



LETÍCIA LIMA MILANI RODRIGUES

**ANÁLISE DE VARIÂNCIA EM SÉRIES
TEMPORAIS: UMA ABORDAGEM USANDO
ONDALETAS**

LAVRAS - MG

2015

LETÍCIA LIMA MILANI RODRIGUES

**ANÁLISE DE VARIÂNCIA EM SÉRIES TEMPORAIS: UMA
ABORDAGEM USANDO ONDALETAS**

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

Orientadora
Dra. Thelma Sáfydi

Coorientador
Dr. Augusto Ramalho de Moraes

**LAVRAS - MG
2014**

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

Rodrigues, Letícia Lima Milani.

Análise de variância em séries temporais : uma abordagem
usando ondaletas / Letícia Lima Milani Rodrigues. – Lavras :
UFLA, 2015.

100 p. : il.

Tese(doutorado)–Universidade Federal de Lavras, 2014.

Orientador(a): Thelma Sáfadi.

Bibliografia.

1. Análise de Variância. 2. Ondaletas. 3. Séries Temporais. I.
Universidade Federal de Lavras. II. Título.

LETÍCIA LIMA MILANI RODRIGUES

**ANÁLISE DE VARIÂNCIA EM SÉRIES TEMPORAIS: UMA
ABORDAGEM USANDO ONDALETAS**

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

APROVADA em 12 de dezembro de 2014.

Dr. Denismar Alves Nogueira	UNIFAL-MG
Dr. Lucas Monteiro Chaves	UFLA
Dr. Paulo César Lima	UFLA
Dr. Renato Ribeiro de Lima	UFLA

Dra. Thelma Sáfydi
Orientadora

Dr. Augusto Ramalho de Morais
Coorientador

**LAVRAS - MG
2014**

*À minha família que participou desta conquista,
Wilson, Mateus e Júlia,
amo vocês!*

DEDICO

AGRADECIMENTOS

A Deus, por me amparar nos momentos difíceis, me dar força interior para superar as dificuldades, mostrar os caminhos nas horas incertas e me suprir em todas as minhas necessidades.

À minha orientadora, Profa. Dra. Thelma Sáfadi, pela sua orientação, dedicação, apoio, confiança e amizade.

Ao meu marido Wilson, que muito contribuiu para a realização deste trabalho. Por suprir minhas ausências junto aos nossos filhos, para que eu pudesse me dedicar ao doutorado com tranquilidade. Agradeço também, pelo amor, carinho, paciência e dedicação, que nesse momento conclusivo, foi de suma importância.

Aos meus filhos, Mateus e Júlia, por estarem sempre ao meu lado, me apoiando simplesmente com um sorriso e um abraço carinhoso. Mamãe ama vocês!

Aos meus pais, Dirceu e Maria, pela dedicação, amor e apoio incondicional.

Aos meus irmãos, Fabricio e Gustavo, pelo apoio e carinho que sempre me dedicaram, por estarem sempre à disposição e torcerem pelo meu sucesso.

Às minhas cunhadas, Stefânia e Elisa, por toda torcida e carinho.

À minha amiga Gislene, pelo incentivo e companheirismo durante todo este período.

À Universidade Federal de Alfenas-UNIFAL-MG (Campus Varginha), em especial ao Instituto de Ciências Sociais Aplicadas-ICSA, pela oportunidade concedida, considerando o afastamento de minhas funções.

Aos colegas Professores do ICSA do grupo da Estatística, pelo apoio e colaboração, os quais foram fundamentais para que eu pudesse desenvolver este trabalho.

À Universidade Federal de Lavras (UFLA), por intermédio do Departamento de Ciências Exatas, pela oportunidade.

Por fim, agradeço a todos que de forma direta ou indireta, contribuíram para que eu concluísse esta etapa.

RESUMO

A análise de variância (ANAVA) é amplamente utilizada para verificar se existe diferença significativa entre tratamentos. Entretanto, para que os resultados sejam confiáveis, é necessário que se verifique os pressupostos da ANAVA, que são aditividade dos efeitos presentes no modelo e a independência, normalidade e homocedasticidade dos erros. Contudo, com certa frequência, constata-se que uma ou mais dessas suposições não se verifica. E é o que acontece quando se trabalha com dados de séries temporais, que leva a problemas como a não independência entre as observações consecutivas. Diante disso, este trabalho objetivou propor uma metodologia para a realização da ANAVA em dados de séries temporais. Como a análise tradicional não é adequada, a proposta é “transformar” os dados, eliminando a dependência entre eles, de forma a atender as pressuposições exigidas. A transformação é realizada por meio da transformada discreta de ondaleta decimada para séries temporais estacionárias e a transformada discreta de ondaleta não decimada para séries temporais não estacionárias, usando em ambos os casos as bases de Haar e Daubechies. Dessa forma, realizou-se um estudo de simulação, considerando séries temporais estacionárias e não estacionárias. Também foi feito um estudo com dados reais. Concluiu-se que a análise de variância usando ondaletas pode ser usada com sucesso, quando se trabalha com dados de séries temporais. Para as séries temporais consideradas categóricas não estacionárias, deve-se priorizar o uso da transformada discreta de ondaleta não decimada na base de Daubechies.

Palavras-chave: Análise de variância. Séries temporais. Ondaletas.

ABSTRACT

The analysis of variance (ANAVA) is widely used to check for significant differences among treatments. However, for the results to be reliable, it is necessary to check the assumptions of ANAVA, which are additivity of effects in the model and independence, normality and homoscedasticity of errors. However, quite frequently it turns out that one or more of these assumptions does not occur. And that's what happens when working with time series data, which leads to problems such as non-independence between consecutive observations. Thus, this study was conducted with the objective to propose a methodology for conducting the ANAVA in time series data. As the traditional analysis is not adequate, the proposal is “ transform ” the data, eliminating the dependency between them, in order to meet the required assumptions. The transformation occurs by means of the discrete wavelet transform decimated for non-stational temporal series and the discrete wavelet transform not decimated for non-stational temporal series, using, in both cases, the Haar and Daubechies bases. Thus, a simulation study was conducted considering stational and non-stational temporal series. We also conducted a study with real data. We concluded that the analysis of variance using wavelets could be successfully used when working with temporal series data. For the temporal series considered non-stational categorical, we must prioritize the use of not decimated discrete wavelet transform on the Daubechies basis.

Keywords: Analysis of variance. Temporal series. Wavelets.

LISTA DE ILUSTRAÇÕES

Figura 1	Ondaleta de Haar	27
Figura 2	Representação do segundo e do terceiro membro da família de ondaletas de Daubechies	28
Figura 3	Transformada de ondaleta: decomposição em tempo escala com tamanhos de janelas variáveis.	29
Figura 4	Esquema do processo de decomposição da Transformada Discreta Ondaleta	29
Figura 5	Esquema do processo de decomposição da Transformada Discreta Ondaleta Não Decimada	32
Figura 6	Resultado da decomposição de um sinal em diferentes níveis de resolução, conhecido como periodograma de ondaletas	40
Figura 7	Gráfico das séries AR(1) geradas para os três tratamentos e suas três repetições estatisticamente iguais	55
Figura 8	Gráfico da decomposição (Haar) para a primeira repetição do primeiro tratamento (ruído branco), considerando a transformada discreta de ondaleta decimada	56
Figura 9	Gráfico da decomposição (Daubechies) para a primeira repetição do primeiro tratamento (ruído branco), considerando a transformada discreta de ondaleta decimada	57
Figura 10	Gráfico da primeira repetição dos dois tratamentos, considerando as séries categóricas simuladas	60
Figura 11	Gráfico da decomposição (Haar) para a primeira repetição do primeiro tratamento considerando a série categórica simulada, usando a transformada discreta de ondaleta decimada	61

Figura 12	Gráfico da decomposição (Daubechies) para a primeira repetição do primeiro tratamento considerando a série categórica simulada, usando a transformada discreta de ondaleta decimada	62
Figura 13	Gráfico das séries com tendência e Ar(1) simuladas para os três tratamentos	65
Figura 14	Gráfico da decomposição (Haar) para a primeira repetição do primeiro tratamento (tendência e AR(1)), considerando a transformada discreta de ondaleta não-decimada	66
Figura 15	Gráfico da decomposição (Haar) para a primeira repetição do segundo tratamento (tendência e AR(1)), considerando a transformada discreta de ondaleta não-decimada	67
Figura 16	Gráfico da decomposição (Haar) para a primeira repetição do terceiro tratamento (tendência e AR(1)), considerando a transformada discreta de ondaleta não-decimada	68
Figura 17	Gráfico da decomposição (Daubechies) para a primeira repetição do primeiro tratamento (tendência e AR(1)), considerando a transformada discreta de ondaleta não-decimada	69
Figura 18	Gráfico da decomposição (Daubechies) para a primeira repetição do segundo tratamento (tendência e AR(1)), considerando a transformada discreta de ondaleta decimada	70
Figura 19	Gráfico da decomposição (Daubechies) para a primeira repetição do terceiro tratamento (tendência e AR(1)), considerando a transformada discreta de ondaleta decimada	71
Figura 20	Gráfico da primeira repetição dos dois tratamentos, considerando séries categóricas diferentes	73

Figura 21	Gráfico da decomposição (Haar) para a primeira repetição do primeiro tratamento (série categórica simulada), considerando a transformada discreta de ondaleta não-decimada	74
Figura 22	Gráfico da decomposição (Haar) para a primeira repetição do segundo tratamento (série categórica simulada), considerando a transformada discreta de ondaleta não-decimada	75
Figura 23	Gráfico da decomposição (Daubechies) para a primeira repetição do primeiro tratamento (série categórica), considerando a transformada discreta de ondaleta não-decimada	76
Figura 24	Gráfico da decomposição (Daubechies) para a primeira repetição do segundo tratamento (série categórica), considerando a transformada discreta de ondaleta não-decimada	77
Figura 25	Gráfico da primeira repetição das faixas etária relativa ao número diário de óbitos por causas respiratória	80
Figura 26	Gráfico do correlograma primeira repetição das faixas etária relativa ao número diário de óbitos por causas respiratória . . .	81
Figura 27	Gráfico da decomposição (Haar) para o número de óbitos por causas respiratórias para o primeiro tratamento (faixa etária < 12 anos), considerando a transformada discreta de ondaleta não-decimada	82
Figura 28	Gráfico da decomposição (Haar) para o número de óbitos por causas respiratórias para o segundo tratamento (faixa etária ≥ 12 e < 35 anos), considerando a transformada discreta de ondaleta não-decimada	83

Figura 29	Gráfico da decomposição (Daubechies) para o número de óbitos por causas respiratórias para o primeiro tratamento (faixa etária < 12 anos), considerando a transformada discreta de ondaleta decimada	84
Figura 30	Gráfico da decomposição (Daubechies) para o número de óbitos por causas respiratórias para o segundo tratamento (faixa etária ≤ 12 e > 35 anos), considerando a transformada discreta de ondaleta decimada	85
Figura 31	Gráfico do valor-p para o número de óbitos por causas respiratórias, considerando a transformada discreta de ondaleta não-decimada na base de Haar. Linha horizontal representa o valor de 0,05 (5% de significância)	86
Figura 32	Gráfico do valor-p para o número de óbitos por causas respiratórias, considerando a transformada discreta de ondaleta não-decimada na base de Daubechies. Linha horizontal representa o valor de 0,05 (5% de significância)	86
Figura 33	Gráfico da primeira repetição do no número de óbito por causas cardíacas entre as faixas etárias	87
Figura 34	Gráfico do correlograma primeira repetição das faixas etárias relativa ao número diário de óbitos por causas cardíacas	88
Figura 35	Gráfico da decomposição (Haar) para o número de óbitos por causas cardíacas para o primeiro tratamento (<12 anos), considerando a transformada discreta de ondaleta não-decimada	89
Figura 36	Gráfico da decomposição (Haar) para o número de óbitos por causas cardíacas para o segundo tratamento (≥ 12 e < 35 anos), considerando a transformada discreta de ondaleta não-decimada	90

Figura 37	Gráfico da decomposição (Daubechies) para o número de óbitos por causas respiratórias para o primeiro tratamento (faixa etária < 12 anos), considerando a transformada discreta de ondaleta decimada	91
Figura 38	Gráfico da decomposição (Daubechies) para o número de óbitos por causas respiratórias para o segundo tratamento (faixa etária ≥ 12 e < 35 anos), considerando a transformada discreta de ondaleta decimada	92
Figura 39	Gráfico do valor-p para o número de óbitos por causas cardíacas, considerando a transformada discreta de ondaleta não-decimada na base de Haar. Linha horizontal representa o valor de 0,05 (5% de significância)	93
Figura 40	Gráfico do valor-p para o número de óbitos por causas cardíacas, considerando a transformada discreta de ondaleta não-decimada na base de Daubechies. Linha horizontal representa o valor de 0,05 (5% de significância)	93

LISTA DE TABELAS

Tabela 1	Proporção de valores-p não significativos em cada nível (j), para séries AR(1), usando a transformada discreta de ondaleta decimada nas bases de Haar e Daubechies	58
Tabela 2	Taxa de erro tipo I em cada nível (j), para séries AR(1), usando a transformada discreta de ondaleta decimada nas bases de Haar e Daubechies	59
Tabela 3	Proporção de valores-p não significativos em cada nível (j), para séries AR(1), usando a transformada discreta de ondaleta decimada nas bases de Haar e Daubechies	63
Tabela 4	Taxa de erro tipo I em cada nível (j), para séries AR(1), usando a transformada discreta de ondaleta decimada nas bases de Haar e Daubechies	64
Tabela 5	Proporção de valores-p significativo em cada nível (j), para séries de tendência e AR(1), usando a transformada discreta de ondaleta não decimada nas bases de Haar e Daubechies . . .	72
Tabela 6	Poder do teste F usando a transformada discreta de ondaleta não-decimada, para séries categórica simuladas, usando a ondaleta de Haar e Daubechies	78
Tabela 7	Proporção de valor-p significativo em cada nível (j), para séries categóricas estatisticamente diferentes, usando a transformada discreta de ondaleta não-decimada nas bases de Haar e Daubechies	79

SUMÁRIO

1	INTRODUÇÃO	16
2	REFERENCIAL TEÓRICO	18
2.1	Análise de Variância e as Pressuposições Fundamentais	18
2.2	Séries temporais e processos estacionários	21
2.3	Ondaleta	23
2.3.1	Transformada discreta de ondaleta decimada	28
2.3.2	Transformada discreta de ondaleta não decimada	31
2.3.3	O espectro de ondaleta	33
2.3.4	O periodograma de ondaletas	40
2.4	Erro tipo I e poder de teste	44
3	MATERIAIS E MÉTODOS	45
3.1	Simulação de séries estacionárias	45
3.2	Séries não estacionárias	46
3.3	Análise de Variância usando Ondaleta com dados reais	47
3.3.1	Estudo da série do número diário de óbitos por causa respiratória no período de 01 de janeiro de 1994 a 31 de dezembro de 1997, em São Paulo-SP	47
3.3.2	Estudo da série do número diário de óbitos por causas cardíacas no período de 01 de janeiro de 1994 a 31 de dezembro de 1997, em São Paulo-SP	47
4	RESULTADOS E DISCUSSÕES	49
4.1	Análise de variância usando ondaleta	49
4.1.1	Processo estacionário	50
4.1.2	Processo não estacionário	52
4.2	Simulação	53
4.2.1	Análise de variância usando ondaleta para séries temporais estacionárias	54
4.2.2	Análise de variância usando ondaleta para séries temporais não estacionárias	64
4.2.3	Análise de variância usando ondaleta com dados reais	79
5	CONCLUSÕES	95
	REFERÊNCIAS	95

1 INTRODUÇÃO

A análise de variância (ANAVA) é um método amplamente difundido, que visa verificar se existe diferença significativa entre tratamentos e se os fatores em estudo exercem influência em alguma variável dependente. Contudo, para que se tenha resultados confiáveis, é necessário que se verifique os pressupostos da ANAVA: aditividade dos efeitos presentes no modelo e a independência, normalidade e homocedasticidade dos erros. Caso um dos pressupostos não seja atendido, as conclusões obtidas não são confiáveis e os testes realizados são apenas aproximados.

Com certa frequência, constata-se que uma ou mais dessas pressuposições não se verifica. É o que acontece quando se trabalha com dados de séries temporais, cujos valores experimentais são observados em diferentes pontos no tempo, o que leva a problemas como a não independência entre as observações consecutivas, afetando a modelagem estatística e inferências.

A correlação existente devido à amostragem de pontos adjacentes no tempo, limita severamente a aplicabilidade dos diversos métodos estatísticos, que dependem da suposição de que as observações adjacentes sejam independentes e identicamente distribuídas.

Uma alternativa para resolver a limitação de se trabalhar com ANAVA, cujas medidas respostas são séries temporais, seria o uso de ondaletas que, por meio da transformada discreta de ondaleta, tem-se que as pressuposições exigidas para se realizar a ANAVA são satisfeitas.

A transformada discreta de ondaleta é uma técnica que teve um grande avanço em processamento de sinais, e tem chamado muita atenção desde seu desenvolvimento teórico em 1984, por Grossmann e Morlet (1984). Seu uso tem aumentado rapidamente como uma técnica alternativa à transformada de Fourier em preservar os fenômenos locais, não-periódico e de multiescala.

A análise em ondaletas tem sido formalizada extensivamente graças aos esforços de Físicos e Matemáticos, constituindo um núcleo de ideias partilhadas também por outros pesquisadores. A transformada discreta de ondaleta pode ser usada em séries estacionárias ou não. Quando se tem séries estacionárias,

a transformada discreta de ondaleta decimada é a indicada. Nessa transformada tem-se uma decimação diádica. Nos casos de séries temporais que não são estacionárias, a transformada discreta de ondaleta não decimada é a apropriada, não ocorrendo a decimação, como na transformada discreta de ondaleta decimada.

A análise de ondaleta para séries temporais estacionária foi apresentada por Chiann e Morettin (1998), no qual os autores comprovaram que a transformada discreta de ondaleta decimada atende às pressuposições da ANAVA, no que diz respeito à independência, normalidade e homocedasticidade. Nason, Sachs e Kroisandt (2000) apresentaram um estudo para as séries não estacionárias usando a transformada discreta de ondaleta não decimada, neste caso, tem-se o espectro de ondaleta evolucionário.

Tendo por base os estudos de Chiann e Morettin (1998), Chiann e Tolo (1998) e Nason, Sachs e Kroisandt (2000), propõe-se neste trabalho, a realização da análise de variância (ANAVA) para estudar as possíveis variações apresentadas nos dados onde as variáveis respostas são séries temporais. Como a análise tradicional não é adequada, pois os dados oriundos de uma série temporal são correlacionados, ou seja, apresentam dependência, o objetivo é “transformar” os dados, eliminando a dependência entre eles, de forma a atender as pressuposições exigidas, utilizando as técnicas tradicionais de análise de variância. A transformação proposta é o uso da transformada discreta de ondaleta decimada para séries temporais estacionárias e a transformada discreta de ondaleta não-decimada para séries temporais não estacionárias. Dessa forma, o método proposto será validado por meio do erro tipo I e poder do teste F.

2 REFERENCIAL TEÓRICO

Na seção 2.1 a seguir são apresentados os conceitos necessários sobre análise de variância e os pressupostos para que se tenha resultados confiáveis. Algumas definições envolvendo processo estacionário serão apresentadas na seção 2.2. A seção 2.3 aborda alguns tópicos fundamentais de ondaletas, incluindo a transformada discreta de ondaleta decimada e não decimada e os teoremas que garantem que com o uso de ondaletas, os pressupostos da análise de variância são atendidos. A seção 2.4 apresenta a análise de variância usando ondaleta para séries temporais estacionárias e não estacionárias.

2.1 Análise de Variância e as Pressuposições Fundamentais

Segundo Memória (2004), as pesquisas científicas de natureza experimental exigiam tratamento adequado às pequenas amostras, com objetivo inferencial, conforme revelaram os trabalhos pioneiros de Gosset, que se tornou conhecido pelo pseudônimo de Student. Esses trabalhos foram continuados por R. A. Fisher *, a figura mais representativa da fase da experimentação, considerado o criador dos métodos modernos da Análise e Delineamentos de experimentos. Assim, a Análise de Variância (ANOVA) foi introduzida por R. A. Fisher na década de 20, como um método de análise de dados experimentais. A maioria das aplicações foi feita nas áreas de agricultura e biologia, mas atualmente, constitui uma das principais técnicas utilizadas em todas as áreas do conhecimento. Segundo Banzatto e Kronka (2006) a ANAVA consiste na decomposição dos graus de liberdade (G.L.) e da variância total de um material heterogêneo, em partes, atribuídas a causas conhecidas e independentes (fatores controlados), e a uma porção residual de origem desconhecida e de natureza aleatória (fatores não controlados).

Cada tipo de delineamento experimental é regido por um modelo estatístico com restrições na variável resposta. Para validar os testes de hipótese

*Para maiores detalhes de sua vida e obra, segundo Memória (2004), tem-se a biografia escrita por sua filha Joan Fisher Box (1978) .

e outras inferências, os modelos estatísticos devem ter suas pressuposições atendidas. A qualidade de um experimento pode ser avaliada pela magnitude do erro experimental e pelo atendimento das pressuposições do modelo estatístico.

As pressuposições fundamentais da ANAVA são tratadas por vários autores, tais como: Cochran e Cox (1957), Dean e Voss (1999) e Federer (1955), dentre outros.

A utilização da análise de variância exige a verificação das seguintes pressuposições básicas:

a) Os efeitos de tratamentos e erros devem ser aditivos (aditividade):

De acordo com Dean e Voss (1999), a não aditividade resulta na heterogeneidade do erro e afeta o nível de significância para comparações entre os tratamentos. Há perda de precisão porque o erro experimental é acrescido do componente de não aditividade.

Os efeitos de dois ou mais fatores são ditos não-aditivos quando, na associação de um ou mais desses fatores, em vez de se somarem, estes se multiplicam, de tal forma que o efeito resultante pode ser ampliado ou reduzido. É o que comumente ocorre nas chamadas interações entre dois ou mais fatores (MONTGOMERY, 2001).

