

Capítulo 5: Introdução às Séries Temporais e aos Modelos ARIMA

Neste capítulo faremos uma introdução às séries temporais. O nosso objetivo aqui é puramente operacional e estaremos mais preocupados com as definições básicas e alguns exemplos simples, já que o estudo de séries temporais é tão extenso que justifica um curso em separado.

Séries Temporais não são comumente estudadas em cursos básicos de Estatística. Por que? Pois no estudo de Séries Temporais a noção de dependência entre as observações é crucial. Ao contrário, na maioria dos procedimentos estudados em cursos elementares, supõe-se que as observações formam uma amostra aleatória, isto é, são independentes e identicamente distribuídas.

5.1. Introdução - algumas considerações sobre séries temporais

Definição 5.1.1. (Série Temporal)

Uma *Série Temporal* é um conjunto de observações ordenadas no tempo (não necessariamente igualmente espaçadas), e que apresentam dependência serial (isto é, dependência entre instantes de tempo). A notação usada aqui para denotar uma Série Temporal é $Z_1, Z_2, Z_3, \dots, Z_T$, que indica uma série de tamanho T . O instante T geralmente indica o último instante disponível.

De uma maneira um pouco mais formal, dizemos que uma série temporal é uma **realização de um processo estocástico**.

Definição 5.1.2. (Processo ergódico)

Um **processo estocástico** é dito **ergódico** se uma única realização do processo é o suficiente para caracterizá-lo. Na análise de séries temporais existe **apenas uma** realização do processo disponível e portanto precisamos supor que o processo subjacente é ergódico, pois iremos usar apenas uma de suas realizações para caracterizá-lo.

Em geral, ao estudarmos uma Série Temporal estaremos interessados em dois aspectos:

- Análise e Modelagem** da Série Temporal - descrever a série, verificar suas características mais relevantes e suas possíveis relações com outras séries;
- Previsão** da Série Temporal - a partir de valores passados da série (e talvez de outras séries também) encontrar boas previsões (de curto prazo) de valores futuros da série. A previsão da série no instante $T + k$ será denotada por \hat{Z}_{T+k} . O número de instantes à frente para o qual é feita a previsão (neste caso, k) é chamado de **horizonte de previsão**. Por exemplo, a previsão de Z_{T+1} é denotada por \hat{Z}_{T+1} .

A dependência serial entre os valores da série é um aspecto *essencial*, pois nos permite gerar previsões de valores futuros da série. Estas previsões seriam puro “chute” se não houvesse dependência serial. Também, diferentes séries possuem diferentes “graus” de previsibilidade; por exemplo, é freqüentemente mais fácil prever uma série de temperaturas médias mensais do que a taxa mensal de inflação. Logo, não se pode garantir que a previsão encontrada por este ou aquele método será sempre “boa”, tudo depende das características da série que está sendo estudada! No entanto, um aspecto deve ser levado em conta ao fazermos previsões de séries temporais: o nível de incerteza aumenta com o horizonte de previsão – *quanto mais longe no futuro, maior é a incerteza associada à previsão*. Isto é intuitivamente razoável, é sempre mais difícil prever o futuro distante, e a nossa previsão estará cercada de incertezas!

Uma medida do “acerto” das nossas previsões é o erro de previsão k-passos à frente, definido a seguir.

Definição 5.1.3. (Erro de Previsão k passos à frente)

O erro de previsão k passos à frente no instante t (onde k é um inteiro maior ou igual a um) é definido como a diferença entre o valor real da série no instante t e a previsão deste valor feita k instantes antes, isto é:

$$e_k(t) = Z_t - \hat{Z}_{t-k}(k)$$

Um caso particular importante é o erro de previsão um passo à frente, dado por:

$$e_1(t) = Z_t - \hat{Z}_{t-1}(1)$$

Um “bom” modelo de previsão produz previsões com erro pequeno, e assim é interessante acompanhar quantidades como a soma dos quadrados dos erros de previsão, ou a soma dos valores absolutos dos erros de previsão.

E como funcionam estas ferramentas quantitativas que nos permitem prever o futuro de uma série temporal?

- Vamos utilizar o passado (dados históricos) para descrever a trajetória mais provável da série no futuro.
- **Isso não é uma bola de cristal!**
- Na maioria dos problemas o passado traz informações relevantes sobre o que irá ocorrer no futuro, pois existe “correlação” entre as variáveis em diversos instantes.
- É claro que o conhecimento do passado não nos diz exatamente como será o futuro, e então sempre existe incerteza associada às nossas previsões !
- Mas, podemos ter uma boa idéia de quais serão os valores mais prováveis no futuro !
- Ou seja, **podemos especificar previsões futuras e limites de confiança.**

Por que usar previsões quantitativas, baseadas em modelos estatísticos?

As previsões quantitativas, baseadas em modelos estatísticos, são importantes pois podem ser reproduzidas, isto é: dois analistas diferentes chegam às mesmas conclusões se utilizarem o mesmo modelo.

Também, as previsões obtidas por métodos estatísticos têm uma justificativa teórica - o resultado encontrado não é baseado apenas em argumentos do tipo : “eu acho que...”, “baseado na minha experiência ...” , etc...

Finalmente, as previsões são obtidas a partir de um modelo, que é uma simplificação da realidade. O processo de modelagem nos leva naturalmente a considerar apenas os aspectos “essenciais” do nosso problema e por si só costuma trazer bastante informação sobre o problema que está sendo tratado.

Afinal, o que queremos ao modelar uma série temporal ?

- Capturar “toda” a estrutura de dependência existente na série;
- Logo, nos resíduos não deve “sobrar” estrutura, pois ela já foi captada pelo modelo. Nota: o resíduo é apenas a diferença entre o valor real e o ajustado por um modelo qualquer. Por exemplo, seja Z_t o valor real da série no instante t, e \hat{Z}_t seu valor ajustado pelo modelo. Então, o resíduo no instante t é apenas $Z_t - \hat{Z}_t$.

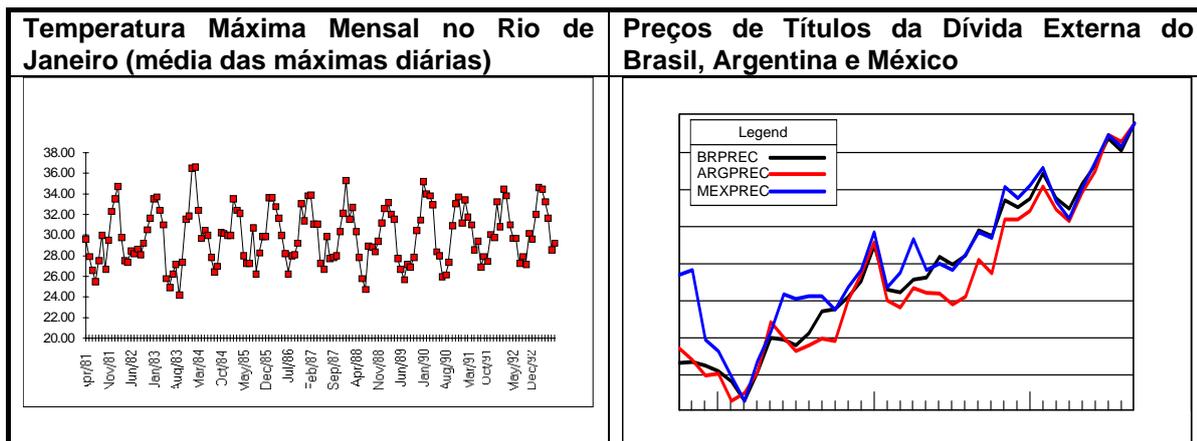
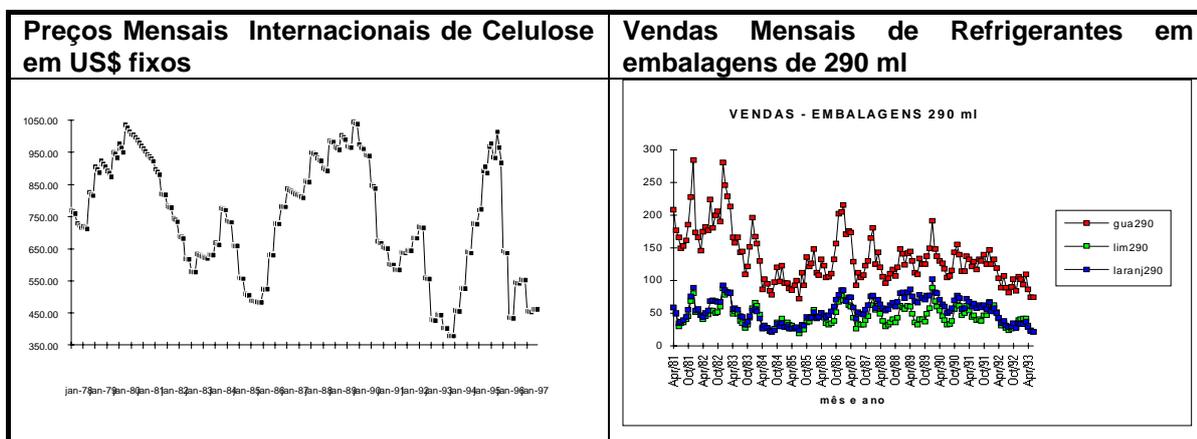
- Em particular, se o modelo é bom, os resíduos não devem apresentar correlação serial (isto é, correlação entre os resíduos em diferentes instantes de tempo);
- Explicar o comportamento da série com o menor número de parâmetros (parcimônia).

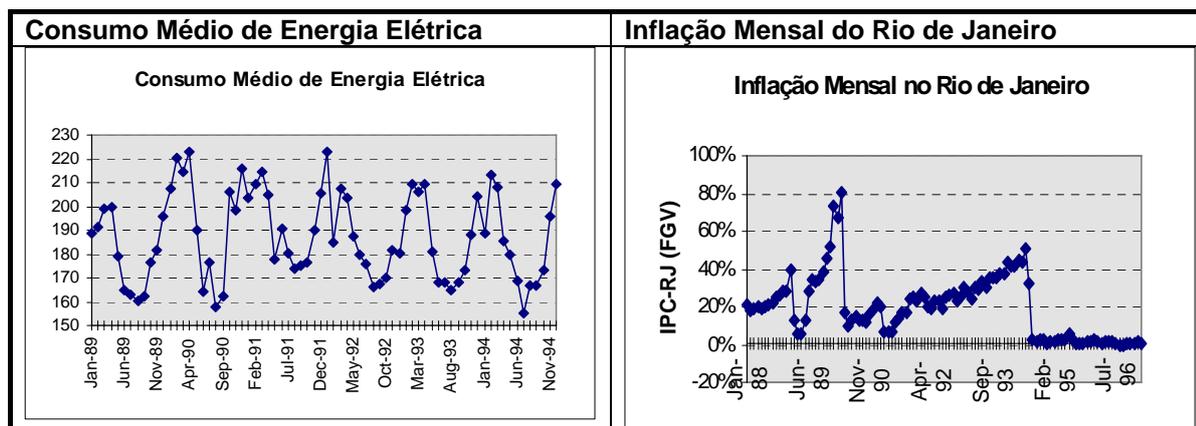
Dica Prática Por onde começar

Em geral, a primeira coisa que fazemos ao estudar uma série temporal é construir um gráfico para mostrar a sua evolução ao longo do tempo. Este procedimento simples costuma ser bastante esclarecedor, e nos permite identificar como evolui a tendência da série, se existe ou não sazonalidade, se ocorrem observações aberrantes, etc ... Logo, a minha sugestão é: sempre comece a sua análise da maneira mais simples – faça o gráfico da série de interesse!

