OECOLOGIA BRASILIENSIS
Volume II: Tópicos em Tratamento de Dados Biológicos
PERES-NETO, P.R.; VALENTIN, J.L. & FERNANDEZ, F.A.S. (editores), 1995, p.91-118.
Programa de Pós-Graduação em Ecologia - Instituto de Biologia - UFRJ, Rio de Janeiro - RJ.

ANÁLISES DE SÉRIES TEMPORAIS

GARBER, R.

Resumo:

O principal objetivo da aplicação de modelos de séries temporais é a obtenção de projeções para os valores futuros de uma variável em estudo. A escolha do modelo mais adequado depende de uma análise dos valores observados para a série. No presente capítulo, é introduzida a utilização dos modelos de série temporal conhecidos como "modelos de amortecimento exponencial". Estes modelos, do ponto de vista da estimativa de parâmetros, são muito mais simples e de fácil compreensão que os demais. São apresentados os detalhes para o procedimento de modelagem estatística para séries estacionárias (amortecimento exponencial simples), séries de tendência linear (amortecimento exponencial duplo) e séries sazonais (série sazonal multiplicativa e aditiva). Em adição é apresentada uma comparação entre os modelos de séries temporais e os de regressão linear simples.

Abstract:

"Time series analysis."

The main goal of applying time series models is to obtain projections for the future values of the variable being studied. The choice of the most suitable model depends on an analysis of the values observed in the series. The present chapter introduces the use of time series models known as "exponetial smoothing models". Such models, as far as parameter estimation is concerned, are much simpler and easier to understand than other models. The chapter presents the methods for statistical modelling of static series (simple exponential smoothing), series with a linear trend (double exponential smoothing) and series with seasonal trend (multiplicative and additive seasonal series). A comparison between the models of time series and of linear regression is also presented.

Introdução

A partir da possibilidade de mensuração do tempo, podemos estabelecer algumas relações entre a passagem do tempo e a ocorrência de fenômenos biológicos. Ao observarmos um fenômeno (variável) na natureza, podemos perceber que este é estimulado por uma infinidade de outros fenômenos (variáveis) correlacionados a ele, uns com interferência maior e outros com menor. Uma das variáveis que podem estar interferindo no desenrolar é o transcurso do tempo. O que se coloca, neste momento, é o nível de importância da passagem do tempo para as alterações que o processo vem sofrendo. Falar que um elemento da espécie humana tem setenta anos de idade nos faz, imediatamente, imaginar um senhor com cabelos brancos, dificuldade de locomoção, uma grande experiência de vida e etc. Note que o exemplo apresentado descreve um grupo de variáveis altamente associadas à passagem do tempo, logo o tempo pode ser uma boa variável para expressar o desenvolvimento destas variáveis. Um contra-exemplo para a relação com o tempo seria a situação de se estimar o sexo de um cão a partir de sua idade, ou de se estimar a cor dos olhos de um gato também a partir de sua idade. Note que a dificuldade das estimações apresentadas nasce do fato de não existir qualquer ligação entre o sexo do cão ou a cor dos olhos do gato e o tempo. O principal elemento que se coloca para o uso do tempo como uma variável auxiliar no entendimento de um fenômeno é que o desenvolvimento deste esteja *intimamente* relacionado com o passar do tempo. O uso da palavra "intimamente" não deve ser entendido como sendo o tempo o único condutor do fenômeno, mais sim, o principal condutor. Uma infinidade de outras variáveis sempre estarão interferindo no processo, porém, o tempo é o *principal* condutor. Se voltarmos ao exemplo do homem de setenta anos, poderíamos argumentar que o seu envelhecimento pode ser afetado por questões étnicas, tipo de alimentação, dificuldades para a obtenção da sobrevivência; porém, inegavelmente, estamos falando de um homem que apresenta, claramente, os sinais de envelhecimento causados pelo tempo.

Previsão ou projeção ?

É muito comum se observar nos livros textos sobre séries temporais a palavra *previsão* ao se referir às estimativas futuras feitas a partir do estudo de uma série temporal (ST), como se o modelo utilizado tivesse a capacidade de fornecer ao analista uma pré visão (ver antes mesmo do fato acontecer), o que é absolutamente falso. Neste capítulo, adotaremos a palavra *projeção* para as estimativas dos valores futuros da série. A escolha da palavra projeção parece mais adequada no sentido de que, na verdade, o que proporemos para valores futuros é o provável desdobramento do passado, ou seja, projetaremos o passado para o futuro, e, jamais, traremos o futuro para o presente, até porque este procedimento está vedado aos seres humanos.

Objetivo dos modelos de séries temporais

O principal objetivo do estudo das séries temporais é a obtenção de uma estimativa (projeção) para os valores futuros. De posse das projeções podemos nos

planejar melhor para o futuro. Outro ponto importante é o estudo da tendência do fenômeno, se este está crescendo, diminuindo, ou mesmo permanece estável no período em estudo. Uma outra utilidade da análise das séries temporais é o estudo dos fenômenos sazonais.

Se tivéssemos que dizer numa única frase para que servem os modelos de séries temporais, diríamos que eles servem para "Administrar a incerteza que tem o homem em relação ao futuro". Um exemplo que pode ser dado é a relação entre a projeção da produção de alimentos e o crescimento populacional. Poderíamos pensar ainda em estudar a evolução no tempo da devastação de uma floresta com os seus animais, ou, a evolução da contaminação de uma população por um certo tipo de vírus.

A filosofia de funcionamento dos modelos de séries temporais

A idéia básica é a de que o passado é o melhor elemento para se estimar o futuro (isto é uma hipótese básica), neste sentido os valores observados da série no passado serão projetados para o futuro, como decorrência natural da hipótese de trabalho, todos os parâmetros dos modelos de séries temporais serão funções matemáticas dos valores passados da própria série. A escolha do modelo mais adequado dependerá de uma análise dos valores observados para a série. Note que os parâmetros do modelo não dependem diretamente do tempo, e sim, dos valores passados da série, e estes sim, supõe-se correlacionados com o tempo. Embora pareça uma diferença insignificante, este fato vai diferenciar os modelos de séries temporais de outros modelos também associados ao tempo.

Outro ponto importante na estimação dos parâmetros dos modelos de séries temporais é que, geralmente, as observações do passado não entram com os mesmos pesos para compor as estimativas. Os modelos de séries temporais exploram a idéia de que à medida que as observações se afastam no tempo do momento presente, estas perdem importância no processo de construção das projeções. A idéia aqui é a de que uma boa projeção para amanhã deve estar mais associada ao hoje do que ao ontem, que por sua vez é mais importante que o antes de ontem e assim por diante.

Cuidado no uso dos modelos de séries temporais

Conforme dito, a hipótese fundamental para o uso dos modelos de séries temporais é que o tempo é o principal condutor do fenômeno em estudo. Se por ventura tivermos um outro elemento interveniente no processo, não relacionado ao tempo, os modelos de ST não se aplicarão. Um exemplo que poderíamos apresentar é a aplicação de um modelo de ST para o crescimento da população de um certo animal na floresta e no período do estudo, o homem provocou uma grande devastação. Neste exemplo qualquer modelo de ST seria inútil, porque neste período houve uma forte intervenção independente do tempo.

A diferença entre os modelos de ST e os de regressão

Os modelos de regressão exploram a relação existente entre um grupo de variáveis para prever o comportamento de outra. Por exemplo, pode-se propor um modelo onde a população de jacarés é explicada em função do número de predadores, disponibilidade de alimentos, questões climáticas e etc. Nada impede que uma das variáveis explicativas seja também o tempo. Podemos voltar ao exemplo da população de jacarés e elegermos o tempo como um explicador da população, porém, isto não transforma o modelo de regressão num modelo de séries temporais. Lembre-se que o que utilizamos nos modelos de séries temporais não é o tempo, e sim, os valores passados da própria série. Se tentássemos explicar a evolução da população de jacarés somente em função do tempo não teríamos um modelo de séries temporais, mesmo usando a hipótese de que o tempo é o principal condutor do processo.

Existe porém um ponto onde os modelos de séries temporais e os modelos de regressão se encontram: são os modelos auto-regressivos. Neste tipo de modelo é explorada a idéia da ST e da regressão linear simultaneamente; sua formulação é tal que os valores da série são explicados como uma função linear (regressão) dos valores passados (ST). Todas as ferramentas estatísticas desenvolvidas para os modelos de regressão como teste de hipóteses, intervalos de confiança e etc, são diretamente aplicadas neste modelo que é um misto de regressão e ST. A única diferença é que o teste de auto correlação serial de Durbin-Watson apresenta problemas (Morettin & Toloi, 1987).