No emprego da estatística paramétrica, a aditividade talvez seja a pressuposição menos rigorosa, devido ao fato de se referir às interações entre os fatores de variação, e não aos próprios fatores em si. Mas, a não-aditividade não deve ser negligenciada, pois pode modificar o valor de erro não controlado. Segundo Montgomery (2001), esse inconveniente deve ter ocorrido muitas vezes no passado, quando o efeito das interações era sistematicamente incorporado ao erro não-controlado do experimento. Dessa forma, os efeitos das interações só podem ser incorporados no resíduo (o erro não controlado), quando a interação for estatisticamente não significativa. No caso das interações serem significativas, a sua variância deve ser isolada, e tratada como se fosse um fator de variação.

b) Os erros devem ser independentes (independência):

A independência dos erros existe quando os erros controlados de um fator de variação não interferem com os erros controlados de outros fatores. Os erros

devem ser independentes, ou seja, a probabilidade de erro de uma observação ter certo valor, não depende dos valores dos erros de outras observações (BANZATTO; KRONKA, 2006).

A causa mais provável da dependência do erro, segundo Dean e Voss (1999), é a similaridade de unidades experimentais juntas no tempo ou no espaço. No caso da dependência, o verdadeiro nível de significância do teste de hipótese pode ser muito maior que o indicado, e o verdadeiro nível de confiança e poder do teste pode ser muito mais baixo que o esperado.

c) Os erros devem ser normalmente distribuídos (normalidade):

De acordo com Dean e Voss (1999), quando essa hipótese não é satisfeita, além da introdução de erro no nível de significância do teste F e de outros, há uma perda de eficiência na estimação dos efeitos de tratamento e uma correspondente perda de poder dos testes.

A falta de normalidade dos erros, de acordo com Conagin et al. (1993), é um fator que pode distorcer os resultados de testes empregados para verificar a hipótese de igualdade de médias ou das variâncias, apesar da robustez do teste F.

d) Os erros devem apresentar variâncias comuns (homocedasticidade):

A falta de homogeneidade de variância é uma das mais graves quebras de suposição básica, principalmente para os modelos não balanceados e os modelos de efeitos aleatórios (BANZATTO; KRONKA, 2006).

Segundo Conagin et al. (1993), a falta de homogeneidade pode surgir quando os erros têm distribuição assimétrica.

Quando uma dessas pressuposições não é satisfeita, a análise de variância pode não ter validade como técnica de análise estatística. Nesse sentido, Conagin et al. (1993), consideram que a verificação dessas pressuposições melhora a qualidade da análise dos experimentos, permitindo ao pesquisador melhor conhecimento da natureza de seus dados.

Quando as pressuposições não são satisfeitas, a análise paramétrica por meio de algum teste de comparações múltiplas de médias ou pelo ajuste de modelos de regressão ficam prejudicadas, podendo levar a falsas conclusões (MARQUES et al., 2000).

Com relativa frequência, constata-se que uma ou mais dessas hipóteses básicas, não se verifica, e nesse caso, antes de se proceder a análise de variância, uma alternativa é utilizar uma transformação dos dados experimentais, de tal forma que as suposições básicas sejam satisfeitas (BANZATTO; KRONKA, 2006).

Se todos os tratamentos tiverem o mesmo número de repetições e se a desigualdade das variâncias não for muito acentuada, a análise da variância e os testes de comparações de médias podem ser feitos sem sérios prejuízos (BOX, 1954), mas, sabendo que os testes estatísticos aplicados são aproximados.

2.2 Séries temporais e processos estacionários

A partir da possibilidade de se medir no tempo, foi possível estabelecer algumas relações entre a passagem do tempo e a ocorrência de determinados fenômenos. Dessa forma, segundo Morettin e Tolo (2006), define-se que uma série temporal é um conjunto de observações ordenadas em intervalos de tempo equidistantes, e que podem apresentar uma dependência serial entre elas. Entretanto, a variável tempo pode ser substituída por qualquer outra variável, como espaço, profundidade etc. E, desse modo, pode-se obter séries temporais discretas e contínuas. As quantidades anuais de chuva de uma determinada cidade e valores diários do preço das ações de uma empresa, são exemplos de séries temporais discretas. As séries temporais contínuas podem ser o registro de um eletrocardiograma de uma pessoa, ou as alturas de marés no porto de Santos, obtidas por meio de um marégrafo.

Em séries temporais, os processos estacionários desempenham um papel fundamental. Segundo Chiann (1997), a suposição básica para desenvolver uma teoria assintótica é a estacionariedade, devido ao fato de que esta garante que as estruturas do processo no presente, contêm as informações do futuro do mesmo.

As definições a seguir foram retiradas de Morettin (2014).

Definição 1: Um processo estocástico $X = X(t, \omega), t \in T, \omega \in \Omega$ é uma família de variáveis aleatórias definidas em um espaço de probabilidade (ω, Λ, P) e indexadas pelos elementos de um conjunto de parâmetros T .

Assim, um processo estocástico é uma função de dois argumentos, $X(t, \omega)$, $t \in T$, $\omega \in \Omega$. Para cada $t \in T$, a função $X(t, \cdot)$ é mensurável relativamente a Λ . Por outro lado, para cada $\omega \in \Omega$, fixado, obtem-se uma função $X(\cdot, \omega)$ de $t: T \rightarrow \mathbb{R}$, que é chamada trajetória, realização, função amostral do processo ou série temporal.

Para facilitar a notação, $X(t, \omega)$ será representado por $X(t)$.

Definição 2: Dado um processo estocástico $X(t)$, $t \in T$, define-se suas distribuições finito-dimensionais por

$$F(x_1, \dots, x_n; t_1, \dots, t_n) = P\{X(t_1) \leq x_1, \dots, X(t_n) \leq x_n\},$$

para todo $n \geq 1$ e $t_1, \dots, t_n \in T$.

Definição 3: Um processo estocástico $X(t)$, $t \in T$ é estritamente estacionário se todas as distribuições finito-dimensionais permanecem as mesmas sob uma translação do tempo, ou seja,

$$F(x_1, \dots, x_n; t_1 + \tau, \dots, t_n + \tau) = F(x_1, \dots, x_n; t_1, \dots, t_n),$$

para quaisquer $t_1, \dots, t_n, \tau \in T$.

Segundo Morettin (2014), isso significa, em particular, que todas as distribuições unidimensionais são invariantes sob translações do tempo. Logo, a média $\mu(t)$ e a variância $\sigma^2(t)$ são constantes, isto é,

$$E\{X(t)\} = \mu(t) = \mu, \text{ para todo } t \in T,$$

$$Var\{X(t)\} = \sigma^2(t) = \sigma^2, \text{ para todo } t \in T.$$

A covariância de um processo $X(t)$ é dada por

$$Cov[X(t_1), X(t_2)] = E[X(t_1)X(t_2)] - E[X(t_1)]E[X(t_2)] = \gamma_{xx}(t_1, t_2).$$

Em particular, se $t_1 = t_2$, tem-se a variância do processo $Var\{X(t)\} = \gamma(t, t)$.

Definição 4: Um processo estocástico $\{X(t), t \in T\}$ é estacionário de segunda ordem, ou fracamente estacionário, ou ainda estacionário em sentido amplo se, e somente se,

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

Segue-se que $Cov\{X(t), X(t + u)\}$ depende somente de u e assim

$$Cov\{X(t), X(t + u)\} = \gamma(u), \quad u = 0, \pm 1, \pm 2, \dots,$$

e $\gamma(u)$ é chamada a função de autocovariância de $\{X(t)\}$.

2.3 Ondaleta

Ondaletas, de acordo com Morettin (2014), são funções que satisfazem certas propriedades, podendo ser suaves ou não, simétricas ou não, e podem ter expressões matemáticas simples ou não.

Segundo Chiann e Morettin (1998), ondaleta é uma técnica útil para lidar com as séries temporais não-estacionárias e com os sinais contaminados com ruídos, devido ao fato de que as funções ondaletas são localizadas no tempo, ou seja, elas permitem analisar uma série no tempo e escala simultaneamente.

Uma função ondaleta é a interpretação de uma onda de curta duração, com crescimento e decrescimento rápidos. Sua teoria baseia-se na representação de funções em diferentes escalas e diferentes resoluções (tempo-escala), considerando assim, uma das suas principais características (DAUBECHIES, 1992).

Segundo Percival e Walden (2006) ondaletas são ferramentas matemáticas, para analisar séries temporais ou imagens, sendo em alguns casos, complementares às técnicas já existentes (correlação e análise espectral) e em outros, capazes de resolver problemas para os quais foram feitos poucos progressos antes de se introduzir ondaletas.

Uma ondaleta, como o próprio nome sugere, é uma pequena onda, que cresce e decai em um período de tempo limitado. Uma grande onda, em contraste, seria por exemplo, a função seno, que se mantém oscilando em todo o domínio real (PERCIVAL; WALDEN, 2006).

A análise de ondaleta é uma técnica matemática muito útil para análise numérica e manipulação de sinais discretos uni ou multidimensionais. E tem se tornado muito popular em diversas áreas, sendo uma ferramenta com diversos usos, com um rico conteúdo matemático e uma grande aplicabilidade. Pode-se encontrar estudos envolvendo ondaletas em várias áreas, como convecção tropical (WENG; LAU, 1994), frentes frias atmosféricas (GAMAGE; BLUMEN, 1993), estruturas coerentes em fluxos turbulentos (FARGE, 1992), processamento e codificação de sinais e imagens (MALLAT, 1989).

Meyer (1993) apresentou uma retrospectiva histórica das ondaletas, desde Fourier (1807), passando por Haar (1910) e Grossmann e Morlet (1980). Seu uso tem crescido exponencialmente, especialmente pelo fato de representar a síntese de técnicas antigas, associadas a resultados matemáticos robustos e algoritmos computacionais eficientes (DAUBECHIES; MALLAT; WILLSKY, 1992).

As ondaletas no enfoque estatístico foram utilizadas na estimação de densidades (DONOHO, 1993; HALL; PATIL, 1995), na análise de regressão não paramétrica (DONOHO; JOHNSTONE, 1998; HALL; PATIL, 1996) e na estimação da densidade espectral (GAO, 1997; MOULIN, 1993).

A análise de ondaleta tem vantagem sobre as análises espectrais clássicas, porque permite analisar a periodicidade de eventos em diferentes escalas da variabilidade temporal e não necessita de uma série estacionária. Assim, a ferramenta é apropriada para analisar eventos irregularmente distribuídos e séries temporais que contenham potências não estacionárias, em diferentes frequências. Por isso, a transformada de ondaleta está se tornando uma ferramenta comum para analisar variações locais de potência dentro da série (SANTOS; FREIRE; TORRENCE, 2013).

A ideia, tanto na análise de Fourier, quanto na análise de ondaletas, é aproximar uma função por uma combinação linear de senos e cossenos, ou ondaletas, respectivamente (MORETTIN, 2014).

Na análise de Fourier, toda função periódica, de período 2π , de quadrado integrável [†], ou seja, de $L^2(0, 2\pi)$, é gerada por uma superposição de exponenciais complexas, $w_n(t) = e^{int}$, $n = 0, \pm 1, \dots$, obtidas por dilatações da função $w_t = e^{it}$. Essa ideia foi estendida para $L^2(\mathbb{R})$, isto é, esse espaço é gerado a partir de uma única função ψ , chamada de *ondaleta mãe*. Isto é possível por dilatações (ou compressões) e translações de ψ , gerando uma família de funções $\psi_{a,b}$ (VETTERLI; KOVACEVIC, 1995),

$$\psi_{a,b}(t) = |a|^{-1/2} \psi\left(\frac{t-b}{a}\right), b \in \mathbb{R}, a > 0. \quad (2.1)$$

Quando se trabalha com sinais discretos, usualmente se utiliza valores especiais para a e b : $a = 2^j$ e $b = k2^j$, $j, k \in \mathbb{Z}$. Dessa forma, considerando um caso particular da equação (2.1), tem-se a expressão para a ondaleta:

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k), j, k \in \mathbb{Z}, \quad (2.2)$$

ou seja, $\psi_{j,k}(t)$ é obtida de $\psi(t)$ por uma dilatação binária 2^j e uma translação diádica $k2^j$. Na análise de ondaletas, segundo Morettin (2014), os parâmetros j e k representam escala e localização temporal, respectivamente. Assim, de forma intuitiva, escala pode ser pensada como como “inverso da frequência”. As funções $\{\psi_{j,k}(t), j, k \in \mathbb{Z}\}$ formam uma base que não precisa ser necessariamente ortogonal. Contudo, ao se trabalhar com bases ortogonais é possível a reconstrução perfeita do sinal original, a partir dos coeficientes da transformada. Daí o interesse em considerar bases de ondaletas que sejam ortogonais (MORETTIN, 2014).

Há diversos tipos de ondaletas, e algumas possuem características mais adequadas para determinadas aplicações. A descoberta de bases ortonormais da forma da equação (2.2) e de suporte compacto [‡] (não nulas em um intervalo finito) se deve a Ingrid Daubechies. Daubechies (1992), inspirada nos trabalhos de Mallat, estendeu o trabalho de Haar, sintetizando essa família de ondaletas ortonormais e possibilitando uma análise e síntese mais eficiente do que a obtida

[†]Uma função f é do tipo quadrado integrável sobre um intervalo (a, b) se a integral $\int_{a,b} |f(x)|^2 dx$ existe.

[‡]Segundo Souza (2008), uma ondaleta é dita de suporte compacta, se toda a energia dessa ondaleta está restrita a um intervalo finito, ou seja, se a função é exatamente zero fora desse intervalo.

com outros sistemas.

Existem várias famílias de ondaletas, sendo que as mais utilizadas na prática, têm suporte compacto, tornando-se ideais para aplicações em séries temporais (NEGRI, 2012).

As ondaletas mais utilizadas são (MORETTIN, 2014):

a) Haar:

A ondaleta de Haar é uma onda quadrática, foi introduzida pelo matemático Haar em 1910 e forneceu a primeira representação em série de ondaleta ortogonal. A ondaleta de Haar tem suporte compacto, sendo a única ondaleta ortogonal compacta, mas não é contínua e possui uma forma analítica:

$$\psi(t) = \begin{cases} 1, & 0 \leq t < 1/2 \\ -1, & 1/2 \leq t < 1 \\ 0, & \text{caso contrário} \end{cases}$$

e

$$\psi_{j,k}^{(H)}(t) = \begin{cases} 2^{j/2}, & 2^{-j}k \leq t < 2^{-j}(k + 1/2) \\ -2^{j/2}, & 2^{-j}(k + 1/2) \leq t < 2^{-j}(k + 1) \\ 0, & \text{caso contrário} \end{cases}$$

Na Figura 1, é ilustrada a ondaleta de Haar.

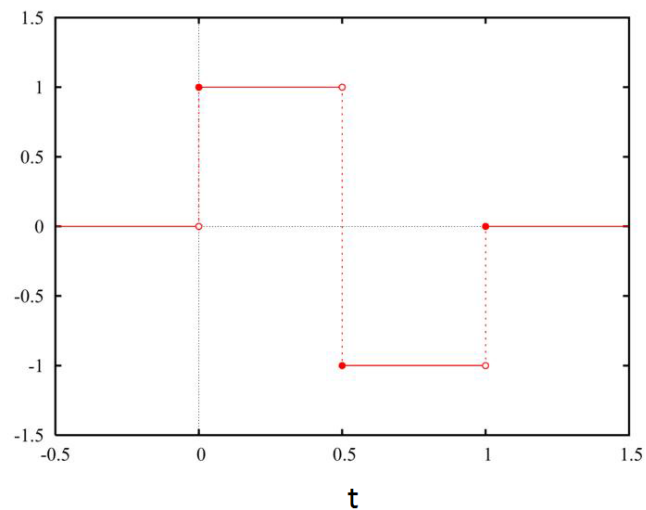


Figura 1 Ondaleta de Haar

Fonte: Negri (2012)

a) Daubechies:

As ondaletas de Daubechies, não têm uma forma explícita, e são calculadas de forma iterativa. São ortonormais de suporte compacto, o que faz a análise discreta de ondaletas mais praticável. São numeradas de acordo com o número de momentos nulos (db1, db2, etc.). A ondaleta db1 equivale à ondaleta de Haar e possui um momento nulo. Na Figura 2, pode-se observar o segundo e terceiro membros dessa família.

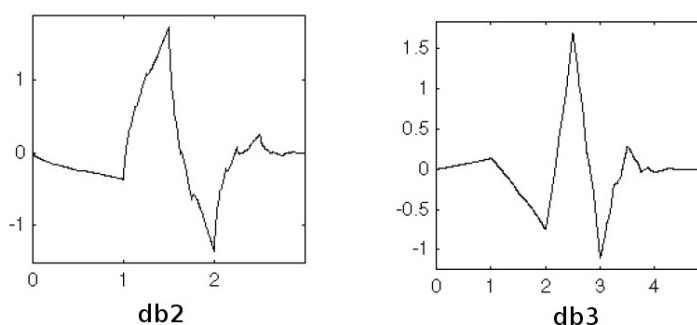


Figura 2 Representação do segundo e do terceiro membro da família de ondaletas de Daubechies

Fonte: Negri (2012)

2.3.1 Transformada discreta de ondaleta decimada

A transformada de ondaleta é uma ferramenta conhecida pela característica de decompor funções e de reconstruí-las, apresentando uma resolução razoavelmente boa. Por exemplo, a reconstrução de um sinal que apresente coeficientes com valores próximos de zero requer um trabalho árduo, além de um alto custo computacional.

Tem-se duas formas da transformada, a contínua e a discreta. A transformada contínua de ondaleta é de grande interesse teórico, principalmente para a derivação e compreensão das propriedades matemáticas das funções ondaletas. Porém, a sua discretização é necessária para aplicações práticas, por exemplo, ao descrever um sinal unidimensional em uma representação bidimensional, ou quando se tem a necessidade de se inverter a operação. O problema dos parâmetros a e b em (2.1) variarem continuamente é resolvido com a operação de discretização, surgindo dessa maneira, a transformada discreta de ondaleta, com a finalidade de propor mais eficiência ao trabalho (CHUI, 1992).

A transformada de ondaletas tem qualidades atraentes que a fazem um método muito útil para séries temporais, exibindo características que poderiam variar tanto em tempo como em frequência (ou escala). Ver Figura 3.

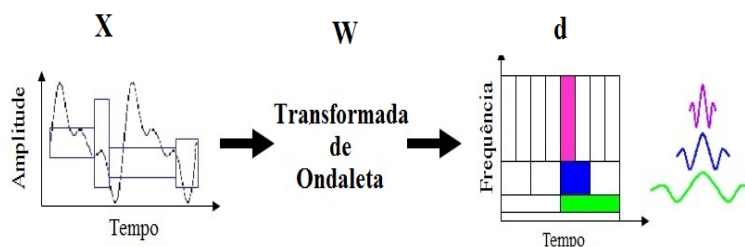


Figura 3 Transformada de ondaleta: decomposição em tempo escala com tamanhos de janelas variáveis.

Fonte: Misiti et al. (2014)

A Figura 3, mostra que os tamanhos de janelas para localização de frequências são ajustáveis a cada trecho do sinal, de forma que elas conseguem localizar altas e baixas frequências, simultaneamente, em um mesmo trecho temporal.

Para Daubechies (1992), a função ondaleta mãe, do ponto de vista matemático, gera as funções ondaletas filhas que irão se decompor hierarquicamente, conforme esquema apresentado na Figura 4. Para cada decomposição são gerados dois subsinais de comprimento $n/2$, onde n representa o sinal inteiro. O primeiro subsinal corresponde aos sinais escalas A_1 , enquanto que o segundo, aos sinais ondaletas D_1 .

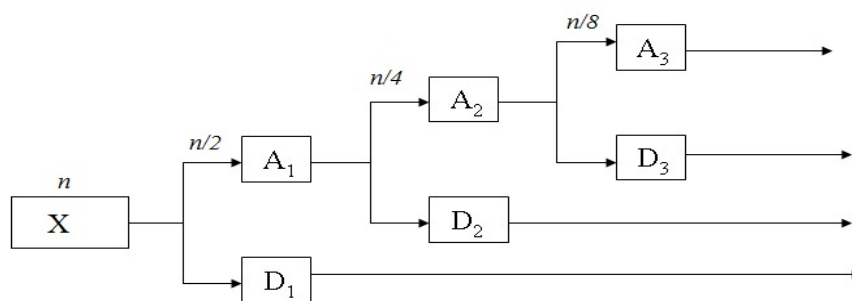


Figura 4 Esquema do processo de decomposição da Transformada Discreta Ondaleta

Fonte: Negri (2012)

Para criar as funções ondaletas-filhas, além de escolher o tipo de ondaletas, é necessário utilizar um banco de filtros, compostos por coeficientes especiais (DAUBECHIES, 1992). Toda transformada discreta de ondaleta decompõe um sinal discreto, em dois subsinais com a metade do seu tamanho (MALLAT, 2008). Para tal, são empregados filtros digitais, de passa-baixa, os quais produzem os coeficientes de aproximação do sinal A1, e de passa-alta, que produzem os coeficientes de detalhamento do sinal D1. O Anexo 1, apresenta informações sobre os filtros passa-baixa e de passa-alta.

Processando os coeficientes de aproximação A1, novamente como dados de entrada com o mesmo banco de filtros, gera-se assim, outro conjunto de coeficientes de aproximação e de detalhe, com um nível de resolução inferior. Porém, é desejável que o sinal que se pretende compactar, esteja com dimensão 2^n , com n pertencendo aos números naturais ou de valor par, já que este sinal é sempre reduzido à metade, em cada nível de compactação (DAUBECHIES, 1992).

A transformada discreta de ondaleta de $X(t)$ com relação a ondaleta (2.1) pode ser definida como

$$d_{j,k}^{(\psi)} = \sum_{t=0}^{T-1} X(t)\psi_{j,k}(t). \quad (2.3)$$

em que $X = (X_0, X_1, \dots, X_{T-1})'$ e $T = 2^M$ é o número de observações. A equação (2.3) é calculada para $j = 1, 2, \dots, M$ e $k = 0, 1, \dots, (2^{(M-j)} - 1)$, com $M \in \mathbb{Z}$ representando a escala mais fina, percorrendo T coeficientes d .

É possível escrever a transformada discreta de ondaleta (2.3) na forma matricial

$$\mathbf{d} = \mathbf{W}\mathbf{X}, \quad (2.4)$$

o que facilita a definição de sua inversa, sendo que \mathbf{W} , apresentada na Figura 3, representa a matriz dos coeficientes de ondaleta e \mathbf{d} o vetor contendo os $d_{j,k}$.