Séries temporais ocorrem com enorme freqüência na prática. No quadro a seguir exibimos os gráficos de algumas séries temporais reais.

Quadro 5.1.1. - algumas séries temporais





Podemos fazer uma distinção básica entre duas grandes classes de modelos:

- **Modelos Univariados:** a série temporal é explicada (prevista) apenas pelos seus valores passados;
- **Modelos Multivariados ou Causais:** a série temporal é explicada (prevista) pelos seus valores passados e também pelos valores passados de outras variáveis.

Neste capítulo consideraremos apenas modelos univariados.

Exemplo 5.1.3. (Alguns Modelos Univariados)

i) Naive (ingênuo)

A previsão de Z_{T+1} (valor da série no instante $T + 1$) é apenas a última observação (Z_T). É claro que você não precisa de um “software” para ajustar isso e, em alguns casos, é o único “método” disponível. Um exemplo clássico é a previsão do preço de uma ação - geralmente a melhor previsão para o preço de amanhã é o preço de hoje, o que certamente é frustrante!

ii) Médias móveis de tamanho N

A cada instante, a previsão é apenas a média das últimas N observações. Um dos problemas com este método é a escolha de N, o **tamanho da janela** a ser utilizado. Quanto maior o valor de N, mais “suave” é a previsão. Ao contrário, se N é pequeno, a previsão tende a ser meio “nervosa”, isto é, oscila muito. Uma característica importante do método de médias móveis é: **todas as observações** usadas para o cálculo **têm o mesmo peso** (que é $1/N$). Mas, na prática é razoável supor que as observações mais recentes sejam mais relevantes para a previsão dos próximos valores da série, e portanto deveriam receber um peso maior que as observações mais antigas. Esta idéia de pesar ou ponderar as observações de acordo com as suas “idades” leva aos diversos métodos de amortecimento exponencial.

iii) Amortecimento Exponencial (“Exponencial Smoothing”)

Existem inúmeras variações destes métodos, para séries sazonais e não sazonais. A idéia geral é parecida com a do método de médias móveis, mas os pesos das observações decrescem à medida que as observações estão mais longe no passado. A taxa de decréscimo do(s) peso(s) é

determinada por uma ou mais constantes de amortecimento. A maior dificuldade na aplicação é escolher a(s) constante(s) de amortecimento, mas alguns softwares já ajustam os modelos de amortecimento automaticamente com constantes de amortecimento otimizadas.

Na prática, os métodos de amortecimento são os métodos de previsão mais usados no dia a dia das empresas, o que é em parte explicado pela sua facilidade de implementação e capacidade em gerar boas previsões.

iv) Modelos ARIMA de Box e Jenkins

São modelos mais sofisticados, que usam as correlações entre as observações em diversos instantes. A ideia por trás dos modelos ARIMA envolve filtros lineares e algum conhecimento de Teoria de Sistemas é útil. A identificação da estrutura do modelo é um pouco complicada, mas alguns softwares já identificam automaticamente a estrutura do modelo ARIMA, evitando o passo mais complicado da análise. Como casos particulares dos modelos ARIMA temos os processos AR (Autoregressivo) e MA (Médias móveis ou “moving average”), já estudados brevemente no capítulo 2.

Os modelos ARIMA costumam apresentar melhores resultados que os métodos de amortecimento quando a série é relativamente longa e “bem comportada”.

Se a série é muito irregular, os resultados são, geralmente, inferiores aos obtidos por métodos de amortecimento.

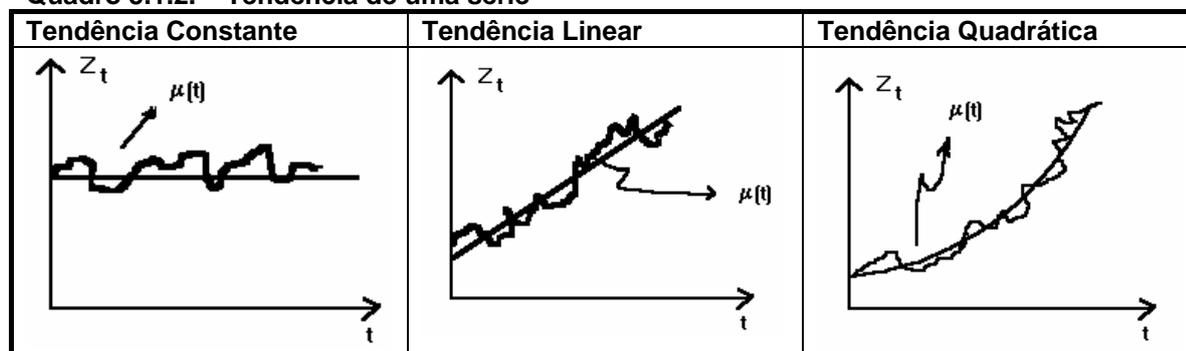
Existem também modelos ARIMA multivariados, geralmente chamados de “modelos de função de transferência”, mas eles não serão estudados aqui.

Decomposição de uma série temporal

A maneira tradicional de analisar uma série temporal é através da sua decomposição nas componentes de **tendência**, **ciclo** e **sazonalidade**.

A **tendência** de uma série indica o seu comportamento “de longo prazo”, isto é, se ela sobe, desce, ou permanece estável, e qual a velocidade destas mudanças. Nos casos mais comuns trabalhamos com tendência constante, linear ou quadrática, como mostrado na próxima figura.

Quadro 5.1.2. – Tendência de uma série



A **sazonalidade** indica a repetição de um padrão na série dentro do período de um ano. Por exemplo, vendas de sorvete são altas no verão e baixas no inverno. No quadro 5.1.1., as séries de temperatura, consumo de energia elétrica e vendas de refrigerantes são claramente sazonais, mas não se pode dizer o mesmo das outras séries.

Os **ciclos** indicam padrões que se repetem na série em períodos superiores a um ano. Por exemplo, ciclos relacionados à atividade econômica ou ciclos meteorológicos.

Características de uma série temporal e estimadores

Um processo estocástico estacionário de 2ª ordem pode ser descrito (no domínio do tempo) por sua média, variância e função de autocovariância (ou de autocorrelação). Na prática não conhecemos totalmente o processo estocástico que gerou a série que está sendo observada, e estas quantidades (média, variância, autocorrelação) devem ser estimadas a partir da série temporal, como indicado no capítulo 2.

5.2. Métodos de Amortecimento Exponencial (séries não sazonais)

Existem inúmeras variações de métodos de amortecimento exponencial. Todas elas, na verdade, têm uma característica comum: a informação atual (representada pelo último valor observado) é ponderada com a informação contida nos instantes anteriores. A existência de uma ou mais constantes de amortecimento irá determinar como funciona este mecanismo de ponderação ou, em outras palavras, o quão rapidamente decai a influência das observações passadas. Os parâmetros do modelo sofrem uma atualização seqüencial, e a chegada de uma nova observação resulta numa nova estimativa dos parâmetros.

Nesta seção estudamos diversos métodos de amortecimento exponencial para séries não sazonais. A extensão destes métodos para séries sazonais é mostrada na próxima seção.

Os métodos de amortecimento exponencial são formas até certo ponto arbitrárias de representar uma série temporal, mas são muito importantes na prática, pois geralmente produzem previsões razoáveis a partir de um esforço computacional baixo. Na verdade, estes métodos são freqüentemente citados como os “favoritos” dos não especialistas em séries temporais, pela qualidade das previsões geradas e facilidade de implementação. Os métodos de amortecimento mais famosos são: Brown, Holt e Winters.

Um ponto potencialmente complicado em todos estes métodos é a escolha da(s) constante(s) de amortecimento, mas na prática isso não é mais problema, já que a maioria dos softwares de séries temporais ajusta modelos com a(s) constante(s) otimizadas.

Seja $Z_T = (Z_1, Z_2, \dots, Z_T)$ uma série temporal de tamanho T. Desejamos escrever um modelo geral para explicar o comportamento da série a cada instante e permitir a produção de previsões futuras.

Suponha que o modelo geral para uma observação qualquer no instante t genérico é: $Z_t = \mu(t) + \varepsilon_t$ onde ε_t é um ruído aleatório com média zero e variância constante σ_ε^2 . O termo $\mu(t)$ representa o **nível médio da série**, que é função apenas do instante de tempo.

Iremos estudar três casos particulares: $\mu(t)$ é constante, uma função linear e uma função quadrática do tempo.

A partir das propriedades da média e da variância podemos concluir que:

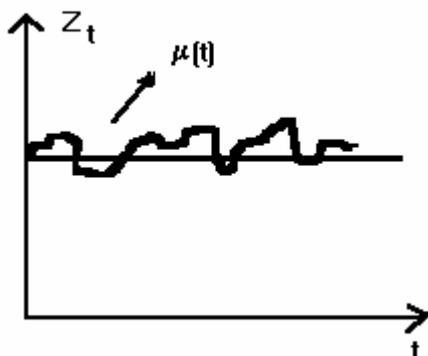
$$\begin{aligned} E(Z_t) &= \mu(t) \quad \text{para todo } t = 1, 2, 3, \dots \\ \text{VAR}(Z_t) &= \sigma_\varepsilon^2 = \text{constante} \end{aligned} \quad (5.2.1)$$

A seguir derivamos estimadores para os parâmetros desconhecidos dos modelos de nível constante, linear e quadrático e encontramos as respectivas equações de previsões.