ST e os modelos de sobrevivência

O objetivo deste modelos é estimar a probabilidade de um elemento de uma população sobreviver a um certo período de tempo, por exemplo: a probabilidade de uma cobaia sobreviver a quinze horas após a inoculação de um vírus, ou de um indivíduo numa população qualquer na natureza estar vivo daqui a um mês. Repare que em todos os problemas propostos estamos estudando algo que depende do tempo, porém, de uma forma geral, os modelos de sobrevivência pressupõe uma certa função de probabilidade de sobrevivência no tempo. Neste tipo de estudo no tempo também existem modelos de sobrevivência que dependem, além do tempo, de outras variáveis explicativas como num modelo de regressão. Este é o caso do modelo de sobrevivência de Cox (Bolfarine *et al.*, 1991). Note que a forma como foi definido, o modelo de sobrevivência é uma função estocástica do tempo, e não função dos valores passados da função de sobrevivência, logo não é uma ST, no sentido do conceito deste texto.

Os modelos de amortecimento exponencial

Os modelos de séries temporais evoluiram muito nos últimos anos, várias teorias foram construidas para estudar fenômenos que se desenvolvem no tempo: modelos de amortecimento exponencial, modelos bayesianos, modelos de Box & Jenkins, etc (Morettin & Toloi, 1987). Notadamente os modelos de Box & Jenkins,

bem como os modelos bayesianos, se destacam dos outros pela sua formulação estatística rebuscada, oferecendo vários testes estatísticos, intervalos de confiança, distribuição de probabilidades dos estimadores dos parâmetros. Porém, estes modelos exigem que o analista tenha uma sólida formação estatística para ser capaz de identificar o modelo adequado, questionar a sua validade, fazer projeções e etc. Como o objetivo deste texto é introduzir os modelos de séries temporais, apresentaremos os modelos conhecidos como "Modelos de amortecimento exponencial". Estes modelos, do ponto de vista da estimação de parâmetros, são mais simples e de fácil compreensão, não exigindo mais que a fixação do conceito do que é ST. Não obstante a sua simplicidade matemática, os modelos de amortecimento exponencial podem, em muitas situações, gerar projeções tão boas quanto os modelos mais sofisticados, embora sejam formulados, do ponto de vista estatístico, com menos rigor (Bowerman & O'Connell, 1979).

Definição de séries temporais

Uma série temporal é uma sequência de valores observados no tempo de forma equiespaçada, ou seja, o espaço de tempo entre as observações deve ser igual.

A questão do espacejamento no tempo imposta na definição, é um problema de modelagem estatística disponível hoje; o que não impede de se desenvolverem no futuro modelos que dispensem esta restrição. Esta restrição se configura num problema intransponível quando se tenta utilizar dados existentes não planejados para ST. Porém se ao iniciarmos o experimento já tivermos em mente a utilização de modelos de séries temporais, o espacejamento igual no tempo aparece como um caminho natural no plano amostral.

Notação para os modelos de séries temporais

Z_t - Valor observado da ST no momento t

h - Horizonte de projeção (número de passos a ser projetado)

$\hat{Z}_t(h)$ - Projeção feita no ponto de projeção t para o horizonte h

$\hat{Z}_t$ - Projeção para o momento t feita no momento anterior, o mesmo que $\hat{Z}_{t-1}(1)$

ε_t - Variável aleatória que representa o erro de projeção no momento, suporemos ainda que esta tem média 0 e variância constante σ^2

A generalização de modelo de ST seria:

$$Z_t = f(Z_{t-1}; Z_{t-2}; Z_{t-3}; \ldots; Z_1) + \varepsilon_t \qquad \text{ou ainda} \qquad Z_t = \hat{Z}_{t-1}(1) + \varepsilon_t$$

Séries estacionárias

Diremos que uma ST apresenta um comportamento estacionário se esta, ao longo do período em estudo, se comporta como uma flutuação aleatória em torno de um nível constante que independe da passagem do tempo (Fig. 1), em termos de modelo temos:

$$Z_t = \mu + \varepsilon_t$$

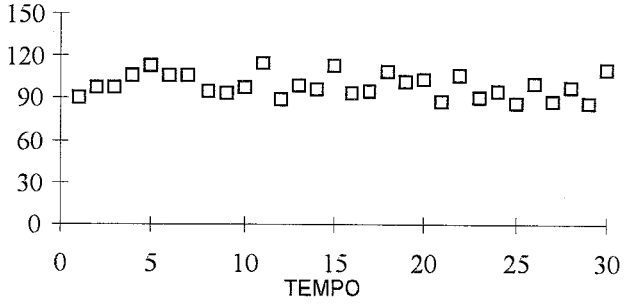


Fig. 1. Representa-ção gráfica de uma série estacionária. O eixo das ordenadas representa uma variável qualquer de interesse.

No caso da Fig. 1 temos:

$$Z_t = 100 + \varepsilon_t \quad \text{com} \quad E(\varepsilon_t) = 0\,,\ V(\varepsilon_t) = 75 \text{ e } COV(\varepsilon_i;\varepsilon_j) = 0 \text{ para } i \neq j$$

Observe que o caso apresentado acima é totalmente fantasioso pois sabemos, a priori, que a forma funcional da série, em casos práticos, é vedado a lei de formação da série, porém, ao traçarmos um gráfico da mesma, podemos ter uma boa noção se estamos estudando uma ST estacionária ou não. Apresentaremos mais tarde um teste estatístico de significância que nos ajudará a decidir sobre a presença de estacionaridade ou tendência da ST.

Amortecimento exponencial simples (AES)

O modelo AES apresenta a idéia de que o valor de uma ST no ponto t (Z_t) pode ser expresso como uma média ponderada dos valores anteriores ($Z_{t-1}; Z_{t-2}; Z_{t-3}; \ldots; Z_1$), onde os pesos dos valores decrescem de forma exponencial a medida que nos afastamos para trás do ponto t. Deste modo temos:

$$\hat{Z}_t = \alpha\ Z_{t-1} + (1-\alpha)\,\hat{Z}_{t-1} \quad 0 \leq \alpha \leq 1$$

onde α é chamada de constante de amortecimento

O modelo acima pode ser expresso da seguinte forma :

$$\begin{aligned}\hat{Z}_t &= \alpha Z_{t-1} + (1-\alpha)\hat{Z}_{t-1} \\ &= \alpha Z_{t-1} + (1-\alpha)[\alpha Z_{t-2} + (1-\alpha)\hat{Z}_{t-2}] \\ &= \alpha Z_{t-1} + \alpha(1-\alpha) Z_{t-2} + (1-\alpha)^2 \hat{Z}_{t-2} \\ &= \alpha Z_{t-1} + \alpha(1-\alpha) Z_{t-2} + (1-\alpha)^2[\alpha Z_{t-3} + (1-\alpha)\hat{Z}_{t-3}] \\ &= \alpha Z_{t-1} + \alpha(1-\alpha) Z_{t-2} + \alpha(1-\alpha)^2 Z_{t-3} + (1-\alpha)^3 \hat{Z}_{t-3}\end{aligned}$$

$$\hat{Z}_t = \alpha Z_{t-1} + \alpha(1-\alpha) Z_{t-2} + \alpha(1-\alpha)^2 Z_{t-3} + \alpha(1-\alpha)^3 Z_{t-4} + \ldots + \alpha(1-\alpha)^{t-1} Z_2 + (1-\alpha)^t \hat{Z}_1$$

Da forma expandida apresentada por último podemos perceber que:

1. A projeção para o momento t é uma ponderação dos valores anteriores.
2. O peso decresce exponencialmente a medida que no afastamos do momento t.
3. Se temos uma constante de amortecimento exponencial perto de 1, estamos valorizando os pontos mais perto do ponto de projeção no modelo de projeção. Já se a constante está perto de 0, estamos valorizando os pontos mais antigos da série para a elaboração das projeções.
4. Nos resta o problema de fazer uma projeção para o momento 1 ($\hat{Z}_1$) para começarmos o processo iterativo.
5. Temos também o problema de arbitrar a constante de amortecimento exponencial α.

A primeira projeção ($\hat{Z}_1$) aparece no modelo de projeção multiplicada por $(1-\alpha)^t$, como $0 \leq \alpha \leq 1$ e o tamanho da amostra (t) é razoavelmente grande, $(1-\alpha)^t$ deve estar bem perto de zero. Por exemplo, se $\alpha=0,5$ e $t=40$, $(1-\alpha)^t=0.000000000000909$; em suma, não importa muito o valor que arbitraremos para $\hat{Z}_1$ pois quando este for multiplicado por $(1-\alpha)^t$ o resultado será praticamente zero, porém, continuamos na situação de ter que arbitrar um valor qualquer. Um bom candidato a ser $\hat{Z}_1$ é Z_1, pois já se encontra perto do valor de μ se a série é estacionária.