É importante destacar que (2.4) não é obtida, na prática, pela multiplicação

matricial, mas pelo algoritmo piramidal de Mallat, ver Morettin (2014), com complexidade $O(n)$, consistindo numa sequência de filtros passa-baixo e passa-alto.

2.3.2 Transformada discreta de ondaleta não decimada

A transformada discreta de ondaleta (decimada), pode ser vista como um processo de filtragem seguido por uma decimação (diádica) por 2, em que são aproveitados apenas os elementos pares ou ímpares, implicando em selecionar uma base diferente (NEGRI, 2012). A transformada discreta de ondaleta seleciona os valores filtrados, de modo que em cada escala, o número de coeficientes é a metade do número de coeficientes da escala anterior. Em contraste, a transformada discreta de ondaletas não decimada, é não ortogonal e redundante, onde cada um dos vetores de coeficientes de ondaletas conterá n elementos, se o conjunto de dados $X(t)$ contiver n valores (MORETTIN, 2014)..

Assim, como a transformada discreta de ondaleta decimada, que é obtida por meio de um algoritmo piramidal com complexidade $O(n)$, há um algoritmo para calcular a transformada discreta de ondaleta não decimada, com uma complexidade um pouco menor, da ordem de $O(\log_2 n)$, que também é computacionalmente eficiente.

Segundo Morettin (2014), uma outra diferença entre a transformada discreta de ondaleta decimada e a transformada discreta de ondaleta não decimada é que a transformada discreta de ondaleta decimada é dependente da escolha da origem dos dados. Por exemplo, na transformada discreta de ondaleta não decimada, uma translação dos dados não resulta em um conjunto distinto de coeficientes de ondaletas, o que acontece com a transformada discreta de ondaleta.

Assim, segundo Negri (2012), uma translação na entrada dos dados pode resultar em um conjunto de coeficientes de ondaletas completamente diferente, se comparado ao conjunto de dados obtidos com a entrada original. Em algumas aplicações, pode ser necessário que o método não seja sensível à origem, ou seja, é preferível que seja invariante à translação, o que é conseguido com a ondaleta não decimada, pois esta leva em consideração todos os elementos: pares e ímpares.

A expressão para a ondaleta não decimada, correspondente à equação (2.2), para $\psi_{j,k}(t)$ se torna

$$\psi_{j,k}(t) = 2^{-j/2} \psi \{2^{-j}(t - k)\}, j, k \in \mathbb{Z}. \quad (2.5)$$

A transformada de ondaleta não decimada é obtida eliminando-se a decimação na transformada discreta de ondaleta decimada. A fórmula (2.3) permanece a mesma, substituindo $\psi_{j,k}(t)$ pela expressão (2.5).

De acordo com Morettin (2014), a ideia básica da transformada discreta de ondaleta não decimada é conservar ambas as decimações, pares e ímpares, em cada nível, e manter esse procedimento ao longo do algoritmo piramidal.

A Figura 5, mostra o cálculo da transformada discreta de ondaleta não decimada. Nesse caso, não se realiza a decimação e modificam-se os filtros a cada escala.

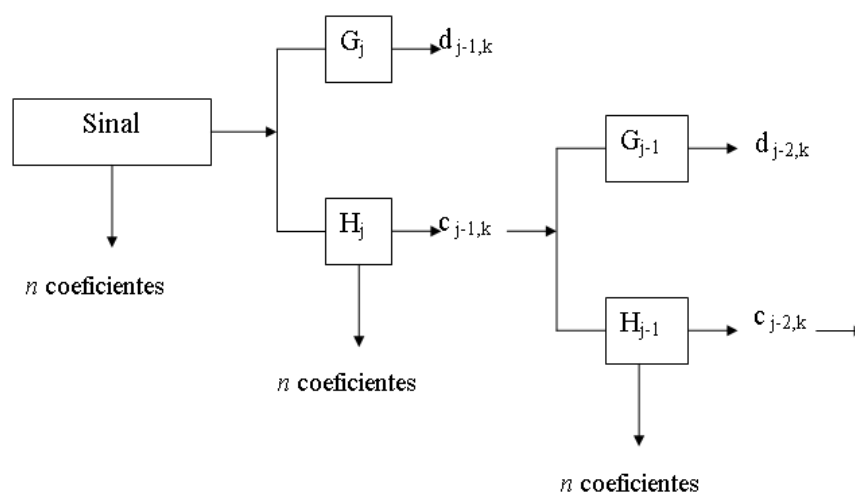


Figura 5 Esquema do processo de decomposição da Transformada Discreta Ondaleta Não Decimada

Fonte: Negri (2012)

2.3.3 O espectro de ondaleta

Uma das maneiras de descrever séries temporais é por meio de modelos não-paramétricos, ou seja, modelos com um número infinito de parâmetros. Um dos modelos não-paramétricos mais utilizados é a análise espectral via exponenciais (Fourier), no domínio da frequência, ou a análise da função de autocovariância, no domínio do tempo (CHIANN, 1997). Com o surgimento das ondaletas, foi desenvolvida uma técnica de análise espectral via ondaletas. Essa teoria será apresentada para processos estacionários e processos não-estacionários.

A análise espectral via ondaleta é feita em tempo-escala, tendo então, dois parâmetros: tempo e escala.

Nesta seção será considerada a análise espectral para o caso de um processo estacionário e de um processo não estacionário.

a) Processo estacionário

Chiann e Morettin (1998) desenvolveram uma análise espectral via ondaletas para um processo estacionário discreto.

Suposição 1: Assumindo que $X(t), t = 0, \pm 1, \dots$ é uma série temporal discreta, com média zero, estacionária de segunda ordem. Seja

$$\gamma(u) = E \{X(t)X(t + u)\},$$

$u = 0, \pm 1, \dots$, a função de autocovariância de $X(t)$. Sendo $X(0), X(1), \dots, X(T - 1)$ com $T = 2^M$, uma amostra da série $X(t), t \in Z$, sendo M um inteiro positivo.

Suposição 2: Para $u = 0, \pm 1, \dots$ e $\gamma(u) = E \{X(t)X(t + u)\}$, tem-se

$$\sum_{u=-\infty}^{\infty} [1 + |u|]|\gamma(u)| < \infty.$$

Supondo que as suposições 1 e 2 estão satisfeitas, então o espectro de ondaletas da série $X(t), t = 0, \pm 1, \dots$ em (j, k) , com respeito a ψ , é definido como

$$\eta_{(j,k)}^{(\psi)} = \sum_{u=-\infty}^{\infty} \gamma(u) \Psi_{j,k}(u), \quad j, k \in \mathbb{Z},$$

em que

$$\Psi_{j,k}(u) = \sum_{t=0}^{\infty} \psi_{j,k}(t) \psi_{j,k}(t + |u|),$$

é chamada a função de autocorrelação de ondaletas, em (j,k) , e $\psi_{j,k}(\cdot)$ é definida em (2.2).

A seguir serão apresentados teoremas que foram extraídos de Chiann e Morettin (1998). As demonstrações dos teoremas encontram-se no referido artigo.

Teorema 1: Sob as suposições 1 e 2, $\eta_{j,k}^{(\psi)}$ é limitado e não-negativo.

Teorema 2: Admitindo a suposição 1, 2 e $T = 2^M$. Seja

$$\eta_{(j,j')(k,k')}^{(\psi)} = \sum_{u=-\infty}^{\infty} \sum_{t=0}^{\infty} \gamma(u) \psi_{j,k}(t + |u|I_{\{u>0\}}) \psi_{j',k'}(t + |u|I_{\{u\leq 0\}}),$$

para $j, j', k, k' \in \mathbb{Z}$, então

- i) $E \{d_{j,k}^{(\psi)} d_{j',k'}^{(\psi)}\} \rightarrow \eta_{(j,j')(k,k')}^{(\psi)}$, quando $T \rightarrow \infty$;
- ii) Se $j = j', k = k'$, então $\eta_{(j,j')(k,k')}^{(\psi)} = \eta_{j,k}^{(\psi)}$;
- iii) $E \{d_{j,k}^{(\psi)} d_{j',k'}^{(\psi)}\} = O(1)$, quando $T \rightarrow \infty$ e $\forall (j,k), (j',k')$.

Observação: $\eta_{(j,j')(k,k')}^{(\psi)}$ é chamado a covariância (assintótica) da transformada de ondaleta com respeito à ψ .

Pelo Teorema 2, item (iii), tem-se que a covariância assintótica das

transformadas finitas de ondaletas em dois pares distintos (j, k) e (j', k') não é necessariamente zero, diferente do caso da transformada de Fourier, onde as covariâncias são assintoticamente zero, sob algumas suposições. Entretanto, em algumas situações a independência pode existir.

Teorema 3: Se $\psi(t)$ tem suporte $[K_1, K_2]$, onde $K_1 \geq 0$ e $K_2 \geq 0$ e $X(t)$ é uma série temporal discreta, com média zero e estacionária de segunda ordem com $\gamma \neq 0$ para $|u| \leq U$, $U \ll T$, então $E \left\{ d_{j,k_1}^{(\psi)} d_{j,k_2}^{(\psi)} \right\} = 0$ para $|k_1 - k_2| > K_2 - K_1 + \frac{U}{2^j}$, $j = 1, \dots, M$ e $k = 0, \dots, (2^{M-j} - 1)$.

Isto é, escolhendo uma ondaleta com valor grande para K_2 e valor pequeno para K_1 , poucos coeficientes de ondaleta satisfazem a condição $|k_1 - k_2| > K_2 - K_1 + \frac{U}{2^j}$ e, então as $d_{j,k}^{(\psi)}$ são independentes somente sobre um conjunto menor.

Chegou-se a um teorema limite central para a transformada de ondaletas. O teorema 4 e a demonstração do mesmo, seguem Hannan (1973).

Teorema 4: Suponha que $X(t)$ seja estritamente estacionário e seja M_t a σ -álgebra gerada por $(X(j), j \leq t)$. Defina

$$\alpha_j = \left[E \left\{ E(X(t)/M_{t-j}) - E(X(t)/M_{t-j-1}) \right\}^2 \right]^{1/2}, \quad j \geq 0.$$

Assume-se que $M_{-\infty}$ é trivial e que

$$\sum_{j=0}^{\infty} \alpha_j < \infty.$$

i) Para $(j_1, k_1), \dots, (j_p, k_p)$ e $m, n = 1, \dots, p$, suponha que

$$E \left\{ d_{j_m, k_m}^{(\psi)} d_{j_n, k_n}^{(\psi)} \right\} \rightarrow \eta_{(j_m, j_n), (k_m, k_n)}^{(\psi)}, \quad \text{quando } T \rightarrow \infty$$

Então,

$$\Phi \xrightarrow{D} N_p(0, \Delta),$$

em que

$$\Delta = \left(\eta_{(j_m, j_n), (k_m, k_n)}^{(\psi)} \right),$$

e

$$\Phi = (d_{j_1, k_1}^{(\psi)}, \dots, d_{j_p, k_p}^{(\psi)})'.$$

ii) Em particular, tem-se

$$d_{j,k}^{(\psi)} \xrightarrow{D} N(0, \eta_{j,k}^{(\psi)}).$$

Observação: O teorema 4 é provado sob uma forma de “strong mixing condition”, isto é, uma independência assintótica.

As demonstrações dos teoremas 1, 2, 3 e 4 são apresentadas por Chiann e Morettin (1998).

b) **Processo não estacionário**

Não existe uma generalização natural da definição de processo estacionário para não estacionário. Um dos maiores problemas é como estabelecer uma teoria assintótica adequada. Se temos observações X_1, X_2, \dots, X_T de um processo não-estacionário, o conceito assintótico para processos estacionários tem que ser modificado, no sentido que, quando o processo é estendido para o futuro ($T \rightarrow \infty$), não se pode obter nenhuma informação do comportamento do processo utilizando-se as observações do mesmo. Para contornar esse problema, estabeleceu-se um conceito assintótico diferente (CHIANN, 1997).

No que se refere a processos não estacionários, Morettin (2014) afirma que várias tentativas foram feitas para tratar formas especiais de não estacionariedade no domínio da frequência, chegando-se ao espectro dependente do tempo. Dessa forma, tem-se um espectro dependente do tempo e da frequência.

O espectro dependente do tempo, apresenta uma dificuldade no estudo de processos não estacionários, pois não se pode estabelecer uma teoria assintótica adequada, que permita obter vieses, variâncias e distribuições assintóticas. Nos processos estacionários, o aumento do tamanho da amostra, conduz a mais informações do mesmo tipo sobre o processo, pois a estrutura probabilística não se altera, o que não acontece com os processos não estacionários. Por este motivo, Dahlhaus (1997) introduziu a classe dos processos localmente estacionários.

Uma classe de processos, denominados de processos de ondaleta localmente estacionários, foi introduzida por Nason, Sachs e Kroisandt (2000), que é uma extensão da teoria assintótica para ondaletas proposta por Chiann e Morettin (1998). A ideia é obter uma representação de séries temporais que tenham a estrutura de segunda ordem, variando no tempo. Uma possibilidade é considerar os processos localmente estacionários propostos por Dahlhaus (1997), onde o enfoque é usar um conjunto de ondaletas discretas não decimadas. Segundo Morettin (2014), um fato importante é que as ondaletas discretas não decimadas, podem ser transladadas para qualquer posição definida pelas ondaletas de resolução mais fina, no algoritmo Piramidal de Mallat, e não somente por translações de 2^{-j} como na transformada discreta de ondaleta decimada.

Definindo $\psi_{j,k}$ como o k -ésimo elemento do vetor ψ_j , e $\psi_{j,k}(\tau)$ como o k -ésimo elemento do vetor $\psi_{j(k-\tau)}$, ou seja, $\psi_{j,k}(\tau) = \psi_{j,k-\tau}$ para cada escala τ (MORETTIN, 2014).

Definição 5: Os processos localmente estacionários são uma sequência duplamente indexada $\{X_{t,T}, t = 0, \dots, T - 1\}$, $T = 2^J \geq 1$, tendo a representação no sentido quadrático médio

$$X_{t,T} = \sum_{j=-J}^{-1} \sum_k w_{j,k;T} \psi_{j,k}(t) \xi_{j,k}, \quad (2.6)$$

em que $\xi_{j,k}$ é uma sequência aleatória ortonormal, $\{\psi_{j,k}\}_{j,k}$ é uma família de ondaleta discreta não decimada para $j = -1, -2, \dots, -J$, $k = 0, \dots, T - 1$ com base em uma ondaleta mãe $\psi(t)$ de suporte compacto e $w_{j,k;T}$ são as amplitudes, com as propriedades:

i) $E \{\xi_{j,k}\} = 0$, para quaisquer j, k . Logo, $E \{X_{t,T}\} = 0$ para quaisquer t e T .

$$\text{ii) } \text{Cov}(\xi_{j,k}, \xi_{l,m}) = \delta_{j,l} \delta_{k,m}.$$

iii) Existe para cada $j \leq -1$ uma função contínua Lipschitz $W_j(z)$ sobre $(0,1)$ satisfazendo:

$$\text{a) } \sum_{j=-\infty}^{-1} |W_j(z)|^2 < \infty, \text{ uniformemente em } z \in (0, 1);$$

b) as constantes Lipschitz L_j são uniformemente limitadas em j e $\sum_{j=-\infty}^{-1} 2^{-j} L_j < \infty$;

c) existe uma sequência de constantes C_j tais que, para cada T , $\sup_k \left| w_{j,k;T} - W_j\left(\frac{k}{T}\right) \right| \leq \frac{C_j}{T}$, onde para cada j , o sup é sobre $k = 0, \dots, T-1$ e $\sum_{j=-\infty}^{-1} C_j < \infty$.

O espectro de ondaleta evolucionário, mede o poder local (a contribuição para a variância) em um processo de ondaleta localmente estacionário a um particular tempo (reescalonado) z e escala j (escala pode ser vagamente interpretada como o lag de série temporal usual). O espectro de ondaleta evolucionário é análogo ao espectro para processo estacionário η (NASON; SACHS; KROISANDT, 2000).

Definição 6: A sequência $\{X_{t,T}\}$, $t = 0, \dots, T-1$, para a sequência infinita $T \geq 1$, tem espectro de ondaleta evolucionário definido por

$$S_j(z) = |W_j(z)|^2, \quad j = -1, \dots, -J(T), z \in (0,1) \quad (2.7)$$

com respeito à base $\{\psi_{j,k}\}$.

Segue-se, da definição (5) item (c), que

$$S_j(z) = \lim_{T \rightarrow \infty} |w_{j,2^j[zT];T}|^2,$$

com $\sum_{j=-\infty}^{-1} S_j(z) < \infty$, para todo $z \in (0,1)$.

Definição 7: A autocovariância de $X_{t,T}$ é definida por

$$c_T(z, \tau) = \text{Cov}(X_{[zT],T}, X_{[zT]+\tau,T})$$

e esta, quando $T \rightarrow \infty$, tende para

$$c(z, \tau) = \sum_{j=-\infty}^{-1} S_j(u) \Psi_j(\tau), \quad (2.8)$$

para $\tau \in \mathbb{Z}$, $0 < z \leq 1$, chamada função de autocovariância local. Em (2.8), $\Psi_j(\tau)$ é a função de autocorrelação de ondaletas, definida por

$$\Psi_j(\tau) = \sum_k \psi_{j,k}(0) \psi_{j,k}(\tau), \quad j < 0, \tau \in \mathbb{Z}.$$

Para processos estacionários, a dependência em z em (2.8) desaparece (MORETTIN, 2014). Segundo Nason, Sachs e Kroisandt (2000) o mais importante são $\Psi_j(0) = 1$, $\sum_{\tau} \Psi_j(\tau) = 0$ para todo j e $\sum_j 2^j \Psi_j(\tau) = \delta_{\tau 0}$.

Proposição 1: Como $T \rightarrow \infty$, uniformemente em $\tau \in \mathbb{Z}$ e $z \in (0,1)$, $|c_T(z, \tau) - c(z, \tau)| = O(T^{-1})$.

Definição 8: Defina o operador $A = (A_{jl})_{j,l < 0}$ por

$$A_{j,l} = \langle \Psi_j, \Psi_l \rangle = \sum_{\tau} \Psi_j(\tau) \Psi_l(\tau),$$

e a matriz de dimensão J é dada por $A_J = (A_{jl})_{j,l = -1, \dots, -J}$.

Teorema 5: A família $\{\Psi_j(\tau)\}_{j=-\infty}^{-1}$ é linearmente independente. Então

a) o espectro de ondaleta evolucionário é definido de forma única, dado o processo de ondaleta localmente estacionário e,

b) o operador A é invertível, e para cada J a norma $\|A_J^{-1}\|$ é limitada por alguma constante C_J .

A prova do Teorema (5) está apresentada no artigo de Nason, Sachs e Kroisandt (2000).

A ligação, apresentada por Nason, Sachs e Kroisandt (2000), entre processos estacionários e processos de ondaletas localmente estacionários está

representada pela seguinte proposição:

Proposição 2:

- a) Todo processo estacionário com função de autocovariância absolutamente somável é um processo de ondaleta estacionário.
- b) Reciprocamente, todo processo de ondaleta localmente estacionário, com a condição adicional $\sum_j 2^{-j} S_j < \infty$, é estacionário com autocovariância absolutamente somável.

2.3.4 O periodograma de ondaletas

A ideia central da análise de ondaletas, consiste em decompor um sinal em diferentes níveis de resolução. A 6, apresenta uma representação dessa decomposição em diferentes níveis de resolução e fornece um diagrama conhecido como periodograma de ondaleta, que permite uma interpretação da série temporal.

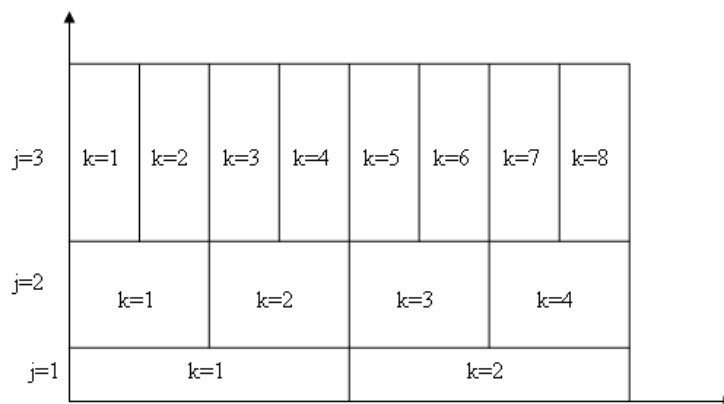


Figura 6 Resultado da decomposição de um sinal em diferentes níveis de resolução, conhecido como periodograma de ondaletas

Fonte: Bolzan (2006)

a) **Caso estacionário**

Tem-se uma sequência observada de valores $X(0), \dots, X(T-1)$ e baseados nesses valores deseja-se estimar o espectro de ondaleta $\eta_{j,k}^{(\psi)}$.

Em primeiro lugar, calcula-se a transformada discreta de ondaletas decimada com respeito a ψ :

$$d_{j,k}^{(\psi)} = \sum_{t=0}^{T-1} X(t)\psi_{j,k}(t).$$

Foi provado, por meio do Teorema 4, que $d_{j,k}^{(\psi)}$ é assintoticamente normal $N(0, \eta_{j,k}^{(\psi)})$, e sugere-se que $\eta_{j,k}^{(\psi)}$ pode ser estimado pela estatística

$$I_{j,k}^{(\psi)} = \left(d_{j,k}^{(\psi)} \right)^2 = \left[\sum_{t=0}^{T-1} X(t)\psi_{j,k}(t) \right]^2. \quad (2.9)$$

Nota-se que um espectro de ondaleta mede a potência local na decomposição da estrutura de variância-covariância do processo $X(t)$ numa certa escala j e uma localização de tempo $2^j k$. Para $j \geq 0$ fixo, essa quantidade é constante sobre a localização de tempo $2^j k$, isto é, a distribuição de $\eta_{(j,k)}^{(\psi)}$ é independente de k .

A estatística $I_{j,k}^{(\psi)}$ em (2.9) é chamada periodograma de segunda ordem de ondaleta, ou simplesmente periodograma de ondaleta dos valores $X(0), \dots, X(T-1)$.

Observação: Define-se a “energia” em uma série temporal como

$$\sum_{t=0}^{T-1} X^2(t),$$

então, usando o teorema de “Parseval” §, pode-se decompor a energia em uma série temporal com respeito à escala j e tempo k por $I_{j,k}^{(\psi)}$ para $j = 1, \dots, M$,

§ A potência média de um sinal periódico é igual à soma das potências médias das suas componentes de Fourier:

$k = 0, 1, \dots, 2^{(M-j)} - 1$ e $T = 2^M$.

