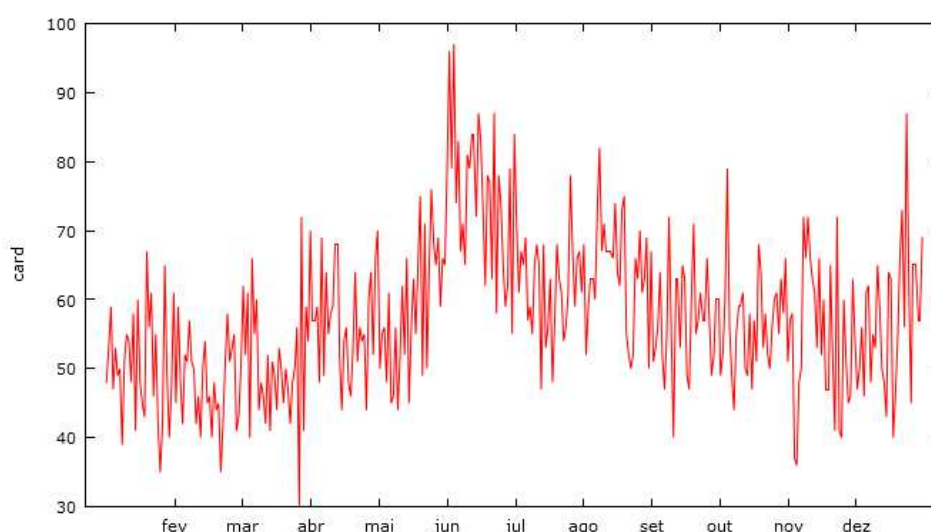


Figura 8: Número de óbitos por problemas cardiovasculares na cidade de São Paulo, de 1º de janeiro de 1997 a 31 de dezembro de 1997

O gráfico da série temporal de temperaturas mínimas na cidade de São Paulo encontra-se na Figura 9.

Na Figura 9 é possível verificar que as maiores temperaturas mínimas, entre 20 °C e 22 °C, são encontradas nos extremos do ano de 1997, ou seja, nos meses de janeiro e dezembro e as mais baixas temperaturas mínimas, entre 4 °C e 8 °C, nos meses junho e agosto. A série também não é classificada como estacionária, decorrente do decrescimento e do crescimento da mesma.

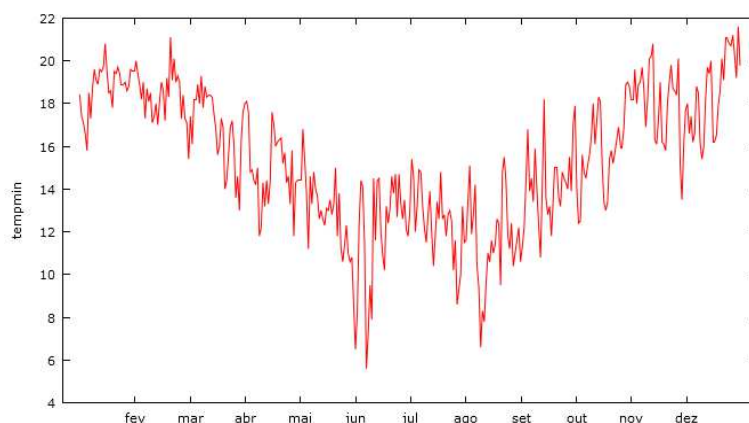


Figura 9: Temperatura mínima da cidade de São Paulo, de 1º de janeiro de 1997 a 31 de dezembro de 1997

Para a análise da série de material particulado (pm10) foi gerado o gráfico da Figura 10. Verifica-se que o maior índice de pm10, por volta de $140 \mu\text{g}/\text{m}^3$, é encontrado no mês de setembro e vários meses mantêm o menor índice de pm10, entre 10 e $20 \mu\text{g}/\text{m}^3$, o qual é visto em todo o ano de 1997. A série não é estacionária.

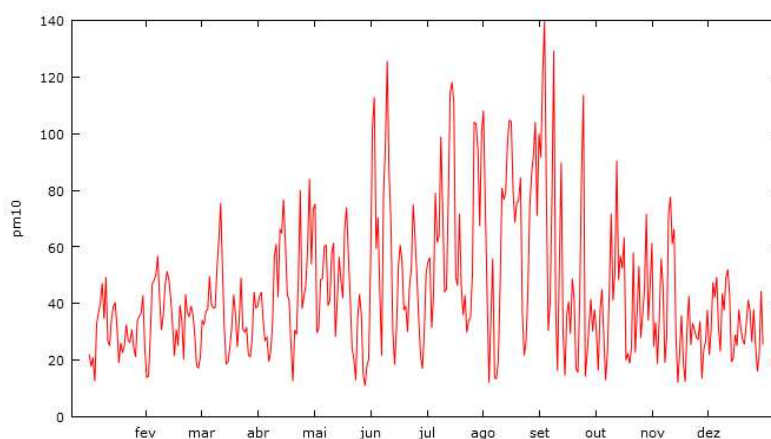


Figura 10: Índice de PM10 da cidade de São Paulo, de 1º de janeiro de 1997 a 31 de dezembro de 1997.

No gráfico da Figura 11 é possível constatar que as maiores taxas de monóxido de carbono (co), em torno de 12 ppm, ocorreram nos meses de julho e agosto. Não é possível identificar a menor taxa, já que a mesma é repetida no decorrer do ano analisado. A série temporal não é estacionária, pois a variações de co vão de 1 ppm a 13 ppm.

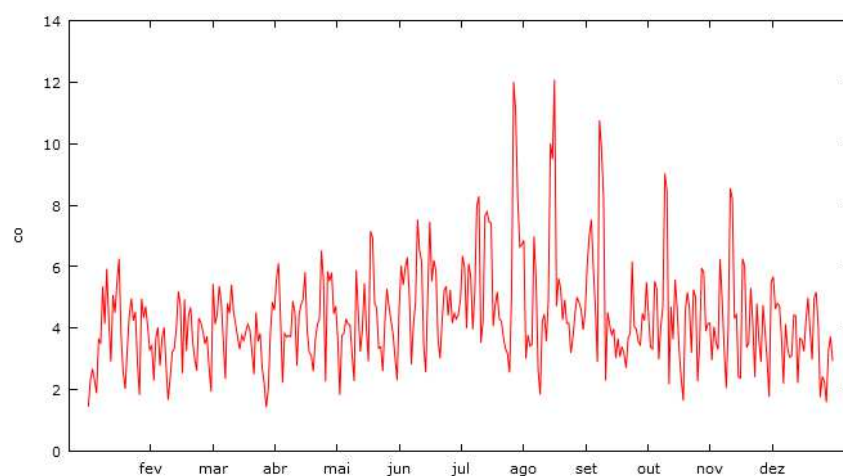


Figura 11: Índices de CO da cidade de São Paulo, de 1º de janeiro de 1997 a 31 de dezembro de 1997

No gráfico da Figura 12 é possível observar que os menores índices de umidade foram encontrados no mês de setembro, em torno de 55%. Já os maiores índices de umidade se encontram em diferentes meses do ano, os quais giram em torno de 95%. A série temporal não é estacionária, pois há crescimento e decrescimento da mesma no decorrer do ano.

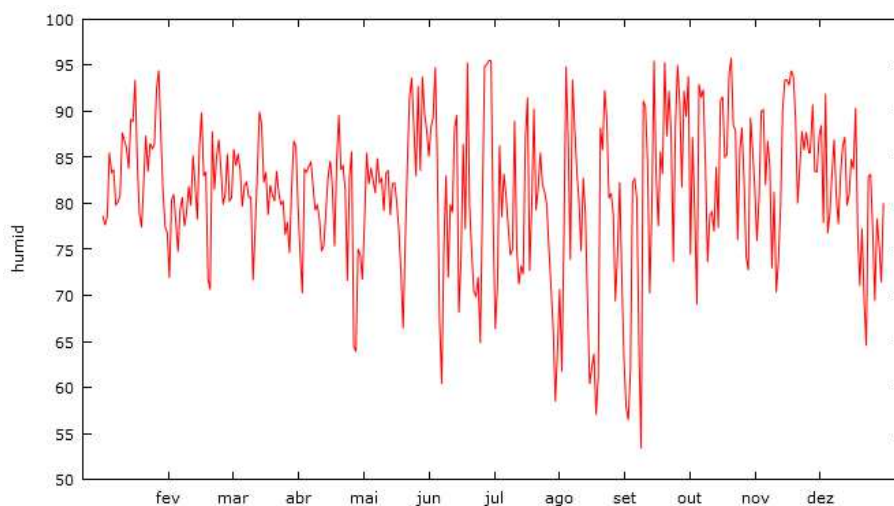


Figura 12: Índices de umidade da cidade de São Paulo, de 1^o de janeiro de 1997 a 31 de dezembro de 1997.