Freqüentemente iremos supor que os erros $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T$ são independentes com uma distribuição de probabilidade conhecida (por exemplo, a Normal) com média zero e variância constante. Das equações acima é fácil ver que a estimação da variância de Z_t (que é igual à variância de ε_t) não apresenta grandes dificuldades, basta utilizar os estimadores usuais da variância de uma amostra aleatória. Por isso, na discussão a seguir, nós nos concentraremos na estimação dos parâmetros que definem o nível médio da série.

5.2.1. Modelo Constante (Horizontal)

Este é um modelo apropriado quando a série temporal apenas oscila em torno de um valor constante, sem exibir tendência de crescimento ou queda. A função **nível médio** é dada por: $\mu(t) = a_1$ para todo instante $t = 1, 2, 3, \dots, T$.



Neste modelo é necessário estimar apenas 2 parâmetros:

- a_1 (o nível médio da série) e
- σ^2 (a variância da série).

A **previsão k passos à frente** obtida no instante T é:

$$\hat{Z}_T(k) = E\{Z_{T+k} | \underline{Z}_T\} = E\{a_1 + \varepsilon_{T+k} | \underline{Z}_T\} = \hat{a}_1(T) \quad (5.2.2)$$

Logo, se conhecermos um estimador do parâmetro a_1 poderemos encontrar as previsões pontuais para qualquer horizonte de tempo k .

E como podemos estimar a constante a_1 (nível médio da série)? Existem diversas possibilidades, mostradas a seguir.

Modelo Constante - Estimação do nível médio da série

1) Estimador ingênuo (“naive”)

$\hat{a}_1(T) = Z_T$ onde Z_T indica a última observação disponível.

2) Estimador Médias Móveis de tamanho N

O nível médio da série num certo instante é estimado como a média móvel de tamanho N usando as observações $Z_T, Z_{T-1}, \dots, Z_{T-N+1}$, isto é:

$$\hat{a}_1(T) = M_T = \frac{Z_{T-N+1} + \dots + Z_{T-1} + Z_T}{N} \quad (5.2.3)$$

Uma questão crucial neste caso é definir qual o tamanho da “janela” a ser usada, isto é, quantas observações serão utilizadas no cálculo da média móvel. Note que, se $N = 1$, apenas a última observação é usada, e o método se reduz ao método ingênuo. Por outro lado, se $N = T$ (tamanho da série), temos o chamado “método conservador”, em que todas as observações disponíveis são empregadas na estimação do nível médio da série, e $\hat{a}_1(T) = \bar{Z}$, a média amostral da série.

O tamanho da janela utilizada (N) pode ser escolhido como aquele inteiro positivo que minimiza a soma de quadrados dos erros de previsão 1 passo à frente, isto é, escolhe-se N de forma a minimizar:

$$S(N) = \sum_{t=1}^T e_1^2(t)$$

3) Estimador pelo Método de Amortecimento Exponencial

Note que o estimador do nível da série produzido pelo método de médias móveis pode ser escrito como:

$$\hat{a}_1(T) = M_T = \frac{Z_{T-N+1} + \dots + Z_{T-1} + Z_T}{N} = M_{T-1} + \frac{Z_T - Z_{T-N}}{N} \quad (5.2.4)$$

Onde M_{T-1} é a média móvel de tamanho N calculada no instante anterior. Em outras palavras, o estimador do nível envolve uma informação “atual” (representada por Z_T , o último valor observado da série) e uma informação “passada” (M_{T-1} e Z_{T-N}).

O método de amortecimento exponencial estende esta idéia de ponderação da informação presente e passada. Na equação (5.2.4) substitua Z_{T-N} por M_{T-1} . Isso nos leva ao seguinte estimador:

$$\hat{a}_1(T) = M_{T-1} + \frac{Z_T}{N} - \frac{M_{T-1}}{N} = \left(\frac{1}{N}\right)Z_T + \left(1 - \frac{1}{N}\right)M_{T-1}$$

Ou seja, este novo estimador tem a forma: (peso)*(informação atual) + (1 – peso)*(informação passada). Note que $\hat{a}_1(T)$ é a previsão do nível da série efetuada no instante atual T e podemos reescrever esta última equação como:

$$\hat{a}_1(T) = \left(\frac{1}{N}\right)Z_T + \left(1 - \frac{1}{N}\right)M_{T-1} = \left(\frac{1}{N}\right)Z_T + \left(1 - \frac{1}{N}\right)\hat{a}_1(T-1)$$

onde $\hat{a}_1(T-1)$ é a média móvel de tamanho N calculada até (inclusive) a observação T-1, ou seja, é a estimativa prévia do nível da série, Em outras palavras: a previsão atual do nível (dada por $\hat{a}_1(T)$) é baseada na ponderação do valor mais recente da série (Z_T) e da última previsão para o nível ($\hat{a}_1(T-1)$).

Uma forma mais geral de definir um estimador com esta característica de ponderação das informações passadas e presentes é:

$$\Rightarrow \hat{a}_1(T) = M_T = \alpha Z_T + (1-\alpha)M_{T-1} = \alpha Z_T + (1-\alpha)\hat{a}_1(T-1) \quad (5.2.5)$$

onde $\alpha \in [0, 1]$ é a **constante de amortecimento**, que controla a taxa de decaimento da informação.

A equação (5.2.5) apresenta algumas características interessantes:

- 1) é bastante econômica do ponto de vista computacional, pois é preciso guardar, a cada instante, apenas o último valor observado e a estimativa anterior do nível, isto é, não precisamos guardar todos os valores anteriores da série temporal para gerar a previsão do nível;
- 2) o método funciona bastante bem, desde que a série não apresente uma tendência.

Ao rearranjar os termos da equação (5.2.5) e encontramos:

$$\hat{a}_1(T) = \alpha Z_T + (1-\alpha)M_{T-1} = \hat{a}_1(T-1) + \alpha(Z_T - \hat{a}_1(T-1)) \quad (5.2.6)$$

A expressão (5.2.6) é freqüentemente conhecida como “*forma de correção do erro*”. Nela, o mecanismo de atualização da previsão do nível fica explícito: o nível estimado num instante T é igual ao nível estimado no instante anterior mais um múltiplo do erro de previsão um passo à frente cometido no instante anterior (que é $Z_T - \hat{a}_1(T-1)$).

Seja e_{T-1} o erro de previsão um passo à frente no instante T-1, isto é:

$$e_{T-1} = Z_T - \hat{a}_1(T-1) \text{ e então podemos rearranjar os termos e escrever: } Z_T = e_{T-1} + \hat{a}_1(T-1).$$

Se aplicarmos a equação (5.2.6) à previsão do nível no instante T + 1 encontramos:

$$\hat{a}_1(T+1) = \alpha Z_{T+1} + (1-\alpha)M_T = \hat{a}_1(T) + \alpha(Z_{T+1} - \hat{a}_1(T)) = \hat{a}_1(T) + \alpha e_T$$

Dáí:

$$Z_{T+1} - Z_T = (e_{T+1} + \hat{a}_1(T)) - (e_T + \hat{a}_1(T-1)) = e_{T+1} - e_T + (\hat{a}_1(T) - \hat{a}_1(T-1)) = e_{T+1} - e_T + \alpha e_T = e_{T+1} + (1-\alpha)e_T$$

Esta expressão não é útil do ponto de vista preditivo, pois o lado direito contém os “outputs” (erros de previsão um passo à frente) enquanto o lado esquerdo contém os “inputs” (observações). Mas, esta expressão é importante em termos estruturais, pois se os e 's (erros de previsão um passo à frente) são substituídos por choques aleatórios, o modelo resultante é um ARIMA (0,1,1), que será estudado na próxima seção. Num modelo ARIMA(0,1,1) o método de amortecimento exponencial simples estudado aqui fornece previsões ótimas.

Por substituições sucessivas na equação (5.2.5) encontramos:

$$M_T = \alpha Z_T + \alpha(1-\alpha) Z_{T-1} + \alpha(1-\alpha)^2 Z_{T-2} + \dots + \alpha(1-\alpha)^{T-1} Z_1 \quad (5.2.7)$$

Esta última equação mostra claramente como funciona o método de amortecimento exponencial: cada estimativa do nível (dada por M_T) é uma soma ponderada da observação atual (Z_T) e das observações passadas. Os pesos decrescem exponencialmente, e a taxa de decaimento dos pesos depende da constante α .

Qual o efeito da constante de amortecimento α ? Note que, da equação (5.2.7), o peso da observação Z_{T-k} é $\alpha(1-\alpha)^k$ (para $k = 0, 1, 2, \dots$). Logo, a razão entre os pesos da observação Z_{T-k} e da observação atual (Z_T) é $(1-\alpha)^k$. Se α é grande (próximo de 1) os pesos decaem rapidamente, e uma observação que está “longe” no passado não tem qualquer influência sobre a estimativa atual do nível, que é estimado usando quase exclusivamente as observações mais recentes. Ao contrário, se α é pequeno (por exemplo, menor que 0.3), uma observação “longe” no passado pode ainda ter uma influência considerável na estimação do nível feita hoje.

A tabela a seguir apresenta alguns valores de α e a “meia vida” das observações, isto é, quantas defasagens são necessárias até que o peso da observação se reduza à metade. A meia vida das observações é encontrada resolvendo a seguinte equação para k (o “lag”):

$$\frac{\text{peso de } Z_{T-k}}{\text{peso de } Z_T} = \frac{1}{2} \Leftrightarrow \frac{\alpha(1-\alpha)^k}{\alpha} = \frac{1}{2} \Leftrightarrow k \cdot \log(1-\alpha) = -\log(2) \Leftrightarrow k = \frac{-\log(2)}{\log(1-\alpha)} \quad \text{onde } \log(.)$$

indica o logaritmo natural. Como k indica uma defasagem, deve-se escolher o valor k inteiro não negativo que satisfaz (aproximadamente) esta equação.

Tabela 5.2.1. – Meia Vida das Observações como Função de α

α	número de “lags” até o peso cair pela metade
0,99	0,15
0,95	0,23
0,90	0,30
0,80	0,43
0,70	0,58
0,60	0,76
0,50	1,00
0,45	1,16
0,40	1,36
0,35	1,61
0,30	1,94
0,25	2,41
0,20	3,11
0,15	4,27
0,10	6,58
0,05	13,51
0,02	34,31
0,01	68,97

Por exemplo, se $\alpha = 0,02$, a observação situada há 34 “lags” contém aproximadamente metade da informação do dado atual. Se $\alpha = 0,05$, são necessárias cerca de 14 defasagens até que o peso da observação caia pela metade.