Teorema 6: Seja a suposição 2 satisfeita, então

$$E \left[I_{j,k}^{(\psi)} \right] = \eta_{j,k}^{(\psi)} + O(T^{-1}).$$

A estrutura de covariância do periodograma é dada a seguir.

Teorema 7: Sob as suposições 1 e 2 e

$$\sum_{l,m,n=-\infty}^{\infty} |C_x(l,m,n)| < \infty$$

onde $C_x(l,m,n) = \text{Cum}(X(0), X(l), X(m), X(n))$ é o cumulante de quarta ordem de $X(t)$, tem-se

$$\text{Cov} \left\{ I_{j,k}^{(\psi)}, I_{j',k'}^{(\psi)} \right\} = 2 \left\{ \eta_{(j,j'),(k,k')}^{(\psi)} \right\}^2 + O(1), \text{ quando } T \rightarrow \infty.$$

Em particular se $j = j', k = k'$, obte-se a variância

$$\text{Var} \left[I_{j,k}^{(\psi)} \right] = 2 \left\{ \eta_{j,k}^{(\psi)} \right\}^2 + O(1)$$

Baseado nos Teoremas 1, 2, 3, 4 e 6, tem-se que as pressuposições da análise de variância, usando as transformadas de ondaletas são atendidas no que diz respeito à independência, normalidade e homocedasticidade para séries estacionárias.

$$\frac{1}{2\pi} \int_0^{2\pi} |f(x)|^2 = \sum_{n=-\infty}^{\infty} |C_n|^2$$

com

$$\sum_{n=-\infty}^{\infty} |C_n|^2 < \infty.$$

a) **Caso não estacionário**

A estimação do espectro evolucionário de ondaletas (2.7) é baseada na transformada de ondaletas discreta não decimada de $X_{1,T}, \dots, X_{T,T}$.

Definição 9: Os coeficientes de ondaletas empíricos de um processo de ondaleta localmente estacionário $X_{t,T}$ é dado por

$$d_{j,k;T} = \sum_{t=1}^T X_{t,T} \psi_{j,k}(t). \quad (2.10)$$

Para j fixo o número de somas na equação (2.10) não muda com T , pois a ondaleta é de suporte compacto (NASON; SACHS; KROISANDT, 2000).

A estatística básica, análoga ao periodograma usual no caso estacionário, é definida a seguir.

Definição 10: O periodograma de ondaleta de um processo de ondaleta localmente estacionário $X_{t,T}$ é dado por

$$I_{k,T}^j = |d_{j,k;T}|^2. \quad (2.11)$$

Segundo Morettin (2014), assumindo que as variáveis $\xi_{j,k}$ na equação (2.6) sejam normais, então o processo $X_{t,T}$ também será gaussiano. O periodograma de ondaleta tem propriedades que são análogas à do periodograma clássico (NASON; SACHS; KROISANDT, 2000).

Proposição 3: Para o estimador (2.11), tem-se (MORETTIN, 2014):

- a) $E \{I_{j;[zT],T}\} = \sum_l A_{j,l} S_l(z) + O(T^{-1}), \forall z \in (0,1)$.
- b) $Var \{I_{j;[zT],T}\} = 2 \left(\sum_l A_{j,l} S_l(z) \right)^2 + O(2^{-j} T^{-1}), \forall z \in (0,1)$.
- c) A correlação entre dois periodogramas $I_{j;k,T}$ e $I_{l;m,T}$ decai com o aumento da distância entre a translação k na escala j e a translação m na escala l aumenta. Se $j = l$, a correlação é zero quando $|k - m|$ excede a intersecção dos suportes das ondaletas.

d) Se $I(u) = [I_{i,[uT],T}]_{-1}^{-J}$ e $L(u) = [L_{j,[uT],T}]_{-1}^{-J}$, com $L(u) = A_J^{-1}I(u)$, então $E(L(u)) = S(u) + O(T^{-1})$, para todo $u \in (0,1)$, com $S(u) = [S_{l(u)}]_{-1}^{-J}$.

A parte (b) da Proposição 3 mostra que o periodograma de ondaleta não é consistente.

Dessa forma, assim como no caso estacionário, tem-se que as pressuposições da análise de variância são atendidas para o caso não estacionário, de acordo com Teorema 5 e a preposição 3.

2.4 Erro tipo I e poder de teste

O erro tipo I e poder do teste pode ser usado com critério de avaliação, para validar o método proposto. Ao realizar a análise de variância em dados de séries temporais, usando a transformada de ondaleta, o interesse está em tomar a decisão de aceitar ou rejeitar determinada afirmação (hipótese). Dessa forma, há um risco de tomar decisões incorretas que deverá ser considerado. Segundo Mood, Graybill e Boes (1974), os erros podem ser classificados de duas formas:

a) Erro tipo I: É o erro cometido ao se rejeitar a hipótese nula verdadeira, quando a hipótese deveria ser aceita. O erro tipo I é controlável pelo pesquisados e a probabilidade de se cometer esse erro é dada por

$$P[\text{Erro tipo I}] = P[\text{rejeitar } H_0 | H_0 \text{ verdadeira}] = \alpha;$$

b) Erro tipo II: É o erro cometido ao não se rejeitar a hipótese nula falsa. O erro tipo II não é controlável pelo pesquisados e a probabilidade de se cometer esse erro é dada por

$$P[\text{Erro tipo II}] = P[\text{não rejeitar } H_0 | H_0 \text{ falsa}] = \beta.$$

A probabilidade de que seja rejeitada uma H_0 falsa é denominada poder do teste e é dado por $1 - \beta$. Assim, o poder do teste corresponde à capacidade do teste de detectar diferenças.

3 MATERIAIS E MÉTODOS

A análise de variância exige que os pressupostos de aditividade, independência, normalidade e homocedasticidade, sejam atendidos. Como os dados de séries temporais são correlacionados no tempo, implica que o pressuposto da independência não é atendido. Assim, aplicando a transformada discreta de ondaletas (decimada e não decimada) nos dados, tem-se que os pressupostos são atendidos assintoticamente.

Uma solução foi proposta para o problema do não cumprimento dos pressupostos exigidos pela ANAVA, usando a transformada de ondaleta. O desempenho do método proposto neste trabalho foi avaliado por meio das taxas de erro tipo I e poder. Para isso, foram realizadas simulações sob a hipótese nula (H_0) e sob a hipótese alternativa (H_1), em situações onde se tem séries temporais estacionárias e séries temporais não estacionárias.

Quando se está trabalhando com séries temporais estacionárias, a transformada discreta de ondaleta decimada foi usada. No caso, onde se tem séries temporais não estacionárias, se propõe o uso da transformada discreta de ondaleta não decimada. Nos dois casos foram usadas as bases de Haar e Daubechies.

Além das simulações, foram realizadas análises de variância, usando ondaletas em dados reais, que se referem à série do número diário de óbitos por causas respiratórias e cardíacas, no período de 01 de janeiro de 1994 a 31 de dezembro de 1997, em São Paulo - SP.

Para a realização da análise de variância usando a transformada discreta de ondaleta (decimada e não decimada) será utilizado o programa R e o pacote waveslim (WHITCHER, 2010).

3.1 Simulação de séries estacionárias

A análise de variância para dados de séries temporais estacionárias, foi feita considerando-se a transformada discreta de ondaleta decimada nas bases de Haar e Daubechies.

A simulação foi feita de forma que os dados tivessem comportamentos

iguais. Duas situações foram consideradas. Na primeira, a simulação foi realizada considerando o modelo auto-regressivo de ordem 1 ($AR(1)$), dado por $X_t = \phi_1 X_{t-1} + a_t$, onde $\phi_1 = 0,35$ e $a_t \sim N(0,1)$. Nesta primeira situação, considerou-se três tratamentos, sendo que cada tratamento teve três repetições. Cada repetição foi representada por uma série temporal, com 512 observações, para que fosse possível a realização da decimação, onde $T = 512 = 2^9$.

Uma segunda situação foi realizada com a simulação de séries temporais categóricas, com comportamentos semelhantes usando a função `sample` (Uniforme discreta) do software R, cujo valor das observações variam de 1 a 6. Para essa segunda situação, foram considerados dois tratamentos, sendo que cada tratamento teve doze repetições. Em cada repetição considerou-se 128 observações, onde $T = 128 = 2^7$.

3.2 Séries não estacionárias

A análise de variância para dados de séries temporais não estacionária, foi feita considerando a transformada discreta de ondaleta não decimada, usando as bases de Haar e Daubechies.

Foram simuladas duas situações que definem comportamentos diferentes.

No primeiro estudo foram considerados 3 tratamentos, cada tratamento com três repetições, sendo que cada repetição foi representada por séries com tendência crescente usando um modelo com tendência e $AR(1)$ com 512 observações, dada por $X_t = \beta_0 + \beta_1 t + Y_t$, sendo que $Y_t = \phi_1 Y_{t-1} + a_t$, $a_t \sim N(0,1)$. Para o primeiro tratamento, tem-se $\beta_0 = 100$, $\beta_1 = 0,01$ e $\phi_1 = 0,35$, para o segundo tratamento $\beta_0 = 500$, $\beta_1 = 0,1$ e $\phi_1 = 0,60$ e para o terceiro tratamento, tem-se $\beta_0 = 1000$, $\beta_1 = 0,02$ e $\phi_1 = -0,35$.

Na segunda situação foi considerada a simulação de séries temporais categóricas, onde um comportamento sazonal foi implementado, com comportamento diferentes ($X(t)$ assumindo os valores 1, 2, 3, 4, 5 ou 6). Para essa segunda situação, foram considerados dois tratamentos, sendo que cada tratamento teve doze repetições. Em cada repetição considerou-se 128 observações, onde $T = 128 = 2^7$.

3.3 Análise de Variância usando Ondaleta com dados reais

Considerando-se dados reais, foram estudados dois casos:

3.3.1 Estudo da série do número diário de óbitos por causa respiratória no período de 01 de janeiro de 1994 a 31 de dezembro de 1997, em São Paulo-SP

Foram considerados quatro grupos (tratamentos) referentes ao número diário de óbitos por causas respiratórias. O primeiro grupo continha informações do número diário de óbitos em indivíduos até 12 anos de idade, o segundo grupo, informações do número diário de óbito por causa respiratória de indivíduos de 12 até 35 anos, o terceiro grupo foi o número diário de óbito por causa respiratória de indivíduos de 35 até 65 anos, e o quarto grupo, será o número diário de óbito por causa respiratória de indivíduos maiores de 65 anos de idade. Nesta situação, os anos foram considerados como as repetições (4 repetições). O objetivo foi verificar se existem diferenças entre os grupos de idade por óbito por causas respiratórias.

Os dados foram obtidos por meio do Projeto PASSO - Poluição Atmosférica e Saúde Social (PROJETO..., 2014).

3.3.2 Estudo da série do número diário de óbitos por causas cardíacas no período de 01 de janeiro de 1994 a 31 de dezembro de 1997, em São Paulo-SP

Foram considerados quatro grupos (tratamentos) do número diário de óbito por causas cardíacas, referentes a diferentes faixas etárias, definidas anteriormente (3.3.1). Nessa situação, na repetição foi considerada os anos (4 repetições). O problema tratado foi o de verificar se existem diferenças entre os grupos de idade por óbito por causas cardíacas.

Os dados foram obtidos por meio do Projeto PASSO - Poluição Atmosférica e Saúde Social (PROJETO..., 2014).

Para a realização da análise de variância usando a transformada discreta de ondaleta (decimada e não decimada) foi utilizado o software R e o pacote waveslim (WHITCHER, 2010).

4 RESULTADOS E DISCUSSÕES

4.1 Análise de variância usando ondaleta

Como resultado, será apresentado a análise de variância (ANAVA) usando as transformadas discretas de ondaletas decimada e não decimada que, conforme visto anteriormente, atendem, assintoticamente, aos pressupostos da análise de variância. Serão analisadas as variações nos dados cujas medidas respostas são séries temporais. As análises que serão apresentadas podem ser entendidas como extensões de procedimentos tradicionais usados na ANAVA. Dessa forma, o objetivo principal é a apresentação de uma versão completa da ANAVA aos componentes das transformadas discretas de ondaletas (decimada e não decimada) das séries observadas.

Para a análise será apresentado o caso estacionário e entre parenteses o não estacionário.

Assim, suponha que $X_{lq}(t)$, $l = 1, \dots, L$, $q = 1, \dots, Q$, $N = LQ$, $t = 0, 1, \dots, T - 1$, é a variável resposta observada na l -ésima repetição do tratamento q , no instante t , e que pode ser modelado por

$$X_{lq}(t) = \theta_{lq} + \Omega_q(t) + \epsilon_{lq}(t) \quad (4.1)$$

em que

$\theta_{lq} = E(X_{lq}(t))$: constante para todo t , quando $Y_{lq}(t)$ for estacionária;

$\Omega_q(t)$: processos estacionários com média zero e espectro de ondaleta $\eta_{\Omega_q}(S_{\Omega_q})$;

$\epsilon_{lq}(t)$: processos estacionários independentes com média zero e espectro de ondaleta $\eta_{\epsilon}(S_{\epsilon})$, tal que $\Omega_q(t)$ e $\epsilon_{lq}(t)$ são não correlacionados e $\epsilon_{ih}(t)$ e $\epsilon_{jz}(t)$ são mutuamente independentes, $i, j = 1, \dots, L$, $h \neq z$ e $i \neq j$.

Dada uma ondaleta mãe $\psi(t)$, constrói-se uma sequência de ondaletas por translações e dilatações de ψ , dada por (2) no caso da transformada discreta de ondaleta decimada e (5) para a transformada discreta de ondaleta não decimada. Seja $X = (x_0, x_1, \dots, x_{T-1})$ uma série temporal de tamanho $T = 2^M$. Para $j =$

$0, 1, \dots, M-1$ e $k = 0, 1, \dots, 2^{M-j} - 1$, define-se a transformada discreta de ondaleta com respeito a ψ por (3). Assim, seja \mathbf{d} o vetor que contém os $d_{j,k}$. Devido a ortogonalidade e linearidade de \mathbf{W} uma transformada discreta de ondaleta para o modelo apresentado em (4.1) é dado por

$$d_{X,lq}(j,k) = U_q(j,k) + Z_{lq}(j,k) + O_{q.c.}(1), \quad (4.2)$$

para $j = 0, 1, 2, \dots, M-1$ e $k = 0, 1, 2, \dots, 2^j - 1$, define a transformada discreta de ondaleta com respeito a ψ por (2.3), $U_q(j,k)$ são variáveis $N(0, \eta_{j,k\Omega_q})$ ($N(0, S_{j,k\Omega_q})$) independentes, $Z_{lq}(j,k)$ são variáveis $N(0, \eta_{j,k\epsilon})$ ($N(0, S_{j,k\epsilon})$) independentes e independente de $U_q(j,k)$. Ainda, $O_{a.s.}(1) \rightarrow 0$ quase certamente, quando $T \rightarrow \infty$.

O problema é detectar se existe diferença entre os sinais $\Omega_1(t), \Omega_2(t), \dots, \Omega_Q(t)$.

Para isso, $\Omega_q(t)$ é decomposto da seguinte forma

$$\Omega_q(t) = \mu(t) + \alpha_q(t),$$

$q = 1, \dots, Q$, em que

$\mu(t)$: sinal comum que tem espectro de ondaleta $\eta_\mu(j,k)$ ($S_\mu(j,k)$);

$\alpha_q(t)$: sinal particular para o grupo q , independente e com espectro de ondaleta $\eta_\Omega(S_\Omega)$.

Para o sinal ser comum, significa testar que $\alpha_q(t) = 0$, conforme

$$H_0 : \alpha_1(t) = \alpha_2(t) = \dots = \alpha_Q(t) = 0,$$

4.1.1 Processo estacionário

Assim, para o caso estacionário, tem-se

$$\eta_{\Omega_q}(j,k) = \eta_\mu(j,k) + \eta_\Omega(j,k)$$

e (4.2) pode ser reescrito da seguinte forma

$$d_{X,lq}(j,k) = U_{j,k} + V_q(j,k) + Z_{lq}(j,k) + O_{q.c.}(1) \quad (4.3)$$

na qual cada j e k , U , V e Z são variáveis independentes com distribuições $N(0, \eta_\mu(j,k))$, $N(0, \eta_\Omega(j,k))$ e $N(0, \eta_\epsilon(j,k))$, respectivamente. As comparações dos sinais dos grupos é realizada pelo teste

$$H_0 : \eta_\Omega(j,k) = 0, \quad (4.4)$$

para todo (j,k) .

Para fazer o teste considerou-se as seguintes somas de quadrados, chama-se atenção para o somatório em k e graus de liberdade em cada nível j :

$$SQH_0(j,k) = \sum_{q=1}^Q \sum_{l=1}^L \sum_{k=1}^{2^{M-j}} [\bar{d}_{X,q}(j,k) - \bar{d}_{X,..}(j,k)]^2 \quad (4.5)$$

e

$$SQRes(j,k) = \sum_{q=1}^Q \sum_{l=1}^L \sum_{k=1}^{2^{M-j}} [d_{X,lq}(j,k) - \bar{d}_{X,q}(j,k)]^2 \quad (4.6)$$

em que

$$\bar{d}_{X,q}(j,k) = L^{-1} \sum_{l=1}^L d_{X,lq}(j,k) \quad (4.7)$$

$$\bar{d}_{X,..}(j,k) = N^{-1} \sum_{q=1}^Q d_{X,lq}(j,k). \quad (4.8)$$

As somas de quadrados (4.5) e (4.6), devido ao modelo (4.3), convergem quase certamente para

$$[L\eta_\Omega(j,k) + \eta_\epsilon(j,k)] \chi_{(Q-1)(2^{M-j})}^2$$

e

$$\eta_\epsilon(j,k) \chi_{(N-Q)(2^{M-j})}^2$$

os quais são duas distribuições de χ^2 independentes.

O teste para (4.4) foi realizado, para cada j fixado, onde no caso estacionário (Transformada discreta de ondaleta decimada), utilizou-se a estatística

$$F_0(j,k) = \frac{SQH_0(j,k)/(Q-1)(2^{M-j})}{SQRes(j,k)/(N-Q)(2^{M-j})} \quad (4.9)$$

que, sob H_0 , tem uma distribuição $F((Q-1)(2^{M-j}), (N-Q)(2^{M-j}))$ para cada j , onde $j = 0, 1, \dots, M-1$ e $k = 1, \dots, 2^{M-j}$.

4.1.2 Processo não estacionário

Para o caso não estacionário, tem-se

$$S_{\Omega q}(j,k) = S_{\mu}(j,k) + S_{\Omega}(j,k)$$

e (4.2) pode ser reescrito da seguinte forma

$$d_{X,lq}(j,k) = U_{j,k} + V_q(j,k) + Z_{lq}(j,k) + O_{q.c.}(1) \quad (4.10)$$

na qual cada j e k , U , V e Z são variáveis independentes com distribuições $(N(0, S_{\mu}(j,k)))$, $(N(0, S_{\Omega}(j,k)))$ e $(N(0, S_{\epsilon}(j,k)))$, respectivamente. As comparações dos sinais dos grupos é realizada pelo teste

$$H_0 : S_{\Omega}(j,k) = 0, \quad (4.11)$$

para todo (j,k) .

Para fazer o teste considera-se as seguintes somas de quadrados:

$$SQH_0(j,k) = \sum_{q=1}^Q \sum_{l=1}^L \sum_{k=1}^{2^M} [\bar{d}_{X,q.}(j,k) - \bar{d}_{X,..}(j,k)]^2 \quad (4.12)$$

e

$$SQRes(j,k) = \sum_{q=1}^Q \sum_{l=1}^L \sum_{k=1}^{2^M} [d_{X,lq}(j,k) - \bar{d}_{X,q}(j,k)]^2 \quad (4.13)$$

em que $d_{X,q}(j,k)$ e $d_{X,..}(j,k)$ são dados em (4.7) e (4.8), respectivamente.

As somas de quadrados (4.12) e (4.13), devido ao modelo (4.3), convergem quase certamente para, respectivamente,

$$[LS_{\Omega}(j,k) + S_{\epsilon}(j,k)]\chi_{(Q-1)(2^M)}^2$$

e

$$S_{\epsilon}(j,k)\chi_{(N-Q)(2^M)}^2$$

sendo que as duas distribuições de χ^2 são independentes.

No caso não estacionário (Transformada discreta de ondaleta não decimada), o teste para (4.11) é realizado, para cada j fixado, utilizando a estatística

$$F_0(j,k) = \frac{SQH_0(j,k)/(Q-1)(2^M)}{SQRes(j,k)/(N-Q)(2^M)} \quad (4.14)$$

que, sob H_0 , tem uma distribuição $F((Q-1)(2^M), (N-Q)(2^M))$ para cada j , com $j = 0, 1, \dots, M-1$ e $k = 0, 1, \dots, 2^M - 1$.

O raciocínio apresentado, para a realização da análise de variância usando ondaleta, pode ser estendido para outros tipos de modelo.

4.2 Simulação

Seguindo a metodologia proposta, serão apresentados os gráficos das séries temporais (simuladas e reais). Foi aplicada a transformada discreta de ondaleta decimada (caso estacionário) e não-decimada (caso não estacionário), nas

bases de Haar e Daubechies, com a construção de alguns gráficos representando as decomposições e posteriormente a aplicação da análise de variância. O objetivo foi o de detectar se existe igualdade ou não entre os grupos.

4.2.1 Análise de variância usando ondaleta para séries temporais estacionárias

Nesta seção tem-se séries temporais estacionárias. Foram estudadas duas situações, cujos tratamentos foram gerados para que as séries apresentem somente variação ao acaso:

a) Modelo auto-regressivo de primeira ordem (AR(1))

Na Figura 7, apresenta-se o gráfico dos tratamentos e de suas respectivas repetições. Nota-se uma similaridade no comportamento dos tratamentos.