4.1.2 Séries temporais transformadas

Nas próximas subseções serão obtidas as séries de médias através da Equação 4, para valores de h iguais a 1, 7 e 28.

4.1.2.1. Séries temporais transformadas com $h = 1$

Para esta transformação (Figuras 13 e 14), utiliza-se a fórmula $z_{t,1} = z_t$ (Equação 4), ou seja, a série continua com as mesmas observações e características.

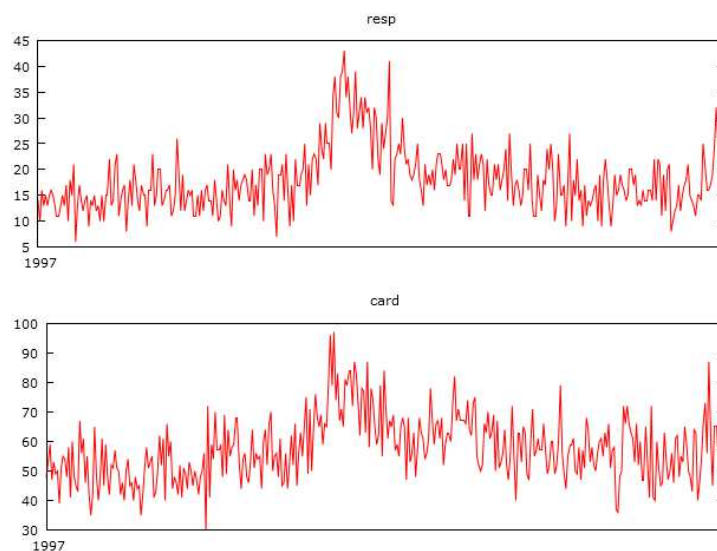


Figura 13: Séries do número de obtidos para $h=1$ do número.

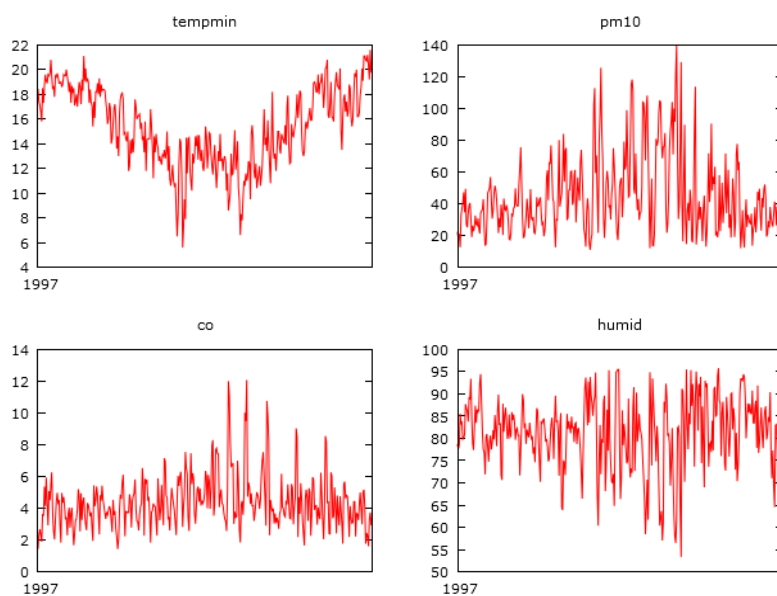


Figura 14: Séries climáticas para $h=1$.

4.1.2.2. Séries temporais transformadas com $h = 7$

Para esta transformação (Figuras 15 e 16), utiliza-se a fórmula

$$z_{t,7} = \frac{z_t + z_{t-1} + z_{t-2} + \dots + z_{t-7+1}}{7} \text{ (Equação 4), ou seja, a série tem (365-6)}$$

observações e cada uma delas passa a ser a média dos sete dias anteriores à observação.

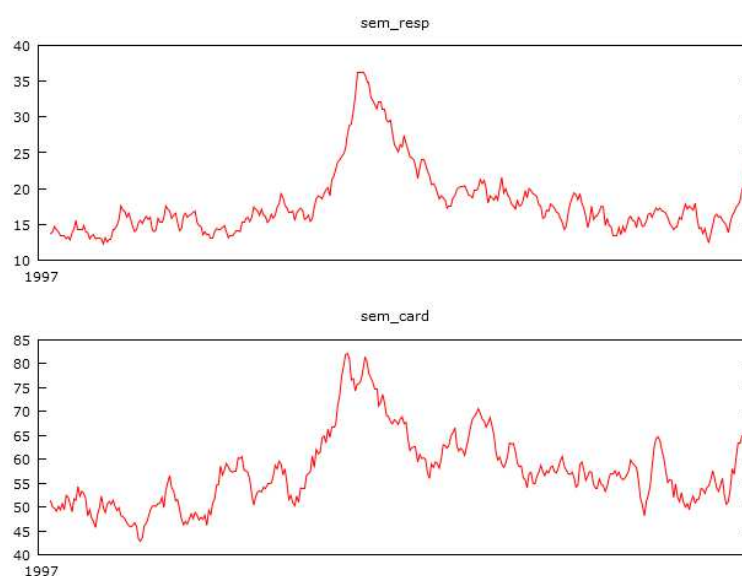


Figura 15: Série das médias semanais do número de óbitos por problemas respiratórios e cardíacos ($h=7$).

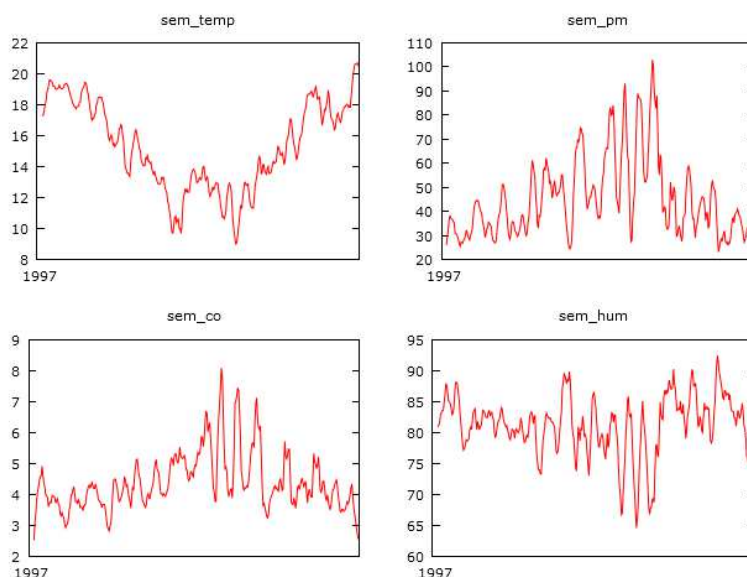


Figura 16: Séries climáticas transformadas com $h=7$.

Após as transformações (Figuras 15 e 16), conforme apresentado na Equação (4) com $h = 7$, o resultado obtido para as séries foi a redução da variância.

4.1.2.3. Séries temporais transformadas com $h = 28$

Para esta transformação (Figuras 17 e 18), utiliza-se a fórmula $z_{t,28} = \frac{z_t + z_{t-1} + z_{t-2} + \dots + z_{t-28+1}}{28}$, ou seja, a série tem (365-27) observações e cada uma delas passa a ser a média dos 28 dias anteriores à observação.

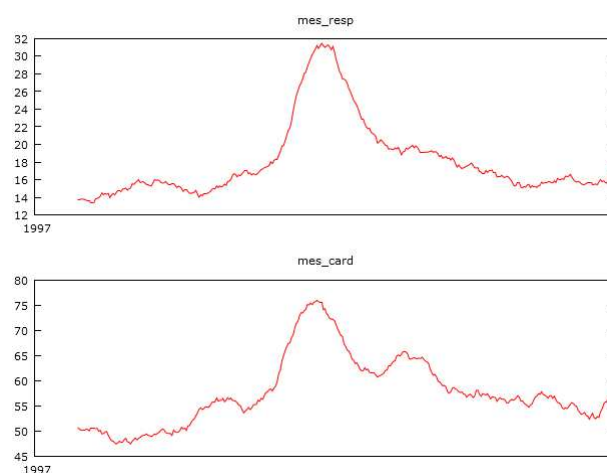


Figura 17: Série das médias mensais do número de óbitos por problemas respiratórios e cardíacos ($h=28$).