Nota: reparametrização

Em alguns livros encontra-se o método de amortecimento escrito de maneira diferente. Seja $\theta = 1 - \alpha$ na equação (5.2.7). Em termos de θ esta equação torna-se:

$$M_T = (1-\theta)Z_T + \theta(1-\theta) Z_{T-1} + \theta^2(1-\theta) Z_{T-2} + \dots + \theta^T(1-\theta)Z_1 \quad (5.2.8)$$

Logo, o peso da observação atual é $(1-\theta)$ e o da k -ésima observação no passado é $\theta^k (1-\theta)$. Nesta parametrização, valores altos de θ indicam que a memória do processo é relativamente longa, ou seja, a influência das observações passadas demora a decair.

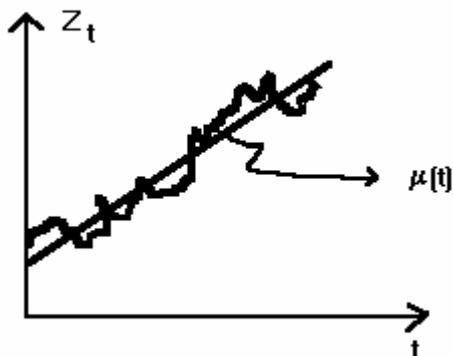
Exemplo 5.2.1.

A tabela a seguir exibe os primeiros valores da série temporal: $Z_t = 2 + e_t$ onde os e_t 's são ruídos iid com distribuição Normal de média 0 e desvio padrão $\frac{1}{2}$. Supomos que em $t = 0$, $Z_0 = 0$ (apenas para permitir a inicialização do processo), e o objetivo aqui é demonstrar a estimação seqüencial de a_1 por amortecimento exponencial e médias móveis (de tamanho 2, 3 e 4). Note que, dependendo da escolha da constante de amortecimento (ou da janela utilizada na média móvel), a estimativa do nível é mais ou menos “nervosa” (e oscila mais ou menos à medida que recebemos cada valor da série). Verifique os valores estimados para o nível da série. Faça (em casa) o gráfico da evolução no nível estimado ao longo do tempo para todos os métodos indicados.

Instante t	Erros (e(t))	Série (Z(t))	Estimador MM(2)	Estimador MM(3)	Estimador MM(4)	Estimador Amort. Expo. (Alfa = 0.2)	Estimador Amort. Expo. (Alfa = 0.3)
1	-1,512	0,488	0,244			0,098	0,147
2	0,080	2,080	1,284	0,856		0,494	0,692
3	-0,433	1,567	1,824	1,379	1,034	0,709	0,816
4	0,437	2,437	2,002	2,028	1,643	1,054	1,227
5	0,107	2,107	2,272	2,037	2,048	1,265	1,370
6	-0,025	1,975	2,041	2,173	2,021	1,407	1,478
7	-0,192	1,808	1,891	1,963	2,082	1,487	1,527
8	0,629	2,629	2,219	2,137	2,130	1,716	1,830
9	0,463	2,463	2,546	2,300	2,219	1,865	1,940
10	0,332	2,332	2,398	2,475	2,308	1,958	2,005
11	-0,469	1,531	1,931	2,109	2,239	1,873	1,830
12	0,538	2,538	2,034	2,134	2,216	2,006	2,072
13	0,277	2,277	2,408	2,115	2,170	2,060	2,087
14	0,017	2,017	2,147	2,277	2,091	2,052	2,047
15	-0,256	1,744	1,880	2,013	2,144	1,990	1,959

5.2.2. Modelo Linear

Esta estrutura é adequada quando existe uma tendência clara de crescimento (ou queda) ao longo do tempo, como indicado na próxima figura. A função **nível médio** é dada por: $\mu(t) = a_1 + a_2.t$ para todo instante $t = 1, 2, 3, \dots, T$.



O modelo linear pode ser escrito como:

$Z_t = a_1 + a_2 \cdot t + \varepsilon_t$ onde $t = 1, 2, \dots, T$ e os ε_t são os erros (choques) aleatórios, supostos independentes e identicamente distribuídos com média zero e variância constante σ^2 . Desta expressão (e usando as propriedades da média e variância) segue direto que:

$$E(Z_t) = a_1 + a_2 \cdot t \quad \text{e} \quad \text{VAR}(Z_t) = \sigma^2 = \text{VAR}(\varepsilon_t) \quad (5.2.9)$$

Neste modelo é necessário estimar três parâmetros: a_1 (coeficiente linear da série), a_2 (coeficiente angular da reta, ou seja, a taxa de crescimento da série) e σ^2 (a variância da série).

A **previsão k passos à frente** obtida no instante T é:

$$\hat{Z}_T(k) = E\{Z_{T+k} \mid \underline{Z}_T\} = E\{a_1 + a_2(T+k) + \varepsilon_{T+k} \mid \underline{Z}_T\} = \hat{a}_1(T) + \hat{a}_2(T) \cdot (T+k) \quad (5.2.10)$$

A previsão dada pela equação (5.2.10) pode ser escrita de uma forma ligeiramente diferente, desde que façamos uma mudança da origem.

Modelo Linear - Estimação dos parâmetros a_1 e a_2

Analogamente ao que fizemos para o modelo constante, a seguir apresentaremos diversos estimadores para os parâmetros desconhecidos que descrevem a evolução do nível da série.

A partir dos valores estimados de a_1 e a_2 podemos encontrar a previsão de uma observação futura qualquer, apenas aplicando a equação (5.2.10).

1) Estimadores de Mínimos Quadrados Ordinários (MQO)

Nosso objetivo é encontrar a_1 e a_2 tais que a seguinte soma seja minimizada:

$$\text{SSE} = \sum_{t=1}^T \underbrace{[Z_t - (a_1 + a_2 t)]^2}_{\varepsilon_t} \quad (5.2.11)$$

SSE é chamado de soma dos quadrados dos erros (ou soma do quadrado dos resíduos). O método de mínimos quadrados é bastante geral, e se aplica à estimação de um número qualquer de parâmetros a_1, a_2, \dots, a_k (supõe-se que k é menor que o tamanho da série, para permitir a estimação).

Considere, de maneira mais geral, um modelo com a seguinte estrutura: $Z_t = f(a_1, a_2, \dots, a_k, t) + \varepsilon_t$ onde $f(\cdot)$ é uma função (possivelmente não linear) do tempo e dos parâmetros desconhecidos. Então a soma dos quadrados dos erros é dada por:

$$SSE = \sum_{t=1}^T (Z_t - f(a_1, a_2, \dots, a_k, t))^2 = \sum_{t=1}^T \varepsilon_t^2$$

O método de mínimos quadrados procura encontrar estimadores dos parâmetros desconhecidos de forma a minimizar a soma do quadrado dos erros (SSE).

Note que os estimadores por mínimos quadrados encontrados pela minimização da soma do quadrado dos erros são calculados usando a série inteira, ou seja, usamos todos os dados disponíveis de uma só vez. Ao contrário, nos métodos médias móveis e de amortecimento, a atualização dos parâmetros é seqüencial, e o recebimento de uma nova observação resulta em novos valores estimados para a_1 e a_2 .

2) Estimadores pelo Método de Médias Móveis Duplas

Defina a média móvel dupla de tamanho N como:

$$M_T^{[2]} = \frac{M_T + M_{T-1} + \dots + M_{T-N+1}}{N} \quad (5.2.12)$$

Onde M_T é a média móvel (simples) de tamanho N calculada usando todas as observações até o instante T (inclusive).

Por que usar médias móveis duplas?

Se os dados exibem uma tendência linear, o uso de médias móveis simples para a previsão dos valores da série induz a erros sistemáticos, pois a média móvel simples segue a tendência com um certo atraso, e este efeito é amplificado quando tentamos prever valores futuros. O método de médias móveis duplas procura diminuir este efeito sistemático.

Pode-se mostrar que:

$$E\{M_T\} = a_1(T) + T.a_2(T) - \frac{N-1}{2}.a_2(T)$$

$$E\{M_T^{[2]}\} = a_1(T) + T.a_2(T) - (N-1)a_2(T)$$

Resolvendo o sistema anterior para a_1 e a_2 leva a:

$$a_1(T) = 2.E\{M_T\} - E\{M_T^{[2]}\} - T.a_2(T)$$

$$a_2(T) = \frac{2}{(N-1)} \{E\{M_T\} - E\{M_T^{[2]}\}\} \quad (5.2.13)$$

Se substituirmos os valores esperados de M_T e $M_T^{[2]}$ na expressão (5.2.13) pelos valores efetivamente observados M_T e $M_T^{[2]}$ encontramos os estimadores de a_1 e a_2 com base na série observada até o instante T , dados por:

$$\begin{aligned}\hat{a}_1(T) &= M_T + (M_T - M_T^{[2]}) - T \cdot \hat{a}_2(T) = 2 \cdot M_T - M_T^{[2]} - T \cdot \hat{a}_2(T) \\ \hat{a}_2(T) &= \frac{2}{(N-1)} \{M_T - M_T^{[2]}\}\end{aligned}\quad (5.2.14)$$

O estimador $\hat{a}_1(T)$ é encontrado corrigindo-se a média móvel simples (M_T) pela diferença entre ela e a média móvel dupla. Se, na verdade, não existisse uma tendência linear, o fator de correção seria pequeno.

A previsão k passos à frente no instante T é:

$$\begin{aligned}\hat{Z}_{T+k} &= \hat{a}_1(T) + \hat{a}_2(T) \cdot (T+k) = 2 \cdot M_T - M_T^{[2]} - T \cdot \hat{a}_2(T) + T \cdot \hat{a}_2(T) + k \cdot \hat{a}_2(T) = \\ &= 2 \cdot M_T - M_T^{[2]} + k \cdot \hat{a}_2(T) = 2 \cdot M_T - M_T^{[2]} + k \cdot \left\{ \frac{2}{N-1} \{M_T - M_T^{[2]}\} \right\}\end{aligned}$$

Expressão alternativa para as equações de atualização do método de médias móveis duplas

DeLurgio (1998) propõe o uso de uma versão modificada das equações de atualização (5.2.14). As novas (e mais simples) equações de atualização são:

$$\begin{aligned}\hat{a}_1(T) &= M_T + (M_T - M_T^{[2]}) = 2 \cdot M_T - M_T^{[2]} \\ \hat{a}_2(T) &= \frac{2}{(N-1)} \{M_T - M_T^{[2]}\}\end{aligned}\quad (5.2.15)$$

E a equação de previsão k passos à frente torna-se:

$$\hat{Z}_{T+k} = \hat{a}_1(T) + \hat{a}_2(T) \cdot (k) = 2 \cdot M_T - M_T^{[2]} + k \cdot \hat{a}_2(T) = 2 \cdot M_T - M_T^{[2]} + k \cdot \left\{ \frac{2}{N-1} \{M_T - M_T^{[2]}\} \right\}$$

Vantagens e Desvantagens deste método

- a implementação é trivial;
- os estimadores obtidos pelo método não são muito influenciados por outliers;
- a escolha do número ótimo de "lags" a serem usados nas médias móveis simples e duplas não é trivial.

