

MÉTODOS PARAMÉTRICOS E NÃO PARAMÉTRICOS NA ESTIMAÇÃO DA TENDÊNCIA

Acássio Valente Feliciani (UFSM) acassiofeliciani@gmail.com
Adriano Mendonça Souza (UFSM) amsouza.sm@gmail.com
Francisca Mendonça Souza (ISCTE/IUL) kikamatcom@gmail.com

Resumo:

O objetivo desta pesquisa é de facilitar a compreensão de como estimar a tendência de uma série temporal através de métodos paramétricos e não paramétricos. Para estimação da tendência por meio dos diferentes métodos, utilizou-se o software Excel, onde os dados em análise referem-se a séries mensais de preços da carne ovina no Rio Grande do Sul, período de dezembro de 2000 a fevereiro de 2009, obtidos via sistema interno de dados EMATER/RS, sendo meramente para fins práticos. Conclui-se que a série original apresentou-se a componente tendência, a qual foi estimada e removida por meio dos diferentes métodos aplicados.

Palavras chave: Séries Temporais, estimação de tendência, modelos paramétricos, modelos não paramétricos.

PARAMETRIC AND NONPARAMETRIC METHODS IN THE ESTIMATION OF TREND

Abstract

The objective of this research is to facilitate the understanding of how to estimate the trend of a time series by parametric and nonparametric methods. For trend estimation by means of different methods, we used the Excel software, where the data in question relate to monthly series of sheepmeat prices in Rio Grande do Sul, period December 2000 to February 2009, obtained via internal system data EMATERRS, being merely for practical purposes. It is concluded that the original series showed the trend component, which was estimated and removed by means of different methods applied.

Key-words: Temporal series, trend estimation parametric models, nonparametric models.

1. Introdução

A imposição de novos desafios e necessidades impostas pelo mercado consumidor causou uma mudança organizacional na maioria das empresas, sendo necessária a inclusão de novas técnicas e métodos com o fim de auxiliar na produção de produtos ou serviços com eficiência e efeitos imediatos, sendo oportuno assim alcançar valorização e resultados financeiros.

O desenvolvimento das informações expressas pelos diferentes métodos quantitativos vem ao encontro das organizações, pois, por meio destas informações podem-se ser tomadas decisões ao escopo de um segmento baseado em avaliações objetivas e fundamentadas em conhecimentos científicos.

Há muitos anos o estudo da estatística esta envolvida nas mais diversas áreas do conhecimento, corroborando-se como uma ferramenta significativa no auxilio a tomada de decisões nas mais diversas áreas de estudo, todavia um dos grandes desafios encontrados para estudantes e pesquisadores é como trabalhar com esta ferramenta.

Dada à esta necessidade procura-se elencar nesta pesquisa diferentes métodos de estimação da componente tendência em um série temporal, onde utilizou-se como exemplo prática a série histórica de preços da carne ovina no Rio Grande do Sul.

Conforme Morettin & Toloi (2004), uma das suposições mais frequentes que se faz a respeito de uma série cronológica é que ela seja estacionária, ou seja, desenvolva ao longo do tempo aleatoriamente ao redor de uma média constante, refletindo alguma forma de equilíbrio estável. Entretanto a maioria das séries temporais apresenta alguma forma de não estacionariedade, associada a quatro componentes: T_t = tendência; S_t = sazonalidade; C_t = ciclo e I_t = irregularidade ou erro aleatório.

Contudo o objetivo principal desta pesquisa é de facilitar a compreensão de como estimar a tendência de uma série temporal por meio de métodos paramétricos e não paramétricos.

2. Desenvolvimento teórico

Nesta seção, são apresentados os métodos utilizados para a estimação da tendência, informações estas necessárias para o desenvolvimento e compreensão do estudo.

2.1 Séries temporais

Segundo Souza e Souza (2011) uma série temporal consiste em uma sequência de dados obtidos em intervalos regulares de tempo e que apresentam dados equidistantes. Se a série histórica for denominada como Z , o valor da série no momento t pode ser escrito como Z_t ($t = 1, 2, \dots, n$).

Como os dados estão ordenados no tempo e possui uma distância regular, dia a dia, mês a mês, ano a ano, estas apresentam-se autocorrelacionadas.

