



Revista de Estatística e Sistemas de Gestão

ISSN: 0972-0510 (Impresso) 2169-0014 (Online) Página inicial da revista: <https://www.tandfonline.com/loi/tsms20>

Revisão metodológica estatística para dados de séries temporais

Deusa Rahardja

Para citar este artigo: Dewi Rahardja (2020): Revisão metodológica estatística para dados de séries temporais, Journal of Statistics and Management Systems, DOI: 10.1080/09720510.2020.1727618

Para acessar este artigo: <https://doi.org/10.1080/09720510.2020.1727618>



Publicado on-line: 26 de maio de 2020.



Envie seu artigo para esta revista [↗](#)



Ver artigos relacionados [↗](#)



Ver dados de marca cruzada [↗](#)

Revisão metodológica estatística para dados de séries temporais

Dewi Rahardja

*Departamento de Defesa dos
EUA Fort Meade Maryland*

20755
cervo

Resumo

Numerosas literaturas sobre métodos estatísticos para dados de séries temporais (TS) foram publicadas. Neste artigo, uma literatura dos métodos de análise de dados TS é revisada. Organizamos a revisão com base na categoria básica de três famílias de modelos TS: a família Exponential Smoothing Model (ESM), a família de modelos Auto-Regressive Integrated Moving Average (ARIMA) e a família Unobserved Component Model (UCM). Um roteiro é fornecido em um formato de diagrama para esses métodos TS que são traduzíveis em instruções de computação atuais. Além disso, a execução desses métodos em comandos SAS (como um dos pacotes de software estatísticos mais populares atualmente) também é apresentada. Este artigo será muito benéfico para profissionais, previsores e pesquisadores em diversas áreas de estudo (como negócios, administração, finanças, economia, etc.) usar.

Classificação de Assuntos: (2010) 37M10 *ÿ* Análise de Séries Temporais; 62M10 – Séries temporais, autocorrelação, regressão, etc.; 91B84 *ÿ* Análise de Séries Temporais Econômicas.

Palavras-chave: Autocorrelação, Previsão, Séries Temporais, Modelo de Suavização Exponencial (ESM), Média Móvel Integrada Auto Regressiva (ARIMA), Modelo de Componentes Não Observados (UCM).

1. Introdução

Vários profissionais, incluindo Previsores Federais, Industriais e Acadêmicos e/ou Pesquisadores, participam da Conferência Federal de Previsão (FFC) semestralmente em Washington, DC.

Declaração de isenção de responsabilidade: Esta pesquisa representa o próprio trabalho e opinião do autor. Não reflete nenhuma política nem representa a posição oficial do Departamento de Defesa dos EUA nem de qualquer outra agência federal dos EUA.

E-mail: rahardja@gmail.com

uma série de conferências que começaram há três décadas em 1988 e trouxeram amplo reconhecimento à importância da previsão como uma importante atividade estatística dentro do Governo Federal, entre suas organizações parceiras e inúmeros profissionais (industriais, acadêmicos e federais). Portanto, essas atividades da conferência mostraram as demandas das necessidades de previsão (em armazenamento de dados baseado em nuvem, previsões orientadas por dados, disponibilidade de métodos, disponibilidade de implementação de software de computação, etc.) desde os últimos 30 anos, [ou seja, como os profissionais implementam o Métodos de Previsão existentes computacionalmente por meio de pacotes de software existentes, especialmente SAS, como o software de propósito orientado a dados mais popular atualmente (em oposição ao software de simulação baseado em simulação) e software estatístico comercial comumente usado]. Assim, à medida que praticantes, analistas, pesquisadores ou previsores orientados a dados exploram a disponibilidade de métodos de previsão e software de computação, há questões importantes sobre a generalização, especificidade e capacidade computacional atual para implementar as previsões ou projeções. Portanto, este artigo será útil para vários profissionais que precisam fazer previsões e projeções válidas e justificáveis em seus campos de aplicação (como ciências sociais, ciências físicas, economia, finanças, negócios, etc.)

Em aplicações práticas, quando as suposições de observações independentes e identicamente distribuídas (*iid*) são violadas, então os métodos básicos **de 3 tipos de dados** mais comuns discutidos em Rahardja *et. al.* (2016; 2017; 2018) não são adequados para serem usados para tais dados ou observações *não iid*. Por exemplo, nos cenários em que os dados exibem correlações temporais (relacionadas ao tempo ou dependentes do tempo) ou espaciais (relacionadas ao espaço ou dependentes do espaço). Este artigo discute o primeiro cenário: quando os dados exibem correlações temporais (ou seja, observações ou dados não são independentes ou autocorrelacionados no tempo, mas são distribuídos de forma idêntica), que é referido na literatura como **Dados de Séries Temporais (TS)**. Observe que vários exemplos de casos em que observações ou dados são distribuídos de forma independente, mas não idêntica, são discutidos em Rahardja (2005), como o quarto caso do artigo; e outro exemplo está no artigo de Zhou (2017). No entanto, esse cenário de dados independentes, mas não distribuídos de forma idêntica, não será discutido neste documento, pois está fora do escopo desta discussão sobre dados de TS.