A constante de amortecimento α

A forma adequada de se obter o melhor valor para α é através de método numérico, não existindo uma forma analítica de otimizar a constante de amortecimento. O que faremos é uma varredura no intervalo [0,1] e, para cada ponto desta varredura calcularemos uma medida de qualidade de ajustamento do modelo. No momento em que a medida de qualidade passar pelo seu ponto ótimo, então o valor de α que gerou este ajustamento, será eleito como o melhor valor para a constante de amortecimento exponencial α.

Estimador para o erro quadrático médio(EQM) das projeções

Assim como ocorre nos modelos de regressão linear simples, a estimativa do EQM será dada pela média dos quadrados da diferença entre os valores observados e os valores projetados, ou seja, a projeção um passo a frente:

$$EQM = \frac{(Z_1 - \hat{Z}_1)^2 + (Z_2 - \hat{Z}_2)^2 + \ldots + (Z_N - \hat{Z}_N)^2}{N} = \frac{1}{N}\sum_{i=1}^{N}(Z_i - \hat{Z}_i)^2$$

Observe que o próprio EQM apresentado pode ser uma medida de qualidade para o ajustamento. Outro indicador de qualidade do ajustamento é o erro porcentual absoluto médio (EPAM). Este indicador expressa o quanto, em média, estamos errando percentualmente ao utilizarmos este modelo. O EPAM possui uma grande vantagem sobre a variância pelo fato das pessoas estarem mais acostumadas a tratar com percentuais. Ao se afirmar por exemplo, que as estimativas de um modelo trazem 3% de erro, isto é mais fácil de ser assimilado do que afirmar que um modelo para prever população de roedores possui uma variância de 103 indivíduos ao quadrado. O EPAM é definido por:

$$EPAM = 100\frac{\sum_{t=1}^{N}\frac{|Z_t - \hat{Z}_t|}{|Z_t|}}{N}$$

Outra medida de qualidade para o ajustamento é o erro absoluto médio (EAM). A desvantagem do EAM em relação ao EPAM, é que o EAM é influenciado pela unidade de medida da série temporal. Por exemplo, um EAM de 100 Kg pode ser muito grande se estamos estimando o peso de seres humanos, porém, se a estimação trata da produção de cereais de um país das dimensões do Brasil, um erro de 100 Kg é infinitamente pequeno. Embora o EAM tenha menos atrativos que o EPAM ele será de grande valia na construção de intervalos de confiança para as projeções. O EAM será definido como :

$$EAM = \frac{\sum_{t=1}^{N}|Z_t - \hat{Z}_t|}{N} = \Delta$$

Intervalo de confiança para as projeções

O erro absoluto na projeção feita no tempo t para h passos a frente, erro de $\hat{Z}_t(h)$, a um nível de confiança (100-γ), será dado por:

$$EP_N^{100-\gamma}(h) = z_{\gamma/2} d_h \Delta$$

Onde : $z_{\gamma/2}$ é o escore da distribuição normal padrão para $\mathbf{P}(N(0,1) > z_{\gamma/2}) = \gamma / 2$

Δ é o erro absoluto médio e $d_h = 1{,}25$

Logo o intervalo de confiança para uma projeção feita no tempo t para h passos a frente será dado por :

$$IC_N^{(100-\gamma)}(h) = \{\hat{Z}_N(h) - EP_N^{(100-\gamma)}(h); \hat{Z}_N(h) + EP_N^{(100-\gamma)}(h)\}$$

Para maiores detalhes ver Bowerman & O'Connell (1979).

A Tabela 1 apresenta um exemplo de aplicação do modelo de AES e os erros associados. A Tabela 2 mostra os intervalos de confiança das projeções.

Tabela 1. Exemplo de aplicação de um modelo AES.

SÉRIE TEMPO	AMORTEC	ERRO 1 PASSO	ERRO QUAD	ERRO ABSOL	ERR % ABS
Z_t	$\hat{Z}_{t-1}(1)$	$\hat{Z}_{t-1}(1) - Z_t$	$\left(\hat{Z}_{t-1}(1) - Z_t\right)^2$	$\left\|\hat{Z}_{t-1}(1) - Z_t\right\|$	$100 * \frac{\left\|\hat{Z}_{t-1}(1) - Z_t\right\|}{\left\|Z_t\right\|}$
300	300,00	0,00	0,00	0,00	0,00
282	300,00	18,00	324,00	18,00	6,38
315	299,82	-15,18	230,43	15,18	4,82
290	299,97	9,97	99,44	9,97	3,44
302	299,87	-2,13	4,53	2,13	0,70
309	299,89	-9,11	82,93	9,11	2,95
280	299,98	19,98	399,38	19,98	7,14
315	299,78	-15,22	231,51	15,22	4,83
287	299,94	12,94	167,36	12,94	4,51
300	299,81	-0,19	0,04	0,19	0,06
318	299,81	-18,19	330,90	18,19	5,72
310	299,99	-10,01	100,18	10,01	3,23
320	300,09	-19,91	396,36	19,91	6,22
288	300,29	12,29	151,05	12,29	4,27
291	300,17	9,17	84,04	9,17	3,15
298	300,08	2,08	4,31	2,08	0,70
306	300,06	-5,94	35,34	5,94	1,94
321	300,11	-20,89	436,20	20,89	6,51
290	300,32	10,32	106,57	10,32	3,56
282	300,22	18,22	331,97	18,22	6,46
285	300,04	15,04	226,14	15,04	5,28
310	299,89	-10,11	102,26	10,11	3,26
299	299,99	0,99	0,98	0,99	0,33
288	299,98	11,98	143,49	11,98	4,16
300	299,86	-0,14	0,02	0,14	0,05
312	299,86	-12,14	147,37	12,14	3,89
288	299,98	11,98	143,56	11,98	4,16
320	299,86	-20,14	405,54	20,14	6,29
282	300,06	18,06	326,28	18,06	6,41
279	299,88	20,88	436,09	20,88	7,48

No nosso exemplo da Tabela 1 os parâmetros do modelo e erros associados são:

$$\alpha=0,01;\ \hat{Z}_1 = 300;\ z_{\gamma/2}=1,96\ (\gamma=5\%);$$

EQM=181,61; EAM=11,17; EPAM=3,93;

Tabela 2. Intervalo de confiança par a as projeções do exemplo de AES da Tabela 1.

Horizonte projetado	Limite inferior	Projeções	Limite superior
31 (h=1)	271,20	299,88	328,56
32 (h=2)	271,20	299,88	328,56
33 (h=3)	271,20	299,88	328,56
34 (h=4)	271,20	299,88	328,56

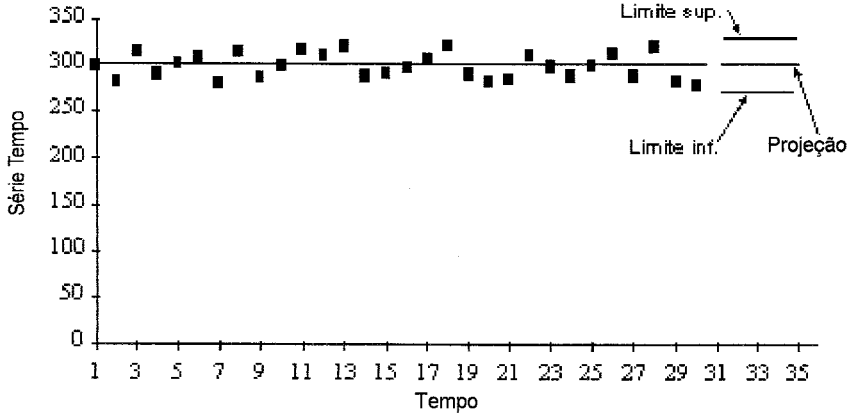


Fig. 2. Projeções do exemplo apresentado na Tabela 1. Série Tempo representa a variável em estudo Z_t.

Como já era esperado as projeções representam uma linha paralela ao eixo do tempo, lembre-se que este modelo supõe que a ST é composta por um nível constante e uma flutuação aleatória (Fig. 2). A proção representa o nível, e o intervalo de confiança considera a flutuação aleatória.

Séries que apresentam tendência linear

Estudaremos agora as séries que apresentam tendência linear crescente ou decrescente. O ambiente onde se vive este tipo de série é aquele onde o nível da série cresce ou decresce com o passar do tempo com incremento aproximadamente constante. Do ponto de vista de modelagem estatística este tipo de série pode ser expressa como:

$$Z_t = \beta_0 + \beta_1 t + \varepsilon_t$$

Note que o modelo apresentado acima em nada difere do modelo usado em regressão linear simples, a grande diferença se fará na forma de se estimar os parâmetros β_0 e β_1. No modelo de regressão estes parâmetros independem do tempo, já no modelo de série temporal as estimativas de β_0 e β_1 se atualizarão a cada ponto no tempo. Sobre a v.a. ε_t são feitas as mesmas hipóteses do modelo AES. Graficamente, temos a imagem apresentada na Fig. 3.