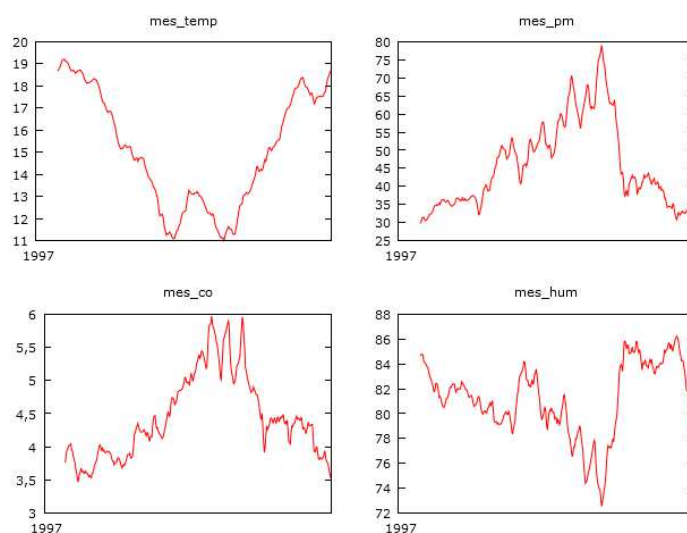


Figura 18: Série das médias mensais dos índices climáticos, temperatura, PM10, CO e umidade ($h=28$).

Nas Figuras 17 e 18 é possível verificar a diminuição da variabilidade, comparada com as séries de diárias e semanais.

A análise descritiva dos dados para as séries de médias diárias, semanais e mensais é apresentada na Tabela 2.

Tabela 2: Análise descritiva das séries de médias do número de óbitos, índices de temperatura, PM10, CO e umidade em São Paulo no ano de 1997.

	Média	Mediana	Mínimo	Máximo
resp. $h = 1$	18,041	17,000	6,0000	43,000
resp. $h = 7$	18,021	16,714	12,286	36,286
resp. $h = 28$	18,145	16,500	13,357	31,464
card. $h = 1$	57,616	57,000	30,000	97,000
card. $h = 7$	57,632	56,857	42,857	82,143
card. $h = 28$	57,807	56,536	47,464	75,964
temp. $h = 1$	15,355	15,400	5,6000	21,600
temp. $h = 7$	15,295	15,229	9,0143	20,700
temp. $h = 28$	15,069	15,052	11,004	19,211
pm10. $h = 1$	44,870	38,970	10,990	139,71
pm10. $h = 7$	45,203	40,333	23,121	102,88
pm10. $h = 28$	46,016	42,460	29,898	79,007
co. $h = 1$	4,3707	4,1500	1,4300	12,070
co. $h = 7$	4,4013	4,2057	2,5471	8,0857
co. $h = 28$	4,4431	4,3196	3,4700	5,9711
hum. $h = 1$	81,152	81,970	53,340	95,790
hum. $h = 7$	81,207	81,713	64,613	92,457
hum. $h = 28$	81,208	81,385	72,544	86,225

4.1.3. Séries temporais estacionárias

Foi necessária apenas uma diferença para fazer com que as séries temporais se tornassem estacionária. Após a diferenciação das séries temporais, por meio do teste de Dickey Fulley Aumentado (ADF), foi possível constatar que as mesmas se desenvolvem ao redor de uma média constante, ou seja, o processo se tornou fracamente estacionário, os resultados seguem na Tabela 3.

Tabela 3: Teste da raiz unitária (ADF), sem constante, com uma diferença

Séries	p-valor
Óbitos por problemas respiratórios $h = 1$	9,065e-042
Óbitos por problemas respiratórios $h = 7$	0,0007083
Óbitos por problemas respiratórios $h = 28$	0,0249
Óbitos por problemas cardiovasculares $h = 1$	8,87e-042
Óbitos por problemas cardiovasculares $h = 7$	2,503e-018
Óbitos por problemas cardiovasculares $h = 28$	0,01509
Temperatura mínima $h = 1$	9,343e-036
Temperatura mínima $h = 7$	2,494e-012
Temperatura mínima $h = 28$	0,03496
Material particulado $h = 1$	5,495e-039
Material particulado $h = 7$	5,831e-014
Material particulado $h = 28$	5,819e-012
Monóxido de carbono $h = 1$	9,625e-040
Monóxido de carbono $h = 7$	7,788e-013
Monóxido de carbono $h = 28$	2,854e-020
Umidade relativa do ar $h = 1$	3,73e-040
Umidade relativa do ar $h = 7$	2,164e-015
Umidade relativa do ar $h = 28$	0,003144

Por meio da função de autocorrelação (fac) e autocorrelação parcial (facp), é verificada a estacionariedade de todas as séries com $h = 1$, $h = 7$ e $h = 28$, que se encontra no ANEXO, através das Figuras (22 - 45).

4.2. Modelagem heterogênea autorregressiva das séries temporais em estudo

Nomeando as séries para modelar o HAR(h):

z_{1t} : número de óbitos por problemas respiratórios; z_{2t} : número de óbitos por problemas cardiovasculares; x_{1t} : temperaturas mínimas; x_{2t} : material particulado (pm10); x_{3t} : monóxido de carbono (co); x_{4t} : umidade relativa do ar (hum).
--

Utilizando a função autorregressiva vetorial do software Gretl e trabalhando com as séries z_{1t} , z_{2t} , x_{1t} , x_{2t} , x_{3t} e x_{4t} já estacionárias, obtêm-se os modelos considerando $h = 1$, $h = 7$ e $h = 28$.

- Modelo para $h = 1$:

O modelo HAR(1) é estimado por

$$z_{1t} = -0,77z_{1t-1} - 0,63z_{1t-2} - 0,5z_{1t-3} - 0,32z_{1t-4} - 0,2z_{1t-5} - 0,14z_{1t-6} \\ + 0,43x_{1t-1} + 0,02x_{2t} + \varepsilon_t$$

Tabelas com os valores dos parâmetros e desvio padrão encontram-se no ANEXO.

Foi observado que, para retardo de 0 dia para pm10, retardo de 1 dia para temperatura mínima e para retardos de 1, 2, 3, 4, 5 e 6 dias para os regimes de números de óbitos por problemas respiratórios, houve diferença significativa para o número de óbitos por problemas respiratórios na cidade de São Paulo, entre janeiro e dezembro de 1997.

A função de autocorrelação e autocorrelação parcial do resíduo segue na Figura 19, garantindo adequabilidade do modelo HAR (1), pois é ruído branco.

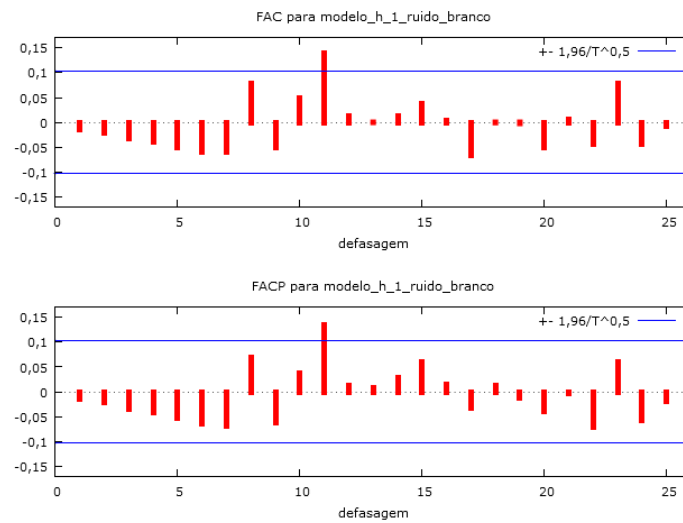


Figura 19: Função de autocorrelação e autocorrelação parcial do resíduo do modelo HAR(1).

- Modelo para $h = 1$ e $h = 7$

O modelo HAR (1, 7) é determinado por

$$z_{2t} = -0,82 z_{2t-1} - 0,63 z_{2t-2} - 0,47 z_{2t-3} - 0,35 z_{2t-4} - 0,29 z_{2t-5} - 0,17 z_{2t-6} - 4,33 x_{1t-2,7} + \varepsilon_t$$

Tabelas com os valores dos parâmetros e desvio padrão encontram-se no ANEXO.

Foi observado que, para retardos de duas semanas para temperaturas mínimas e para retardos de 1, 2, 3, 4, 5 e 6 dias para os regimes de números de óbitos por problemas cardiovasculares, houve diferença significativa para o número de óbitos por problemas cardiovasculares na cidade de São Paulo, compreendido entre janeiro e dezembro de 1997.