Geralmente, no escopo da discussão de dados TS, existem 3 razões principais pelas quais as suposições de dados *iid* podem ser violadas: 1) o sorteio de um ponto de dados influencia o resultado de um sorteio subsequente (interdependências ou autocorrelacionado), 2) a distribuição muda em algum ponto (não estacionaridade),

e 3) os dados não são gerados por uma distribuição (adversarial). Este artigo se concentra em (1) e (2), que é o escopo do TS Data.

Para dados TS, a variável de *resultado* (ou *resposta* ou *dependente*) de interesse não é independente, mas identicamente distribuída, séries autocorrelacionadas de tempo igual (como por hora, diariamente, semanalmente, mensalmente, anualmente, década, etc.) . Análise de previsão é outra terminologia comumente usada para análise de dados TS.

Métodos estatísticos tradicionais (padrão/clássicos) que se baseiam em suposições do *iid* , por exemplo, os métodos de análise de dados **contínuos** discutidos em Rahardja (2017), os métodos de análise de dados **categóricos** discutidos em Rahardja, *et. al.* (2016), e os métodos **de tempo até o evento (TTE)** ou Análise de Sobrevivência/Falha/Confiabilidade discutidos em Rahardja e Wu (2018) dependem das suposições do *iid* e, portanto, não são adequados para incluir os métodos autocorrelacionados suposição de dados como o resultado (ou resposta) no cenário TS. A metodologia de análise deve contabilizar corretamente essas observações autocorrelacionadas, no tempo. Portanto, esses conjuntos de dados só podem ser analisados por meio de métodos TS. Até o momento, não há literatura atual atualizada que resuma de forma concisa e metódica a revisão dos métodos TS para o tipo de variável de resposta (ou medida de resultado) de dados autocorrelacionados (ou TS); e junto com como implementá-los em softwares de computação atuais (atualizados), como o SAS. Observe que essa implementabilidade atualizada do SAS nem sempre está disponível por meio da documentação on-line do SAS. Muitas vezes, alguns procedimentos SAS só podem ser aprendidos por meio de cursos de treinamento SAS pagos no local. Mesmo assim, muitas vezes, o instrutor nem sempre consegue responder a todas as perguntas, pois normalmente não são eles que escrevem as sub-rotinas e às vezes não conseguem encontrar documentações/referências internas. Portanto, muitas vezes, usuários ou profissionais precisam recorrer a trabalhos de pesquisa ou qualquer uma das formas acima, como o autor tem feito e, portanto, está prescrevendo os resultados aqui para que outros reinventar as rodas.

Neste artigo, apresentamos a categoria básica de três famílias de modelos TS juntamente com a disponibilidade de procedimentos de implementação (PROC's) no software SAS, o software estatístico mais utilizado. A revisão inclui a categoria básica de três famílias de modelos TS: a família Exponential Smoothing Model (ESM), a família de modelos Auto-Regressive Integrated Moving Average (ARIMA) e a família Unobserved Component Model (UCM). Sypsas (1989) e De Gooijer e Hyndman (2006), entre muitas outras revisões de literatura mais antigas, discutiram apenas os modelos da família ESM e da família ARIMA, e mencionaram brevemente o Modelo Estrutural Básico (BSM) como outra representação dos modelos da família UCM; mas não

discussão detalhada nem tradução para implementação do modelo via software estatístico (como SAS) procedimento (PROC) para a viabilidade/implementação da computação. Portanto, em nossa linha de prática (estatística), muitas vezes encontramos muitos estatísticos e não estatísticos, previsores e pesquisadores que ficam confusos/misturados sobre o método, presos no modelo, mas não têm certeza de qual procedimento de computação de software para traduzir/implementar/executar. Para fechar essa lacuna de tradução, este artigo será uma orientação / roteiro básico muito prático e útil para estatísticos e não estatísticos em vários campos de estudo.

2. Modelos TS de três famílias

Fundamentalmente, existem três categorias básicas de modelos TS: a família Exponential Smoothing Model (ESM), a família de modelos Auto-Regressive Integrated Moving Average (ARIMA) e a família Unobserved Component Model (UCM). A maior parte do ESM também pode ser expressa em termos do modelo ARIMA.