Para a observância da autocorrelação das observações no tempo, analisa-se por meio da função de autocorrelação e a função de autocorrelação parcial a qual indica o tipo e a ordem do modelo probabilístico capaz de representar a série original. Uma série temporária das vezes apresenta efeitos estilizados como: tendência, sazonalidade e/ou ciclo, os quais devem ser estimados, tratados ou eliminados para obtermos uma boa representação da série, possibilitando assim a interpretação correta da mesma.

Ao se modelar uma série temporal, deve-se observar se o conjunto de dados em estudo apresenta efeitos aditivos ou multiplicativos. Segundo Fávero, Oliveira, Angelo (2003), uma maneira prática de observarmos este efeito e através da construção de um gráfico da amplitude sazonal em função da tendência anual conforme as Figura 1 e 2.

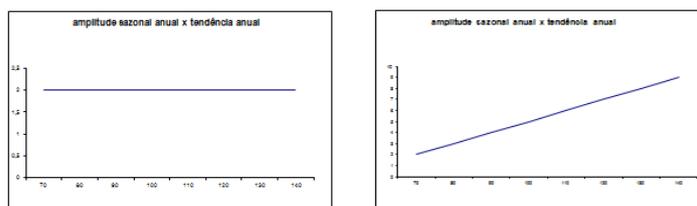


Figura 1- Aditivo; Figura 2- Multiplicativo

Na Figura 1, observa-se que não existe uma dependência da sazonalidade sobre a tendência, o que confirma a utilização de um modelo aditivo, pois há um deslocamento paralelo ao eixo X. Em contrapartida a Figura 2 a reta ajustada apresenta uma inclinação diferente de zero em relação ao eixo das abscissas, indicando uma dependência da sazonalidade sobre a tendência, confirmando a adequabilidade de um modelo multiplicativo para o relacionamento dos componentes da série, conforme Arita e Dias (2000).

Classificada a série temporal em aditiva ($Z_t = T_t + S_t + C_t + e_t$) ou multiplicativa ($Z_t = T_t \times S_t \times C_t \times e_t$), deve-se ter o cuidado no momento da eliminação da componente desejada. Se a série é multiplicativa a retirada da componente será dada pela divisão da série original pela

tendência estimada. Todavia, se a série for aditiva, bastaria subtrair o efeito da componente estimada pela série original.

Lembrando que uma série multiplicativa pode ser transformada em aditiva, através da aplicação de logaritmos.

3. Estimação da tendência

Há vários métodos de estimar a componente de tendência T_t de uma série temporal sendo: tendência polinomial, suavização e o método das diferenças.

Para verificar se existe tendência em um conjunto de dados devem-se aplicar alguns testes antes da estimação, como exemplificado a seguir.

3.1 Antes da estimação de T_t

Segundo Morettin e Tolo (2004), existem diferentes testes para estimar a tendência, dentre os quais: Teste de Wald-Wolfowitz, teste do sinal Cox-Stuart e teste baseado no coeficiente de Spearman.

Nesta pesquisa será empregado o teste de Wald-Wolfowitz, no qual é empregado para verificar se há tendência em um conjunto de dados, do mesmo modo para verificar a estacionariedade da série. Utilizam-se as N observações $Z_t, t = 1, \dots, N$, de uma série temporal e seja m a mediana destes valores. Atribui-se a cada valor Z_t o símbolo A , se ele for maior ou igual à mediana (m), e B se ele for menor que m . Teremos, então, $N = (n_1 \text{ pontos } A) + (n_2 \text{ pontos } B)$.

Este teste é baseado em um teste de hipóteses onde os valores são classificados em relação à mediana, como temos n_1 e n_2 maiores que 20 podemos usar a aproximação normal, $T_1 \sim N(\mu, \sigma^2)$, onde.

Consideram-se as seguintes hipóteses:

H_0 : a sequência é aleatória (não existe tendência).

H_1 : a sequência não é aleatória (existe tendência).

A estatística do teste é dada por $Z_{cal} = \frac{T_1 - \mu}{\sigma}$, onde T_1 é o número total de sequências (n° de grupos com símbolos iguais).