A função de autocorrelação e autocorrelação parcial do resíduo segue na Figura 20, garantindo adequabilidade do modelo HAR (1,7), pois é ruído branco.

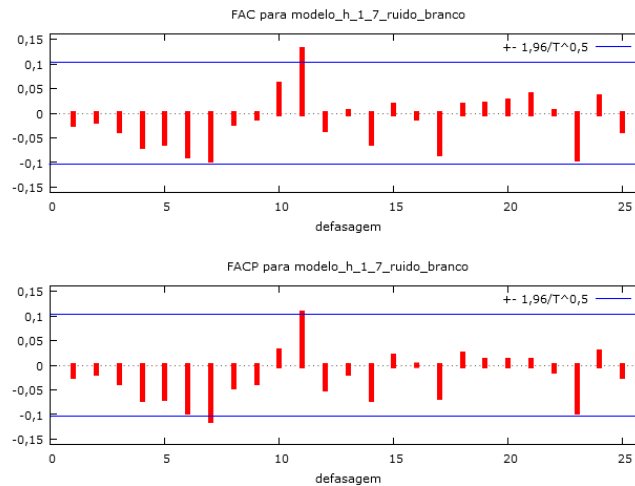


Figura 20: Função de autocorrelação e autocorrelação parcial do resíduo do HAR (1,7)

- Modelo para $h = 1$, $h = 7$ e $h = 28$

O modelo HAR (1, 7, 28)

$$z_{1t} = -0,76z_{1t-1} - 0,62z_{1t-2} - 0,47z_{1t-3} - 0,32z_{1t-4} - 0,2z_{1t-5} - 0,13z_{1t-6} \\ + 0,06x_{4t-9} - 0,25x_{4t,7} - 5,72x_{1t-2,28} + \varepsilon_t$$

Tabelas com os valores dos parâmetros e desvio padrão encontram-se no ANEXO.

Foi observado que, para retardos de 9 dias para umidade relativa do ar, retardo de 0 semana para umidade relativa do ar, retardos de 2 meses para temperaturas mínimas e para retardos de 1, 2, 3, 4, 5 e 6 dias para os regimes de números de óbitos por problemas respiratórios, houve diferença significativa para o número de óbitos por problemas respiratórios na cidade de São Paulo, entre janeiro e dezembro de 1997.

A função de autocorrelação e autocorrelação parcial do resíduo segue na Figura 21, garantindo adequabilidade do modelo HAR (1, 7, 28), pois é ruído branco.

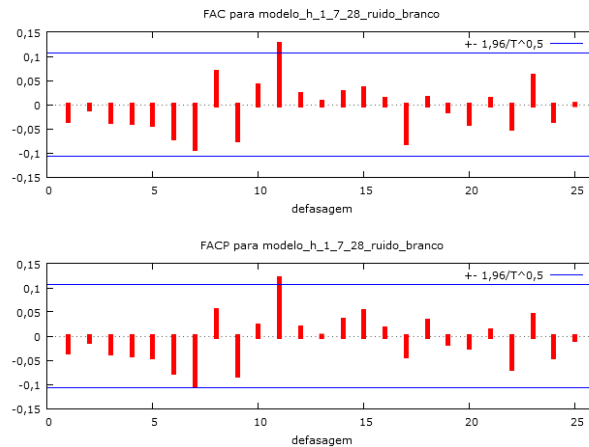


Figura 21: Função de autocorrelação e autocorrelação parcial do resíduo do HAR (1,7,28)

5. CONCLUSÕES

Observou-se que as séries de temperatura mínima, material particulado e umidade relativa do ar influenciaram o número de mortes por problemas respiratórios na cidade de São Paulo, por meio do HAR (1) e HAR (1, 7, 28).

Observou-se que a série de temperatura mínima influenciou o número de mortes por problemas cardiovasculares na cidade de São Paulo, através do HAR(1,7).

A série de monóxido de carbono não influenciou o número de mortes por problemas respiratórios nem mortes por problemas cardiovasculares na cidade de São Paulo, com os resultados do HAR (1), HAR(1,7) e HAR (1, 7, 28).

Através desta dissertação foi possível verificar que a modelagem heterogênea autorregressiva, aparentemente apresentada na literatura para modelagem da volatilidade e dados de finanças é também aplicada aos dados de mortalidade e climáticos, com características de processos de memória longa.

REFERÊNCIAS

- ANDERSEN, T.; BOLLERSLEV, T.; DIEBOLD, F. X. Roughing it up: including jump components in the measurement, modeling and forecasting of return volatility. **Review of Economics and Statistics**, Cambridge, v. 89, n. 4, p. 701-720, Nov. 2006.
- AUDRINO, F.; CORSI, F. Modeling tick-by-tick realized correlations. **Computational Statistics & Data Analysis**, New York, v. 54, n. 11, p. 2372-2382, Nov. 2010.
- BRAGA, A. L. F. et al. Air pollution and pediatric respiratory admissions in São Paulo, Brazil. **Occupational & Environmental Medicine**, London, v. 1, n. 2, p. 95-102, Apr./June 1999.
- CHANG, C.; MCALEER, M. **Aggregation, heterogeneous autoregression and volatility of daily international tourist arrivals and exchange rates**. Disponível em: <<http://eprints.ucm.es/12730/1/1113.pdf>>. Acesso em: 10 out. 2014.
- CONCEIÇÃO, G. M. S.; SINGER, J. M.; SALDIVA, P. H. N. Modelos MLG e MAG para análise da associação entre poluição atmosférica e marcadores de morbi-mortalidade: uma introdução baseada em dados da cidade de São Paulo. **Revista Brasileira de Epidemiologia**, São Paulo, v. 4, n. 3, p. 206-219, Nov. 2001.
- CORSI, F. A simple approximate long-memory model of realized volatility. **Journal of Financial Econometrics**, Cary, v. 7, n. 2, p. 174-196, Feb. 2009.
- CORSI, F.; RENO, R. **HAR volatility modeling with heterogeneous leverage and jumps**. Siena: University of Siena, 2009. 31 p.
- DICKEY, D.; FULLER, W. A. Distribution of the Estimates for Autoregressive Time Series with a Unit Root. **Journal of the American Statistical Association**, n.74, p.427-31, June 1979.
- GOUVEIA, N. et al. Poluição do ar e efeitos na saúde nas populações de duas grandes metrópoles brasileiras. **Epidemiologia e Serviços de Saúde**, Brasília, v. 12, n. 1, p. 29-40, jan./mar. 2003.

GRETTL. **Gnu regression, econometrics and time-series library**. Disponível em: <<http://gretl.sourceforge.net/>>. Acesso em: 1 fev. 2014.

LIN, C. A. et al. Air pollution and respiratory illness of children in São Paulo, Brazil. **Pediatric and Perinatal Epidemiology**, New York, v. 13, p. 475-488, 1999.

MIRAGLIA, S. G. E. K. et al. Analysis of the impact of fuel consumption on mortality rates in São Paulo. In: INTERNATIONAL CONFERENCE ON URBAN TRANSPORT AND THE ENVIRONMENT FOR THE 21ST CENTURY, 3., 1997, Acquasparta. **Proceedings...** Boston: Computational Mechanics, 1997. p. 434-444.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. 2. ed. São Paulo: E. Blucher, 2006. 538 p.

PEREIRA, L. A. A. et al. Association between air pollution and intrauterine mortality in São Paulo, Brazil. **Environmental Health Perspectives**, Research Triangle Park, v. 106, n. 6, p. 325-329, June 1998.

SÁFADI, T.; PEÑA, D. Bayesian analysis of dynamic factor models: an application to air pollution in São Paulo, Brazil. **Environmetrics**, London, v. 19, p. 582-601, Sept. 2008.

SALDIVA, P. H. N. et al. Air pollution and mortality in elderly people: a time-series study in São Paulo, Brazil. **Archives of Environmental Health**, Bethesda, v. 50, p. 159-163, Mar./Apr. 1995.

SALDIVA, P. H. N. et al. Association between air pollution and mortality due to respiratory diseases in children in São Paulo, Brazil: a preliminary report. **Environmental Research**, New York, v. 65, p. 218-225, May 1994.

SALSBURG, D. **Uma senhora toma chá...: como a estatística revolucionou a ciência no século XX**. Rio de Janeiro: J. Zahar, 2009. 286 p.