Primeiro, começamos com os **Modelos de Suavização Exponencial (ESM)** como a primeira família de Modelos TS. A ideia básica do ESM é ajustar as previsões de valores futuros como médias ponderadas de observações passadas. Observações recentes têm mais peso na determinação de previsões do que observações no passado distante. Um método de “suavização” implica essa abordagem de médias ponderadas. O adjetivo “exponencial” deriva do fato de que alguns dos ESMs não apenas têm pesos que diminuem com o tempo, mas o fazem de forma exponencial. Listadas na Tabela 1 estão as versões mais conhecidas do ESM (que também são traduzidas em software de computação SAS). O termo ARIMA, Autoregressive Integrated Moving Average é explicado no próximo parágrafo; no entanto, será usado brevemente neste parágrafo do ESM para incluir a expressão do modelo de equivalência ARIMA. Na Tabela 1, como podemos ver, o ESM Simples (Single) também pode ser expresso como modelo ARIMA (0, 1, 1); o ESM Duplo ou Brown (1959 e 1962) também pode ser expresso como modelo ARIMA (0, 2, 2); o ESM Linear ou Holt (1957) também pode ser expresso como modelo ARIMA (0, 1, 2); o Damped Trend ESM discutido em Gardner e McKenzie (1985) também pode ser expresso como modelo ARIMA (1, 1, 2) e este método “amortece” a tendência do ESM Linear (Holt) para uma linha plana em algum momento no futuro; o ESM Add-Sazonal ou Sazonal Aditivo ou Sazonal pode ser expresso como ARIMA (0, 1, 1) (0, 1, 1)_m onde m é o número de períodos por temporada, ou ARIMA (p, 0, 0) (0, 1, 1) + c onde p = 1, 2, 3 e c é uma constante; o ESM Multi-Sazonal ou Multiplicativo Sazonal não possui Equivalência ARIMA; os invernos

(1960) ou Winters Multiplicative ESM não tem equivalência ARIMA; e o Add-Winters ou Winters Additive ESM também pode ser expresso como ARIMA $(0, 1, [p+1])(0, 1, 0)^p$. Tanto o ESM quanto sua equivalência ARIMA produzirão aproximadamente as mesmas projeções de séries temporais. Fomby (2008) descreve esses modelos ESM em mais detalhes, com base no modelo escrito no SAS Manual (1995), para os 4 primeiros ESM na Tabela 1. Hyndman e Athanasopoulos (2018) generalizaram as fórmulas matemático-estatísticas para os 4 últimos ESM discutido na Tabela 1.

A segunda família TS é o modelo **ARIMA**. ARIMA significa **Média Móvel Integrada Autoregressiva**. Os modelos ARIMA são os modelos gerais de previsão TS que decompõem ou filtram os atributos das séries temporais em termos de *parâmetros* (p, d, q e P, D, Q), usando o **algoritmo de decomposição espectral**. ARIMA expressa valores de uma série temporal dependente com uma combinação de seu componente auto-regressivo (p), diferenciação (d), componente de média móvel (q) e outras séries temporais *preditoras*. Assim, um modelo genérico ARIMA consiste em um termo autoregressivo (AR), um termo de média móvel (MA) e, para uma série temporal que requer diferenciação, um elemento integrado (I). Além disso, um modelo ARIMA abrangente pode ter componentes não sazonais e sazonais, cada um caracterizado por três parâmetros, (p, d, q) para não sazonal e (P, D, Q) s para sazonal. Como tal, um modelo ARIMA é geralmente denotado / abreviado por ARIMA $(p, d, q) (P, D, Q)_s$, onde p, d, q, P, D, Q são inteiros não negativos e s é uma equação de 12 meses; e $s = 4$ meses para uma temporada de dados trimestrais).

Um modelo ARIMA também pode incluir um ou mais preditores. Teoricamente, a significância dos preditores precisa ser estatisticamente testada antes de ajustar uma previsão. Praticamente (em termos de interpretação), também é importante ressaltar que os preditores só devem ser incluídos em um modelo quando houver uma razão sensata para acreditar que o preditor influencia o resultado. Assim, para a inclusão de qualquer preditor em um modelo ARIMA, ambas as perspectivas (teórica e prática) devem concordar.

A terceira família TS é o **Unobserved Components Model (UCM)**. Os UCMs também são chamados de **Modelos Estruturais** na literatura de séries temporais, por exemplo, veja o livro de Harvey (1989). No software SAS, o UCM decompõe ou filtra a série de respostas em “componentes” usando o **algoritmo de filtragem de Kalman** (em vez de expressar as equações do modelo como “parâmetros” como no cenário ARIMA) de uma maneira aditiva conveniente, como tendência (interceptação e inclinação), sazonalidade, ciclos, componentes auto-regressivos, um termo de regressão ou efeitos envolvendo as defasagens das variáveis dependentes, devido a séries preditoras, e o Irregular

Componente, também chamado de Termo de Perturbação, normalmente é assumido como Ruído Branco Gaussiano com sua variância correspondente. A soma dos termos de regressão (envolvidos como preditores) inclui a contribuição das variáveis de regressão com coeficientes de regressão *fixos*.