As equações utilizadas para calcular a média e a variância são mostradas a seguir:

$$\mu = \frac{2n_1n_2}{N} + 1 \quad (1)$$

$$\sigma^2 = \frac{2n_1n_2(2n_1n_2 - N)}{N^2(N - 1)} \quad (2)$$

3.2 Tendência Polinomial

Segundo Neufeld (2003), a regressão linear simples analisa a relação entre duas variáveis. Tem-se uma variável independente (variável X) e uma dependente (variável Y), sendo que a primeira causa mudanças na segunda. Por meio deste método paramétrico onde é ajustado uma função do tempo através de uma regressão, de posse dos coeficientes angular e linear estima-se a tendência.

O modelo estatístico é dado:

$$Y_i = \beta_0 + \beta_1 X_i + e_i \quad (3)$$

Onde:

Y_i – Representa o valor observado para a variável dependente Y, no i-ésimo nível da variável independente X;

β_0 – constante de regressão representa o intercepto da reta com o eixo Y;

β_1 – coeficiente de regressão representa a variação de Y em função da variação de uma unidade da variável X;

X_i – i-ésimo nível da variável independente X(i=1,2, n);

e_i – é o erro que está associado á distância entre o valor observado Y_i e o correspondente ponto na curva do modelo proposto para o mesmo nível de X.

3.3 Suavização por meio de médias móveis

Segundo Morettin (2006) existem vários métodos de suavização, dentre os quais: métodos por médias móveis, medianas móveis e lowess que conforme Carvalho, Kupfer (2007), consiste em um método de modelagem que se baseiam em métodos clássicos (linear e não linear), onde realiza regressões locais para uma variável explicativa (X) sobre uma variável independente (Y), o que implica na realização de uma regressão não paramétrica para cada uma das observações da amostra, através da atribuição de maior peso na regressão aos dados que estão mais próximos de cada observação. Por meio de médias móveis um procedimento não paramétrico utiliza-se de uma transformação linear para modificar a série original Z_t em uma série suavizada Y_t , através da equação:

$$Z_t = Z_t - Z_t^* \quad (6)$$

Calculando a média móvel obtêm os valores de Z_t^* , encontrado estes valores faz-se a diferença entre os valores de Z_t e Z_t^* para a obtenção de Y_t série livre de tendência.

3.4 Diferenças

Conforme Macerau (2009) a maioria dos procedimentos de análises de séries temporais, supõe que estas sejam estacionárias, geralmente séries financeiras não apresentam este comportamento, para tornar a série estacionária umas das transformações clássicas utilizadas e tomar diferenças sucessivas da série original, até se obter uma série estacionária.

Primeira diferença:

$$\Delta Z_t = Z_t - Z_{t-1} \quad (7)$$

Segunda diferença é dada por:

$$\Delta^2 Z_t = \Delta[\Delta Z_t] = \Delta[Z_t - Z_{t-1}] \quad (8)$$

$$\Delta^2 Z_t = \Delta Z_t - \Delta Z_{t-1} \quad (9)$$

$$\Delta^2 Z_t = Z_t - 2Z_{t-1} + Z_{t-2} \quad (10)$$

de modo geral, a n-ésima diferença de Z_t é $\Delta^n Z_t = \Delta [\Delta^{n-1} Z_t]$.

Em situações normais, será suficiente tomar uma ou duas diferenças para que a série se torne estacionária.

Considerando o operador de translação ao passado B , definido por:

$$BZ_t = Z_{t-1}, B^m Z_t = Z_{t-m} \quad (11)$$

A equação (7) pode ser escrita por:

$$\Delta Z_t = (1 - B) Z_t \quad (12)$$

Além disso, da equação (11) segue que $\Delta = I - B$.

4. Metodologia

Nesta etapa da pesquisa apresenta-se por meio de exemplos práticos como e feita à estimação da tendência. Os dados coletados são referentes a séries mensais de preços pagos a carne ovina do Rio Grande do Sul no período de dezembro de 2000 a fevereiro de 2009, obtidos na ASCAR/EMATER/RS, via sistema interno de dados.

4.1 Aplicações práticas

Este estudo está desenvolvido nas seguintes etapas: Testes antes da estimação, estimação da tendência pelos métodos: Tendência polinomial, Suavização e Diferenças.

Passo 1 – Deve-se calcular a média, mediana e variância da série, no qual atribuímos o valor de A como sendo a variável que mostrará valores iguais ou maiores que a mediana e B valores menores do que a mediana que é calculada pela ordenação dos valores e toma-se o valor central. Desta forma consegue estimar a média e a variância da série, conforme mostrado Figura 3 e 4.