SINGER, J. M. **Dados de poluição atmosférica e mortalidade**. Disponível em: <<http://www.ime.usp.br/~jmsinger/doku.php?id=start>>. Acesso em: 1 maio 2014.

WINK JUNIOR, M. V.; PEREIRA, P. L. V. Modelling and forecasting realized volatility: evidence from Brazil. **Brazilian Review of Econometrics**, Rio de Janeiro, v. 31, n. 2, p. 315-337, 2011.

ANEXOS

Séries temporais estacionárias do número de óbitos por problemas respiratórios

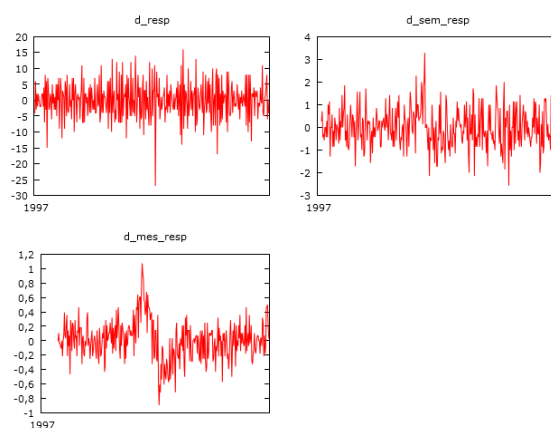


Figura 22: Série temporal com uma diferença do número de óbitos por problemas respiratórios na cidade de São Paulo, sendo resp quando $h=1$, sem_resp quando $h=7$ e mes_resp quando $h=28$.

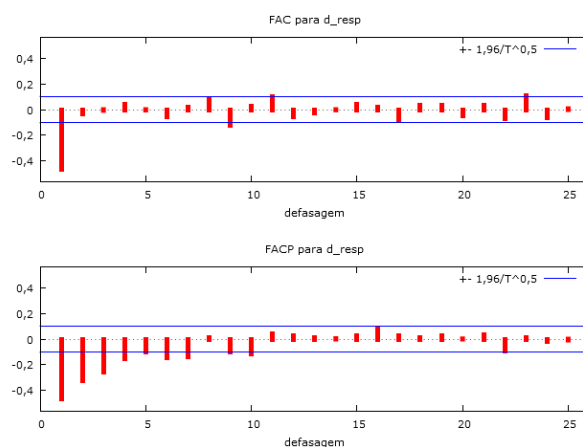


Figura 23: Função de autocorrelação e autocorrelação parcial da série diferenciada do número de óbitos por problemas respiratórios com $h=1$.

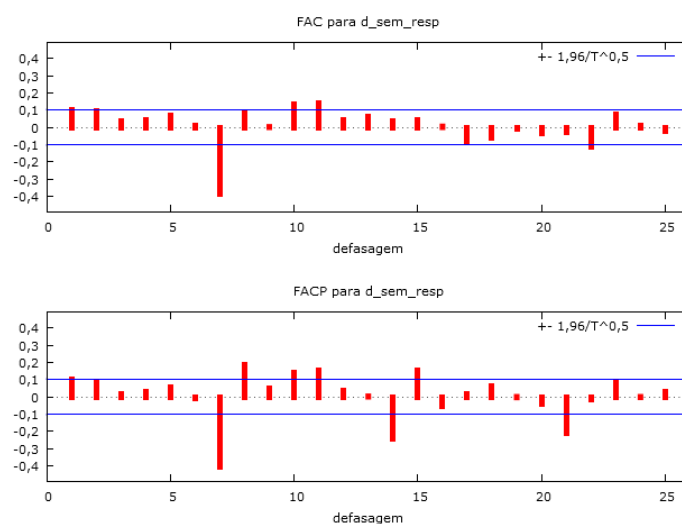


Figura 24: Função de autocorrelação e autocorrelação parcial da série diferenciada do número de óbitos por problemas respiratórios com $h=7$.

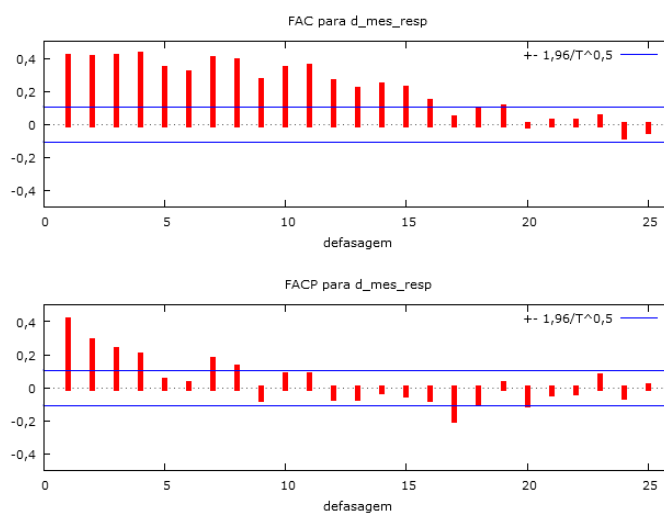


Figura 25: Função de autocorrelação e autocorrelação parcial da série diferenciada do número de óbitos por problemas respiratórios com $h=28$.

Séries temporais estacionárias do número de óbitos por problemas cardiovasculares

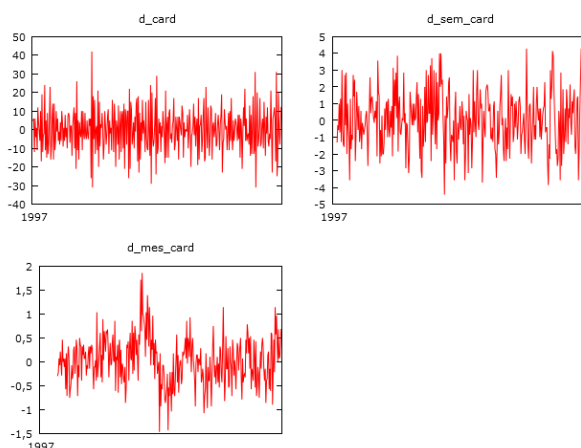


Figura 26: Série temporal com uma diferença do número de óbitos por problemas cardiovasculares na cidade de São Paulo, sendo card quando $h=1$, sem_card quando $h=7$ e mes_card quando $h=28$.

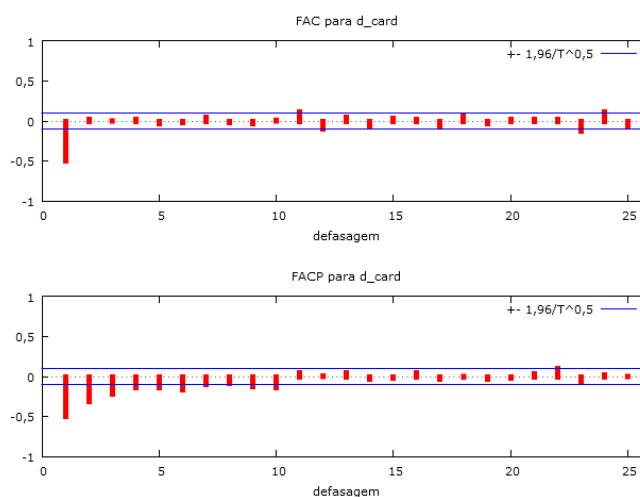


Figura 27: Função de autocorrelação e autocorrelação parcial da série diferenciada do número de óbitos por problemas cardiovasculares com $h=1$.

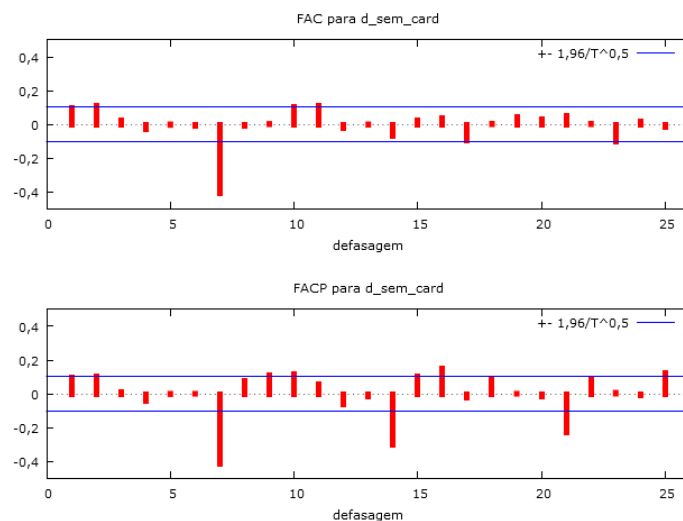


Figura 28: Função de autocorrelação e autocorrelação parcial da série diferenciada do número de óbitos por problemas cardiovasculares com $h=7$.