Para a notação de UCM, o UCM totalmente especificado é escrito como

$$y_t = \mu + \alpha t + \beta t^2 + \gamma_t + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^m b_{jt} x_{jt} + e_t \quad (1)$$

Na equação (1), o componente do lado esquerdo y_t representa a resposta ou série temporal a ser modelada ou prevista. No lado direito da equação (1), o primeiro componente μt representa o componente de tendência – que pode ser dividido em dois componentes: um intercepto (ou nível) e uma inclinação (ou tendência); o segundo componente γ_t representa o componente de sazonalidade; a terceira componente $\sum_{i=1}^p \phi_i y_{t-i}$ representa a componente cíclica; o quarto componente $\sum_{j=1}^m b_{jt} x_{jt}$ representa o componente autorregressivo; o quinto componente representa o i -ésimo termo de regressão (auto) envolvendo as defasagens das variáveis dependentes; o sexto componente representa o j -ésimo termo dos preditores ou regressores; e e_t o componente irregular (ou termo de erro de ruído branco). Todos esses componentes são considerados **não observados** (parâmetros **desconhecidos**) e deve ser **estimado** considerando os dados da série temporal em y_t e x_{jt} , portanto, dê o título de modelo de componentes não observados (UCM). Em nota adicional, a equação (1) permite a inclusão do autorregressivo

termos de regressão $\sum_{i=1}^p \phi_i y_{t-i}$ e os termos explicativos da regressão $\sum_{j=1}^m b_{jt} x_{jt}$,

o primeiro representando o *momento* da série temporal em relação às suas observações passadas e o último representando os fatores causais que se está disposto a supor que afetam a série temporal em questão.

Os componentes do modelo UCM devem capturar as características salientes da série que são úteis para explicar e prever seu comportamento. Esses componentes são então testados sequencialmente para determinar se são **determinísticos (fixos)** ou **estocásticos (aleatórios)** e para inclusão/exclusão ao modelo.

O UCM considerado no software SAS pode ser pensado em um caso especial de modelos mais gerais, o (linear) Gaussian State Space Model (GSSM). A formulação do espaço de estados de um UCM tem muitas vantagens computacionais, como algoritmos convenientes e bastante robustos (em termos de convergência computacional ou dificuldades/falhas). Outra grande vantagem do UCM na equação (1) é sua *interpretabilidade* e o SAS PROC UCM fornece representações gráficas muito agradáveis dessa decomposição.

tabela 1

Listagem de Modelos de Família de Séries Temporais (TS) com as Equações de Modelo apropriadas e o comando SAS correspondente.

Séries Temporais (TS) Modelos de família	Equação do Modelo	Tipo de software SAS e Estimador padrão	SAS (Manual Codificação) Comando
Família ESM ESM = Termo de tendência temporal + Termo de sazonalidade + erro Prazo; (com diferentes tempos de suavização de variação Pesos para cada termo: peso de suavização de <i>nível</i> , peso de suavização de <i>tendência</i> e peso de suavização <i>sazonal</i>)			
SIMPLES (Único) ESM [Predefinição]; ou equivalente, ARIMA (011)	Modelo de tendência de nível: $yt = mt + at$ Pesos de suavização Fórmula: $Zt = w yt + (1 - w) Zt-1$ Onde, yt = a série temporal observada até o t-ésimo tempo; mt = termo médio (nível) variável no tempo; w = constante de suavização; $0 < w < 1$; $w \in [0, 1]$; Zt = valor previsto que é o valor suavizado em t; t = índice de tempo a valor atual; Este modelo deve ser usado quando os dados do TS Sem tendência e Sem sazonalidade.	<ul style="list-style-type: none"> • Codificação Manual SAS: MLE • SAS-ETS: MLE • SAS-FS: somente CLS (ou seja, a opção MLE é N/A). <p>N/A = Não disponível.</p> <ul style="list-style-type: none"> • Além disso, versões transformadas destes modelos também estão disponíveis: v NONE: para nenhuma transformação. Este é o padrão. v LOG: para transformação logarítmica. v SQRT: para transformação de raiz quadrada. v LOGÍSTICA: para transformação logística v BOXCox(n): Caixa Cox. 	PROC ESM com / model=Simple opção

Continua...