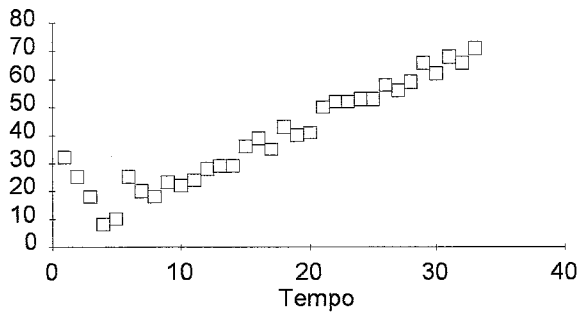


Fig. 3. Exemplo de uma série temporal com tendência linear. O eixo das ordenadas representa uma variável qualquer de interesse.

Repare que na fase inicial da série não fica clara a tendência linear, apenas com o passar do tempo ela aparece e daí por diante se confirma cada vez mais. A simulação foi feita desta forma para que possamos mais tarde entender a diferença entre o modelo de série temporal e uma regressão linear simples, além de entendermos um pouco melhor como funciona a constante de amortecimento exponencial.

Amortecimento exponencial duplo (AED)

O modelo AED será definido como:

$$\hat{Z}_t(h) = \beta_{(0,t)} + \beta_{(1,t)}h$$

Aqui começa a diferença fundamental entre o modelo de ST e o de regressão, perceba que os parâmetros agora são uma função do tempo. Vejamos então o processo de estimação dos parâmetros envolvidos:

$$\beta_{(0,t)} = 2Z_t^{[1]} - Z_t^{[2]}; \quad \beta_{(1,t)} = \frac{\alpha}{1-\alpha}(Z_t^{[1]} - Z_t^{[2]})$$

onde :

$$Z_t^{[1]} = \alpha Z_{t-1} + (1-\alpha)Z_{t-1}^{[1]}; \quad Z_t^{[2]} = \alpha Z_{t-1}^{[1]} + (1-\alpha)Z_{t-1}^{[2]}$$

Note que $Z_t^{[1]}$ representa o amortecimento da ST e $Z_t^{[2]}$ o amortecimento da série amortecida, por isso este modelo é chamado de amortecimento exponencial duplo.

Assim como no modelo AES, quando a constante de amortecimento ($\alpha \in [0,1]$) está próxima de 1, representa uma maior valorização dos pontos vizinhos ao ponto de previsão; valores mais perto de 0, representam uma consideração maior dos valores mais antigos.

Intervalo de confiança para as projeções do modelo de AED

$$IC_{\mathrm{N}}^{(100-\gamma)}(h) = \{\hat{Z}_{\mathrm{N}}(h) - EP_{\mathrm{N}}^{(100-\gamma)}(h); \hat{Z}_{\mathrm{N}}(h) + EP_{\mathrm{N}}^{(100-\gamma)}(h)\}$$

onde:

$$EP_N^{100-\gamma}(h) = z_{\gamma/2} d_h \Delta$$

$z_{\gamma/2}$ é o escore da distribuição normal padrão para $\mathbf{P}(\mathrm{N}(0,1) > z_{\gamma/2}) = \gamma / 2$

$$\Delta = EAM = \frac{\sum_{\mathrm{t}=1}^{\mathrm{N}} |Z_{\mathrm{t}} - \hat{Z}_{\mathrm{t}}|}{N}$$

$$d_h = 1{,}25\left[\frac{1+\frac{\alpha}{(1+\nu)^3}\left[(1+4\nu+5\nu^2)+2\alpha(1+3\nu)h+2\alpha^2h^2\right]}{1+\frac{\alpha}{(1+\nu)^3}\left[(1+4\nu+5\nu^2)+2\alpha(1+3\nu)+2\alpha^2\right]}\right]$$

onde:

$\nu = 1 - \alpha$ e **h** é o horizonte de projeção.

Para maiores detalhes acerca do procedimento de cálculo do intervalo de confiança para as projeções de um modelo de AED ver Bowerman & O'Connell (1979).

A Tabela 3 apresenta um exemplo de aplicação do modelo de AED e os erros associados. A Tabela 4 mostra os intervalos de confiança das projeções.

Tabela 3. Exemplo de aplicação de um modelo de amortecimento exponencial duplo.

T	Z_t	$Z_t^{[1]}$	$Z_t^{[2]}$	$\beta_{(0,t)}$	$\beta_{(1,t)}$	$\hat{Z}_{t-1}(1)$	$\lvert\hat{Z}_{t-1}(1)-Z_t\rvert$	$100*\frac{\lvert Z_{t-1}(1)-Z_t\rvert}{\lvert Z_t\rvert}$
1	32	32,00	32,00	32,00	0,00			
2	25	32,00	32,00	32,00	0,00	32,00	7,00	28,00
3	18	28,71	32,00	25,42	-2,92	32,00	14,00	77,78
4	8	23,68	30,45	16,90	-6,01	22,50	14,50	181,28
5	10	16,31	27,27	5,35	-9,72	10,89	0,89	8,89
6	25	13,34	22,12	4,57	-7,78	-4,37	29,37	117,48
7	20	18,82	17,99	19,65	0,73	-3,21	23,21	116,05
8	18	19,38	18,38	20,37	0,88	20,39	2,39	13,25
9	23	18,73	18,85	18,61	-0,11	21,25	1,75	7,61
10	22	20,74	18,79	22,68	1,72	18,50	3,50	15,90
11	24	21,33	19,71	22,95	1,44	24,40	0,40	1,68
12	28	22,59	20,47	24,70	1,88	24,39	3,61	12,88
13	29	25,13	21,46	28,80	3,25	26,58	2,42	8,36
14	29	26,95	23,19	30,71	3,34	32,05	3,05	10,51
15	36	27,91	24,96	30,87	2,62	34,05	1,95	5,43
16	39	31,71	26,35	37,08	4,76	33,49	5,51	14,12
17	35	35,14	28,87	41,41	5,56	41,84	6,84	19,55
18	43	35,07	31,82	38,33	2,89	46,97	3,97	9,23
19	40	38,80	33,35	44,25	4,84	41,22	1,22	3,05
20	41	39,36	35,91	42,82	3,06	49,09	8,09	19,72
21	50	40,13	37,53	42,73	2,31	45,88	4,12	8,24
22	52	44,77	38,75	50,79	5,33	45,04	6,96	13,39
23	52	48,17	41,58	54,75	5,84	56,12	4,12	7,92
24	53	49,97	44,68	55,26	4,69	60,60	7,60	14,33
25	53	51,39	47,16	55,62	3,75	59,95	6,95	13,12
26	58	52,15	49,15	55,15	2,66	59,37	1,37	2,37
27	56	54,90	50,56	59,24	3,85	57,80	1,80	3,22
28	59	55,42	52,60	58,23	2,50	63,08	4,08	6,92
29	66	57,10	53,92	60,28	2,82	60,73	5,27	7,98
30	62	61,28	55,42	67,15	5,20	63,10	1,10	1,77
31	68	61,62	58,17	65,07	3,06	72,35	4,35	6,40
32	66	64,62	59,79	69,44	4,28	68,12	2,12	3,22
33	71	65,27	62,06	68,47	2,84	73,72	2,72	3,83

No nosso exemplo da Tabela 3 os parâmetros do modelo e erros associados são:

$$\alpha=0{,}01;\ z_{\gamma/2}=1{,}96\ (\gamma=5\%);\ \text{EAM}=5{,}82;\ \text{EPAM}=23{,}86\%$$

Tabela 2. Intervalo de confiança par as projeções do exemplo de AES da Tabela 1.

Horizonte projetado	Limite inferior	Projeção	Limite superior
31 (h=1)	57,06	71,32	85,58
32 (h=2)	58,60	74,16	89,73
33 (h=3)	60,24	77,00	93,77
34 (h=4)	61,96	79,85	97,74

A Fig. 4 apresenta as projeções para o modelo com o intervalo de confiança para as estimativas.