Exemplo 5.2.2.

Um ponto importante a ser lembrado quando tratamos de métodos de médias móveis é: estes procedimentos implicam num conjunto de pesos definidos para as observações passadas. Isso nos permite compará-los com os métodos de amortecimento exponencial.

Por exemplo, considere um modelo linear onde os parâmetros desconhecidos são estimados por médias móveis simples e duplas de tamanho 3. Como podemos escrever os estimadores em termos dos valores observados da série?

Note que, neste caso, para qualquer instante t maior que 3:

$$M_t = \frac{Z_t + Z_{t-1} + Z_{t-2}}{3}$$

$$M_t^{[2]} = \frac{M_t + M_{t-1} + M_{t-2}}{3}$$

E assim a previsão 1 passo à frente no instante T (com a origem transladada) é:

$$\begin{aligned} \hat{Z}_{T+1} &= \hat{a}_1(T) + \hat{a}_2(T)(1) = 2M_T - M_T^{[2]} + \frac{2}{3-1}(M_T - M_T^{[2]}) = \\ &= 2M_T - M_T^{[2]} + M_T - M_T^{[2]} = 3M_T - 2M_T^{[2]} = 3M_T - 2\left(\frac{M_T + M_{T-1} + M_{T-2}}{3}\right) = \\ &= \left(\frac{1}{3}\right)\left\{\frac{7}{3}M_T - 2M_{T-1} - 2M_{T-2}\right\} = \left(\frac{1}{3}\right)\left\{\frac{7(Z_T + Z_{T-1} + Z_{T-2})}{3} - \frac{2(Z_{T-1} + Z_{T-2} + Z_{T-3})}{3} - \frac{2(Z_{T-2} + Z_{T-3} + Z_{T-4})}{3}\right\} = \\ &= \left(\frac{1}{3}\right)\left\{\left(\frac{7}{3}\right)Z_T + \left(\frac{5}{3}\right)Z_{T-1} + \left(\frac{3}{3}\right)Z_{T-2} + \left(\frac{-4}{3}\right)Z_{T-3} + \left(\frac{-2}{3}\right)Z_{T-4}\right\} \end{aligned}$$

3) Método de Amortecimento Exponencial Duplo (Método de Brown)

Um método semelhante ao anterior é o método de amortecimento exponencial duplo, também conhecido como método de Brown. Nele, uma única constante de amortecimento é empregada no processo de atualização de ambos os parâmetros a_1 e a_2 . Analogamente ao método de médias móveis duplas, a componente de tendência é estimada através da diferença entre um valor amortecido “simples” e um “duplo”.

As equações de atualização no método de Brown são:

$$\begin{aligned} M_T &= \alpha Z_T + (1-\alpha)M_{T-1} \\ M_T^{[2]} &= \alpha M_T + (1-\alpha)M_{T-1}^{[2]} \end{aligned} \quad (5.2.16)$$

onde α é a constante de amortecimento.

Os estimadores de a_1 e a_2 encontrados pelo método de Brown são:

$$\begin{aligned} \hat{a}_1(T) &= M_T + (M_T - M_T^{[2]}) = 2M_T - M_T^{[2]} \\ \hat{a}_2(T) &= \frac{\alpha}{(1-\alpha)}(M_T - M_T^{[2]}) \end{aligned} \quad (5.2.17)$$

Ou seja, estes estimadores têm a mesma forma que na equação (5.2.15). No entanto, agora M_T e $M_T^{[2]}$ **não mais** representam as médias móveis descritas no item 5.2.2., e são calculados através das equações (5.2.16).

A previsão k passos à frente no instante T é:

$$\hat{Z}_{T+k} = \hat{a}_1(T) + \hat{a}_2(T).(k) = 2M_T - M_T^{[2]} + \frac{\alpha}{1-\alpha}(M_T - M_T^{[2]})k \quad (5.2.18)$$

Escolha dos valores iniciais

Em todos os métodos de amortecimento é necessário escolher valores iniciais dos parâmetros de forma a inicializar as equações de atualização. Algumas escolhas possíveis são:

$$M_1 = Z_1$$

$$\hat{a}_1(1) = Z_1$$

$$\hat{a}_2(1) = \frac{(Z_2 - Z_1) + (Z_4 - Z_3)}{2}$$

Vantagens e Desvantagens do método de Brown

- modela o nível e a tendência da série;
- é mais eficiente em termos computacionais que as médias móveis duplas;
- é menos flexível que o modelo de Holt (a seguir), que usa duas constantes de amortecimento.

4) Método de Holt (2 parâmetros)

Este método é bastante parecido com o anterior, mas agora existem duas constantes de amortecimento diferentes, responsáveis pela “velocidade” de decaimento da informação na estimação dos coeficientes linear e angular, respectivamente.

O modelo de Holt é implementado através das seguintes equações:

$$M_T = \alpha \cdot Z_T + (1 - \alpha) \cdot (M_{T-1} + \hat{a}_2(T-1))$$

$$\hat{a}_2(T) = \beta \cdot (M_T - M_{T-1}) + (1 - \beta) \cdot \hat{a}_2(T-1) \quad (5.2.19)$$

$$\hat{a}_1(T) = M_T$$

onde α e β são as constantes de amortecimento.

Características importantes do método

- As constantes de amortecimento usadas para cálculo do nível e tendência são diferentes;
- Na equação de atualização do nível, o novo nível é função de 3 variáveis: a observação mais recente (Z_T), a estimativa anterior do nível (M_{T-1}) e a taxa de crescimento estimada no instante anterior ($\hat{a}_2(T-1)$);
- A taxa de crescimento atualizada é função da taxa de crescimento no instante anterior e também depende da diferença entre os níveis nos instantes T-1 e T.

Escolha dos valores iniciais

Possíveis escolhas são:

$$\hat{a}_1(1) = Z_1$$

$$\hat{a}_2(1) = \frac{(Z_2 - Z_1) + (Z_4 - Z_3)}{2}$$

Vantagens e Desvantagens do método de Holt

- modela o nível e a tendência da série;
- é mais flexível que o método de Brown, pois usa constante de amortecimento separadas para o nível e a tendência;
- a contrapartida deste ganho em flexibilidade é a existência de um parâmetro adicional a ser otimizado, o que dificulta o processo de otimização.

Exemplo 5.2.3.

Considere uma série temporal puramente determinística, $Z_t = 2 + 3.t$. A tabela a seguir exibe alguns valores da série e as previsões um passo à frente de Z_t usando: 1) uma média móvel simples de tamanho 3, 2) a média móvel simples e a média móvel dupla, como na equação (5.2.15). Verifique os valores encontrados na tabela, e note que o erro da previsão um passo à frente obtida através da equação (5.2.15) é nulo em todos os instantes.

Instante t	Série Z(t) (puramente determinística)	MM(3)	Previsão 1 passo à frente de Z(t) usando MM(3)	Erro de Previsão 1 passo à frente usando MM(3)	MM Dupla (3)	Diferença entre MM(3) e MM Dupla (3)	Previsão 1 passo à frente = MM(3) + (MM(3)-MMDup(3)) + tendência	Erro de Previsão 1 passo à frente usando MM Dupla
1	5,00							
2	8,00							
3	11,00	8,00						
4	14,00	11,00	8,00	6,00				
5	17,00	14,00	11,00	6,00	11,00	3,00		
6	20,00	17,00	14,00	6,00	14,00	3,00	20,00	0,00
7	23,00	20,00	17,00	6,00	17,00	3,00	23,00	0,00
8	26,00	23,00	20,00	6,00	20,00	3,00	26,00	0,00
9	29,00	26,00	23,00	6,00	23,00	3,00	29,00	0,00
10	32,00	29,00	26,00	6,00	26,00	3,00	32,00	0,00
11	35,00	32,00	29,00	6,00	29,00	3,00	35,00	0,00
12	38,00	35,00	32,00	6,00	32,00	3,00	38,00	0,00
13	41,00	38,00	35,00	6,00	35,00	3,00	41,00	0,00
14	44,00	41,00	38,00	6,00	38,00	3,00	44,00	0,00
15	47,00	44,00	41,00	6,00	41,00	3,00	47,00	0,00

Exemplo 5.2.4.

Considere o exemplo anterior, mas agora adicione um erro aleatório à Z_t . O modelo gerador destes dados é: $Z_t = 2 + 3.t + e_t$ onde os e_t 's são ruídos aleatórios independentes Normais, com média zero e variância $\frac{1}{4}$. Refaça o exemplo anterior e note que os erros de previsão um passo à frente usando médias móveis simples são consideravelmente maiores que os erros encontrados quando usamos a equação (5.2.15) para prever Z_t . Calcule a soma do quadrado dos erros de previsão um passo à frente segundo os dois métodos.