<p>DUPLO (Marrom) ESM; ou equivalente, ARIMA (022)</p>	<p>Modelo de tendência temporal:</p> $y_t = m_t + bt + a_{-t}$ <p>onde, bt</p> <p>= termo de tendência (inclinação) variável no tempo;</p> <p>1 Peso de alisamento Fórmula:</p> $Z_t^{[2]} = wy_t + (1 - w) Z_{t-1}^{[2]}$ <p>Este modelo deve ser usado quando os dados do TS <u>Tendência</u> mas <u>Sem sazonalidade</u>.</p>	<p>PROC ESM com / modelo = Opção DUPLA</p>
<p>LINEAR (Hot) ESM ; ou equivalente, ARIMA (012)</p>	<p>Modelo de tendência temporal:</p> $Z_t = m_t + bt + a_{-t}$ <p>2 Pesos de alisamento Fórmula:</p> $mt = ay_t bt(1-w)(m_{t-1} + bt+1);$ $t - mt-1) + (1 - w) bt+1 ;$ <p>onde, a</p> <p>= outra constante de suavização; $0 < a < 1$; Este modelo deve ser usado quando os dados de TS têm Tendência (2 pesos de suavização), mas Sem Sazonalidade.</p>	<p>PROC ESM com opção / model=LINEAR</p>
<p>TENDÊNCIA DE VAPOR ESM ; [quando $f = 1$ então $bt + a$ é t] equivalente Linear (Holt) ESM] ou equivalente, ARIMA (112)</p>	<p>Modelo de tendência temporal:</p> $Z_t = m_t + [f + f2 + \dots + fh]$ <p>2 Pesos de Suavização para Fórmula:</p> $mt = ay_t bt(1-w)(m_{t-1} + fbt+1);$ <p>onde, f = parâmetro $(1 - w) fbt + 1$; de amortecimento; 0 $< f < 1$; h = horizonte de passo h</p> <p>Este modelo deve ser usado quando os dados TS têm tendência (2 pesos de suavização e um parâmetro de amortecimento), mas sem sazonalidade.</p>	<p>PROC ESM com / modelo = TENDÊNCIA DE VAPOR</p>

<div>ADICIONAR</div> <div>SAZONAL ou SAZONAL (Aditivo Sazonal) ESM;</div> <div>TRIPLO (Hot Invernos) ESM;</div> <div>ou equivalente, ARIMA (011) (011)m onde m é o número de períodos por temporada;</div> <div>ou equivalente, ARIMA (01q) (011) com q=1 ou 2;</div> <div>ou equivalente, ARIMA (p00) (011)+c com p=1,2,3, e c é uma constante.</div> <div>MULTSEASONAL (Multiplicativo Sazonal) ESM;</div> <div>Modelo Híbrido, portanto Sem Equivalência ARIMA.</div> <div>INVERNOS (Invernos Multiplicativo) ESM;</div> <div>Sem ARIMA Equivalência.</div>	<div>Modelo sazonal de tempo:</div> <div>$Z_t = m_t + s_t + hm(k+1) + at$</div> <div>Onde,</div> <div>$m_t = a(y_t - s_t - m)(1-a)(m_t - 1 + bt - 1);$</div> <div>$s_t = g(y_t - m_t - 1 - bt - 1) + (1-g) s_t - m;$</div> <div>k = a parte inteira de (h/1)/m, que garante que as estimativas dos índices sazonais usados para previsões vêm do último ano da amostra;</div> <div>$Z_t + hm(k+1)$ = o termo sazonal variável no tempo para o m estação do ano; e</div> <div>at = termo de erro de ruído branco (e às vezes indicado por et).</div> <div>Este modelo deve ser usado quando o TS tiver <u>Nenhuma tendência</u> , mas <u>sazonalidade</u>.</div> <div>Modelo sazonal de tempo:</div> <div>$Z_t = (m_t) s_t + hm(k+1) + at$</div> <div>$m_t = a(y_t / s_t - m) + (1-a)(m_t - 1 + bt - 1);$</div> <div>$s_t = g[y_t / (m_t - 1 - bt - 1)] + (1-g) s_t - m;$</div> <div>Este modelo deve ser usado quando o TS tiver <u>Nenhuma tendência</u> , mas <u>sazonalidade</u>.</div> <div>Tendência Temporal e Tempo Modelo sazonal:</div> <div>$Z_t = (m_t + h bt) s_t + hm(k+1) + at$</div>	<div>PROC ESM com / modelo = SAZONAL</div> <div>PROC ESM com / modelo = MULTSEASONAL</div> <div>PROC ESM com / modelo = opção INVERNOS</div>
--	--	--

<p>ADICIONAR INVERNOS (Invernos Aditivo) ESM;</p> <p>ou equivalente, ARIMA (01[p+1])(010)p</p>	<p>$mt = a(yt/st-m) + (1-a)(mt-1 +bt-1);$ $bt= w(mt - mt-1) + (1-w) bt-1;$ $st = g[yt/(mt-1 -bt-1)] + (1-g) st-m;$</p> <p>Este modelo deve ser usado quando os dados TS tem tendência e sazonalidade.</p> <p>Tendência Temporal e Tempo Modelo sazonal:</p> <p>$Zt = mt + h bt t + st+hm(k+1) + at$</p> <p>Onde,</p> <p>$mt = a(yt -st-m)(1-a)(mt-1 +bt-1);$ $bt= w(mt - mt-1) + (1-w) bt-1;$ $st = g(yt - mt-1 -bt-1) + (1-g) st-m;$</p> <p>k = a parte inteira de (hŷ1)/m, que garante que as estimativas dos índices sazonais utilizados para previsão sejam provenientes do último ano da amostra;</p> <p>$Zt+hm(k+1)$ = a variação no tempo termo sazonal para o m estação do ano; e at = termo de erro de ruído branco (e às vezes indicado por et).</p> <p>Este modelo deve ser usado quando os dados TS tem tendência e sazonalidade.</p>		<p>PROC ESM com / modelo = ADICIONAR INVERNOS</p>
<p>Família ARIMA</p>			
<p>Genérico / Geral / Básico</p> <p>Modelo ARIMA:</p> <p>ARIMA(pdq) (PDQ)</p> <p>Onde,</p>	<p>$Dem_{t-1} = + \frac{\ddot{y} B ()}{\mu \ddot{y} (B)}_{-m_t}$</p> <p>Onde,</p>	<ul style="list-style-type: none">• Codificação Manual SAS: MLE• SAS-ETS: MLE• SAS-FS: CLS (mas MLE identifica opção está disponível)	<p>PROC ARIMA com o seguinte declarações:</p> <p>estimativa previsão</p>