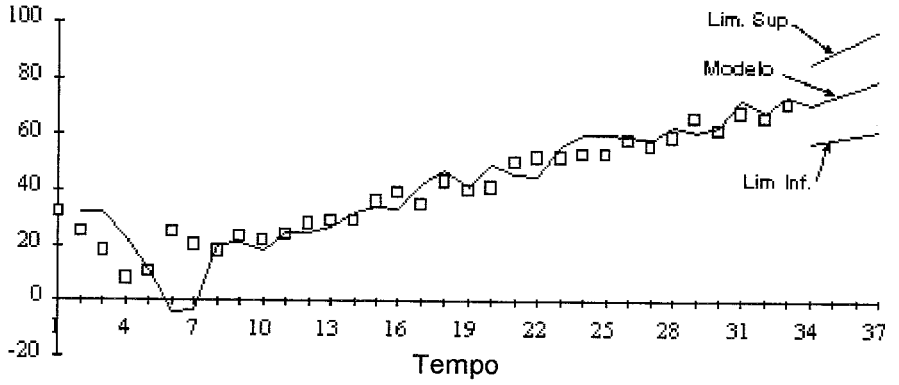


Fig. 4. Projeções do exemplo de série com tendência linear apresentada na Tabela 3. Série tempo representa a variável em estudo Z_t.

Comparando o modelo de ST com a regressão linear simples

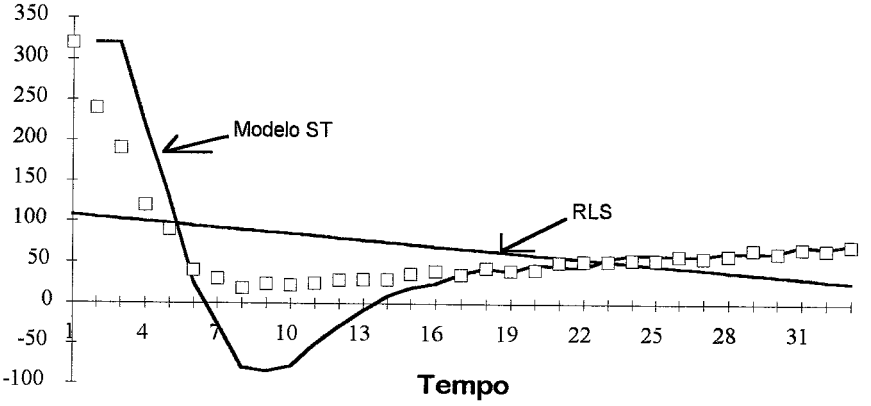


Fig. 5. Comparação entre os modelos de série temporal (ST) e de regressão linear simples (RLS). O eixo das ordenadas representa uma variável qualquer de interesse.

Observe que o modelo de regressão (Fig. 5), por não fazer distinção entre os pontos mais afastados e mais próximos do ponto de projeção, expressa uma inclinação "média" entre a inclinação inicial descendente até o momento 7, e a parte ascendente do momento 8 em diante (Abraham & Ledolter, 1983). Já o modelo de

ST, por dar mais importância aos pontos mais próximos do ponto de projeção, descreve a inclinação descendente inicial e percebe a mudança da tendência ocorrida entre os momentos 7, 8 e 9. No momento 17 o modelo de ST já se confunde com o dado observado, mostrando um rápido reconhecimento de uma nova realidade. Uma projeção feita utilizando o modelo de regressão, no caso acima, apontaria uma tendência descendente quando na verdade, a partir do momento 8, a série tem claramente um comportamento crescente.

Teste para a detecção de tendência

Em muitas situações não fica claro para o analista se a ST apresenta tendência pelo simples exame visual do gráfico. Para estes casos faz-se necessário a aplicação de um teste estatístico de significância, vejamos a seqüência de construção do teste e seu respectivo critério de decisão:

1. Uma ST de tamanho N será dividida em duas partes: a primeira com as primeiras N/2 observações, se N for par, ou (N-1)/2 se N for ímpar.
2. Se a primeira observação do primeiro grupo for menor que a primeira observação do segundo grupo, então $T_1 = 1$, caso contrário $T_1 = 0$. Repetiremos esta comparação entre a segunda observação do primeiro grupo e a segunda do segundo grupo até que se chegue a última comparação, sem perda de generalidade; suporemos N par:

$$Z_i < Z_{(N/2)+1} \Rightarrow T_i = 1 \qquad Z_i \geq Z_{(N/2)+1} \Rightarrow T_i = 0$$

3. Definiremos a estatística teste como sendo a soma dos $\mathbf{T_i}$'s, ou seja, o número de vezes que o primeiro bloco foi menor que o segundo, em termos de fórmula:

$$S = \sum_{i=1}^{N/2} T_i$$

4. Aceitando-se a hipótese nula de que a série não apresenta tendência, a estatística S tem distribuição binomial com n=N/2 e p=0,5.
5. O critério de decisão será tal que se $P(B(N/2;0,5)>S)<\alpha/2$ a série tem tendência crescente, caso $P(B(N/2;0,5)<S)<\alpha/2$ diremos que a série tem tendência decrescente, e se nenhuma dessas duas situações ocorrerem, diremos que não há evidência significativa da presença de tendência.
6. Trabalhar com a distribuição binomial pode ser muito incômodo, principalmente porque não é uma distribuição facilmente encontrada em tabelas. Propomos então trabalhar com a distribuição normal padrão e efetuarmos um teste aproximado, porém mais fácil de se operacionalizar:

$$S = B(N/2;0,5) \Rightarrow N(N/4;N/8) \quad \text{ou} \quad \frac{S - N/4}{\sqrt{N/8}} \Rightarrow N(0;1)$$

7. O critério de decisão, trabalhando com o teste aproximado, será:

a) A série tem tendência crescente se : $S > \frac{N}{4} + z_{\alpha/2}\sqrt{\frac{N}{8}}$;

b) A série tem tendência decrescente se : $S < \frac{N}{4} - z_{\alpha/2}\sqrt{\frac{N}{8}}$;

c) Não há evidência de tendência se: $\frac{N}{4} - z_{\alpha/2}\sqrt{\frac{N}{8}} \leq S \leq \frac{N}{4} + z_{\alpha/2}\sqrt{\frac{N}{8}}$

onde $z_{\alpha/2}$ é o escore da distribuição normal padrão.

Obs.: caso N seja ímpar, basta substituir N por N-1 na fórmula acima.

Para maiores detalhes ver Siegel (1975).

Vejamos um exemplo (Tabela 5):

Tabela 5. Exemplo para um teste de detecção de tendência.

Tempo	Z_t	T_t	Tempo	Z_t	T_t
1	50,80	0	21	50,66	
2	50,64	1	22	51,21	
3	50,30	1	23	50,63	
4	51,04	1	24	51,06	
5	50,43	1	25	50,85	
6	50,47	1	26	51,42	
7	50,40	1	27	50,75	
8	51,03	1	28	51,55	
9	50,32	1	29	51,17	
10	50,64	1	30	51,25	
11	50,95	1	31	51,14	
12	50,51	1	32	51,48	
13	51,18	0	33	50,73	
14	51,22	1	34	51,42	
15	50,31	1	35	50,85	
16	50,41	1	36	51,14	
17	50,91	1	37	51,00	
18	51,20	0	38	50,82	
19	51,11	1	39	51,41	
20	50,74	1	40	51,02	

A forma grafica é apresentada na Fig. 6.

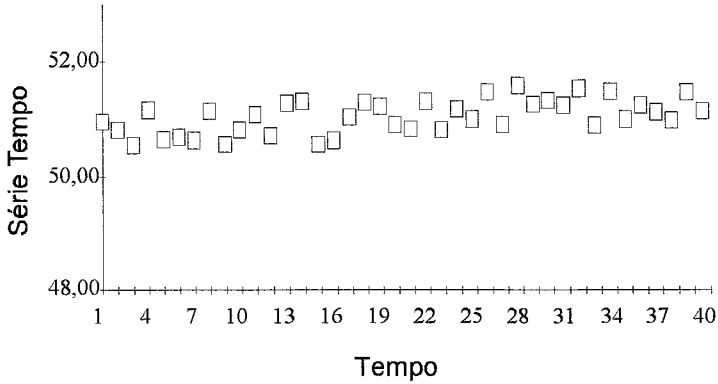


Fig. 6. Projeções do exemplo de série temporal apresentada na Tabela 5. Série tempo representa a variável em estudo Z_t.

S=17, como $S > \frac{40}{4} + 1{,}96\sqrt{\frac{40}{8}} = 14{,}3$ existe evidência estatística da existência de tendência ao nível de significância de 5%, mais precisamente, existe tendência crescente ao nível de significância de 2,5%.

Obs.: Caso N<40 devemos usar o teste original com a distribuição binomial, pois a aproximação da binomial pela normal não é aconselhável para menos de 20 pares.

Séries sazonais

Ao estudarmos o desenvolvimento de um fenômeno no tempo podemos nos deparar com a situação de uma ST que se "repete" em intervalos fixos de tempo, apresentando um padrão de comportamento como por exemplo, a variação de temperatura ao longo do ano. Neste caso temos sempre os dias mais frios no inverno e os mais quentes no verão, e mais, a cada período de um ano (período fixo) o padrão se repete. Além da questão do padrão repetitivo podemos também ter a presença de uma tendência crescente ou decrescente. Voltando ao caso das estações do ano podemos ter as variações de temperatura relativas às mudanças de estações associadas a um aumento da temperatura média anual, que representaria uma tendência crescente. Todas ST que apresentam este padrão repetitivo serão chamadas de séries temporais sazonais (STS), e a este padrão de repetição chamaremos de efeito sazonal.