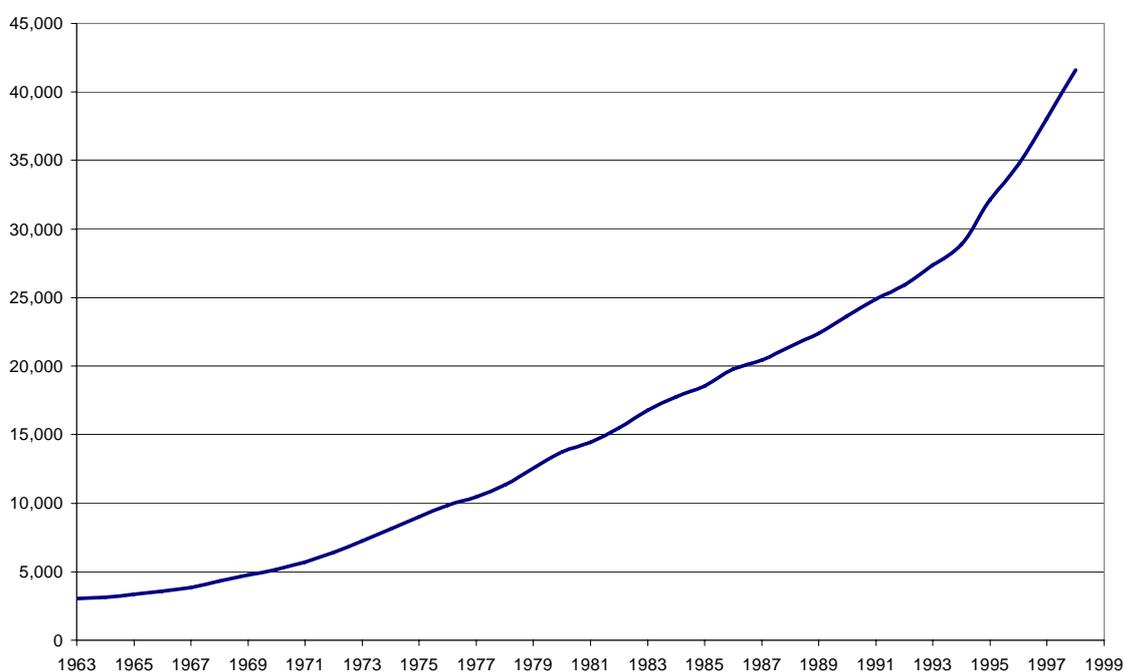
Instante t	Erros (e(t))	Série (Z(t))	MM(3)	Previsão 1 passo à frente de Z(t) usando MM(3)	Erro de Previsão 1 passo à frente usando MM(3)	MM Dupla (3)	Diferença entre MM(3) e MM Dupla (3)	Previsão 1 passo à frente = MM(3) + (MM(3)-MMDup(3)) + tendência	Erro de Previsão 1 passo à frente usando MMDupla
1	-0,150	4,850							
2	-0,639	7,361							
3	0,122	11,122	7,778						
4	0,638	14,638	11,041	7,778	6,861				
5	0,599	17,599	14,453	11,041	6,559	11,090	3,363		
6	0,867	20,867	17,701	14,453	6,413	14,398	3,303	21,179	-0,31
7	-1,092	21,908	20,125	17,701	4,207	17,426	2,698	24,307	-2,40
8	-0,117	25,883	22,886	20,125	5,758	20,237	2,649	25,521	0,36

9	0,548	29,548	25,780	22,886	6,662	22,930	2,850	28,183	1,36
10	-0,543	31,457	28,962	25,780	5,677	25,876	3,086	31,479	-0,02
11	-0,345	34,655	31,886	28,962	5,693	28,876	3,010	35,135	-0,48
12	-0,845	37,155	34,422	31,886	5,268	31,757	2,665	37,907	-0,75
13	-0,923	40,077	37,295	34,422	5,654	34,535	2,761	39,752	0,32
14	-0,489	43,511	40,248	37,295	6,216	37,322	2,926	42,817	0,69
15	-0,387	46,613	43,400	40,248	6,366	40,314	3,086	46,099	0,51

Exemplo 5.2.5.

No gráfico a seguir mostramos o consumo comercial de energia elétrica anual no Brasil no período 1963 a 1998, e os resultados do ajuste de um modelo de Holt à série.

Consumo Comercial (GWh)



Componente	Constante de Amortecimento	Valor Final
Nível	0.8086	41539
Tendência	0.9994	3536

As previsões geradas pelo modelo para o período 1999-2002 são indicadas a seguir, assim como os respectivos erros de previsão em 1999 e 2000.

Ano	Limite Inferior Previsão (2.5%)	Previsão	Limite Superior Previsão (97.5%)	Valor Real	Erro Percentual Absoluto de Previsão
1999	42,949	45,074	47,199	43,583	3.4%
2000	44,570	48,610	52,650	47,365	2.6%
2001	46,842	52,145	57,448		
2002	49,362	55,681	62,000		

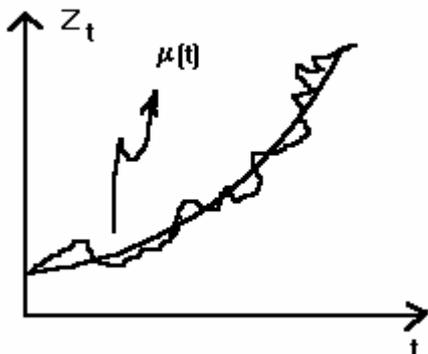
5.2.3. Modelo Quadrático

A função **nível médio** é dada por $\mu(t) = a_1 + a_2t + a_3t^2$ para $t = 1, 2, \dots, T$. Agora existem três parâmetros a serem estimados para o nível, a_1 , a_2 e a_3 . Supondo que o modelo tem a forma:

$Z_t = a_1 + a_2t + a_3t^2 + \varepsilon_t$ onde $t = 1, 2, \dots, T$ e $E(\varepsilon_t) = 0$, $VAR(\varepsilon_t) = \sigma^2$ segue que:

$E(Z_t) = \mu(t) = a_1 + a_2t + a_3t^2$ e $VAR(Z_t) = \sigma^2 = \text{constante}$.

O modelo quadrático se aplica a séries que têm o comportamento indicado na figura a seguir.



A previsão k passos à frente no instante T é dada por:

$$\begin{aligned}\hat{Z}_T(k) &= E\{Z_{T+k} \mid Z_1, Z_2, \dots, Z_T\} = \\ &= E\{a_1 + a_2k + a_3k^2 \mid Z_1, Z_2, \dots, Z_T\} \text{ (com a origem transladada, isto é, fazendo } T = 0) \\ &= \hat{a}_1(T) + k \cdot \hat{a}_2(T) + k^2 \cdot \hat{a}_3(T)\end{aligned}$$

Analogamente aos modelos constante e linear, existem diversos métodos para a estimação dos parâmetros que especificam o nível da série. Aqui, ao contrário dos casos anteriores, não iremos exibir em detalhes as equações de atualização.

1) Estimadores de Mínimos Quadrados Ordinários (MQO)

Devemos encontrar os estimadores de a_1 , a_2 e a_3 que minimizem:

$$SSE = \sum_{t=1}^T [Z_t - \underbrace{(a_1 + a_2t + a_3t^2)}_{\varepsilon_t}]^2$$

2) Médias Móveis Triplas

Defina a média móvel tripla de tamanho N avaliada no instante T como:

$$M_T^{[3]} = \frac{M_T^{[2]} + M_{T-1}^{[2]} + \dots + M_{T-N+1}^{[2]}}{N} \quad (5.2.20)$$

Onde $M_T^{[2]}$ é a média móvel dupla de tamanho N avaliada no instante T .

Os estimadores dos parâmetros do modelo são obtidos como funções das médias móveis simples, dupla e tripla.

$$\hat{a}_i(T) = f_i [M_T, M_T^{[2]}, M_T^{[3]}]; i = 1, 2, 3$$

3) Amortecimento Exponencial

$$M_T = \alpha Z_T + (1 - \alpha)M_{T-1}$$

$$M_T^{[2]} = \alpha M_T + (1 - \alpha)M_{T-1}^{[2]}$$

$$M_T^{[3]} = \alpha M_T^{[3]} + (1 - \alpha)M_{T-1}^{[3]}$$

Onde $\alpha \in [0, 1]$ é a única constante de amortecimento usada para atualizar todas as médias móveis.

Os estimadores dos parâmetros a_1 , a_2 e a_3 (correspondentes à origem transladada para o instante T) no instante T são obtidos através das equações:

$$M_T = \hat{a}_1(T) - \left(\frac{1 - \alpha}{\alpha}\right)\hat{a}_2(T) + \frac{(1 - \alpha)(2 - \alpha)}{2\alpha^2}\hat{a}_3(T)$$

$$M_T^{[2]} = \hat{a}_1(T) - 2\left(\frac{1 - \alpha}{\alpha}\right)\hat{a}_2(T) + \frac{2(1 - \alpha)(3 - 2\alpha)}{2\alpha^2}\hat{a}_3(T)$$

$$M_T^{[3]} = \hat{a}_1(T) - 3\left(\frac{1 - \alpha}{\alpha}\right)\hat{a}_2(T) + \frac{3(1 - \alpha)(4 - 3\alpha)}{2\alpha^2}\hat{a}_3(T)$$

5.3. Métodos de Amortecimento Exponencial (séries sazonais)

Séries sazonais ocorrem com enorme freqüência na prática. Por exemplo, séries de temperatura média mensal e séries de consumo de energia elétrica são claramente sazonais, e exibem características que se repetem ano a ano sempre no mesmo mês (ou semana, ou trimestre, ...). No caso de séries de consumo de energia no Brasil, existe um nítido aumento de consumo nos meses de verão, devido às elevadas temperaturas e ao uso de ar refrigerado. Entender a sazonalidade de uma série é fundamental não apenas para gerar previsões consistentes, mas também para propor políticas governamentais ou empresariais. Por exemplo, o horário de verão no Brasil é conseqüência direta do aumento de consumo observado neste período, e tem o objetivo de reduzir o consumo e evitar "black-outs" cujas conseqüências em termos econômicos poderiam ser dramáticas. A sazonalidade pode ser modelada de duas formas: através de fatores sazonais ou através de funções trigonométricas. Aqui estaremos interessados apenas na modelagem através de fatores sazonais.

As técnicas já estudadas de amortecimento exponencial podem ser estendidas para modelar séries que apresentam tendência e sazonalidade.

Um dos modelos mais utilizados na prática (especialmente em empresas) é o modelo de **amortecimento exponencial de Holt e Winters**. Este procedimento é capaz de modelar séries sazonais e gera, na maioria dos casos, boas previsões, a um custo computacional bastante baixo. O modelo de Holt e Winters utiliza a idéia de amortecimento das informações passadas três vezes para:

- estimar o nível da série;
- estimar a taxa de crescimento da série;
- estimar os fatores sazonais.

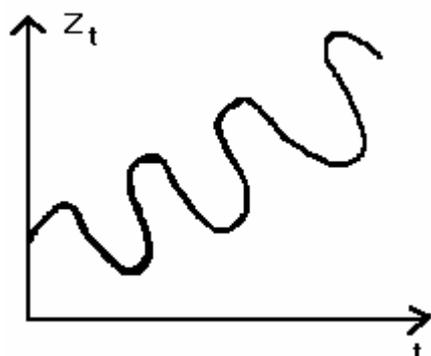
Em cada um dos passos indicados emprega-se uma constante de amortecimento diferente.

Seja S o período sazonal, por exemplo, $S = 12$ no caso de dados mensais, $S = 4$ no caso de observações trimestrais. Na discussão a seguir iremos nos referir, com freqüência, ao “mês” t , mas todos os resultados se aplicam a períodos arbitrários, e não se referem exclusivamente a dados mensais. Denote por ρ_t o **fator sazonal correspondente** ao mês t , onde $t = 1, 2, \dots, S$.

Existem dois tipos de modelos para a sazonalidade: **modelos aditivos e multiplicativos**. Nos modelos aditivos, o padrão sazonal não se altera à medida que o nível da série muda. Ao contrário, nos modelos multiplicativos, a sazonalidade da série é afetada pelo nível da mesma, como indicado nos gráficos a seguir.