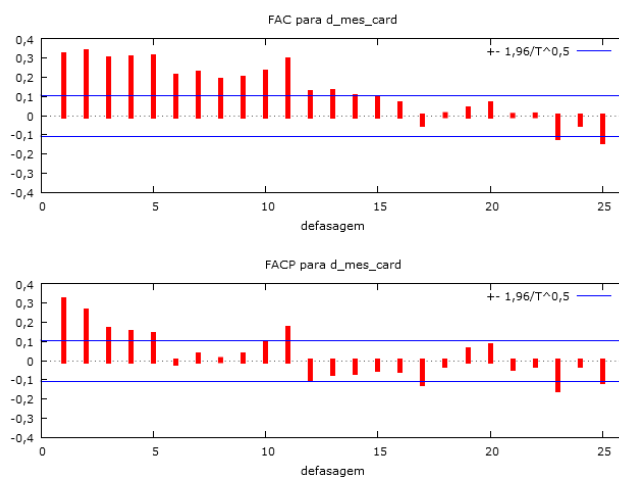


Figura 29: Função de autocorrelação e autocorrelação parcial da série diferenciada do número de óbitos por problemas cardiovasculares com $h=28$.

Séries temporais estacionárias das temperaturas mínimas

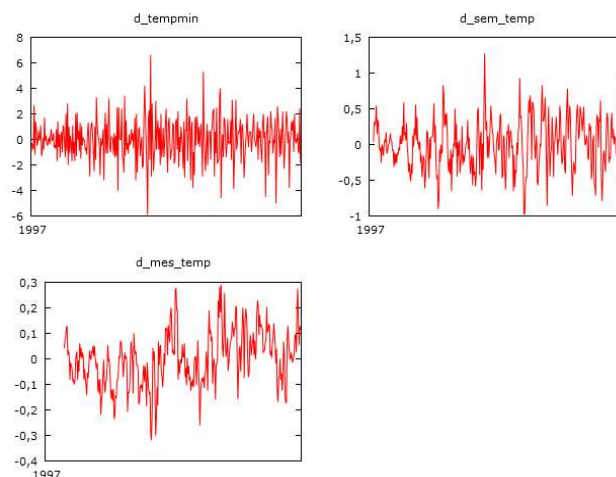


Figura 30: Série temporal com uma diferença da temperatura mínima da cidade de São Paulo, sendo tempmin quando $h=1$, sem_temp quando $h=7$ e mes_temp quando $h=28$.

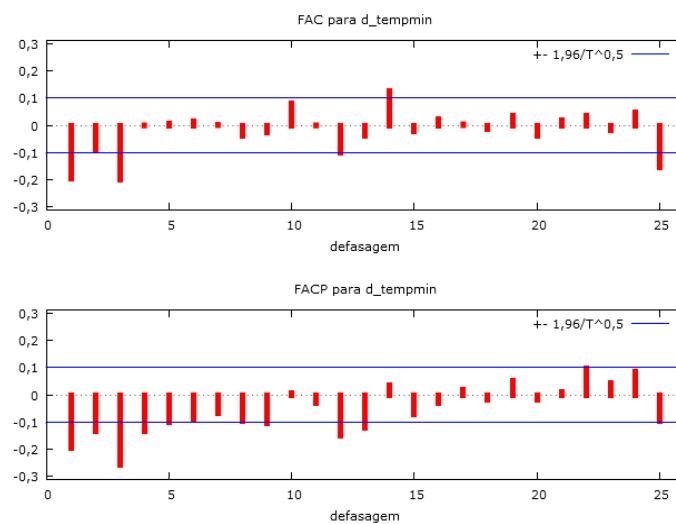


Figura 31: Função de autocorrelação e autocorrelação parcial da série diferenciada da temperatura mínima com $h=1$.

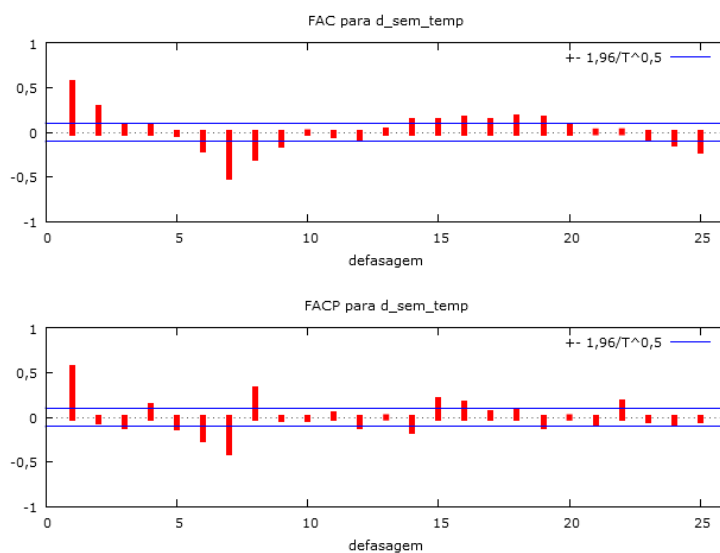


Figura 32: Função de autocorrelação e autocorrelação parcial da série diferenciada da temperatura mínima com $h=7$.

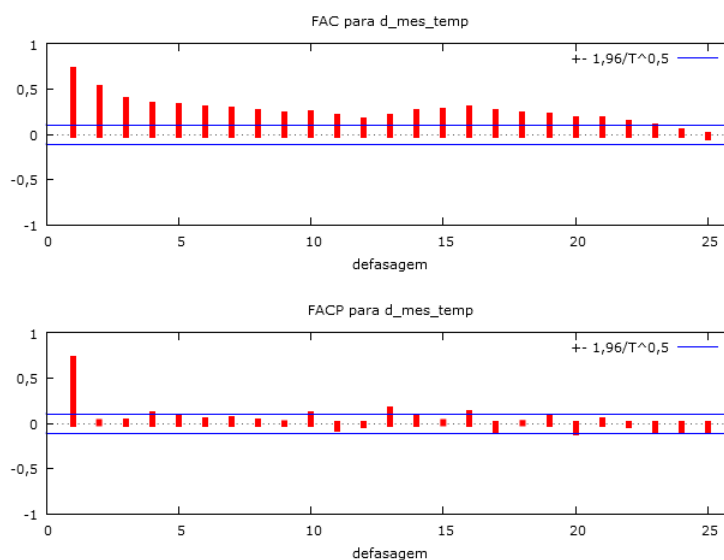


Figura 33: Função de autocorrelação e autocorrelação parcial da série diferenciada da temperatura mínima com $h=28$.

Séries temporais estacionárias do material particulado (pm10)

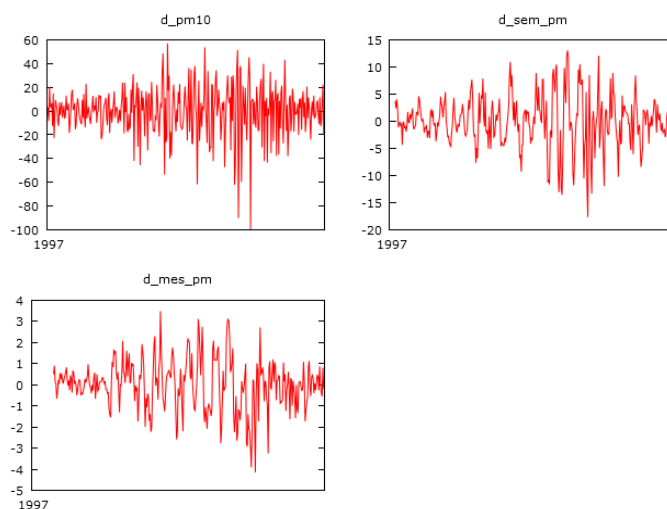


Figura 34: Série temporal com uma diferença de pm10 da cidade de São Paulo, sendo pm10 quando $h=1$, sem_pm quando $h=7$ e mes_pm quando $h=28$.