Figura 3- Teste de sequencias (Wald-Wolfowitz); Figura 4-Cálculo da média

Em relação à mediana obtiveram-se 50 observações iguais ou maiores ao valor 2,2 o que corresponde ao valor da mediana, e 49 observações menores, totalizando assim 99 observações. O cálculo do T_1 foi encontrado o valor 26 (total de sequencias).

Passo 2- Como ilustrada na Figura 4 a média é calculada pela equação 1.

Onde: $n_1 = A$ e $n_2 = B$ e N é o total das observações.

Passo 3- Cálculo da variância utiliza-se da equação 2, conforme Figura 5.

Figura 5-Cálculo da Variância; Figura 6-Cálculo do Z calculado

Passo 4 - Cálculo do Z calculado conforme equação 13 demonstrado na figura 6.

$$Zcal = \frac{T_1 - \mu}{\sigma} \quad (13)$$

Como o valor de Z_{cal} foi de -4,95, a um nível de significância de 0,05, rejeitamos H_0 , assim podemos concluir que existe tendência na série.

4.2 Estimação da tendência

Passo 1- Devem-se ter os dados coletados e revelados em uma planilha do Excel. Esta pesquisa utiliza-se software Microsoft Excel 2007, como mostra figura 7, lembrando que estes dados devem ser fidedignos, para que não haja problemas na caracterização dos resultados.

Passo 2- Deve-se construir um gráfico linear (em linhas), para observar o comportamento da série, ou seja, observar se a série apresenta tendência, sazonalidade ou ciclo. Conforme mostrados nas figuras 8 e 9.

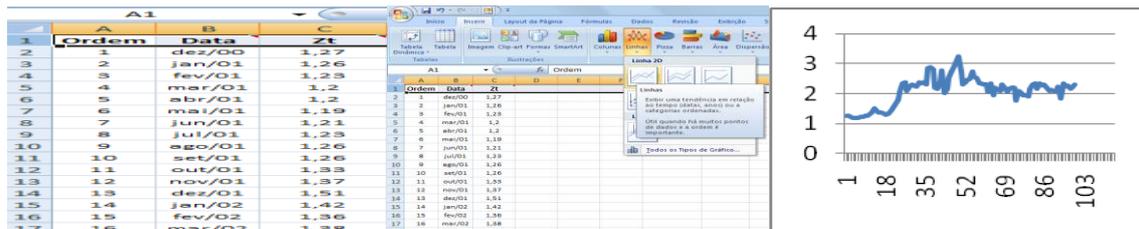


Figura 7- Dados originais; Figura 8-Gráfico linear; Figura 9- Comportamento da série original.

Os valores que foram utilizados equivalem à variável de interesse, neste caso o preço pago a carne ovina no Rio Grande do Sul.

Passo 3- Realizar o cálculo da média e da amplitude, para observância se a série apresenta efeitos aditivos ou multiplicativos. Conforme Figura 10; em seguida grafa-se o gráfico da média x amplitude conforme Figuras 11 e 12.

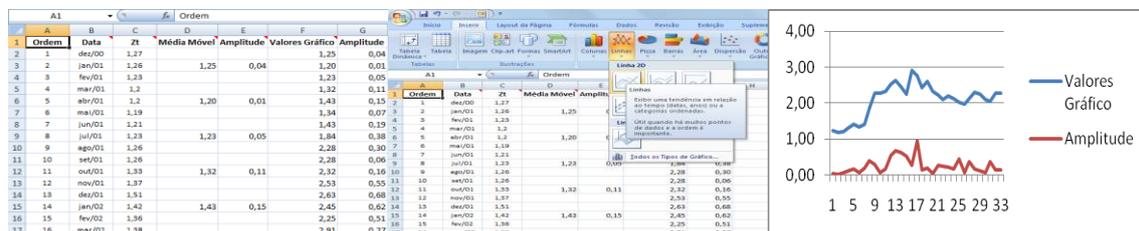


Figura 10- Cálculo da média e da amplitude. Figura 11-Inserir gráfico da média x amplitude; Figura 12-Aditiva x Multiplicativa