<p>p = ordem do Auto t = tempo dos índices; componente regressivo (AR);</p> <p>q = ordem de movimentação Componente média (MA);</p> <p>d = diferenciação; P = ordem de sazonalidade do componente Auto Regressivo (AR); Q = ordem de sazonalidade do componente Média Móvel (MA); D = diferencial de sazonalidade.</p>	<p>W_t = a série de resposta Y_t ou uma diferença da série de respostas;</p> <p>m = o termo médio; B = o operador de retrocesso; isso é $B X_t = X_{t-1}$; $f(B)$ = o operador AR, representado como um polinômio no operador backshift: $f(B) = 1 - f_1B - \dots - f_pB^p$; $q(B)$ = o operador MA, representado como um polinômio no operador backshift: $q(B) = 1 - q_1B - \dots - q_qB^q$; a_t = a perturbação independente, também chamada de erro aleatório e às vezes denotado por e_t e muitas vezes chamado de termo de erro de ruído branco.</p>		
ARIMA com preditores:			
<p>ARIMA(pdq) (PDQ) + Predictor_i onde i não é inteiro negativo.</p>	<p>$\hat{Y}_t = \mu + \frac{\sum_{i=1}^p \hat{Y}_{t-i}}{p}$ + Predictor_i</p> <p>onde, Predictor_i = o i-ésimo preditor ou regressor.</p>		<p>PROC ARIMA com a seguinte opção: crosscorr = (pred); sob a identificação demonstração. Também, entrada=((1)pred); sob a estimativa demonstração.</p>
Família UCM			
<p>Genérico ou Em geral Modelo UCM</p>	<p>Veja a Equação (1). NOTA: A equação (1) já incluiu a possibilidade de componentes preditores ou regressores.</p>	<ul style="list-style-type: none">• Codificação Manual SAS: MLE• SAS-ETS: Não Disponível• SAS-FS: somente CLS (ou seja, a opção MLE é N/A).	<p>PROC UCM</p>

3. Roteiro

A Figura 1 fornece o roteiro para profissionais, previsores e pesquisadores otimizar um modelo adequado (tipo família) para sua análise de dados de TS (medida de resultado ou variável de resposta). Na Figura 1, o método do roteiro é fornecido pela 3 categoria de modelos do tipo família: família ESM, ARIMA ou UCM.

Primeiro, para o tipo de família de **ESM**, o roteiro levará a oito (8) modelos candidatos/iniciais/de linha de base disponíveis do PROC ESM [por codificação manual SAS (2013), SAS-ETS automático (2017) ou SAS- FS (2015)]. Observe que apenas a codificação manual SAS e o SAS-ETS automático têm MLE (Maximum Likelihood Estimation) como padrão e o padrão pode ser alterado para a opção CLS (Conditional Least Squares); para o SAS-FS, o padrão é CLS e não há opção para MLE porque um dos oito modelos ESM (Multiplicative-Winters ESM) não tem formulação de pura verossimilhança; portanto, os MLEs não podem ser disponibilizados de maneira uniforme (e, portanto, de uma vez por todas para a opção MLE automática) para todos os modelos de ESM no SAS-FS. O modelo Multiplicative-Winters ESM é na verdade um modelo híbrido com uma tendência multiplicativa/especificação sazonal e uma especificação de erro aditiva. Não possui formulação de verossimilhança que se encaixe na formulação usual de Levenberg-Marquardt (1944 e 1963) que converte um problema de máxima verossimilhança em um problema de otimização de mínimos quadrados. Observe que tal explicação não está disponível na documentação on-line do SAS e só pode ser obtida por meio de pergunta direta ao Instrutor do SAS que pesquisou a documentação interna do SAS por um longo tempo.