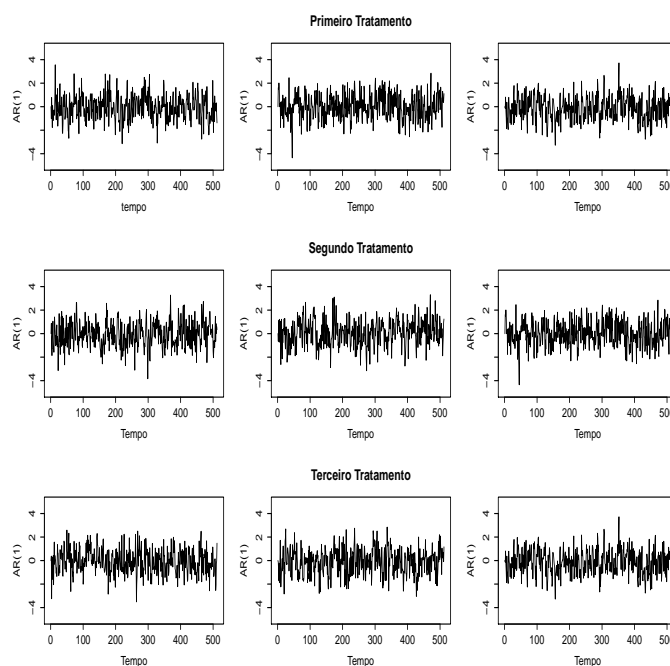


Figura 7 Gráfico das séries AR(1) geradas para os três tratamentos e suas três repetições estatisticamente iguais

As observações foram analisadas de acordo com o modelo apresentado em (4.1). A análise de variância foi conduzida utilizando a transformada discreta de ondaleta decimada, dada por (2.3), nas bases de Haar e Daubechies.

Para exemplificar, as Figuras 8 e 9, apresentam a decomposição relativa à primeira repetição do primeiro tratamento nas bases de Haar e Daubechies, respectivamente. Os d_j são os coeficientes de ondaletas, nos níveis $j = 1, 2, 3$ e 4. É importante observar que o nível $j = 1$ tem uma resolução mais fina.

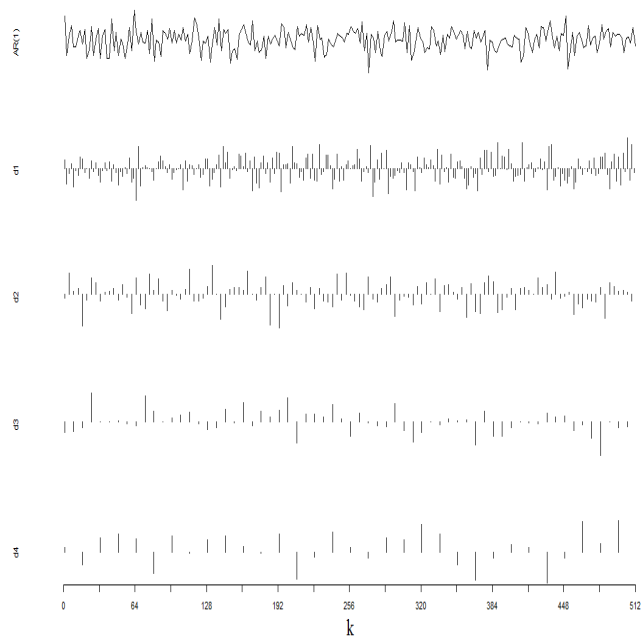


Figura 8 Gráfico da decomposição (Haar) para a primeira repetição do primeiro tratamento (ruído branco), considerando a transformada discreta de ondaleta decimada

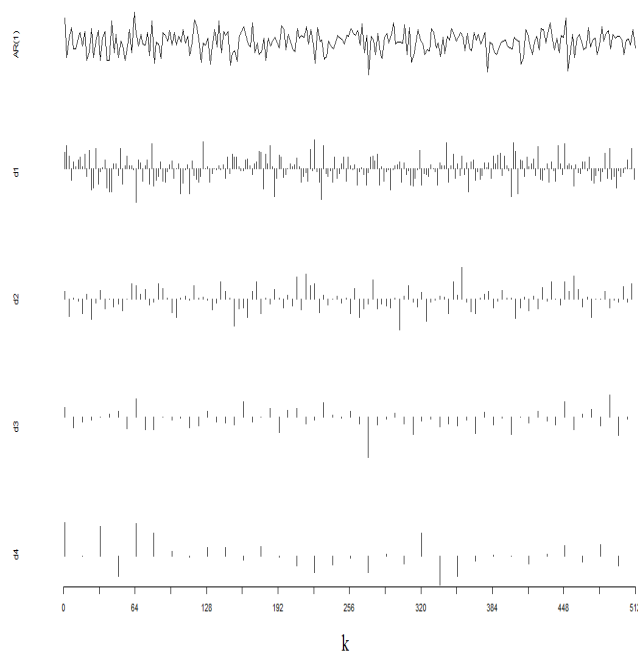


Figura 9 Gráfico da decomposição (Daubechies) para a primeira repetição do primeiro tratamento (ruído branco), considerando a transformada discreta de ondaleta decimada

A análise de variância utilizando as ondaletas de Haar e Daubechies foi realizada considerando-se para $j = 9$ níveis, pois $512 = 2^9$. Considerando $X_{lq}(t)$, onde $l = 1,2,3$, $q = 1,2,3$ e $t = 0, 1, \dots, 511$. Considerou-se 1.000, 5.000 e 10.000 simulações.

Considerando os valores-p do teste F para tratamentos dado em 4.9, tem-se que, tanto para a ondaleta de Haar, quanto para a ondaleta de Daubechies, o teste detectou em pelo menos um nível (j), valor-p maior que 0,05 em 100% das simulações propostas, ou seja, não rejeita-se H_0 , e conclui-se que as médias dos tratamentos não diferem entre si. Como as séries geradas eram AR(1), a análise de variância, usando ondaleta mostrou-se eficiente nesta situação.

Na Tabela 1, pode-se verificar a proporção de valores-p não significativos em cada nível j , considerando a ondaleta de Haar e Daubechies e 1.000, 5.000 e 10.000 simulações.

Tabela 1 Proporção de valores-p não significativos em cada nível (j), para séries AR(1), usando a transformada discreta de ondaleta decimada nas bases de Haar e Daubechies

Nível(j)	Haar			Daubechies		
	1.000	5.000	10.000	1.000	5.000	10.000
1	1	1	1	1	1	1
2	1	1	1	1	1	0,9999
3	1	0,9998	1	1	0,9998	0,9999
4	1	0,9990	0,9994	0,9980	0,9960	0,9991
5	1	0,9974	0,9972	0,9940	0,9964	0,9975
6	0,9940	0,9912	0,9938	0,9910	0,9928	0,9934
7	0,9880	0,9894	0,9908	0,9890	0,9866	0,9902
8	0,9860	0,9868	0,9861	0,9930	0,9858	0,9859
9	0,9880	0,9848	0,9862	0,9840	0,9860	0,9851

Observa-se, na Tabela 1, que a análise de variância, usando a transformada discreta de ondaleta decimada nas bases de Haar e Daubechies, apresenta comportamento semelhante no que diz respeito à não-significância nos níveis iniciais ($j = 1,2,3$). Dessa forma, quando as séries temporais se apresentam de forma semelhante (AR(1)), usando transformada discreta de ondaleta decimada, na base de Haar ou a Daubechies, tem-se que a análise de variância apresenta-se de forma coerente, não rejeitando, portanto, a hipótese de nulidade H_0 dada em (4.4).

Observa-se também que o número de simulações pouco afetou as estimativas dos valores-p; aproximando-se de 0,98 com o aumento de simulações.

Na Tabela 2, são apresentadas as taxas de erro tipo I d teste F dado por (4.9), em função do número de simulações e o nível de decomposição j , sob H_0 . Observa-se que para os níveis (j) iniciais o teste apresentou uma taxa de erro tipo I de 0%. Quando a decomposição da transformada de ondaleta apresenta níveis com menos informação ($j = 9$), tem-se uma taxa de erro de aproximadamente 1%.

Tabela 2 Taxa de erro tipo I em cada nível (j), para séries AR(1), usando a transformada discreta de ondaleta decimada nas bases de Haar e Daubechies

Nível(j)	Haar			Daubechies		
	1.000	5.000	10.000	1.000	5.000	10.000
1	0	0	0	0	0	0
2	0	0	0	0	0	0,0001
3	0	0,0002	0	0	0,0002	0,0001
4	0	0,001	0,0006	0,002	0,004	0,0009
5	0	0,0026	0,0028	0,006	0,0036	0,0025
6	0,006	0,0088	0,0062	0,009	0,0072	0,0066
7	0,012	0,0106	0,0092	0,011	0,0134	0,0098
8	0,014	0,0132	0,0139	0,007	0,0142	0,0141
9	0,012	0,0152	0,0138	0,016	0,014	0,0149

b) Série temporal categórica com comportamentos iguais

Para as séries categóricas, foram geradas 24 séries estacionárias que assumem os valores 1, 2, 3, 4, 5 ou 6, onde foram considerado dois tratamentos, cada um com 12 repetições de tamanho 128. Na Figura 10, tem-se o gráfico da primeira repetição de cada tratamento, onde pode-se observar uma similaridade entre os tratamentos (levando em consideração a primeira repetição de cada tratamento).

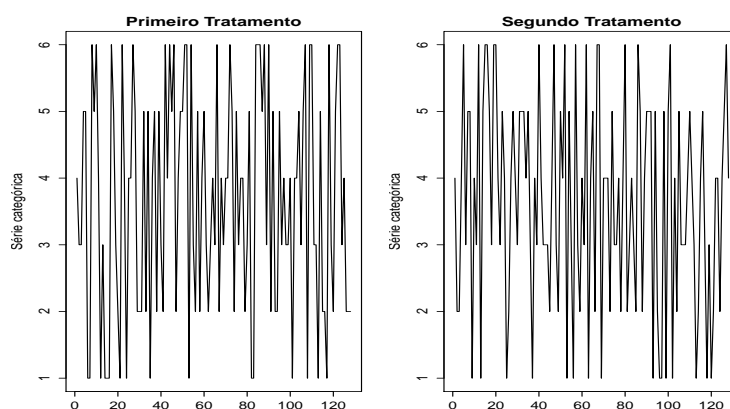


Figura 10 Gráfico da primeira repetição dos dois tratamentos, considerando as séries categóricas simuladas

As observações foram analisadas de acordo com o modelo apresentado em (4.1), considerando $X_{lq}(t)$, onde $l = 1, 2$, $q = 1, 2, \dots, 12$ e $t = 0, 1, \dots, 127$. Para a análise considerou-se 128 observações, tal que tem-se $128 = 2^7$. A análise de variância será considerada para $j = 7$ níveis, pois $128 = 2^7$. Considerou-se 1.000, 5.000 e 10.000 simulações.

Tem-se nas Figuras 11 e 12 as decomposições relativas à primeira repetição do primeiro tratamento, nas bases de Haar e Daubechies, respectivamente. Os d_j são os coeficientes de ondaletas, nos níveis $j = 1, 4, 6$ e 7 . Observa-se que o nível d_7 apresenta-se com menos informações que o nível d_1 .

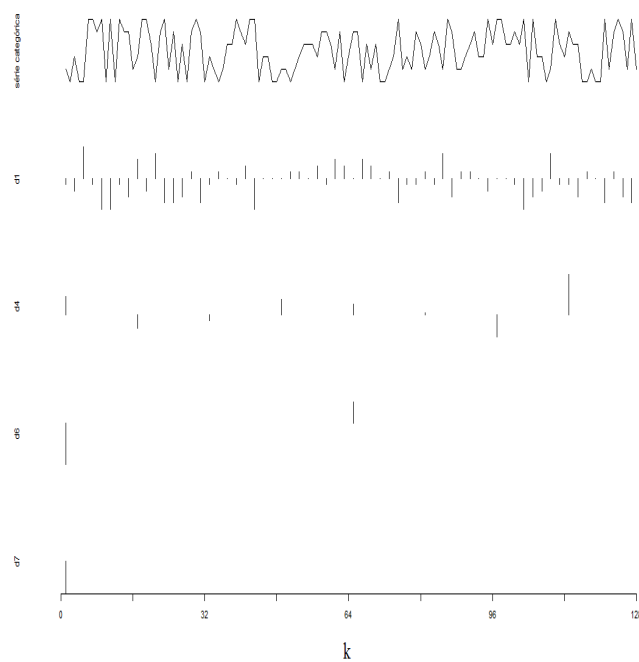


Figura 11 Gráfico da decomposição (Haar) para a primeira repetição do primeiro tratamento considerando a série categórica simulada, usando a transformada discreta de ondaleta decimada

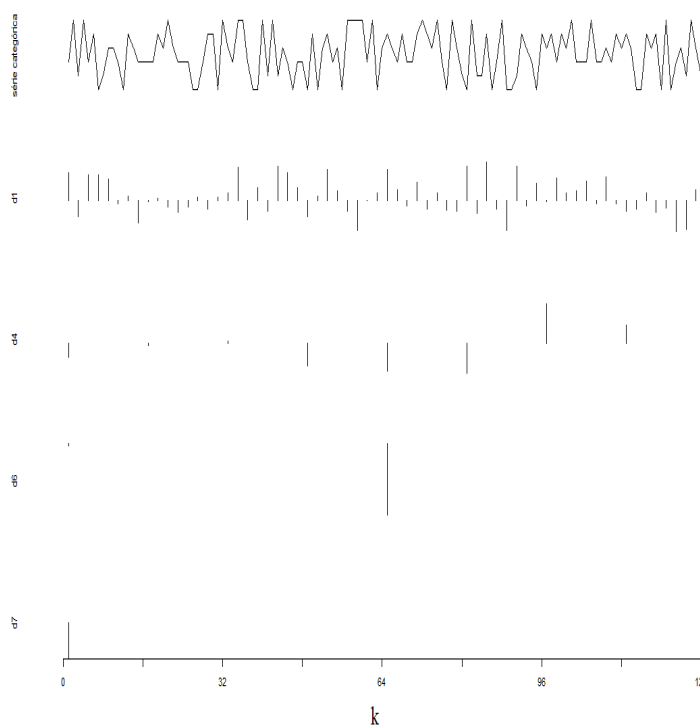


Figura 12 Gráfico da decomposição (Daubechies) para a primeira repetição do primeiro tratamento considerando a série categórica simulada, usando a transformada discreta de ondaleta decimada

Observa-se, por meio das Figuras 11 e 12, tanto para a ondaleta de Haar, quanto para a ondaleta de Daubechies, que os níveis menores de j apresentam mais informações que os níveis maiores de j .

De acordo com os valores-p do teste F para tratamentos dado em (4.9), tem-se que para a ondaleta de Haar e Daubechies, em 100% das vezes, o teste detectou que, em pelo menos um nível (j) para todas as simulações propostas, o valor-p foi maior que 0,05. Dessa forma, não rejeita-se a hipótese nula, o que comprova a eficiência do teste já que as séries foram geradas para que entre elas houvessem similaridade, ou seja, efeito igual a zero entre tratamentos.

Na Tabela 3, pode-se verificar a proporção de valores-p não significativos em cada nível j , considerando a ondaleta de Haar e Daubechies e 1.000, 5.000 e

10.000 simulações.

Tabela 3 Proporção de valores-p não significativos em cada nível (j), para séries AR(1), usando a transformada discreta de ondaleta decimada nas bases de Haar e Daubechies

Nível(j)	Haar			Daubechies		
	1.000	5.000	10.000	1.000	5.000	10.000
1	0,9480	0,9492	0,9520	0,9460	0,9530	0,9484
2	0,9500	0,9496	0,9485	0,9530	0,9472	0,9491
3	0,9410	0,9564	0,9504	0,9540	0,9502	0,9469
4	0,9640	0,9492	0,9508	0,9480	0,9542	0,9503
5	0,9510	0,9504	0,9496	0,9430	0,9442	0,9516
6	0,9560	0,9496	0,9520	0,9500	0,9458	0,9490
7	0,9500	0,9502	0,9459	0,9480	0,9518	0,9519

Observa-se, por meio da Tabela 3, que as ondaletas de Haar e Daubechies apresentaram comportamento semelhante no que diz respeito à não-significância em cada nível (j). A proporção de não rejeição de H_0 ficou próxima de 0,95 em cada nível e, também, quando se aumentou o número de simulações. Dessa forma, quando as séries temporais são categóricas, usando a ondaleta de Haar ou a Daubechies, tem-se que a análise de variância apresenta-se eficiente, não rejeitando, portanto, a hipótese de nulidade H_0 dada em (4.4). Na Tabela 4, observa-se que o teste aplicado está controlando bem a taxa de erro tipo I.

Tabela 4 Taxa de erro tipo I em cada nível (j), para séries AR(1), usando a transformada discreta de ondaleta decimada nas bases de Haar e Daubechies

Nível(j)	Haar			Daubechies		
	1.000	5.000	10.000	1.000	5.000	10.000
1	0,0520	0,0508	0,0480	0,0540	0,0470	0,0516
2	0,0500	0,0504	0,0515	0,0470	0,0528	0,0509
3	0,0590	0,0496	0,0496	0,0410	0,0498	0,0531
4	0,0360	0,0492	0,0492	0,0520	0,0458	0,0497
5	0,0490	0,0504	0,0504	0,057	0,0558	0,0484
6	0,0440	0,0480	0,0480	0,0500	0,0542	0,0510
7	0,0500	0,0541	0,0541	0,0520	0,0482	0,0481

4.2.2 Análise de variância usando ondaleta para séries temporais não estacionárias

Nesta seção tem-se séries temporais não estacionárias. Foram estudadas duas situações, cujos tratamentos foram gerados para que as séries apresentassem comportamentos diferentes:

a) Série temporal com tendência e AR(1)

Nesta seção será discutido o caso onde os tratamentos são diferentes.

Na Figura 13, tem-se os gráficos dos tratamentos e de suas respectivas repetições. Nota-se, por meio da Figura 13 uma similaridade entre as repetições e uma diferença entre os tratamentos.

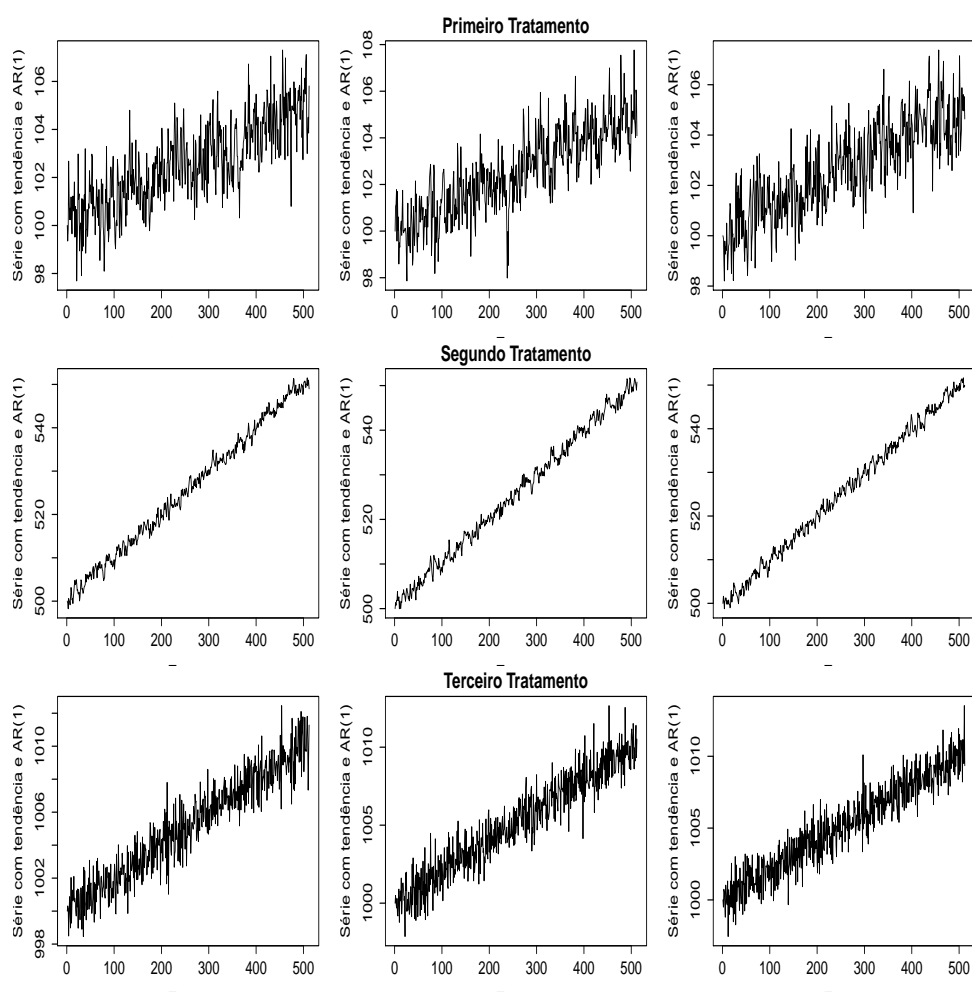


Figura 13 Gráfico das séries com tendência e Ar(1) simuladas para os três tratamentos

Para cada série $X(t)$ composta por 512 observações (2^9), foi aplicada a transformada discreta de ondaleta não-decimada usando a base de Haar e Daubechies. Para exemplificar, as Figuras 14, 15 e 16 apresentam, respectivamente, a decomposição relativa à primeira repetição dos três tratamentos na base de Haar, onde os d_j são os coeficientes de ondaletas, nos níveis $j = 2, 5, 7$ e 8.