A média foi feita de ordem três, ou seja, a soma dos três primeiros meses, dividido por três, isto se deve a sazonalidade observada na Figura 9, onde nota-se uma repetição de 3 em 3 meses. O cálculo da amplitude é feita pela diferença da maior média subtraída da menor. Conforme equação 14.

$$R = X_{max} - X_{min} \quad (14)$$

Para construir o gráfico da média vs amplitude, estas, devem ser selecionadas até o seu final. Uma maneira prática de visualizar se o efeito é aditivo ou multiplicativo e observando a média e a amplitude, no qual pela Figura 12 mostram-se independentes, caracterizando uma série aditiva, visto que sua representação foi paralela ao eixo da abscissa.

Agora parte-se para os métodos de estimação de tendência, sendo os métodos pela regressão (Tendência Polinomial), método por médias móveis(suavização) e por diferenças os mais aplicados.

4.3 Tendência Polinomial

Para a estimação da tendência pelo método da tendência polinomial, utiliza-se a regressão, sendo utilizada uma reta para estimar $T(\text{tempo})$. Inicialmente encontram-se os coeficientes da equação de regressão, em seguida estima-se a tendência para cada observação através da equação obtida. Em sequencia subtrai-se da série original a tendência estimada, obtendo assim a série livre de tendência. Para realizar os cálculos procede-se da seguinte maneira.

Passo 1- Calcular a regressão, caso o suplemento “Análise de dados” não estiver apto em seu aplicativo Excel 2007, proceda da seguinte forma.

Click- botão office- opções do excel- suplementos- suplementos do excel ir- selecionar Análise de dados - ok.

Passo 2- Para estimarmos a tendência pelo procedimento paramétrico constroi-se o gráfico da série original para obtermos os valores α, β como mostrado nas Figuras 13 e 14.

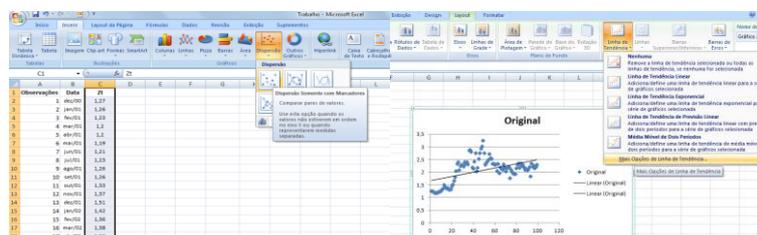


Figura 13- Inserir gráfico de dispersão; Figura 14-Inserir equação o gráfico

Para dispor a equação da Regressão linear deve-se seguir o seguinte procedimento. Seleciona-se o gráfico, em seguida click em layout, linha de tendência, mais opções de linha de tendência, exibir equações no gráfico como mostrado na Figuras 15 e 16.

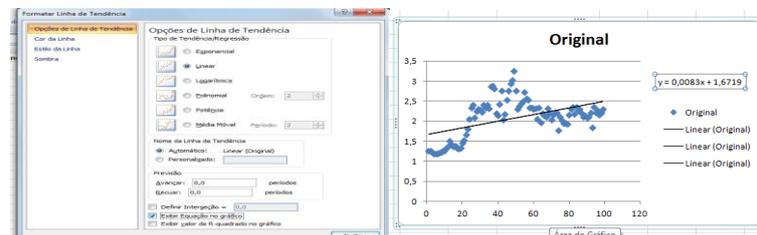


Figura 15-Inserir equação no gráfico. Figura 16-Dados originais, com a equação da regressão

Esta equação deve-se ser utilizada para estimarmos a tendência pelo método da regressão, ou seja, $Y = 1,6719 + 0,0083 * \text{tempo}(1)$, conforme Figura 17.

Estimada a série com tendência, estima-se a série livre de tendência, subtraindo o valor da série original pela tendência estimada pela regressão, conforme Figura 18.