Em segundo lugar, para o tipo de família do modelo **ARIMA**, usamos PROC ARIMA, e o roteiro indica se a opção ou caminho de computação SAS é codificação manual, codificação automática via SAS-ETS candidato/ modelos iniciais/de linha de base de 21 modelos de repositório (ou 42 com sua transformação de log) ou codificação automática via SAS Forecast Studio (SAS-FS) modelos candidatos/iniciais/de linha de base de 28 repositórios de modelo (ou 56 com sua transformação de log). Para a modelagem da família ARIMA, a codificação manual pode utilizar o procedimento X11 para obter cinco (5) modelos candidatos; enquanto o procedimento FORECAST pode ainda ser usado para obter outros doze (12) modelos candidatos diferentes. Em seguida, o procedimento TIMESERIES pode ser usado para obter um gráfico TS rápido [para examinar qualquer tendência ou padrões, visualmente] enquanto o procedimento AUTOREG está disponível para fornecer um resultado de teste de hipótese formal (em termos de teste ADF e/ou PP) para examinar se há diferenciação ou se existe raiz unitária e não é rejeitada. Para ARIMA

modelos, todas essas 3 opções de software SAS podem ser alteradas de MLE para CLS e vice-versa. No entanto, notamos que, quando se trata de selecionar critérios de seleção de modelos, alguns critérios baseados em verossimilhança, como (em oposição aos baseados em distância) Akaike Information Criteria (AIC) e Bayesian Information Criteria (BIC), têm fórmulas diferentes; ou seja, as fórmulas no SAS-FS são diferentes da codificação manual do SAS-ETS e/ou SAS porque são codificadas por diferentes sub-rotinas com codificadores diferentes e em tempos diferentes. Portanto, não é aconselhável compará-los porque a comparação não será justa/sob as mesmas fórmulas. No entanto, esses critérios de seleção de modelos baseados em distância, como erro percentual médio absoluto (MAPE), erro quadrático médio (RMSE), etc., têm as mesmas fórmulas. Portanto, suas saídas (da codificação manual SAS-ETS, SAS-FS e SAS) são sensíveis para serem comparadas.

Terceiro, para o tipo de família de modelo **UCM**, o roteiro levará ao procedimento UCM. Para PROC UCM, o padrão de codificação manual SAS é MLE; enquanto o padrão do SAS-FS é CLS (e sem opção MLE); e o SAS-ETS automático não inclui o UCM como parte do modelo de repositório 21 (ou 42 com sua transformação de log).

Observe que aqui, o padrão SAS para a codificação manual Base SAS 9.4 (SAS Institute Inc. 2013) e a versão automática SAS-ETS 14.3 (SAS Institute Inc. 2017) estão usando a abordagem de Estimativa de Máxima Verossimilhança (MLE); enquanto o SAS-FS automático versão 14.1 (SAS Institute Inc. 2015) está usando a abordagem Conditional Least Square (CLS). Tipicamente, MLE é o melhor estimador e o mais desejado devido às suas 5 boas propriedades assintóticas (que são Assintoticamente ou por Grande Amostra: Estimadores Consistentes, Normais, Imparciais, Eficientes e Invariantes a transformações); no entanto, nem todas as funções de verossimilhança são duas vezes diferenciáveis nem convergem (para sua distribuição limite) nem têm convergência suficientemente rápida; portanto, o MLE agradável e fácil (formato fechado) nem sempre pode ser obtido. Portanto, um estimador CLS pode ser uma solução viável alternativa tangível, uma vez que é uma abordagem baseada na distância (não baseada na probabilidade). No entanto, embora o CLS possa não ter os mesmos problemas que o MLE, também não está livre de problemas. O CLS pode ter outros problemas computacionais, como a existência de matrizes de célula zero que causam matrizes não inversíveis devido à divisão por entradas zero, etc. Além disso, o CLS pode não garantir essas 5 boas propriedades assintóticas. Assim, CLS pode nem sempre ser uma estimativa melhor do que MLE; mas ambos podem servir como estimadores válidos razoavelmente assegurados (isto é, sólidos e justificáveis) e factíveis. Ao decidir quais métodos de estimativa usar, quantas iterações de computação estabilizar, critérios de seleção de modelos a serem usados, adicionar/reduzir o p , d , q , P , D , Q , pulso, etc., uma estatística

julgamento de um especialista no assunto (SME) é necessário e altamente recomendado.