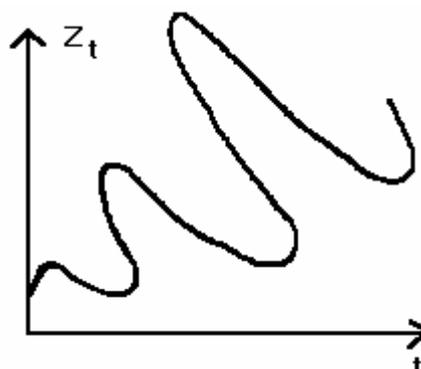
MODELO ADITIVO

$$Z_t = \mu(t) + \rho_t + \epsilon_t$$



MODELO MULTIPLICATIVO

$$Z_t = \mu(t) \cdot \rho_t \cdot \epsilon_t \text{ ou } Z_t = \mu(t) \cdot \rho_t + \epsilon_t$$



5.3.1. Modelo Horizontal (constante) Sazonal Multiplicativo

Neste caso o nível da série ($\mu(t)$) é uma constante, e a estrutura do modelo é:

$$Z_t = a_1 \cdot \rho_t + \epsilon_t \quad (5.3.1)$$

Neste caso precisamos estimar o nível da série (a_1) e os fatores sazonais ρ_t para todos os “meses” t . Note que, no caso de dados mensais, precisamos estimar 12 fatores sazonais e o nível, ou seja, existem 13 parâmetros a serem estimados. No caso de dados trimestrais, existem 4 fatores sazonais (e também o nível), o que resulta na estimação de 5 parâmetros. Na verdade, se S é o período sazonal, existem S fatores sazonais a serem estimados, mas a soma destes fatores está sujeita a uma restrição, e na prática o número de fatores sazonais a serem estimados é $S - 1$. Logo, na prática, no modelo mensal devemos estimar 11 fatores sazonais (pois o remanescente é obtido a partir da restrição).

A previsão k passos à frente feita no instante T é dada por:

$$\hat{Z}_T(k) = E\{Z_{T+k} | \underline{Z}_T\} = \hat{a}_1(T) \cdot \hat{\rho}_{T+k}(T) \quad (5.3.2)$$

É importante conhecer o significado dos estimadores de a_1 e ρ_t na equação (5.3.2).

$\hat{a}_1(T)$ indica o estimador do nível da série usando todas as observações até o instante T (inclusive), ou seja, é a estimativa mais atualizada do nível da série.

$\hat{\rho}_{T+k}(T)$ é o estimador do fator sazonal correspondente ao mês T + k no futuro efetuado usando todas as observações até o instante T (inclusive). Ou seja, $\hat{\rho}_{T+k}(T)$ é a estimativa mais atualizada do fator sazonal que corresponde ao mês T + k.

Os parâmetros do modelo são atualizados seqüencialmente da forma indicada a seguir. Suponha que, para um instante genérico T, as estimativas de a_1 e do fator sazonal para o instante anterior estão disponíveis (isto é, são conhecidas), ou seja, $\hat{a}_1(T-1)$ e $\hat{\rho}_{m(T)}(T-1)$ são dados, onde $m(T)$ é o “mês” correspondente ao instante T. Então, os parâmetros atualizados no instante T são encontrados a partir das seguintes fórmulas:

$$\hat{a}_1(T) = \alpha \left[\frac{Z_T}{\hat{\rho}_{m(T)}(T-1)} \right] + (1-\alpha) \cdot \hat{a}_1(T-1) \quad (5.3.3)$$

$$\hat{\rho}_{m(T)}(T) = \gamma \left[\frac{Z_T}{\hat{a}_1(T)} \right] + (1-\gamma) \cdot \hat{\rho}_{m(T)}(T-1) \quad (5.3.4)$$

Note que, nas equações (5.3.3) e (5.3.4) são usadas duas constantes de amortecimento diferentes, uma para o nível e outra para o fator sazonal.

Além disso, uma característica importante da atualização do fator sazonal é: ele só muda quando recebemos uma informação referente ao “mês” $m(T)$. Por exemplo, o fator sazonal referente ao mês de maio só é atualizado quando recebemos uma informação relativa a este mês. Ou seja:

$$\hat{\rho}_j(T) = \hat{\rho}_j(T-1) \text{ para } j = 1, 2, \dots, S \text{ e } j \neq m(T)$$

Em seguida, os fatores sazonais estimados são normalizados, de tal forma que a sua soma seja igual ao período sazonal (S). Esta condição de normalização é válida para fatores sazonais multiplicativos. No caso de fatores aditivos, sua soma (para todos os meses) é zero.

$$\sum_{j=1}^S \hat{\rho}_j(T) = S \quad (5.3.5)$$

É necessário fornecer estimativas iniciais do nível e dos fatores sazonais. O procedimento de inicialização é descrito brevemente a seguir.

1º PASSO: Cálculo das Médias Anuais

Ano 1	Z_1	Z_2	...	Z_S	Média do Ano 1 →	$\bar{Z}_{(1)}$	J = T/S
Ano 2	Z_{S+1}	Z_{S+2}	...	Z_{2S}	Média do Ano 2 →	$\bar{Z}_{(2)}$	
.....		
Ano J	$Z_{JS-(S-1)}$	$Z_{JS-(S-2)}$...	Z_{JS}	Média do Ano J →	$\bar{Z}_{(J)}$	

2º PASSO: Cálculo dos fatores sazonais grosseiros C_i

$$C_1 = \frac{Z_1}{Z_{(1)}} \quad \dots \quad C_S = \frac{Z_S}{Z_{(1)}}$$

$$\dots \quad \dots \quad \dots$$

$$C_{JS-(S-1)} = \frac{Z_{JS-(S-1)}}{Z_{(J)}} \quad \dots \quad C_{JS} = \frac{Z_{JS}}{Z_{(J)}}$$

O fator sazonal grosseiro para o mês i é apenas a observação Z_i dividida pela média do ano correspondente. Este procedimento simples funciona bem desde que a série não apresente componente de tendência. Do contrário, se existe tendência, as estimativas dos fatores sazonais encontradas por este método irão conter componentes de tendência, e as previsões resultantes serão prejudicadas.

Note que, neste passo, encontramos J (um para cada ano em que existem dados) fatores sazonais para cada mês. No passo seguinte tomamos a média destes J fatores para obter um único fator sazonal para o mês i .

3º PASSO: Cálculo dos fatores sazonais "suaves"

Os fatores sazonais para cada um dos meses são encontrados através da média dos fatores encontrados no passo anterior, isto é:

$$C_1^* = [C_1 + C_{S+1} + \dots + C_{JS-(S-1)}] / J$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$C_S^* = [C_S + C_{2S} + \dots + C_{JS}] / J$$

Por exemplo, o fator sazonal correspondente ao mês 1 é a média dos fatores sazonais previamente estimados para os meses 1, 13, 25, 37,

4º PASSO: Normalização dos fatores sazonais "suaves"

A partir dos fatores sazonais "suaves" encontrados no passo anterior, calcula-se os fatores sazonais normalizados simplesmente aplicando a restrição de que a soma dos fatores sazonais deve ser igual a S (período sazonal). Ou seja, o fator sazonal normalizado correspondente ao mês i é apenas:

$$\hat{\rho}_i(0) = \frac{S}{\sum_{i=1}^S C_i^*} \cdot C_i^* \quad \text{onde } C_i^* \text{ é o fator sazonal "suave" calculado no passo anterior.}$$

5º PASSO: Obtenção da estimativa inicial do nível

A partir dos fatores sazonais normalizados $\hat{\rho}_i(0)$ encontrados no passo anterior, e supondo que o nível inicial estimado seja igual à média do primeiro ano de dados, $\hat{a}_1(0) = \bar{Z}_{(1)}$, pode-se inicializar o processo de estimação. Em seguida usa-se as equações (5.3.3) e (5.3.4) para obter as estimativas do nível e fatores sazonais em cada instante.

5.3.2. Modelo Linear Sazonal Multiplicativo

Neste caso o nível da série ($\mu(t)$) é uma função linear, e a estrutura do modelo é:

$$Z_t = (a_1 + a_2 \cdot t) \cdot \rho_t + e_t \quad (5.3.6)$$

Neste caso precisamos estimar o nível da série (a_1), sua taxa de crescimento (a_2) e os fatores sazonais ρ_t para todos os “meses” t . A atualização de parâmetros é feita seqüencialmente, pelo **método de Holt-Winters** descrito a seguir.

Suponha que no instante $T-1$ conhecemos as estimativas do nível, taxa de crescimento e fatores sazonais dadas por: $\hat{a}_1(T-1)$, $\hat{a}_2(T-1)$ e $\hat{\rho}_j(T-1)$. O procedimento de atualização é bem semelhante ao apresentado antes (modelo com nível constante e sazonalidade multiplicativa). A grande diferença em relação ao modelo da seção 5.3.1. é que agora temos mais uma quantidade a ser atualizada seqüencialmente, a taxa de crescimento a_2 .

- Atualização do nível (a_1)

$$\hat{a}_1(T) = \alpha \left[\frac{Z_T}{\hat{\rho}_{m(T)}(T-1)} \right] + (1-\alpha)[\hat{a}_1(T-1) + \hat{a}_2(T-1)] \quad (5.3.7)$$

Note que, do lado direito da equação (5.3.7), o termo $\frac{Z_T}{\hat{\rho}_{m(T)}(T-1)}$ representa a observação do instante T dessazonalizada.

- Atualização da tendência

$$\hat{a}_2(T) = \beta \cdot [\hat{a}_1(T) - \hat{a}_1(T-1)] + (1-\beta) \cdot \hat{a}_2(T-1) \quad (5.3.8)$$

A equação de atualização da tendência é exatamente igual à do método de Holt (usado para séries com tendência linear e sem sazonalidade), como indicado na expressão (5.2.19). Note que o termo $\hat{a}_1(T) - \hat{a}_1(T-1)$ é a diferença entre os níveis estimados nos instantes T e $T-1$.

- Atualização dos fatores sazonais $\hat{\rho}_j(T)$

$$\hat{\rho}_{m(T)}^* = \gamma \left[\frac{Z_T}{\hat{a}_1(T)} \right] + (1-\gamma) \hat{\rho}_{m(T)}(T-1) \quad (5.3.9)$$

$$\hat{\rho}_j^*(T) = \hat{\rho}_j(T-1); j = 1, 2, \dots, S; j \neq m(T) \quad (5.3.10)$$

A atualização dos fatores sazonais é feita apenas no mês correspondente ao fator sazonal que se procura estimar. Logo, pela equação (5.3.10) segue que o fator sazonal permanece fixo até recebermos a informação relativa ao mês atual. Quando esta informação é recebida (isto é, quando o instante T corresponde ao mês $m(T)$), o fator sazonal é estimado através da expressão (5.3.9). Nesta equação nota-se que o fator sazonal correspondente ao mês $m(T)$ atualizado no instante T é uma

média ponderada da estimativa feita no ano passado para o fator deste mês e também da observação sem o nível (dada por $\frac{Z_T}{\hat{a}_1(T)}$).