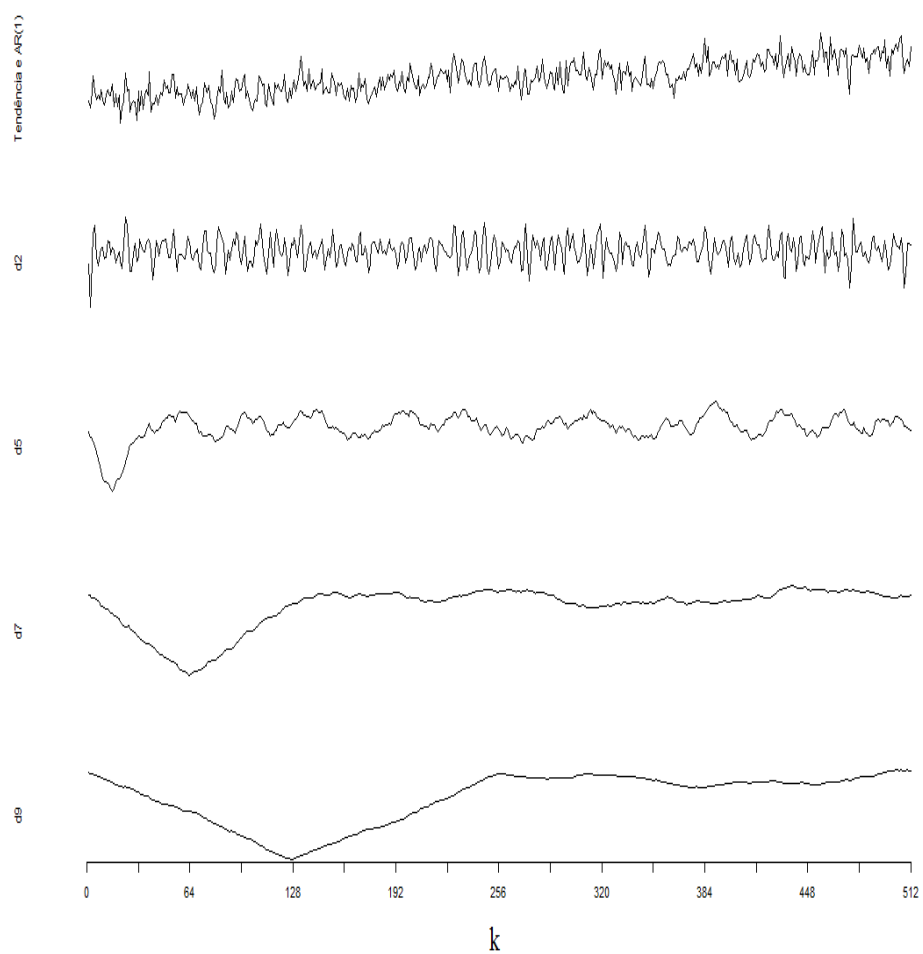


Figura 14 Gráfico da decomposição (Haar) para a primeira repetição do primeiro tratamento (tendência e AR(1)), considerando a transformada discreta de ondaleta não-decimada

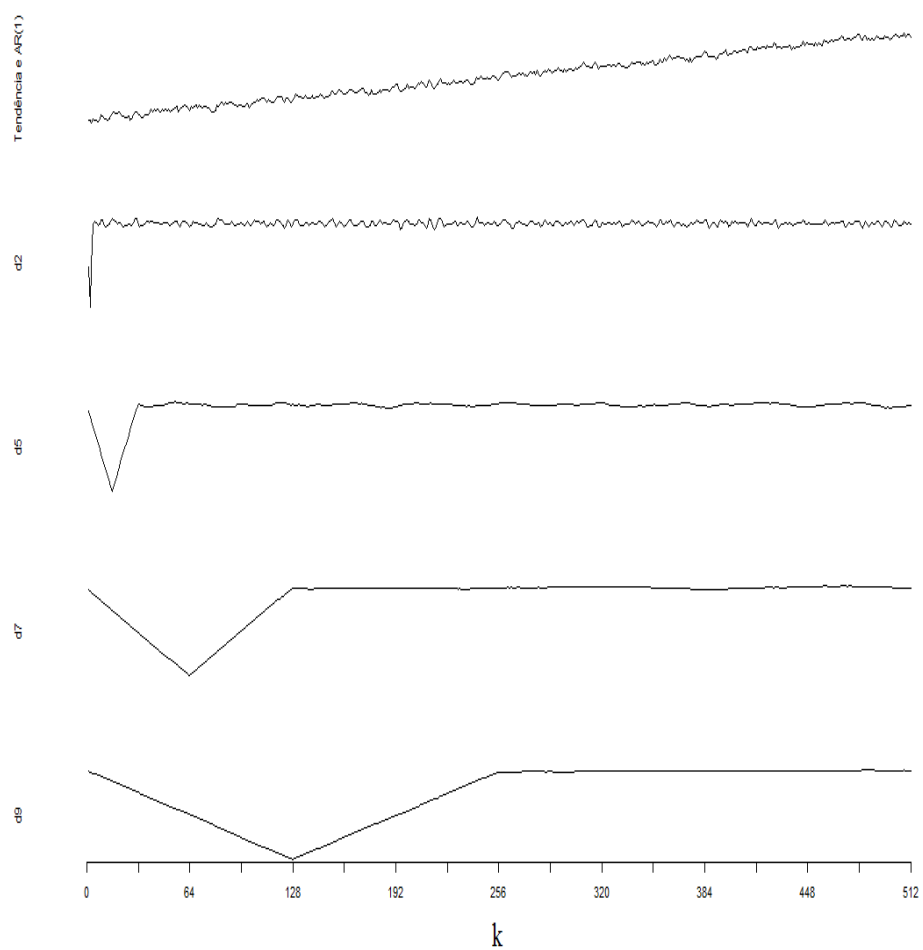


Figura 15 Gráfico da decomposição (Haar) para a primeira repetição do segundo tratamento (tendência e AR(1)), considerando a transformada discreta de ondaleta não-decimada

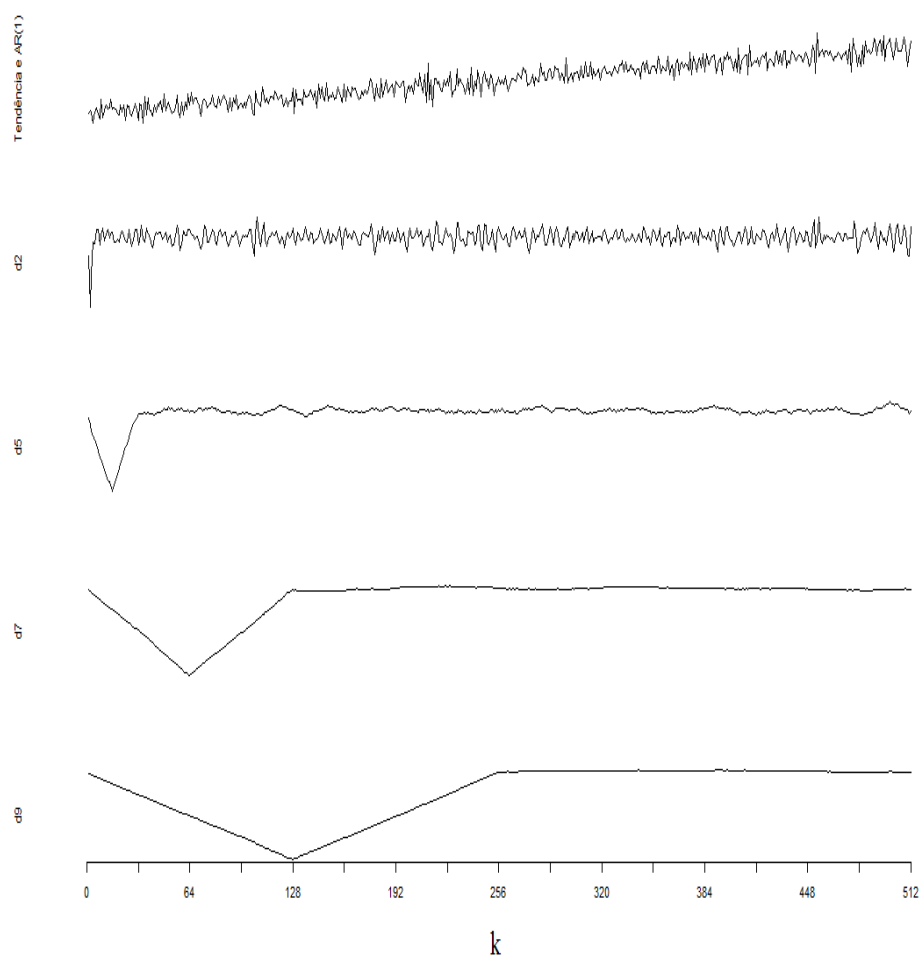


Figura 16 Gráfico da decomposição (Haar) para a primeira repetição do terceiro tratamento (tendência e AR(1)), considerando a transformada discreta de ondaleta não-decimada

As Figuras 17, 18 e 19, apresentam a decomposição relativa à primeira repetição de cada tratamento na base de Daubechies.

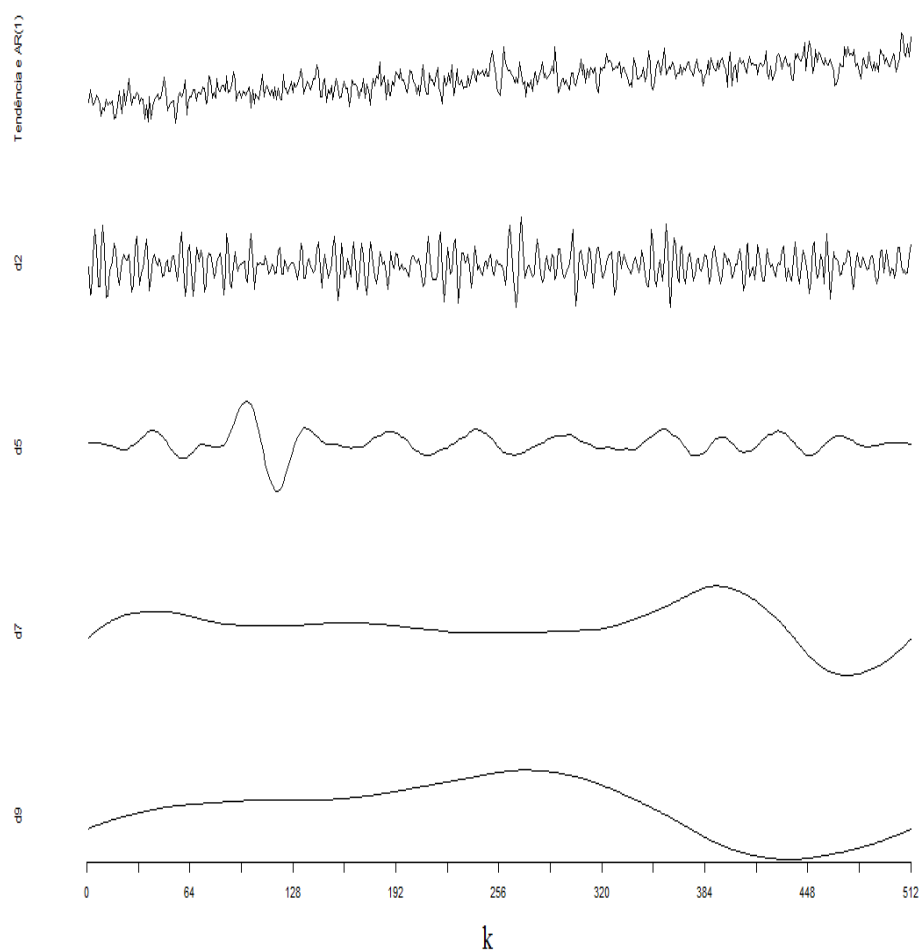


Figura 17 Gráfico da decomposição (Daubechies) para a primeira repetição do primeiro tratamento (tendência e AR(1)), considerando a transformada discreta de ondaleta não-decimada

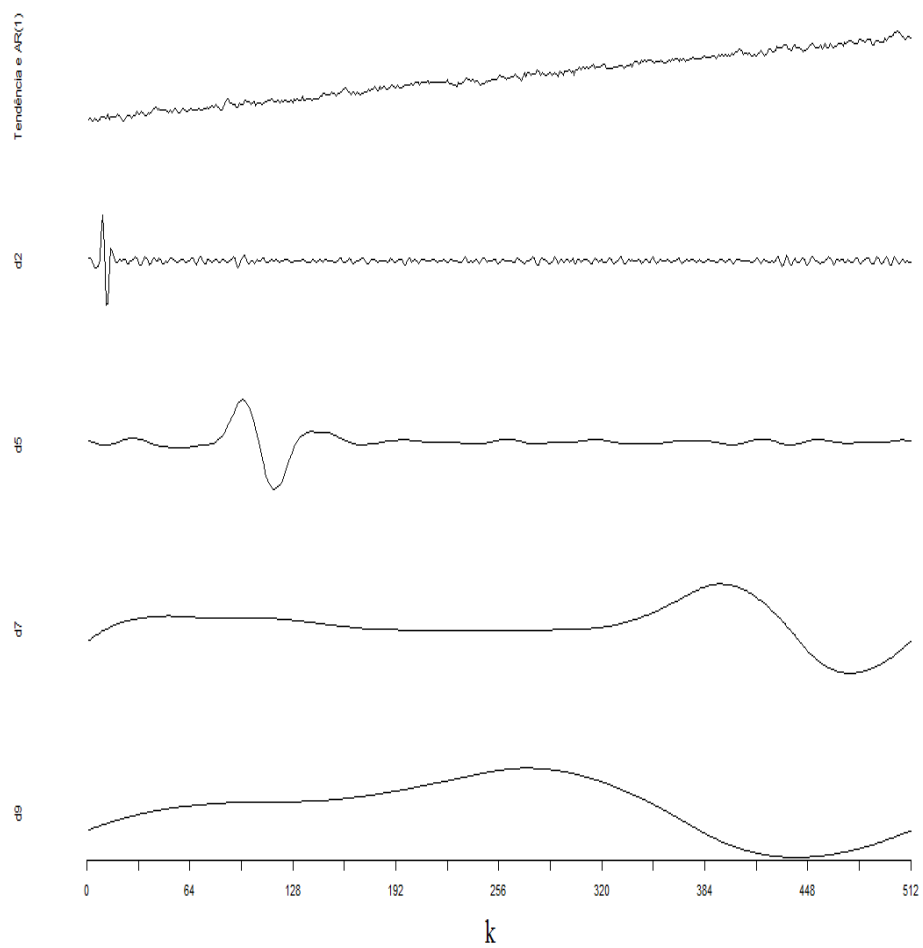


Figura 18 Gráfico da decomposição (Daubechies) para a primeira repetição do segundo tratamento (tendência e AR(1)), considerando a transformada discreta de ondaleta decimada

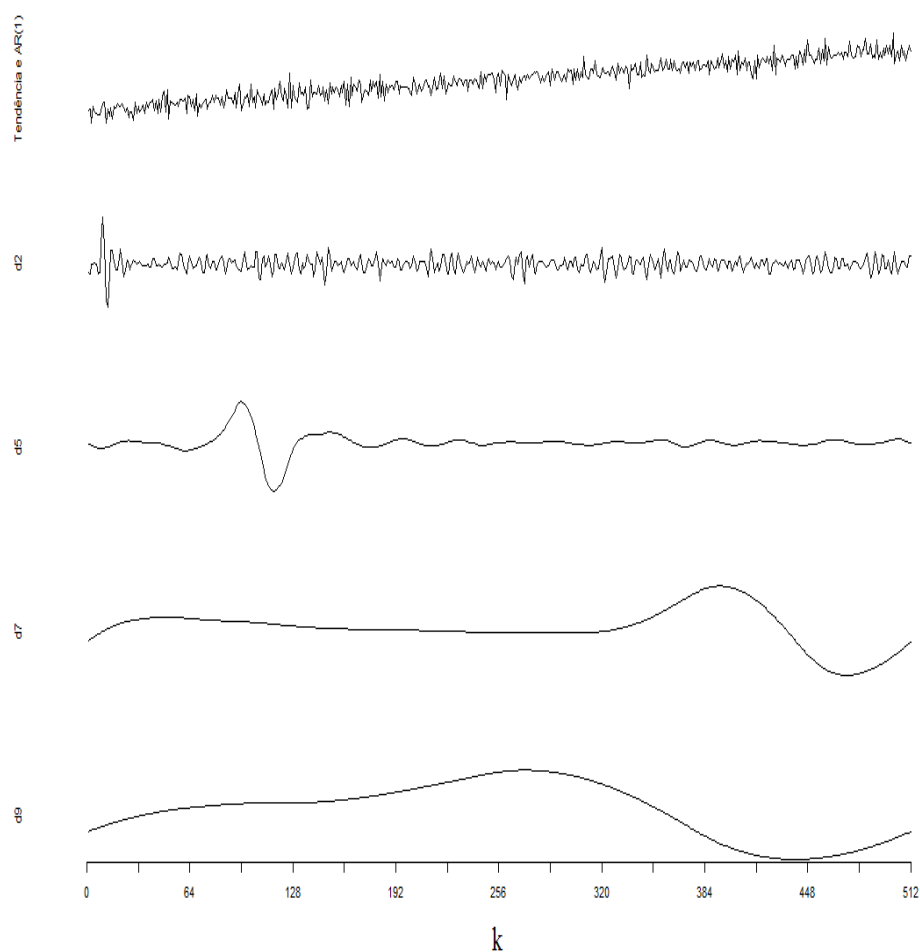


Figura 19 Gráfico da decomposição (Daubechies) para a primeira repetição do terceiro tratamento (tendência e AR(1)), considerando a transformada discreta de ondaleta decimada

A análise de variância utilizando as ondaletas de Haar e Daubechies foi considerada para $j = 9$ níveis, pois $512 = 2^9$. Considerando $X_{lq}(t)$, em que $l = 1, 2, 3$, $q = 1, 2, 3$ e $t = 0, 1, \dots, 511$. Para a análise considerou-se 512 observações, tal que tem-se $512 = 2^9$.

De acordo com o teste F dado em (4.14), tem-se que para a ondaleta de Haar e Daubechies, em 100% das vezes, o teste detectou que em pelo menos

um nível para todas as simulações propostas, o valor-p foi menor que 0,05. Na Tabela 5, pode-se verificar a proporção de valores-p significativo em cada nível j .

Tabela 5 Proporção de valores-p significativo em cada nível (j), para séries de tendência e AR(1), usando a transformada discreta de ondaleta não decimada nas bases de Haar e Daubechies

Nível(j)	Haar			Daubechies		
	1.000	5.000	10.000	1.000	5.000	10.000
1	1	1	1	1	1	1
2	1	1	1	1	1	1
3	1	1	1	1	1	1
4	1	1	1	1	1	1
5	1	1	1	1	1	1
6	1	1	1	1	1	1
7	1	1	1	1	1	1
8	1	1	1	1	1	1
9	1	1	1	1	1	1

Assim, o poder do teste, que corresponde à capacidade do teste de detectar diferenças é de 100%. Dessa forma, rejeita-se a hipótese nula, o que comprova a eficiência do teste, já que as séries foram geradas para que os tratamentos apresentassem diferenças.

b) Séries categóricas simuladas com comportamentos diferentes

Nesta seção será discutido o caso onde os tratamentos são diferentes, representados por séries categóricas diferentes não estacionárias de tamanho 128 (2^7), assumindo os valores de 1 a 6.

Na Figura 20, tem-se a representação gráfica da série simulada da primeira repetição de cada um dos tratamentos. Nota-se uma diferença entre os tratamentos, considerando a primeira repetição.

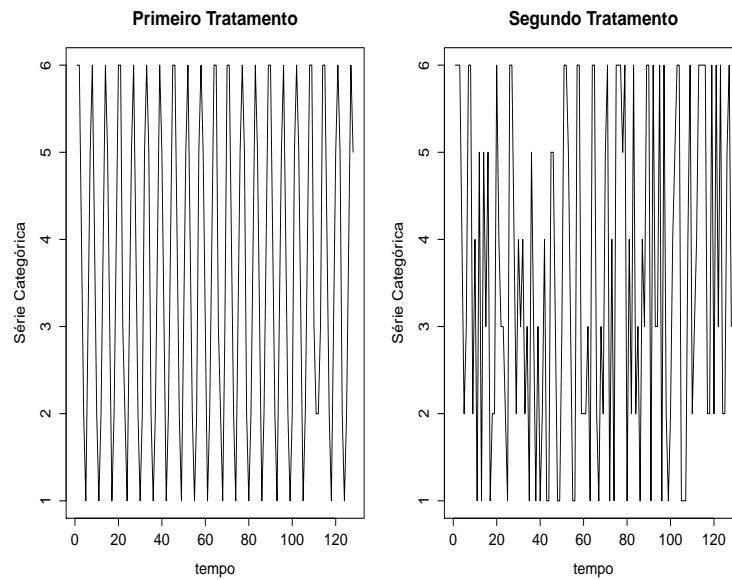


Figura 20 Gráfico da primeira repetição dos dois tratamentos, considerando séries categóricas diferentes

As observações foram analisadas de acordo com o modelo apresentado em (4.1). A análise de variância será conduzida utilizando a transformada discreta de ondaleta não-decimada, dada por (2.3), nas bases de Haar e Daubechies.

Para exemplificar a decomposição tem-se que as Figuras 21 e 22 apresentam a decomposição relativa à primeira repetição de cada tratamento na base de Haar, onde os d_j são os coeficientes de ondaletas, nos níveis $j = 1, 2, 6$ e 7 .

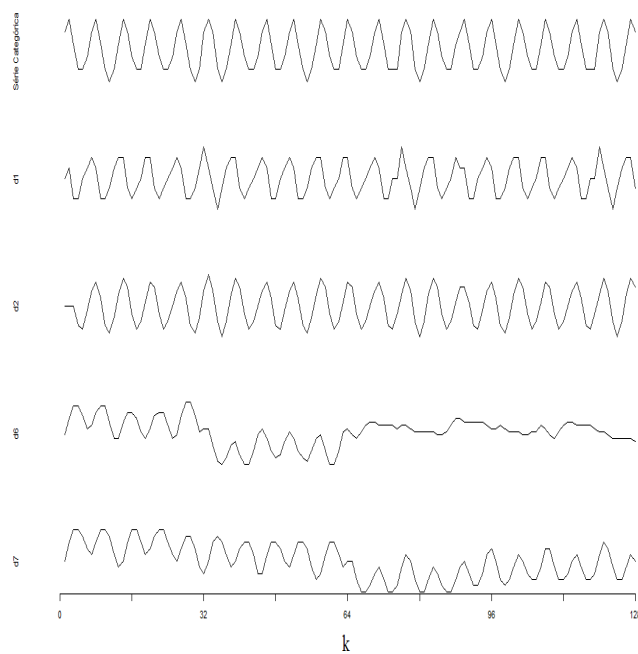


Figura 21 Gráfico da decomposição (Haar) para a primeira repetição do primeiro tratamento (série categórica simulada), considerando a transformada discreta de ondaleta não-decimada

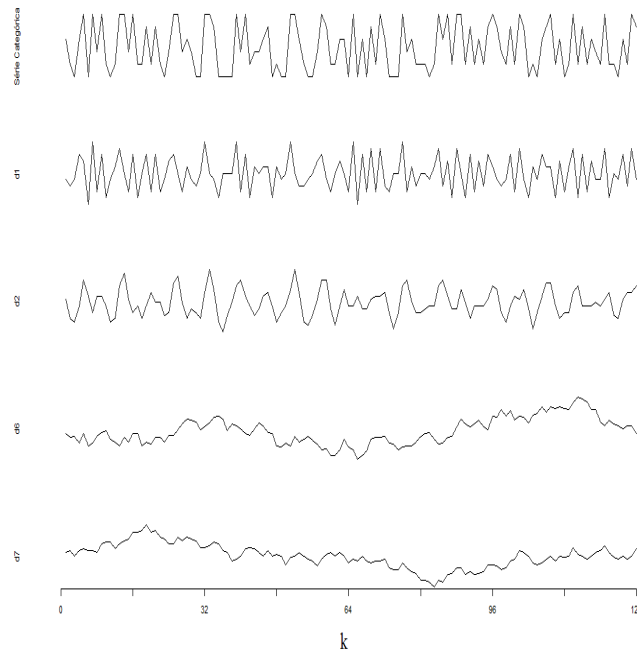


Figura 22 Gráfico da decomposição (Haar) para a primeira repetição do segundo tratamento (série categórica simulada), considerando a transformada discreta de ondaleta não-decimada

As Figuras 23 e 24 apresentam a decomposição relativa à primeira repetição de cada tratamento na base de Daubechies.