Sazonalidade aditiva e sazonalidade multiplicativa

Diremos que a STS é aditiva (STSA) se o efeito sazonal se adiciona ao nível da série. Por exemplo, se temos sempre um aumento de 5°C, por ser dezembro independente do nível da série, então temos uma STS aditiva. Já se o efeito sazonal é

uma proporção do nível, por exemplo, a temperatura sempre aumenta 10% por ser dezembro, então temos uma STS multiplicativa (STSM), vejamos os gráficos apresentados na Fig. 7.

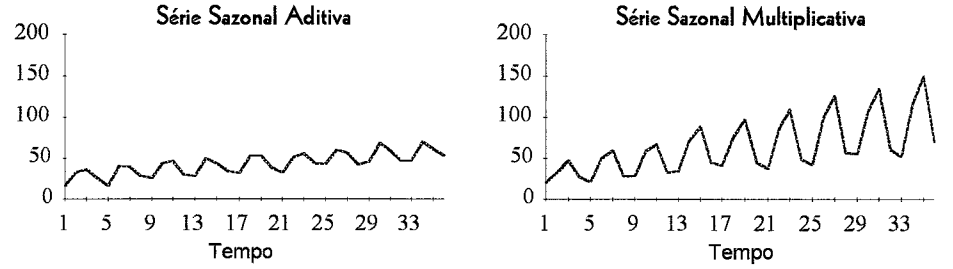


Fig. 7. Exemplo de uma série temporal aditiva e uma multiplicativa. As ordenadas representam uma variável qualquer de interesse.

Definiremos como tamanho do período sazonal a quantidade de momentos onde se forma um padrão sazonal, por exemplo, nos dois gráficos acima podemos perceber que o período sazonal tem tamanho 4. Denotaremos o tamanho do período sazonal pela letra s, $\hat{Z}_t(h) = (\beta_{(0,t)} + \beta_{(1,t)}h)M_{t+h-s}$ no caso acima s=4.

Em termos de modelagem estatística uma STSA pode ser expressa como:

$$Z_t = \beta_0 + \beta_1 t + A_t + \varepsilon_t$$

onde A_t é o efeito sazonal aditivo. Já para a STSM o modelo é definido como:

$$Z_t = (\beta_0 + \beta_1 t)M_t + \varepsilon_t$$

onde M_t é o efeito sazonal multiplicativo.

Método de Winters para séries temporais sazonais multiplicativas (STSM)

O modelo de Winter para séries STSM é definido por :

$$\hat{Z}_t(h) = (\beta_{(0,t)} + \beta_{(1,t)}h)M_{t+h-s}$$

Com os seguintes estimadores :

$$\beta_{(0,t)} = \alpha\frac{Z_t}{M_{t-s}} + (1-\alpha)(\beta_{(0,t-1)} + \beta_{(1,t-1)});$$

$$\beta_{(1,t)} = \lambda(\beta_{(0,t)} - \beta_{(0,t-1)}) + (1-\lambda)\beta_{(1,t-1)}; \quad M_t = \gamma\frac{Z_t}{\beta_{(0,t)}} + (1-\gamma)M_{t-s}$$

Onde α, λ, e γ são as constantes de amortecimento exponencial, e devem ser otimizadas exatamente da mesma forma que nos modelos anteriores. Note que o processo de estimação tem uma seqüência natural, já que os estimadores são interdependentes, vejamos então a seqüência de estimação :

1) Estimar o efeito sazonal para um período sazonal inteiro.
2) Estimar $\beta_{(0,t)}$ e $\beta_{(1,t)}$ para o próximo período sazonal.
3) Estimar o efeito sazonal para o próximo período.
4) Repetir 2) e 3) até o último período sazonal.

Valores iniciais. Como o processo de estimação é iterativo e depende do passado, passamos a ter o problema do ponto de partida, ou seja como obter os primeiros valores para que estes sejam revistos a cada momento no tempo. Suponha que a amostra é composta de k períodos sazonais de tamanho s cada, $\overline{Z}_k$ é a média da ST para o último período sazonal e $\overline{Z}_1$ a média do primeiro período sazonal. Propomos estão a seguinte regra para obter os valores iniciais, que representa uma pequena modificação simplificadora, com relação ao apresentado por Johnson & Montgomery (1976) e Winters (1960):

$$\beta_{(1,s)} = \frac{\overline{Z}_k - \overline{Z}_1}{(k-1)s}\ ;\ \beta_{(0,s)} = \overline{Z}_1 + \frac{s}{2}\beta_{(1,s)}$$

$$M_t = \frac{Z_t}{\beta_{(0,s)} - (s-t)\beta_{(1,s)}} \qquad \text{com} \qquad t=1,2,...,s$$

Intervalo de confiança para as projeções

Usaremos uma aproximação para o intervalo de confiança das projeções proposta em Bowerman & O'Connell (1979). A intuição por trás da fórmula a ser apresenta se resume a retirar o efeito sazonal da série, e estudar o erro existente entre a série analisada e a projeção utilizando apenas o modelo de tendência; a partir deste ponto, o intervalo de confiança recai no caso anterior do modelo com tendência linear.

O intervalo de confiança, aproximado, para as projeções será definido como:

$$IC_{\mathrm{N}}^{(100-\gamma)}(h) = \{\hat{Z}_{\mathrm{N}}(h) - EP_{\mathrm{N}}^{(100-\gamma)}(h); \hat{Z}_{\mathrm{N}}(h) + EP_{\mathrm{N}}^{(100-\gamma)}(h)\}$$

onde:

$$EP_N^{100-\gamma}(h) = z_{\gamma/2} d_h \Delta_{WM}$$

$z_{\gamma/2}$ é o escore da distribuição normal padrão para $\mathbf{P}(N(0,1) > z_{\gamma/2}) = \gamma / 2$

$$\Delta_{WM} = EAM_{ds} = \frac{\sum_{t=s+1}^{N} \left| \frac{Z_t}{M_{t-s}} - (\beta_{(0,t-1)} + \beta_{(1,t-1)}) \right|}{N - s}$$

$$d_h = 1{,}25 \left[\frac{1 + \frac{\theta}{(1+\theta)^3}\left[(1+4\nu+5\nu^2) + 2\theta(1+3\nu)h + 2\theta^2 h^2\right]}{1 + \frac{\theta}{(1+\theta)^3}\left[(1+4\nu+5\nu^2) + 2\theta(1+3\nu) + 2\theta^2\right]} \right]$$

$\nu = 1 - \theta$, h é o horizonte de projeção e $\theta = MAX\{\alpha; \lambda; \gamma\}$

Um exemplo de aplicação do modelo de Winters (modificado) para STSM é apresentado na Tabela 6. O intervalo de confiança para as projeções (4 primeiras) é apresentado na Tabela 7. A Tabela 8 mostra d_h e os erros padrões associados para os 4 primeiros horizontes de projeção.

Tabela 6. Exemplo de aplicação do método de Winters (modificado) para um modelo de STSM.

T	Z_t	$\beta_{(0,t)}$	$\beta_{(1,t)}$	M_t	$\hat{Z}_{t-1}(1)$	$\frac{Z_t}{M_{t-4}} - (\beta_{(0,t-1)} + \beta_{(1,t-1)})$	$100 * \frac{\lvert \hat{Z}_{t-1}(1) - Z_t \rvert}{\lvert Z_t \rvert}$
1	20,31			0,67			
2	32,74			1,02			
3	47,82			1,40			
4	27,91	36,27	2,04	0,77			
5	21,39	38,31	2,04	0,60	25,81	6,56	20,66
6	49,69	40,35	2,04	1,15	41,02	8,52	17,45
7	59,83	42,39	2,04	1,41	59,21	0,44	1,03
8	28,74	44,42	2,04	0,70	34,19	7,08	18,94
9	27,86	46,46	2,04	0,60	28,08	0,36	0,78
10	59,16	48,50	2,04	1,19	55,56	3,15	6,09
11	67,21	50,53	2,04	1,36	71,03	2,72	5,68
12	33,04	52,56	2,03	0,66	36,59	5,09	10,73
13	35,22	54,60	2,04	0,63	32,85	3,94	6,74
14	71,86	56,64	2,04	1,24	67,41	3,74	6,20
15	89,10	58,69	2,04	1,46	79,82	6,82	10,42
16	45,35	60,74	2,04	0,71	39,81	8,45	12,21
17	41,66	62,78	2,05	0,65	39,41	3,60	5,42
18	77,52	64,83	2,04	1,21	80,21	2,18	3,48
19	97,73	66,87	2,04	1,46	97,30	0,30	0,44
20	44,77	68,91	2,04	0,67	48,95	5,88	9,34
21	37,65	70,94	2,04	0,58	46,06	12,96	22,35