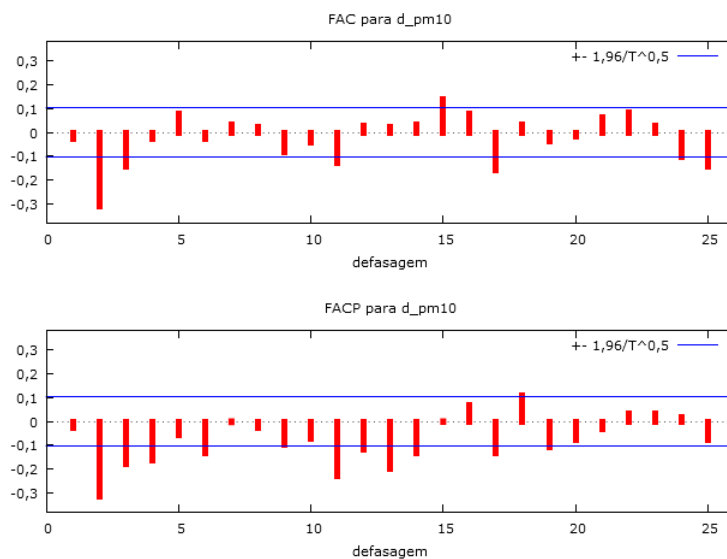


Figura 35: Função de autocorrelação e autocorrelação parcial da série diferenciada de pm10 com $h=1$.

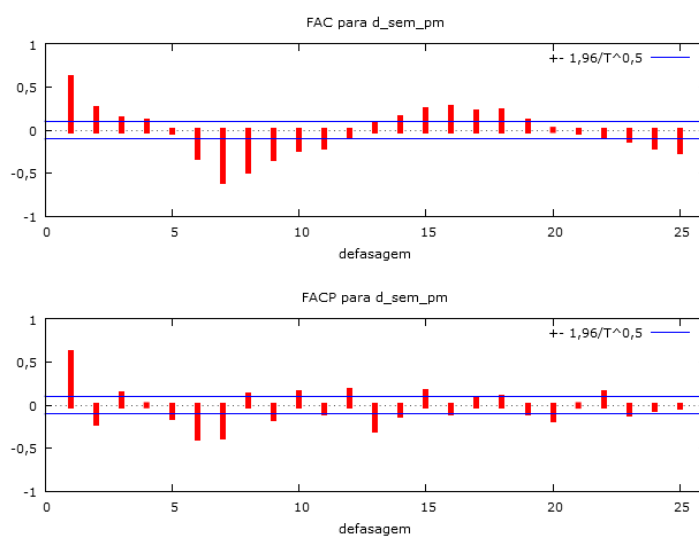


Figura 36: Função de autocorrelação e autocorrelação parcial da série diferenciada de pm10 com $h=7$.

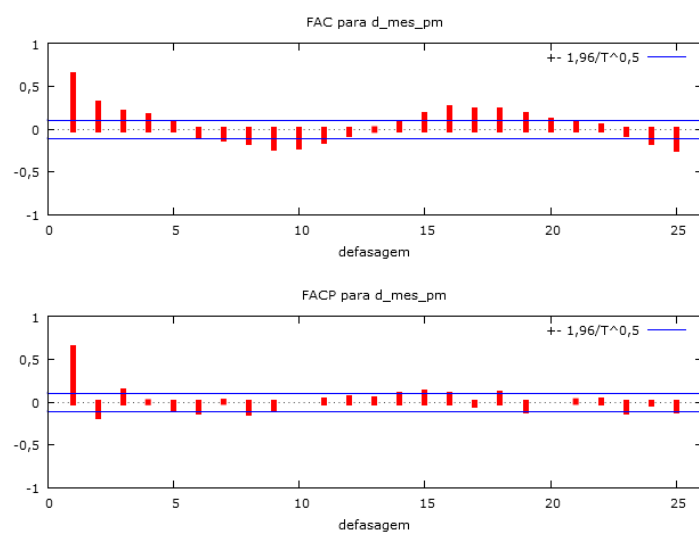


Figura 37: Função de autocorrelação e autocorrelação parcial da série diferenciada de pm10 com $h=28$.

Séries temporais estacionárias de monóxido de carbono (CO)

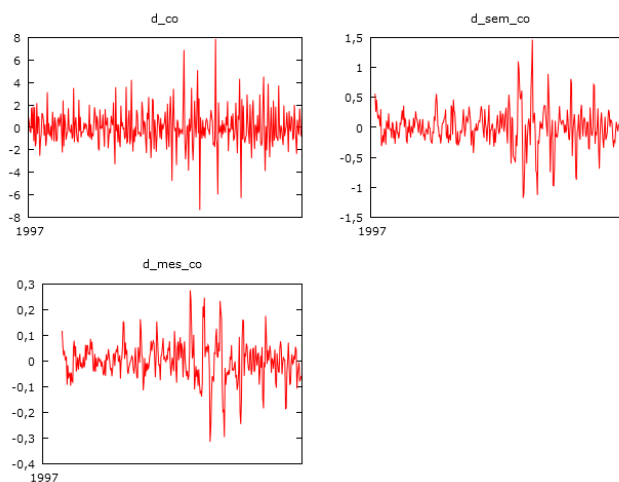


Figura 38: Série temporal com uma diferença de co da cidade de São Paulo, sendo co quando $h=1$, sem_co quando $h=7$ e mes_co quando $h=28$.

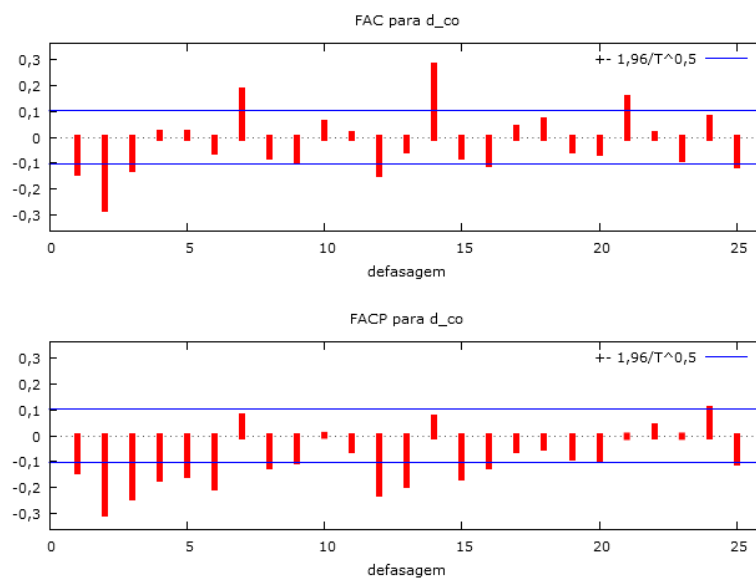


Figura 39: Função de autocorrelação e autocorrelação parcial da série diferenciada de co com $h=1$.

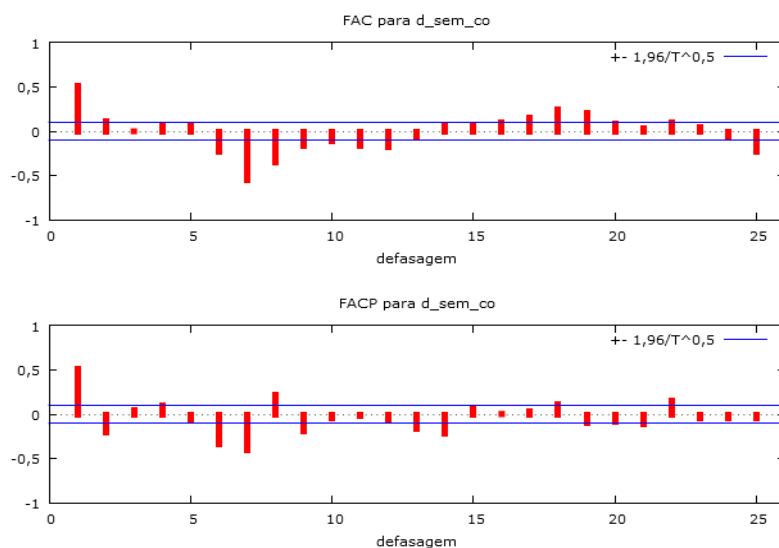


Figura 40: Função de autocorrelação e autocorrelação parcial da série diferenciada de co com $h=7$.