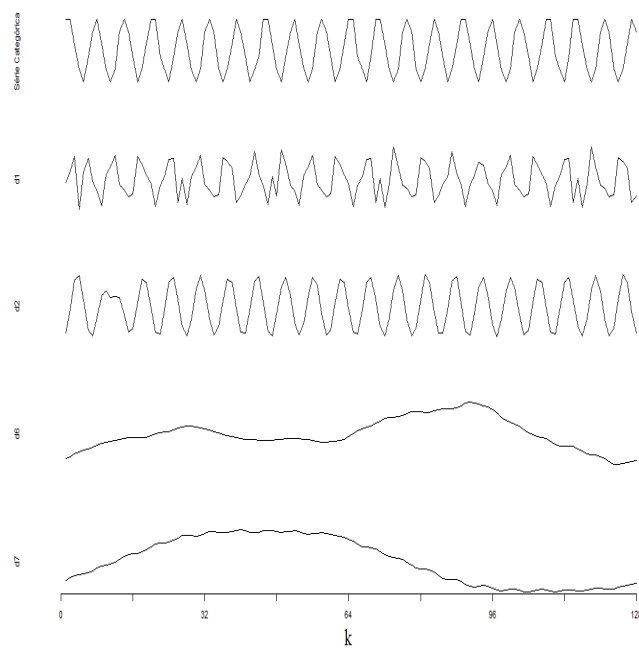


Figura 23 Gráfico da decomposição (Daubechies) para a primeira repetição do primeiro tratamento (série categórica), considerando a transformada discreta de ondaleta não-decimada

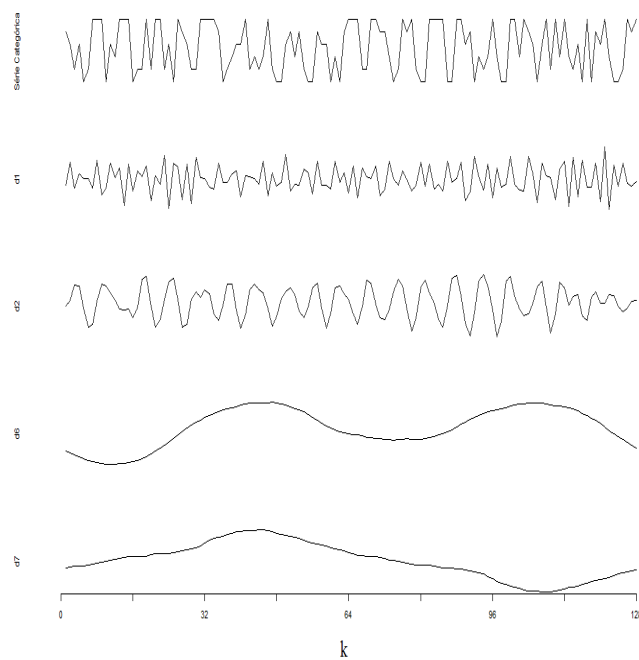


Figura 24 Gráfico da decomposição (Daubechies) para a primeira repetição do segundo tratamento (série categórica), considerando a transformada discreta de ondaleta não-decimada

Observa-se, por meio das Figuras 21, 22, 23 e 24, que os coeficientes de ondaletas apresentam maiores informações nos níveis (j) iniciais.

De acordo com o teste F dado em (4.14), a Tabela 6 apresenta a proporção de valores-p menores que 0,05 em pelo menos um nível para todas as simulações propostas.

Tabela 6 Poder do teste F usando a transformada discreta de ondaleta não-decimada, para séries categórica simuladas, usando a ondaleta de Haar e Daubechies

Número de Simulações	Haar	Daubechies
	Poder do teste	Poder do teste
1.000	0,7180	0,8900
5.000	0,7318	0,8878
10.000	0,7368	0,8895

Na Tabela 6 verifica-se que a ondaleta de Haar apresenta proporção de valores-p significativos mais baixos que a ondaleta de Daubechies, em pelo menos um nível para todas as simulações, indicando um menor poder do teste quando se trabalha com a ondaleta de Haar.

Na Tabela 7, pode-se verificar a proporção de valores-p significativos em cada nível j , considerando a transformada discreta de ondaleta não-decimada nas bases de Haar e Daubechies

Tabela 7 Proporção de valor-p significativo em cada nível (j), para séries categóricas estatisticamente diferentes, usando a transformada discreta de ondaleta não-decimada nas bases de Haar e Daubechies

Nível	Haar			Daubechies		
	1.000	5.000	10.000	1.000	5.000	10.000
1	0,0300	0,0060	0,0060	0,0390	0,0480	0,0461
2	0	0,0004	0	0	0,0002	0,0006
3	0,0400	0,0028	0,0029	0,0430	0,0418	0,0396
4	0,2790	0,2762	0,2865	0,3580	0,3632	0,3672
5	0,5300	0,5490	0,5482	0,7400	0,7454	0,7533
6	0,3730	0,3518	0,3683	0,3070	0,3292	0,3295
7	0,4040	0,4066	0,4131	0,3390	0,3334	0,3348

A Tabela 7, apresenta que nos níveis menores, a proporção de valores-p significativos é menor que em níveis maiores, tanto para a ondaleta de Haar, quanto para a ondaleta de Daubechies. Além disso, percebe-se que, no nível $j = 5$, tem-se uma maior proporção de valores-p significativos, tanto na Haar quanto na Daubechies.

4.2.3 Análise de variância usando ondaleta com dados reais

a) Série do número diário de óbitos por causas respiratórias no período de 01 de janeiro de 1994 a 31 de dezembro de 1997 em São Paulo - SP

Foram considerados quatro grupos (tratamentos) referentes ao número diário de óbitos por causas respiratórias, a primeira repetição de cada tratamento está representada na Figura 25, e percebe-se que o número diário de óbitos por causas respiratórias é maior na faixa etária de indivíduos maiores que 65 anos de idade.

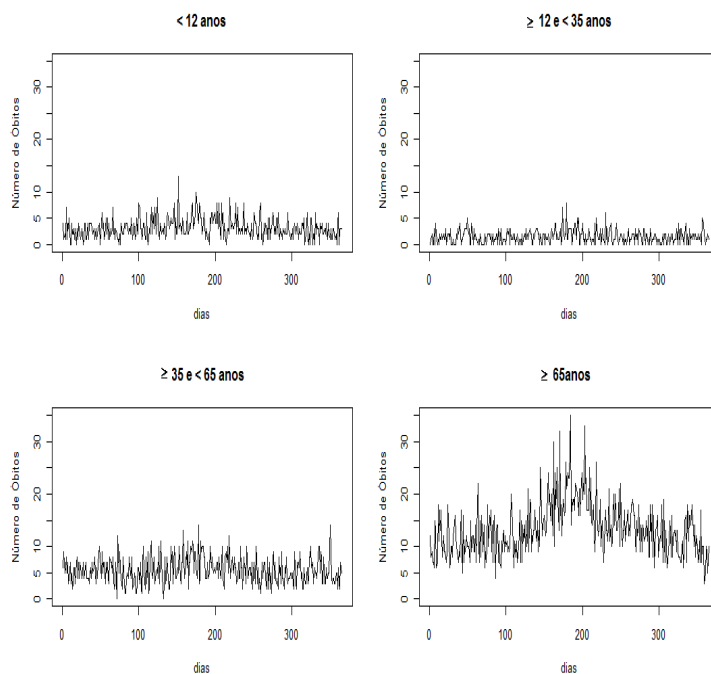


Figura 25 Gráfico da primeira repetição das faixas etária relativa ao número diário de óbitos por causas respiratória

Nota-se, por meio da Figura 25, que as séries não são estacionárias e que existe uma possível diferença no número de óbitos por causas cardíacas entre as faixas etárias. A Figura 26, caracteriza os correlogramas ¶ da primeira repetição de cada tratamento, percebe-se a não estacionariedade. O objetivo nesta situação foi verificar se existem diferenças no número de óbitos por causas respiratórias entre as faixas etárias.

¶Segundo Morettin e Tolo (2006), correlograma é um gráfico com os k primeiros coeficientes de autocorrelação e pode ser uma ferramenta poderosa para identificar características da série temporal.

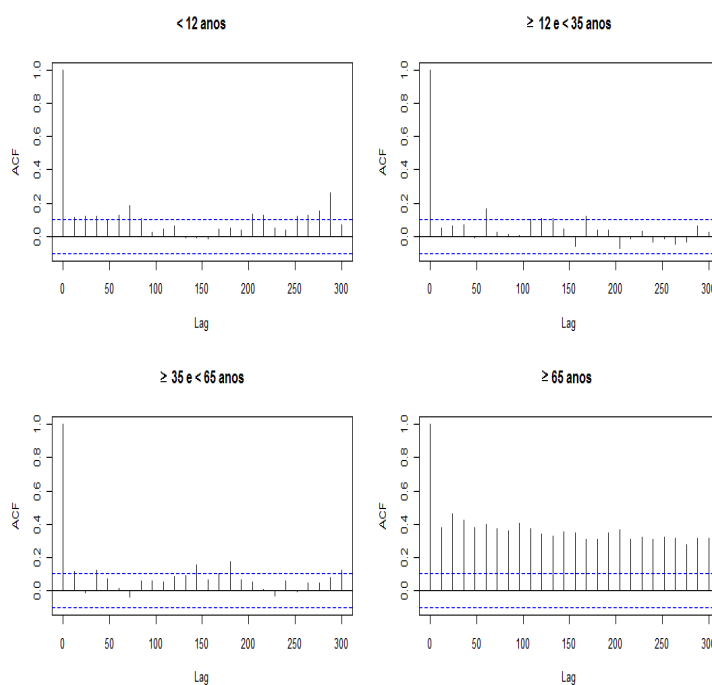


Figura 26 Gráfico do correlograma primeira repetição das faixas etária relativa ao número diário de óbitos por causas respiratória

Como as séries são não estacionárias, para cada série $X(t)$ foi aplicada a transformada discreta de ondaleta não-decimada usando a base de Haar e Daubechies considerando $j = 8$, pois $256 = 2^8$. As Figuras 27 e 28 apresentam a decomposição relativa à primeira repetição dos dois primeiros tratamentos (faixa etária < 12 anos e faixa etária ≥ 12 e < 35 anos) apresentados na base de Haar, onde os d_j são os coeficientes de ondaletas, nos níveis $j = 1, 2, 3$ e 4.

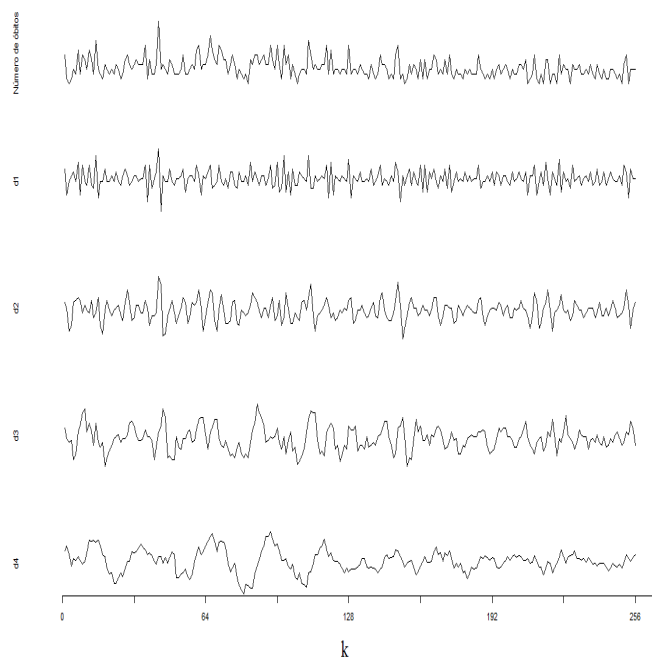


Figura 27 Gráfico da decomposição (Haar) para o número de óbitos por causas respiratórias para o primeiro tratamento (faixa etária < 12 anos), considerando a transformada discreta de ondaleta não-decimada

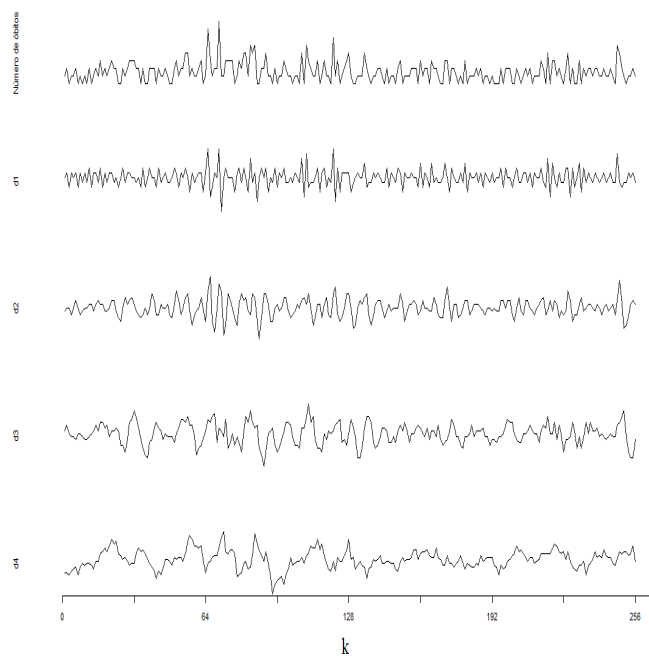


Figura 28 Gráfico da decomposição (Haar) para o número de óbitos por causas respiratórias para o segundo tratamento (faixa etária ≥ 12 e < 35 anos), considerando a transformada discreta de ondaleta não-decimada

As Figuras 29 e 30 apresentam a decomposição relativa à primeira repetição dos dois primeiros tratamentos na base de Daubechies.

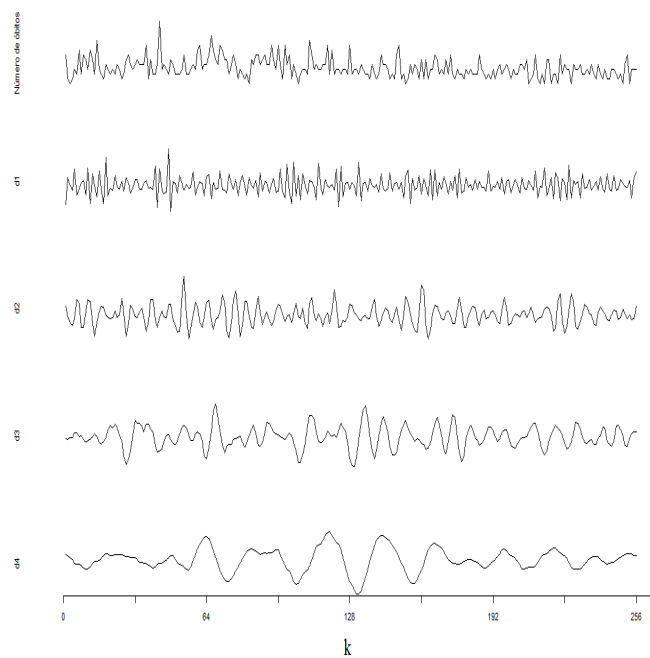


Figura 29 Gráfico da decomposição (Daubechies) para o número de óbitos por causas respiratórias para o primeiro tratamento (faixa etária < 12 anos), considerando a transformada discreta de ondaleta decimada

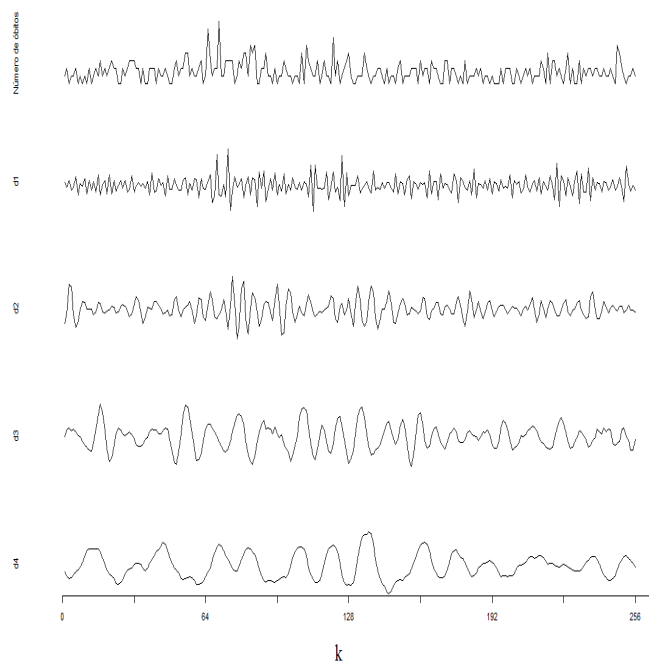


Figura 30 Gráfico da decomposição (Daubechies) para o número de óbitos por causas respiratórias para o segundo tratamento (faixa etária ≤ 12 e > 35 anos), considerando a transformada discreta de ondaleta decimada

Considerando $X_{lq}(t)$, onde $l = 1,2,3,4$, $q = 1,2,3,4$ e $t = 0, 1, \dots, 255$. Para a análise considerou-se 256 observações, tal que tem-se $256 = 2^8$.

As Figuras 31 e 32 apresentam o valor-p de cada nível j para a ondaleta de Haar e Daubechies, respectivamente.

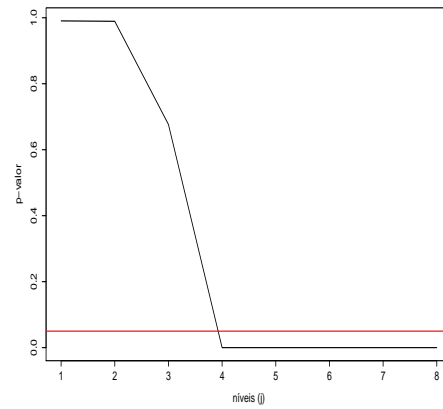


Figura 31 Gráfico do valor-p para o número de óbitos por causas respiratórias, considerando a transformada discreta de ondaleta não-decimada na base de Haar. Linha horizontal representa o valor de 0,05 (5% de significância)

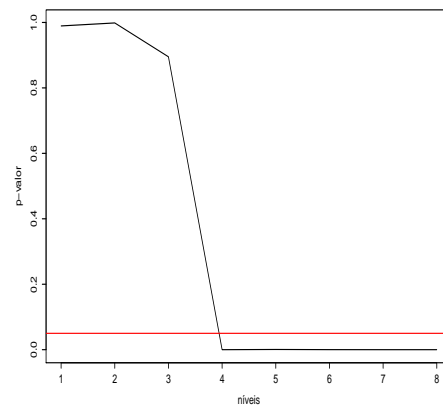


Figura 32 Gráfico do valor-p para o número de óbitos por causas respiratórias, considerando a transformada discreta de ondaleta não-decimada na base de Daubechies. Linha horizontal representa o valor de 0,05 (5% de significância)

Nota-se, nas das Figuras 31 e 32, que para a ondaleta de Haar e Daubechies, nos níveis $j = 4, 5, 6, 7$ e 8 , tem-se valores-p menores que $0,05$. Dessa forma, foram encontradas evidências de diferenças significativas, ao nível de 5% de probabilidade, entre o número de óbitos por causas respiratórias entre as faixas etárias. Rejeita-se, portanto, a hipótese nula, tanto para a ondaleta de Haar, quanto para a ondaleta de Daubechies.

b) Série do número diário de óbitos por causas cardíacas o período de 01 de janeiro de 1994 a 31 de dezembro de 1997 em São Paulo - SP

Foram considerados quatro grupos (tratamentos) referentes ao número diário de óbitos por causas cardíacas, a primeira repetição de cada tratamentos está representada na Figura 33.

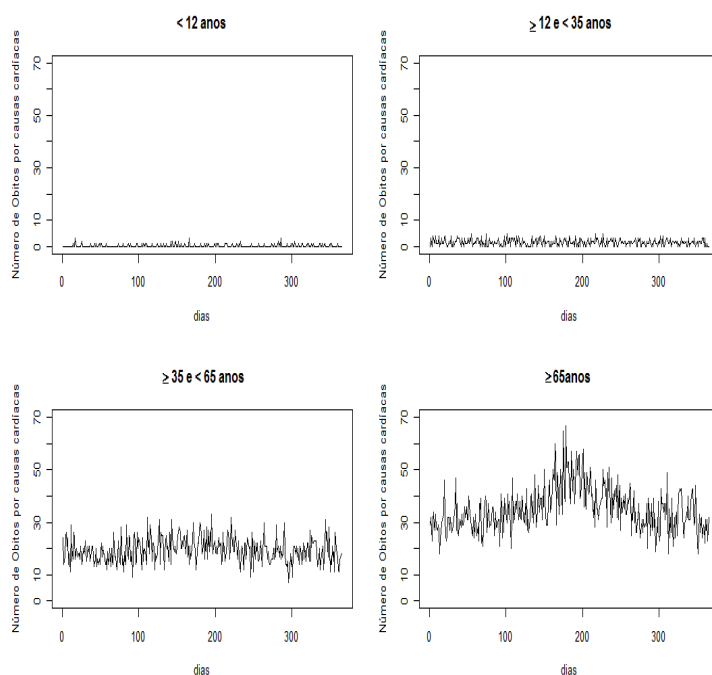


Figura 33 Gráfico da primeira repetição do no número de óbito por causas cardíacas entre as faixas etárias

Nota-se uma possível não estacionariedade e uma diferença entre o número diário de óbitos por causas cardíacas entre as faixas etárias. Percebe-se também, que o número de óbitos por causa cardíaca é maior na faixa etária > 65 anos. A Figura 34, apresenta o correlograma da primeira repetição de cada tratamento, observando uma não estacionariedade. O objetivo foi verificar se existem diferenças no número de óbito por causas cardíacas entre as faixas etárias.