- Normalização (padronização) dos fatores sazonais

$$\hat{\rho}_j(T) = \left[\frac{\hat{\rho}_j^*(T)}{\sum_{j=1}^S \hat{\rho}_j^*(T)} \right] \cdot S$$

O objetivo desta normalização é fazer com que os fatores sazonais encontrados pela equação (5.3.9) tenham soma igual a S (período sazonal).

A previsão k passos à frente no instante T é obtida substituindo-se os parâmetros desconhecidos a_1 , a_2 e ρ por suas estimativas, isto é:

$$\hat{Z}_{T+k}(T) = [\hat{a}_1(T) + k \cdot \hat{a}_2(T)] \hat{\rho}_{m(T+k)}(T) \quad (5.3.11)$$

- Inicialização do Método

A exemplo do que fizemos na seção anterior, aqui encontramos estimativas iniciais para o nível, taxa de crescimento e fatores sazonais.

Sejam $\bar{Z}_{(1)}, \bar{Z}_{(2)}, \dots, \bar{Z}_{(J)}$ as médias anuais, onde J é o último ano disponível.

A estimativa inicial da taxa de crescimento é:

$$\hat{a}_2(0) = \frac{\bar{Z}_{(J)} - \bar{Z}_{(1)}}{(J-1)S}$$

A estimativa inicial do nível é:

$$\hat{a}_1(0) = \bar{Z}_{(1)} - \frac{(S+1)}{2} \hat{a}_2(0)$$

A estimativa inicial dos fatores sazonais é:

$$C_t = \left\{ \frac{Z_t}{\bar{Z}_{(i)} - \left(\frac{S+1}{2} - m(t) \right) \hat{a}_2(0)} \right\} \quad \text{onde } t = 1, 2, \dots, JS \text{ e } i \text{ é o ano correspondente ao instante } t.$$

Em seguida o procedimento é igual ao da seção anterior.

5.3.3. Modelo Linear Sazonal Aditivo

A estrutura do modelo é: $Z_t = a_1 + a_2 t + \rho_t + e_t$.

A atualização seqüencial dos parâmetros é análoga às equações (5.3.7) a (5.3.10), com pequenas modificações, mostradas abaixo.

- Atualização do nível (a_1)

$$\hat{a}_1(T) = \alpha[Z_T - \hat{\rho}_{m(T)}(T-1)] + (1-\alpha)[\hat{a}_1(T-1) + \hat{a}_2(T-1)] \quad (5.3.12)$$

Note que, do lado direito da equação (5.3.12), o termo $Z_T - \hat{\rho}_{m(T)}(T-1)$ representa a observação do instante T dessazonalizada.

- Atualização da tendência

$$\hat{a}_2(T) = \beta.[\hat{a}_1(T) - \hat{a}_1(T-1)] + (1-\beta).\hat{a}_2(T-1) \quad (5.3.13)$$

A equação de atualização da tendência é exatamente igual à do modelo multiplicativo e à do método de Holt (expressões (5.3.8) e (5.2.19)).

- Atualização dos fatores sazonais $\hat{\rho}_j(T)$

$$\hat{\rho}_{m(T)}^* = \gamma[Z_T - \hat{a}_1(T)] + (1-\gamma)\hat{\rho}_{m(T)}(T-1) \quad (5.3.14)$$

$$\hat{\rho}_j^*(T) = \hat{\rho}_j(T-1); j = 1, 2, \dots, S; j \neq m(T) \quad (5.3.15)$$

Note que, em (5.3.14), o termo $[Z_T - \hat{a}_1(T)]$ é uma medida da variação sazonal dos dados, obtida pela diferença entre a observação atual e a estimativa mais recente do nível.

Da equação (5.3.15) conclui-se que as estimativas de cada fator sazonal só são atualizadas uma vez a cada ano, quando observamos o dado correspondente àquele mês.

A previsão 1 passo à frente no instante T é obtida substituindo-se os parâmetros desconhecidos a_1 , a_2 e ρ por suas estimativas, isto é:

$$\hat{Z}_{T+1}(T) = [\hat{a}_1(T) + \hat{a}_2(T)] + \hat{\rho}_{m(T+1)}(T) \quad (5.3.16)$$

É importante lembrar que os fatores sazonais aditivos sofrem uma restrição diferente dos fatores multiplicativos. No caso dos fatores aditivos, a sua soma (para todos os meses) é zero.

Inicialização do método

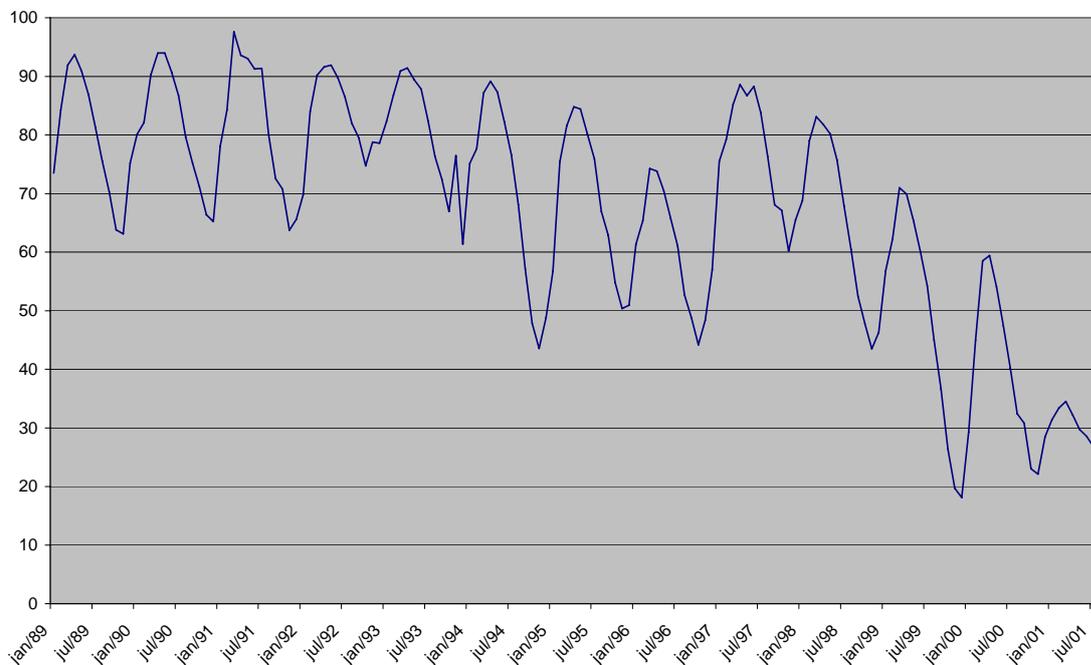
Montgomery e Johnson sugerem obter as estimativas iniciais dos parâmetros através de um procedimento de mínimos quadrados ordinários, que permite a estimação simultânea dos $S + 2$ parâmetros desconhecidos.

Requisitos mínimos para usar o método de Holt e Winters

Para modelar adequadamente a sazonalidade, é necessário pelo menos 3 anos de dados mensais e 4 anos de dados trimestrais.

Exemplo 5.3.1.

O próximo gráfico exibe o nível dos reservatórios (isto é, a combinação das usinas hidroelétricas) na região Sudeste do Brasil entre Janeiro de 1989 e Agosto de 2001.

Nível dos Reservatórios - Sudeste

Os resultados de um modelo Holt-Winters ajustado a esta série são:

Componente	Constante Amortec.	Valor Final
Nível	0.540	25.109
Tendência	0.002	-0.179
Sazonalidade	0.529	

Os fatores sazonais multiplicativos estimados são:

Fatores Sazonais	
Jan	1.028
Fev	1.181
Mar	1.282
Abr	1.241
Mai	1.176
Jun	1.126
Jul	1.052
Ago	0.936
Set	0.877
Out	0.751
Nov	0.727
Dez	0.829

Este modelo foi então ajustado até Dezembro de 2000 e as previsões obtidas para o período entre Janeiro e Agosto de 2001 para testar sua capacidade preditiva. A próxima tabela apresenta os valores reais e ajustados no período, além dos erros percentuais de previsão.

Data	Limite Inferior do IC 95%	Previsão	Limite Superior do IC 95%	Valor Real	Erro de previsão (em %)
jan/01	24.3	36.6	48.9	31.4	-5.2
fev/01	27.4	43.2	59.0	33.4	-9.8
mar/01	28.1	46.6	65.1	34.5	-12.1
abr/01	24.6	44.2	63.7	32.2	-12.0
mai/01	20.7	40.9	61.2	29.7	-11.2
jun/01	17.2	38.1	59.0	28.6	-9.6
jul/01	13.7	35.0	56.3	26.8	-8.2
ago/01	9.4	30.9	52.3	23.5	-7.4

Os erros são bastante grandes, o que certamente é consequência de uma mudança no padrão de evolução dos reservatórios, que se torna nítida a partir de 2001. Note que, neste último ano, a amplitude no nível de reservatórios é bem menor que nos anos anteriores, indicando que não foi possível recuperar o nível dos reservatórios com as chuvas de verão. Portanto, (pelo menos a posteriori) é clara a necessidade de impor o racionamento de energia (iniciado em Junho de 2001).

Se agora consideramos apenas as previsões um passo à frente, a capacidade preditiva do modelo é sensivelmente melhor, como indicado na próxima tabela. Os erros, especialmente nos meses de inverno, são sensivelmente reduzidos.

Previsões um passo à frente					
Data	Limite Inferior do IC 95%	Previsão	Limite Superior do IC 95%	Valor Real	Erro de previsão (em %)
jan/01	24.3	36.6	48.9	31.4	-5.2
fev/01	24.9	39.2	53.5	33.4	-5.8
mar/01	23.3	38.6	53.9	34.5	-4.1
abr/01	20.1	34.7	49.3	32.2	-2.5
mai/01	17.5	31.1	44.8	29.7	-1.4
jun/01	15.7	28.5	41.3	28.6	0.1
jul/01	14.5	26.4	38.2	26.8	0.4
ago/01	13.0	23.6	34.2	23.5	-0.1