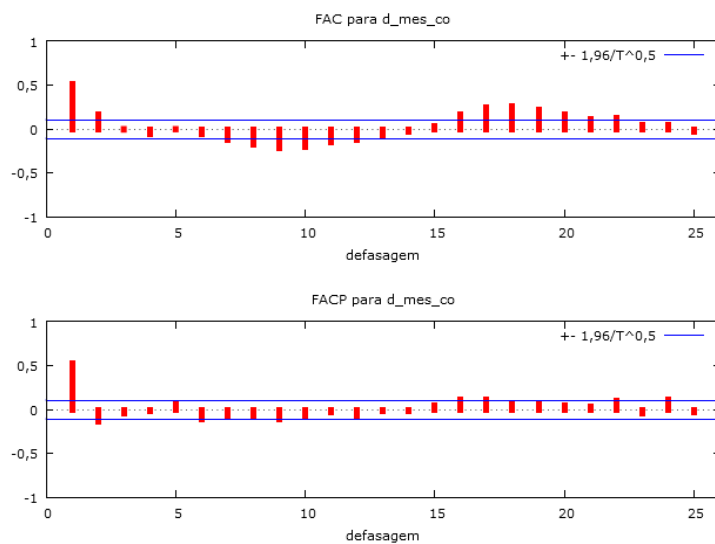


Figura 41: Função de autocorrelação e autocorrelação parcial da série diferenciada de co com $h=28$.

Séries temporais estacionárias de umidade relativa (hum)

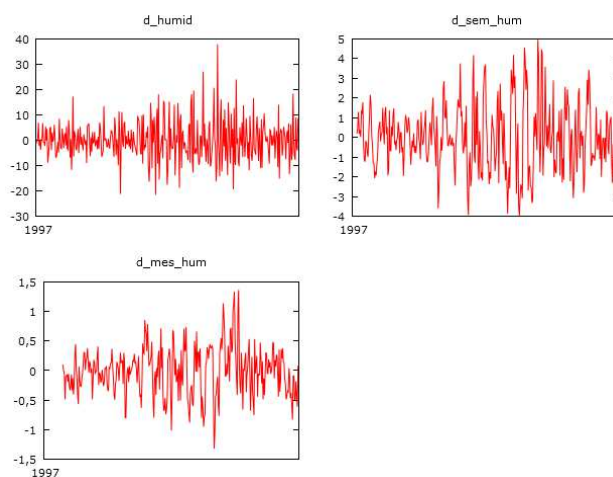


Figura 42: Série temporal com uma diferença da umidade da cidade de São Paulo, sendo humid quando $h=1$, sem_hum quando $h=7$ e mes_hum quando $h=28$.

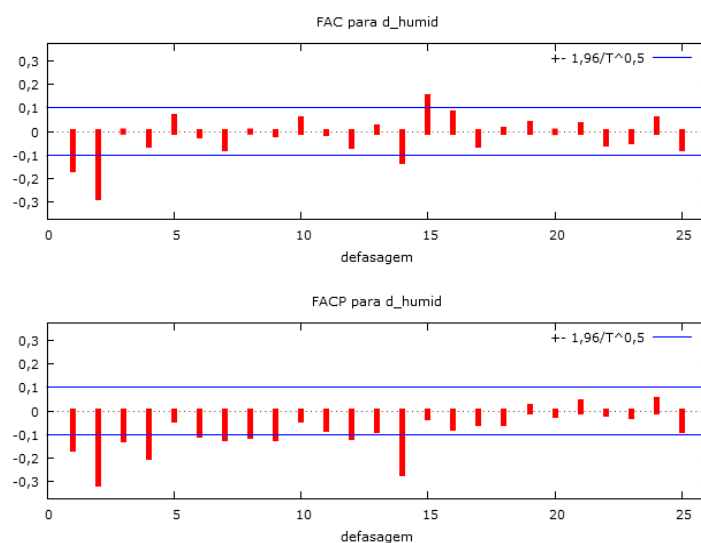


Figura 43: Função de autocorrelação e autocorrelação parcial da série diferenciada de hum com $h=1$.

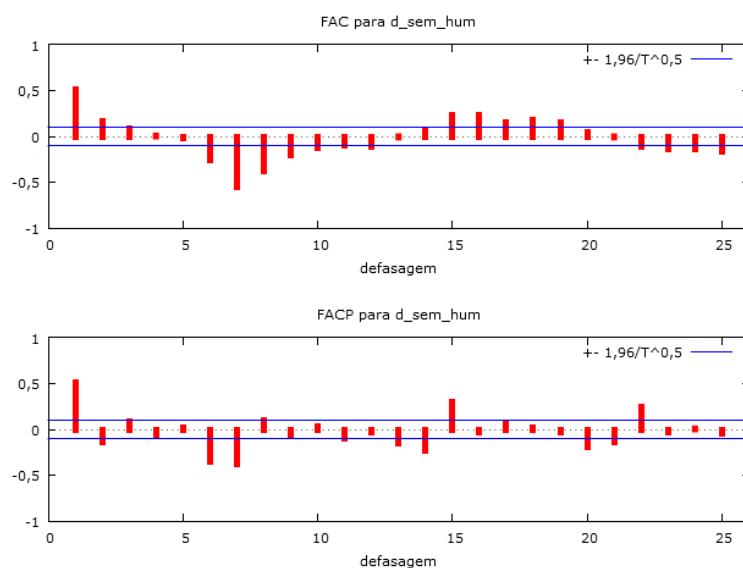


Figura 44: Função de autocorrelação e autocorrelação parcial da série diferenciada de hum com $h=7$.

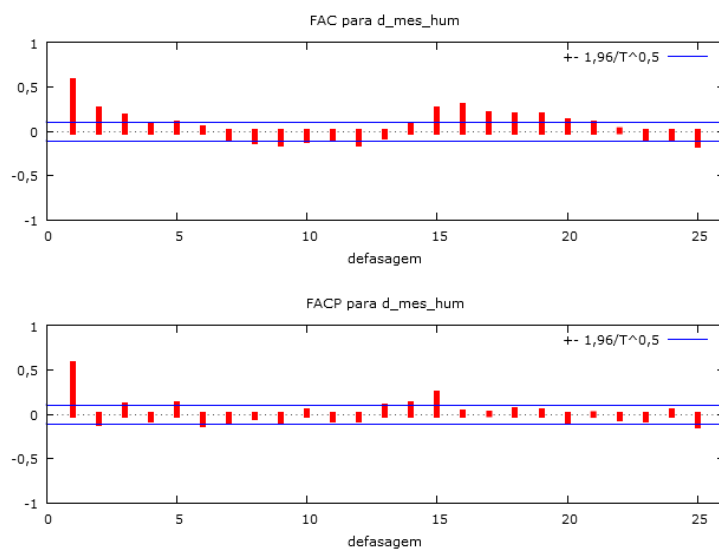


Figura 45: Função de autocorrelação e autocorrelação parcial da série diferenciada de hum com $h=28$.

Tabela 4: Modelo HAR(1) do número de óbitos por problemas respiratórios.

Variáveis	<i>Coeficiente</i>	<i>Erro padrão</i>	<i>p-valor</i>
z_{1t-1}	-0,768483	0,0525333	<0,00001***
z_{1t-2}	-0,633110	0,0654316	<0,00001***
z_{1t-3}	-0,496168	0,0716842	<0,00001***
z_{1t-4}	-0,322413	0,0721045	<0,00001***
z_{1t-5}	-0,209258	0,0662902	0,0017***
z_{1t-6}	-0,144802	0,0527122	0,0063***
x_{1t-1}	0,428416	0,152553	0,0053***
x_{2t}	0,0254168	0,0123258	0,0399**

Tabela 5: Modelo HAR (1,7) do número de óbitos por problemas cardiovasculares

Variáveis	<i>Coeficiente</i>	<i>Erro padrão</i>	<i>p-valor</i>
z_{2t-1}	-0,820357	0,0523598	<0,00001***
z_{2t-2}	-0,630544	0,0659554	<0,00001***
z_{2t-3}	-0,472549	0,0711755	<0,00001***
z_{2t-4}	-0,350073	0,0711147	<0,00001***
z_{2t-5}	-0,288076	0,0662865	<0,00001***
z_{2t-6}	-0,170388	0,0522703	0,0012***
$x_{1t-2,7}$	-4,33006	1,37790	0,0018***

Tabela 6: Modelo HAR (1, 7, 28) do número de óbitos por problemas respiratórios

Variáveis	<i>Coeficiente</i>	<i>Erro padrão</i>	<i>p-valor</i>
z_{1t-1}	-0,762129	0,0539706	<0,00001 ***
z_{1t-2}	-0,624628	0,0670119	<0,00001 ***
z_{1t-3}	-0,476017	0,0733907	<0,00001 ***
z_{1t-4}	-0,321643	0,0738356	<0,0001 ***
z_{1t-5}	-0,208267	0,0679166	0,0023 ***
z_{1t-6}	-0,134915	0,0545772	0,0139 **
x_{4t-9}	0,0572223	0,0322793	0,0772 *
$x_{4t,7}$	-0,253440	0,152174	0,0968 *
$x_{1t-2,28}$	-5,72054	2,39053	0,0173 **