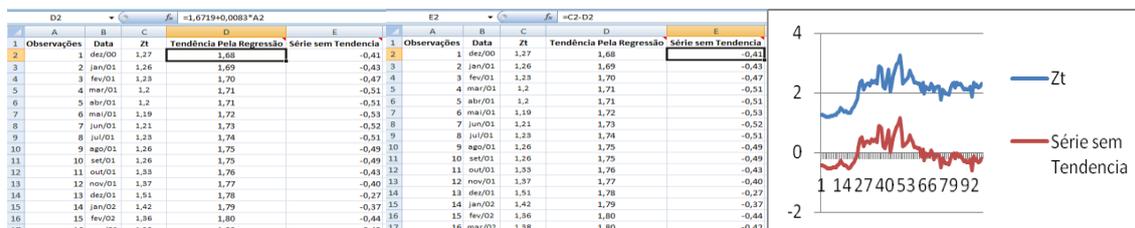


Figura 17-Estimação da tendência por regressão; Figura 18- Série sem tendência; Figura 20-Série original vs série sem tendência

Aplicado este procedimento plota-se o gráfico da série original pela série livre de tendência na busca de verificar se o método de tendência polinomial foi vantajoso.

Passo 3- Grafa-se o gráfico da série original vs série livre de tendência, conforme Figura 20.

Conforme mostra Figura 20, obteve-se uma série livre de tendência, concluindo assim, que o método pela tendência polinomial mostrou-se eficiente.

4.4 Suavização por meio de médias móveis

Consiste em um método não paramétrico, processo este chamado de suavização por médias móveis. Este método não é aconselhável para realizar previsões, pois as primeiras observações e as últimas são perdidas, no entanto o método de médias móveis constitui-se adequado para remover a tendência da série original.

Os valores são dados pelas médias das observações, para realização destes cálculos são tomados k sub-grupo de 4,6,8,ou 12 observações, mas geralmente toma-se um sub-grupo do tamanho da sazonalidade que a série apresentar. O procedimento é feito análogo a Figura 10, ou seja, é feita a média móvel das três primeiras observações, conforme Figura 21. Para dispor a série original livre de tendência, deve-se subtrair a série estimada pela série original Z_t . Conforme Figura 22.

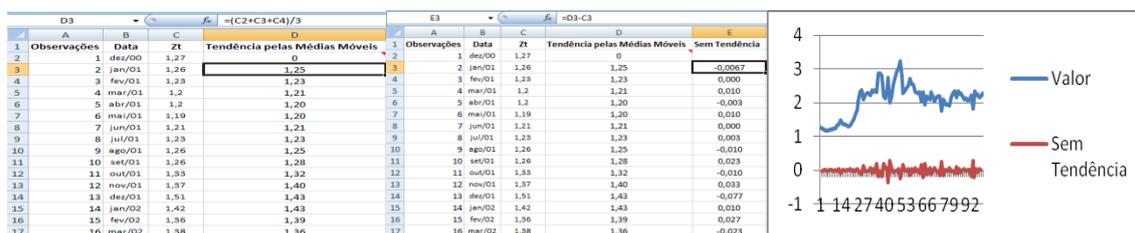


Figura 21-Estimação da tendência por médias móveis. Figura 22 Série sem tendência; Figura 23-Série original vs série livre de tendência

Desta forma obtem-se a série estimada livre de tendência, sendo a construção do gráfico linear necessário para verificassia se o procedimento foi eficiente, conforme Figura 23.

Conforme mostra Figura 23, o método de suavização por médias móveis mostrou-se eficiente na eliminação da tendência.

4.5 Método pela diferenças

Métodos não paramétricos (dados não normais), para obtenção de uma série livre de tendência, desta forma, utilizando um número apropriado de diferenças obteremos uma constante.

Passo 1- O método das diferenças consiste em aplicar a primeira diferença aos dados originais Z_t , sendo a perda da primeira observação fato explicado por se tratar de um método não paramétrico onde utiliza-se uma transformação linear para modificar a série original Z_t , em uma série suavizada Z_t^* livre de tendência. Conforme Figuras 24 e 25.

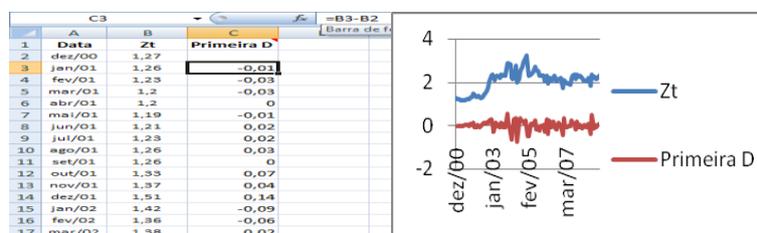


Figura 24-Primeira diferença; Figura 25- Série original x Primeira diferença

Passo 2- Feita esta transformação constroi-se um gráfico linear da série original Z_t , com a série aplicada de primeira diferenças. Conforme Figura 25, observa-se que com a aplicação da primeira diferença foi possível estacionarizar a série, caso isto não acontecesse seria necessário a aplicação da segunda ou da terceira diferença.