Tabela 6. (Continuação)

22	86,21	72,97	2,04	1,19	88,47	1,86	2,62
23	110,52	75,01	2,04	1,47	109,43	0,74	0,98
24	49,39	77,04	2,04	0,65	51,92	3,76	5,12
25	42,67	79,07	2,03	0,56	45,72	5,27	7,13
26	102,16	81,11	2,03	1,23	96,83	4,47	5,22
27	126,82	83,15	2,04	1,50	122,02	3,27	3,79
28	57,39	85,19	2,04	0,67	55,73	2,54	2,90
29	55,21	87,24	2,04	0,60	48,42	12,24	12,30
30	108,43	89,28	2,04	1,22	110,10	1,36	1,55
31	135,30	91,32	2,04	1,49	137,18	1,25	1,39
32	60,85	93,36	2,04	0,66	62,17	1,98	2,16
33	52,70	95,39	2,04	0,57	57,40	7,82	8,93
34	117,04	97,42	2,04	1,21	119,05	1,65	1,72
35	150,03	99,46	2,04	1,50	148,18	1,24	1,24
36	69,95	101,50	2,04	0,68	66,73	4,90	4,61

Tabela 7. Intervalo de confiança para as projeções do exemplo de STSM pelo método de Winters (modificado) apresentado na Tabela 6.

Horizonte projetado	Lim Inferior	Projeção	Lim Superior
37 (h=1)	49,03	59,24	69,45
38 (h=2)	115,53	127,71	139,89
39 (h=3)	147,19	161,54	175,88
40 (h=4)	57,55	74,18	90,81

Tabela 8. 4 primeiros horizontes de projeção com seus d_h e erros padrões associados.

h	d_h	$EP_N^{95\%}(h)$
1	1,25	10,21
2	1,49	12,18
3	1,76	14,35
4	2,04	16,63

Os parâmetros, erros e valores inicais são apresentados abaixo:

$\beta_{(1,4)}$	2,04	k	9
$\beta_{(0,4)}$	36,27	s	4
α	0,001	$\overline{Z}_9$	98,44
λ	0,4	$\overline{Z}_1$	32,20
γ	0,6	$\theta=\gamma$	0,6
EAM_{ds}	4,25		
EPAM	6,86		

A Fig. 8 apresenta as projeções para o modelo com o intervalo de confiança para as estimativas.

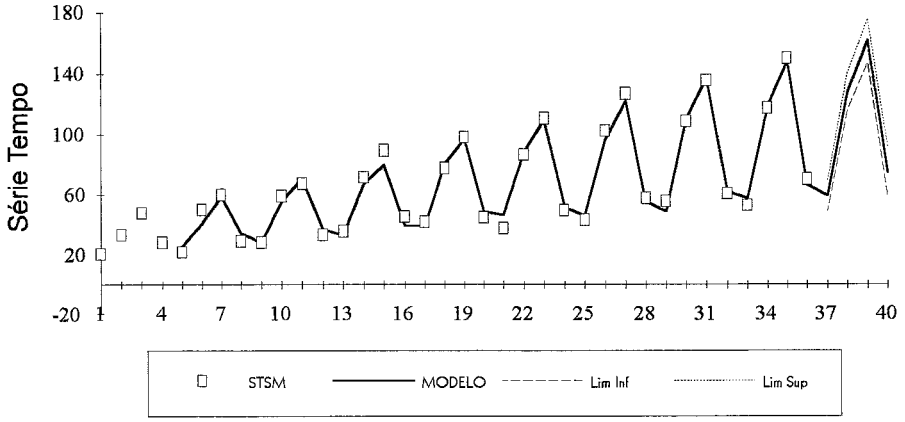


Fig. 8. Projeções do exemplo de série temporal sazonal multiplicativa apresentada na Tabela 6. Série tempo representa a variável em estudo Z_t.

Método de Winters para séries temporais sazonais aditivas (STSA).

O modelo que veremos agora é muito parecido com o modelo STSM, exceto que lá o efeito sazonal multiplica o nível da série, funcionando como uma taxa de crescimento ou desconto. Aqui, nos modelos aditivos, o efeito sazonal independe do nível da série. Por consequência desta diferença conceitual, onde tínhamos a operação de multiplicação passaremos a ter a soma, e, onde tínhamos a divisão passaremos a ter a subtração.

O modelo de Winters para séries STSA é definido por :

$$\hat{Z}_t(h) = \beta_{(0,t)} + \beta_{(1,t)}h + A_{t+h-s}$$

Com os seguintes estimadores :

$$\beta_{(0,t)} = \alpha(Z_t - A_{t-s}) + (1-\alpha)(\beta_{(0,t-1)} + \beta_{(1,t-1)})$$

$$\beta_{(1,t)} = \lambda(\beta_{(0,t)} - \beta_{(0,t-1)}) + (1-\lambda)\beta_{(1,t-1)}$$

$$M_t = \gamma(Z_t - \beta_{(0,t)}) + (1-\gamma)A_{t-s}$$

Valores iniciais:

$$\beta_{(1,s)} = \frac{\overline{Z}_k - \overline{Z}_1}{(k-1)s}\ ;\ \beta_{(0,s)} = \overline{Z}_1 + \frac{s}{2}\beta_{(1,s)}$$

$$A_t = Z_t - [\beta_{(0,s)} - (s-t)\beta_{(1,s)}] \qquad \text{com} \qquad t=1,2,...,s$$

Intervalo de confiança para as projeções.

A única diferença do modelo multiplicativo para o aditivo, no que diz respeito ao intervalo de confiança, é o cálculo de Δ_{WM}, que será chamado de Δ_{WA} e será definido por :

$$\Delta_{WA} = EAM_{ds} = \frac{\sum_{t=s+1}^{N} \left| (Z_t - M_{t-s}) - (\beta_{(0,t-1)} + \beta_{(1,t-1)}) \right|}{N-s}$$

Um exemplo de aplicação do modelo de Winters para STSA é apresentado na Tabela 9. O intervalo de confiança para as projeções (4 primeiras) é apresentado na Tabela 10. A Tabela 11 mostra d_h e os erros padrões associados para os 4 primeiros horizontes de projeção.

Tabela 9. Exemplo de aplicação do método de Winters para um modelo STSA.

T	Z_t	$\beta_{(0,t)}$	$\beta_{(1,t)}$	M_t	$\hat{Z}_{t-1}(1)$	$\left\| \frac{Z_t}{M_{t-4}} - (\beta_{(0,t-1)} + \beta_{(1,t-1)}) \right\|$	$100 * \frac{\left\| \hat{Z}_{t-1}(1) - Z_t \right\|}{\left\| Z_t \right\|}$
1	16,19			-10,80			
2	33,00			5,06			
3	36,95			8,07			
4	25,60	29,82	0,95	-4,23			
5	16,58	30,77	0,95	-13,17	19,97	20,46	3,39
6	40,81	31,72	0,95	7,88	36,78	9,87	4,03
7	40,11	32,66	0,95	7,63	40,73	1,55	0,62
8	28,66	33,61	0,95	-4,73	29,38	2,53	0,73
9	25,75	34,56	0,95	-10,11	21,38	16,98	4,37
10	43,95	35,50	0,95	8,28	43,38	1,29	0,57
11	46,84	36,45	0,95	9,56	44,08	5,89	2,76
12	30,64	37,39	0,95	-6,15	32,66	6,61	2,03
13	28,20	38,34	0,95	-10,14	28,23	0,11	0,03
14	50,27	39,29	0,95	10,17	47,57	5,37	2,70
15	43,82	40,23	0,95	5,38	49,80	13,65	5,98
16	34,31	41,17	0,95	-6,65	35,02	2,09	0,72
17	31,69	42,12	0,95	-10,34	31,98	0,94	0,30
18	53,54	43,07	0,95	10,38	53,23	0,57	0,31
19	53,32	44,01	0,95	8,13	49,39	7,36	3,93

Tabela 9. (continuação)

20	38,05	44,96	0,95	-6,83	38,31	0,69	0,26
21	32,44	45,90	0,95	-12,53	35,56	9,63	3,12
22	51,71	46,84	0,95	6,52	57,23	10,67	5,52
23	56,13	47,79	0,95	8,28	55,92	0,39	0,22
24	43,62	48,74	0,95	-5,63	41,90	3,94	1,72
25	43,16	49,69	0,95	-8,33	37,15	13,92	6,01
26	60,37	50,64	0,95	8,77	57,16	5,31	3,21
27	56,58	51,58	0,95	5,98	59,86	5,81	3,29
28	42,00	52,52	0,95	-9,05	46,89	11,64	4,89
29	46,06	53,47	0,95	-7,68	45,14	1,99	0,92
30	68,92	54,42	0,95	12,78	63,18	8,33	5,74
31	59,54	55,36	0,95	4,72	61,35	3,04	1,81
32	46,78	56,31	0,95	-9,39	47,26	1,02	0,48
33	47,35	57,25	0,95	-9,24	49,57	4,69	2,22
34	70,72	58,20	0,95	12,60	70,98	0,37	0,26
35	61,81	59,14	0,95	3,28	63,86	3,31	2,05
36	52,90	60,09	0,95	-7,84	50,70	4,16	2,20

Tabela 10. Intervalo de confiança para as projeções do exemplo de STSA pelo método de Winters apresentado na Tabela 9.