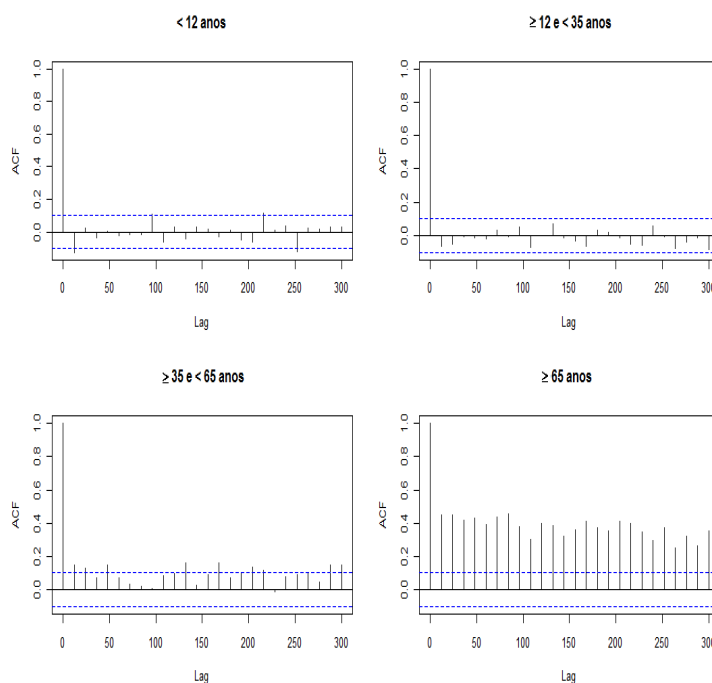


Figura 34 Gráfico do correlograma primeira repetição das faixas etárias relativa ao número diário de óbitos por causas cardíacas

Os coeficientes da decomposição de ondaletas serão analisados de acordo com o modelo apresentado em (4.1). A análise de variância foi conduzida utilizando a transformada discreta de ondaleta não-decimada, dada por (2.3), nas bases de Haar e Daubechies.

As Figuras 35 e 36 apresentam a decomposição relativa à primeira repetição de dois dos tratamentos apresentados na base de Haar, onde os d_j são

os coeficientes de ondaletas, nos níveis $j = 1, 2, 3$ e 4 .

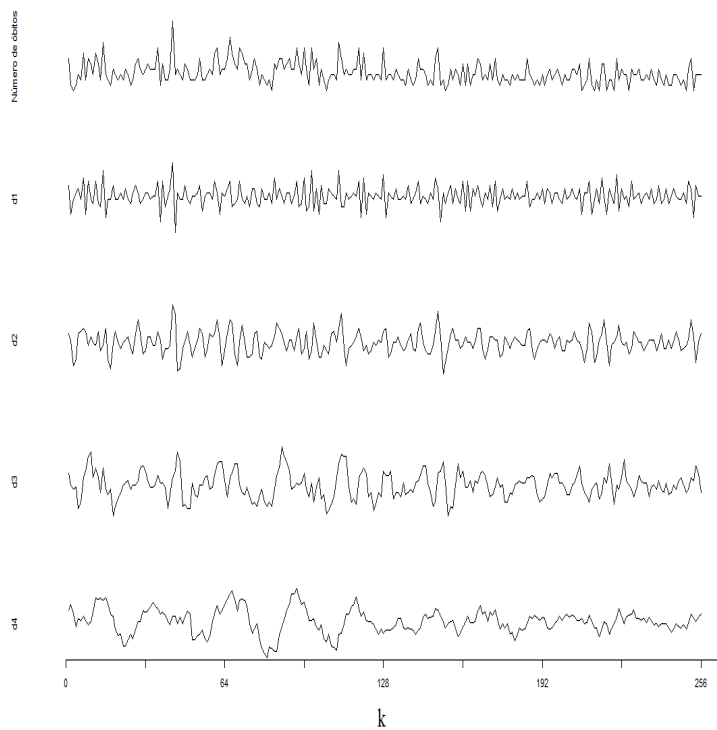


Figura 35 Gráfico da decomposição (Haar) para o número de óbitos por causas cardíacas para o primeiro tratamento (<12 anos), considerando a transformada discreta de ondaleta não-decimada

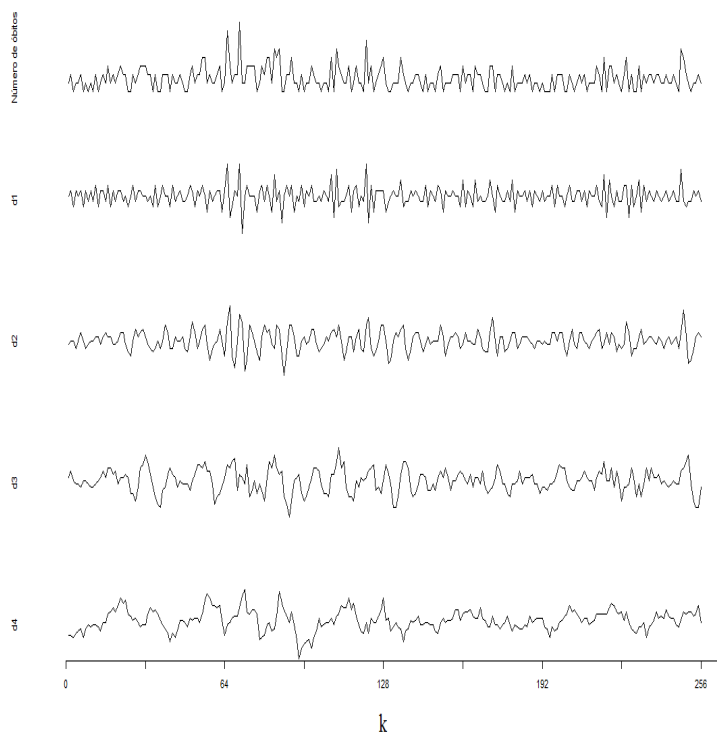


Figura 36 Gráfico da decomposição (Haar) para o número de óbitos por causas cardíacas para o segundo tratamento (≥ 12 e < 35 anos), considerando a transformada discreta de ondaleta não-decimada

As Figuras 37 e 38 apresentam a decomposição relativa à primeira repetição do dois primeiros tratamentos na base de Daubechies.

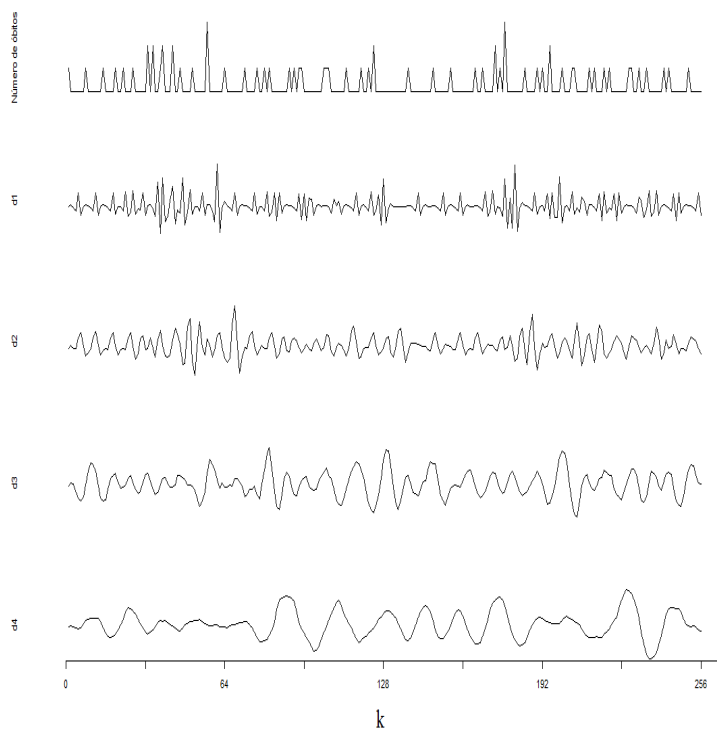


Figura 37 Gráfico da decomposição (Daubechies) para o número de óbitos por causas respiratórias para o primeiro tratamento (faixa etária < 12 anos), considerando a transformada discreta de ondaleta decimada

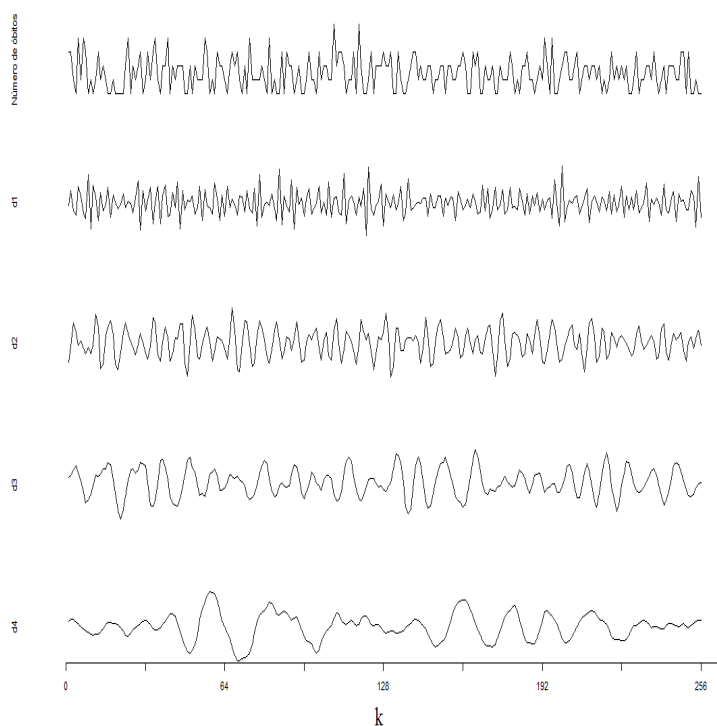


Figura 38 Gráfico da decomposição (Daubechies) para o número de órbitas por causas respiratórias para o segundo tratamento (faixa etária ≥ 12 e < 35 anos), considerando a transformada discreta de ondaleta decimada

A análise de variância utilizando as ondaletas de Haar e Daubechies será considerada para $j = 8$ níveis. Considerando $X_{lq}(t)$, onde $l = 1, 2, 3, 4$, $q = 1, 2, 3, 4$ e $t = 0, 1, \dots, 255$.

As Figuras 39 e 40 apresentam o valor-p de cada nível j para a ondaleta de Haar e Daubechies, respectivamente.

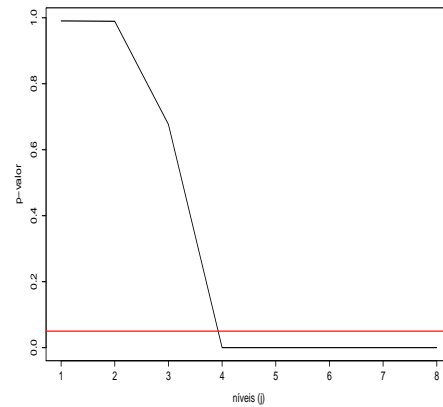


Figura 39 Gráfico do valor-p para o número de óbitos por causas cardíacas, considerando a transformada discreta de ondaleta não-decimada na base de Haar. Linha horizontal representa o valor de 0,05 (5% de significância)

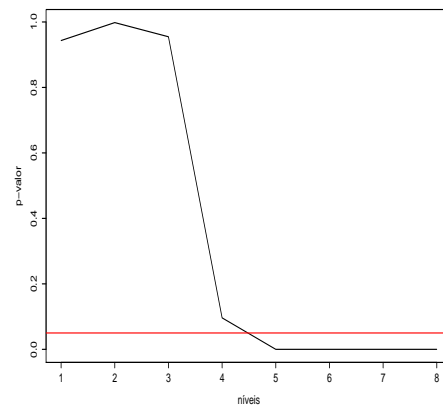


Figura 40 Gráfico do valor-p para o número de óbitos por causas cardíacas, considerando a transformada discreta de ondaleta não-decimada na base de Daubechies. Linha horizontal representa o valor de 0,05 (5% de significância)

Nota-se, por meio da Figura 39, que para a ondaleta de Haar, nos níveis $j = 4, 5, 6, 7$ e 8 , tem-se valores-p menores que $0,05$ e para a ondaleta de Daubechies,

conforme Figura 40, nos níveis $j = 5, 6, 7$ e 8 . Dessa forma, foram encontradas evidências de diferenças significativas, ao nível de 5% de probabilidade, entre o número de óbitos por causas cardíacas entre as faixas etárias. Rejeita-se, portanto, a hipótese nula tanto para a ondaleta de Haar, quanto para a ondaleta de Daubechies.

5 CONCLUSÕES

Com base nos resultados de estudos de simulação e de dados reais, pode-se observar que a análise de variância usando a transformada discreta de ondaleta decimada (para o caso estacionário), e a transformada discreta de ondaleta não-decimada (para o caso não estacionário) nas bases de Haar e Daubechies, pode ser aplicada para a análise de variância em séries temporais, com sucesso. Contudo, os resultados observados de maiores destaques foram:

- a) Para as situações onde as séries eram estacionárias, as ondaletas de Haar e Daubechie apresentaram comportamento semelhantes;
- b) Haar mostra-se com um poder do teste menor que a ondaleta de Daubechies, apresentando níveis menores de valores-p significativos em todos os níveis da simulação;
- c) Para dados reais, a análise de variância usando a transformada discreta de ondaleta não decimada (séries não estacionárias), nas bases de Haar e Daubechies, mostrou-se eficiente nas duas situações estudadas.

Em resumo, conclui-se que a análise de variância usando ondaletas, pode ser usada com sucesso na determinação de diferença entre séries temporais. Para séries categóricas não estacionárias, deve-se priorizar o uso da transformada discreta de ondaleta não decimada na base de Daubechies.

REFERÊNCIAS

- BANZATTO, A. D.; KRONKA, S. N. **Experimentação agrícola**. 4. ed. Jaboticabal: FUNEP, 2006. 249 p.
- BOLZAN, M. J. A. Transformada em ondeleta: uma necessidade. **Revista Brasileira de Ensino de Física**, Porto Alegre, v. 28, n. 4, p. 563-567, 2006.
- BOX, G. E. P. Some theorems on quadratic forms applied in the study of analysis of variance problems: I., effect of inequality of variance in the one-way classification. **The Annals of Mathematical Statistics**, Ann Arbor, v. 25, n. 2, p. 290-302, 1954.
- CHIANN, C. **Análise de ondaletas em séries temporais**. 1997. 137 f. Dissertação (Mestrado em Estatística) - Unversidade de São Paulo, São Paulo, 1997.
- CHIANN, C.; MORETTIN, P. A. A wavelet analysis for time series. **Journal Nonparametric Statistics**, Chicago, v. 10, n. 1, p. 1-46, 1998.
- CHIANN, C.; TOLOI, C. M. C. Alguns modelos de análise de variância em séries temporais utilizando transformada de Fourier: uma aplicação. **Revista Brasileira de Estatística**, Rio de Janeiro, v. 59, n. 211, p. 81-112, 1998.
- CHUI, C. K. **Wavelet analysis and its applications**. San Diego: Academic, 1992. 267 p.
- COCHRAN, W. G.; COX, G. M. **Experimental designs**. 2nd ed. New York: J. Wiley, 1957. 611 p.
- CONAGIN, A. et al. Efeito da falta de normalidade em testes de homogeneidade das variâncias. **Bragantia**, Campinas, v. 52, n. 2, p. 173-180, 1993.
- DAHLHAUS, R. Fitting time series models to nonstationary process. **The Annals of Statistics**, Hayward, v. 25, n. 1, p. 1-37, 1997.
- DAUBECHIES, I. **Ten lectures on wavelets**. Philadelphia: CBMS, 1992. 341 p. (Lecture Notes Series).

DAUBECHIES, I.; MALLAT, S. G.; WILLSKY, A. S. **Special issue on wavelet transforms and multiresolution signal analysis**. New York: IEEE, 1992. (IEEE Transactions on Information Theory). Disponível em: <<http://books.google.com.br/books?id=yTkrGwAACAAJ>>. Acesso em: 10 nov. 2014.

DEAN, A. M.; VOSS, D. T. **Design and analysis of experiments**. New York: Springer, 1999. 741 p.

DONOHO, D. L. Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data. In: _____. **In proceedings of symposia in applied mathematics**. Providence: American Mathematical Society, 1993. p. 173-205.

DONOHO, D. L.; JOHNSTONE, I. M. Minimax estimation via wavelet shrinkage. **The Annals of Statistics**, Hayward, v. 26, n. 3, p. 879-921, 1998.

FARGE, M. Wavelet transforms and their applications to turbulence. **Annual Review of Fluid Mechanics**, Evanston, v. 24, p. 395-457, 1992.

FEDERER, W. T. **Experimental design**. New York: MacMillan, 1955. 591 p.

GAMAGE, N.; BLUMEN, W. Comparative analysis of low-level cold fronts: wavelet, fourier, and empirical orthogonal function decompositions. **Monthly Weather Review**, Boston, v. 121, n. 10, p. 2867-2878, Oct. 1993.

GAO, H. Choice of thresholds for wavelet shrinkage estimate of the spectrum. **Journal of Time Series Analysis**, Clevedon, v. 18, n. 3, p. 231-251, 1997.

GROSSMANN, A.; MORLET, J. Decomposition of hardy functions into square integrable wavelets of constant shape. **Siam Journal on Mathematical Analysis**, Philadelphia, v. 15, p. 723-736, 1984.

HALL, P.; PATIL, P. On the choice of smoothing parameter, threshold and truncation in nonparametric regression by non-linear wavelet methods. **Journal of the Royal Statistical Society. Series B (Methodological)**, London, v. 58, n. 2, p. 361-377, 1996.

- HALL, P.; PATIL, P. On wavelet methods for estimating smooth functions. **Bernoulli, International Statistical Institute**, Voorburg, v. 1, n. 1/2, p. 41-58, 1995.
- HANNAN, E. Central limit theorems for time series regression. **Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete**, Berlin, v. 26, n. 2, p. 157-170, 1973.
- MALLAT, S. **A wavelet tour of signal processing**. 3rd ed. New York: Academic, 2008. 832 p.
- MALLAT, S. G. A theory for multiresolution signal decomposition: the wavelet representation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, New York, v. 11, p. 674-693, 1989.
- MARQUES, D. G. et al. Qualidade dos ensaios de competição de cultivares de milho no estado do Rio Grande do Sul. **Ciência Rural**, Santa Maria, v. 30, n. 3, p. 381-385, maio/jun. 2000.
- MEMÓRIA, J. M. P. **Breve história da estatística**. 21. ed. Brasília: EMBRAPA Informação Tecnológica; EMBRAPA - Secretaria de Gestão e Estratégia, 2004. Disponível em: <<http://www.embrapa.br/publicacoes/tecnico/folderTextoDiscussao/arquivos-pdf/texto21.pdf>>. Acesso em: 10 nov. 2014.
- MEYER, Y. **Wavelets: algorithms and applications**. Philadelphia: SIAM, 1993. 142 p.
- MISITI, M. et al. **Wavelet toolbox users guide: for use with matlab**. Natick: The Math Works, 2014. 572 p.
- MONTGOMERY, D. **Desing and analysis of experiments**. 5th ed. New York: J. Wiley, 2001. 752 p.
- MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of statistics**. 3rd ed. New York: McGraw-Hill, 1974. 480 p.
- MORETTIN, P. A. **Ondas e ondaletas: da análise de fourier à análise de ondaletas de séries temporais**. São Paulo: EDUSP, 2014. 276 p.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de série temporais**. 2. ed. São Paulo: E. Blücher, 2006. 38 p. (ABE-Projeto Fisher).

MOULIN, P. Wavelet thresholding techniques for power spectrum estimation. **IEEE Transactions on Signal Processing**, New York, v. 42, n. 11, p. 3126-3136, 1993.

NASON, G. P.; SACHS, R. V.; KROISANDT, G. Wavelet process and adaptive estimation of the evolutionary wavelet spectrum. **Journal of the Royal Statistical Society, Serie B**, London, v. 62, n. 2, p. 271-292, 2000.

NEGRI, T. T. **Reconhecimento de padrões via Wavelets**: aplicação no monitoramento de estruturas a partir do efeito do multicaminho proveniente de sinais de satélites GPS. 2012. 77 f. Dissertação (Mestrado em Matemática Aplicada e Computacional) - Universidade Estadual Paulista, Registro, 2012.

PERCIVAL, D. B.; WALDEN, A. T. **Wavelet methods for time series analysis**. Cambridge: Cambridge University, 2006. 594 p.

PROJETO PASSO: poluição atmosférica e saúde social. Disponível em: <<http://www.ime.usp.br/jmsinger/doku.php>>. Acesso em: 10 nov. 2014.

SANTOS, C. A. G.; FREIRE, P. K. M. M.; TORRENCE, C. A transformada wavelet e sua aplicação na análise de séries hidrológicas. **Revista Brasileira de Recursos Hídricos**, Porto Alegre, v. 18, n. 3, p. 271-280, 2013.

SOUZA, E. M. **Análise de Wavelets para detecção e correção do multicaminho no posicionamento relativo GNSS estático e cinemático**. 2008. 170 f. Dissertação (Mestrado em Ciências Cartográficas) - Universidade Estadual Paulista, Bauru, 2008.

VETTERLI, M.; KOVACEVIC, J. **Wavelets and subband coding**. Washington: Prentice Hall, 1995. 522 p.

WENG, H. Y.; LAU, K. M. Wavelets, period doubling, and time: frequency localization with application to organization of convection over the tropical western pacific. **Journal of the Atmospheric Sciences**, Boston, v. 51, n. 7, p. 2523-2541, Sept. 1994.

WHITCHER, B. **Waveslim**: basic wavelet routines for one, two- and three-dimensional signal processing. R Package Version 1.6.4. 2010. Disponível em: <<http://CRAN.Rproject.org/package=waveslim>>. Acesso em: 10 nov. 2014.