Conforme mostra Figura 25, o método de por diferenças mostrou-se eficiente na estimação e/ou eliminação da tendência.

5. Conclusão

Ao desenvolver uma pesquisa sobre estimação da tendência, procurou-se fornecer um propriedade a mais para estudantes e pesquisadores, onde a aplicabilidade dos diferentes métodos de estimação de tendência na busca da série livre de tendência. A metodologia empregada para estimação da tendência foi a de métodos paramétricos e não paramétricos, mostrando-se eficientes na eliminação da tendência que a série original apresentava. Também, pode-se observar a aplicação de testes antes da estimação de tendência significativo na verificação de alguma anormalidade presente na série, visto que sem a utilização destes testes o estudante ou pesquisador inspeciona a série original apenas graficamente, exigindo assim, um bom conhecimento de séries temporais.

Desta forma procurou-se, por meio de um ferramental claro e objetivo, fornecer uma pesquisa baseada em dados quantitativos, contribuindo assim para o bom aprendizado de quem possa utilizar estas ferramentas.

Agradecimentos

A “CAPES”, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior entidade governamental brasileira promotora do desenvolvimento científico e tecnológico, pelo auxílio financeiro. Também ao trabalho anônimo dos revisores, por auxiliarem na melhoria da qualidade da pesquisa.

Referências Bibliográficas

ARITA, H. H. e DIAS, M. H. A. Análise comparativa dos métodos de ajustamento sazonal X-11 e X-11-ARIMA: uma aplicação numa série temporal econômica brasileira. *XXII EBE*. Campinas, 2000.

ASCAR/EMATER Séries históricas. Disponível em: <<http://www.emater.tcche.br/site/servicos/informativos.php?#precos>>. Acesso em: 20/12/2011.

BOX, G. E. P.; JENKINS, G. M. Time Series Analysis: forecasting and control. San Francisco: Holden Day, 1976.

BOX, G. E. P.; LUCEÑO, A. *Statistical control by monitoring and feedback adjustment*. New York: Wiley Interscience, 1997.

CARVALHO, L.; KUPFER, D. A transição estrutural da indústria brasileira: Da diversificação para a especialização. XXXV Encontro Nacional de Economia da ANPEC, 2007, Recife.

FÁVERO, L. P. L. ; OLIVEIRA, M. A. ; ANGELO, C. F. . Aplicação de métodos de ajustamento sazonal em séries temporais: VI Semead - Seminários em Administração FEA-USP, 2003, São Paulo.

MACERAU, W. M. Modelos de série temporal para previsão de vendas de bebidas a base de soja da cocamar cooperativa agroindustrial, 2009 Maringá.

MORETTIN, P. A.; TOLOI, C. M. Análise de Séries Temporais. 1ª ed. São Paulo: E.Blücher, 2004.

MORETTIN, P, A. Métodos quantitativos: Séries Temporais. ATL: São Paulo, 1986.

NEUFELD, J. L. Estatística Aplicada à Administração usando Excel. 1ª ed., São Paulo: Prantice Hall, 2003.

NOLAN, D.; SPEED, T. P. *Teaching statistics theory through applications. The American Statistician*, Alexandria, v. 53, n. 4, p. 370-375, Nov. 1999.

REIS, G. M & RIBEIRO JR, J. I. Experimentos com um Fator de Interesse, Capítulo 6, Viçosa, 2007.

SOUZA, F.; ALMEIDA, S.; SOUZA, A.; LOPES, L. Previsão do preço da gasolina para a região sul do Brasil. *IJIE – Iberoamerican Journal of Industrial Engineering*, v. 3, n. 1, p. 234-248, Julho, 2011. Disponível em: <<http://periodicos.incubadora.ufsc.br/index.php/IJIE/article/view/186/pdf>>. Acesso em 02 de setembro de 2012.