Horizonte Projetado	Limite Inferior	Projeção	Limite Superior
37 (h=1)	45,95	51,80	57,64
38 (h=2)	66,25	73,63	81,02
39 (h=3)	55,27	64,32	73,37
40 (h=4)	42,40	53,19	63,98

Tabela 11. 4 primeiros horizontes de projeção com seus d_h e erros padrões associados

h	d_h	$EP_N^{95\%}(h)$
1	1,25	5,85
2	1,58	7,38
3	1,93	9,05
4	2,31	10,79

Os parâmetros, erros e valores iniciais são apresentados abaixo:

$\beta_{(1,4)}$	0,95	k	9
$\beta_{(0,4)}$	29,82	s	4
α	0,001	$\overline{Z}_9$	58,19
λ	0,01	$\overline{Z}_1$	27,93
γ	0,7	$\theta=\gamma$	0,7
EAM_{ds}	2,39		
EPAM	5,76		

A Fig. 9 apresenta as projeções para o modelo com o intrervalo de confiança para as estimativas.

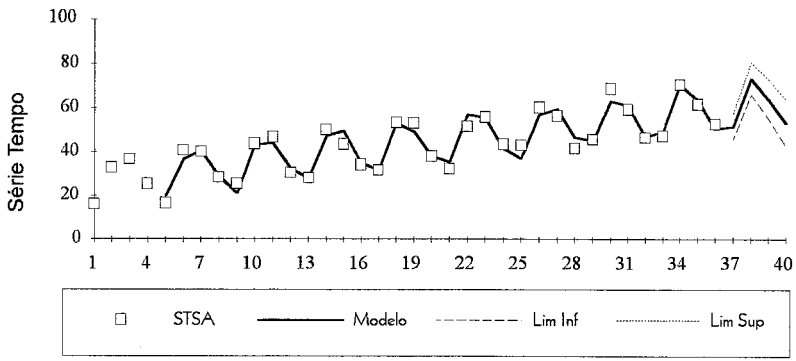


Fig. 9. Projeções do exemplo de série temporal sazonal aditiva apresentada na Tabela 9. Série tempo representa a variável em estudo Z_t.

Teste para a detecção de efeito sazonal

O teste estatístico de significância que apresentaremos é uma aplicação, em séries temporais, do teste não-paramétrico de Kruskal-Wallis (Siegel, 1975). Aqui usaremos o mesmo conceito intuitivo da análise da variância. Suponha que os dados da STSA apresentada no exemplo seja a produção trimestral de um tomate. O teste dirá se há diferença entre os trimestres, se houver é porque existe um efeito sazonal fruto de termos trimestres diferentes. Vejamos uma aplicação utilizando os dados do modelo de STSA.

Construção do teste:

1) Dispor a série temporal numa tabela cujas linhas representam os anos e a linhas os trimestres por exemplo, ou qualquer outra unidade temporal.

	Trim 1	Trim 2	Trim 3	Trim 4
Ano 1	16,19	33,00	36,95	25,60
Ano 2	16,58	40,81	40,11	28,66
Ano 3	25,75	43,95	46,84	30,64
Ano 4	28,20	50,27	43,82	34,31
Ano 5	31,69	53,54	53,32	38,05
Ano 6	32,44	51,71	56,13	43,62
Ano 7	43,16	60,37	56,58	42,00
Ano 8	46,06	68,92	59,54	46,78
Ano 9	47,35	70,72	61,81	52,90

2) Os N valores da ST são substituídos pelos suas posições relativas (postos) e em toma-se a soma dos postos de cada coluna. Chamaremos o posto de cada célula de R_{ij}, onde i representa a linha da célula e j a coluna. À soma dos postos de uma coluna chamaremos de R_{*j}.

	Trim 1	Trim 2	Trim 3	Trim 4
Ano 1	1	10	12	3
Ano 2	2	15	14	6
Ano 3	4	20	23	7
Ano 4	5	25	19	11
Ano 5	8	29	28	13
Ano 6	9	26	30	18
Ano 7	17	33	31	16
Ano 8	21	35	32	22
Ano 9	24	36	34	27
R_{*j}	91	229	223	123

$$R_{*j} = \sum_{i=1}^{k} R_{ij}$$

3) A estatística teste será definida como :

$$H = \frac{12}{N(N+1)} \sum_{j=1}^{s} \frac{R_{*j}^2}{k} - 3(N+1)$$

4) Sob a hipótese nula de que não há sazonalidade, H tem distribuição Qui-Quadrado com s-1 graus de liberdade. Voltando ao exemplo :

	Trim 1	Trim 2	Trim 3	Trim 4
R_{*j}^2	8281	52441	49729	15129
$\frac{R_{*j}^2}{k}$	920.11	5826.78	5525.44	1681.00

$$\sum_{j=1}^{k} \frac{R_{*j}^2}{k} = 13.953,33$$

logo:

$$H = \frac{12}{36(36+1)} 13993,33 - 3(36+1) = 15$$

Como H=15 é maior do que $\chi^2_{(3;5\%)} = 7{,}82$, conclui-se que existe evidência estatística da presença de sazonalidade na ST. Aliás, no exemplo, um simples exame do gráfico já denuncia a presença de sazonalidade, porém, em algumas situações onde o efeito sazonal não é tão flagrante o teste pode ser de grande valia para o analista da série que pode estar na dúvida entre a aplicação do modelo sazonal, ou não.

Considerações finais

Os modelos que apresentamos embora representem os primeiros passos em termos dos modelos ST, conseguem cobrir um grande número de casos. Poderíamos ainda caminhar no sentido de apresentar formas de tendência polinomial de grau maior que dois, como por exemplo, séries que crescem quadraticamente, ou como um polinômio do terceiro grau. Porém, isto já estaria além de uma abordagem introdutória. Todo o tratamento estatístico está baseado na teoria chamada "General Direct Smoothing" apresentada em Bowerman & O'Connell (1979), com adaptações, o que implica em intervalos de confiança aproximados. Para os que desejam continuar no estudo das séries temporais o caminho natural são os modelos de Box & Jenkins (1970). Do ponto de vista da qualidade das projeções não há nenhuma garantia de que estes modelos tenham um desempenho melhor do que os de amortecimento exponencial. Porém, as propriedades estatísticas dos estimadores e construção de intervalos de confiança são, do ponto de vista técnico de estatística, mais elaborados, além de cobrir numa teoria geral um grupo grande de modelos de projeção.

Referências bibliográficas

ABRAHAM, B. & J. LEDOLTER. 1983. Statistical Methods for Forecasting. John Wiley & Sons, New York.

BOLFARINE, H., J. RODRIGUES & J. ACHCAR. 1991. Análise de Sobrevivência. ABE & IM-UFRJ, Rio de Janeiro.

BOWERMAN, B.L. & R.T. O'CONNELL. 1979. Time Series and Forecasting: An Approach. Duxbury Press, Belmont.

BOX, G. E. P. & G.M. JENKINS. 1970. Time Series Analysis, Forecasting and Control. Hoden-Day, Inc., San Francisco.

JOHNSON, L.A. & D.C. MONTGOMERY. 1976. Montgomery, Forecasting and Time Series Analysis. McGraw-Hill Book Company, New York.

MORETTIN, P.A. & C.M. TOLOI. 1987. Previsão de Séries Temporais. Atual Editora Ltda, Rio de Janeiro.

SIEGEL, S. 1975. Estatística não-Paramétrica para as Ciências do Comportamento. McGraw Hill do Brasil, Rio de Janeiro.

WINTERS, P.R. 1960. Forecasting sales by exponentially weighted moving averages. *Management Science*, **6**: 324-342.

Endereço

ROGÉRIO GARBER
Xerox do Brasil - Rua Eduardo Nadruz (Rua N) 491
CEP: 21931-600 - Ilha do Governador, Rio de Janeiro - RJ.