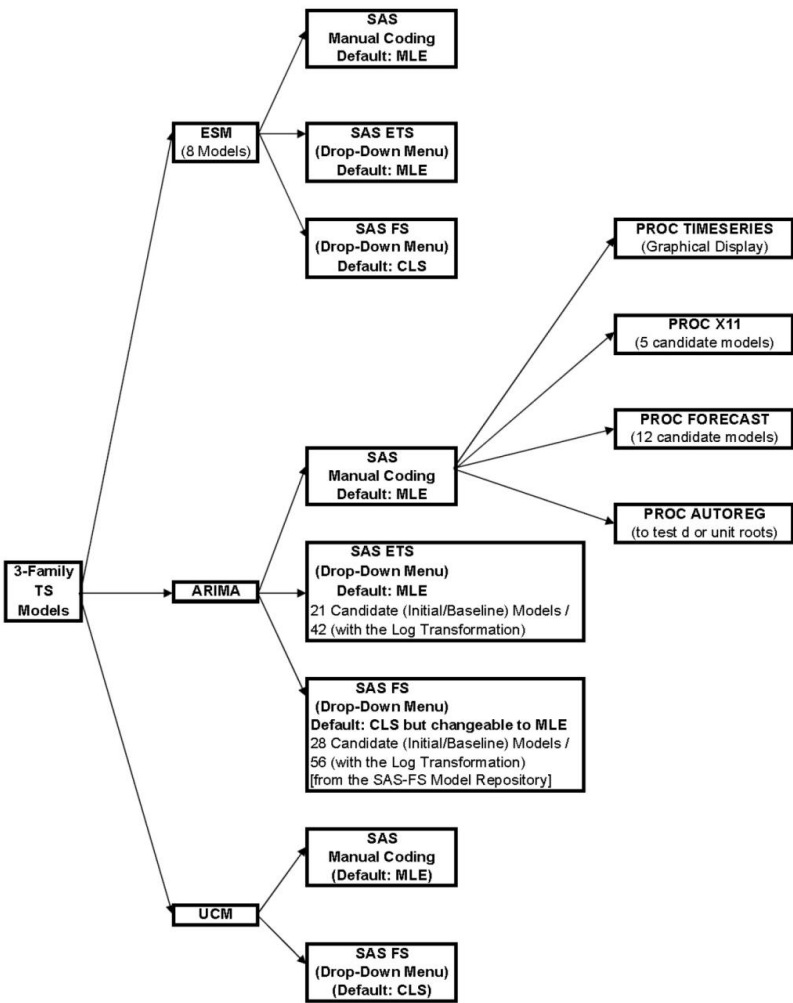


figura 1

**Um roteiro para os modelos de séries temporais de 3 famílias
para o SAS PROC correspondente.**

Em seguida, os procedimentos SAS correspondentes aos dados estatísticos adequados
O método direcionado a partir da Figura 1 pode ser encontrado na prescrição da Tabela 1.

4. Resumo

Os dados de TS (ou resposta ou medida de resultado) são muito comuns em aplicações de dados reais, como sociologia, ciências físicas, economia, dados financeiros, etc. A análise de tais dados de TS (medida de resultado ou variável de resposta) tem uma longa história, condensada ou resumidos na categoria básica de três famílias de modelos TS: ESM, ARIMA e UCM. Neste artigo, fornecemos uma revisão da categoria básica de três famílias de modelos e procedimentos estatísticos TS que são práticos/implementáveis na literatura recente para tais dados TS ou observações autocorrelacionadas (medida de resultado ou variável de resposta), ou seja, não -observações independentes, mas identicamente distribuídas.

Além disso, também fornecemos com os modelos TS, o procedimento SAS correspondente, para as traduções/implementação de computação estatística, o software estatístico profissional mais utilizado para análise de dados. O modelo ideal obtido via automação (SAS-ETS e/ou SAS-FS) pode ser aprimorado manualmente (melhor recomendado para alta precisão) ou através do menu suspenso disponível (do SAS-ETS e SAS-FS). No entanto, existem prós e contras. A codificação manual do SAS é mais difícil de aprender (a dificuldade é tolerada para precisão); enquanto a automação do SAS ou o menu suspenso são mais fáceis de aprender (a precisão é sacrificada pela facilidade de uso).

Em resumo, este artigo será útil para os profissionais, previsores e pesquisadores nas várias áreas de estudo para determinar o método apropriado para seus dados, de acordo com o roteiro fornecido na Figura 1.

Referências

- [1] Brown, RG (1959). *Previsão estatística para controle de estoque*, Mc Graw-Hill: Nova York, NY.
- [2] Brown, RG (1962). *Suavização, Previsão e Previsão de Séries Temporais Discretas*, Prentice Hall: New Jersey, NJ.
- [3] De Gooijer, JG, Hyndman, RJ (2006). 25 anos de elenco Time Series Fore. *International Journal of Forecasting*, Vol. 22, No. 3, pp. 443-473.
- [4] Fomby, TB (2008). Modelos de suavização exponencial. *Notas de aula Versão 6*. Departamento de Economia. Southern Methodist University, Dallas, TX, junho de 2008.

- [5] Gardner, ES, & McKenzie, E. (1985). Previsão de tendências em séries temporais. *Management Science*, 31(10), 1237-1246. <https://doi.org/10.1287/mnsc.31.10.1237>
- [6] Harvey, AC (1989). Previsão, Modelos Estruturais de Séries Temporais e o Filtro de Kalman. Cambridge University Press.
- [7] Holt, CE (1957). *Previsão de sazonalidades e tendências por médias exponencialmente ponderadas* (Memorando ONR No. 52). Carnegie Institute of Technology, Pittsburgh, EUA. <https://doi.org/10.1016/j.ijforelenco.2003.09.015>
- [8] Hyndman, RJ, e Athanasopoulos, G. (2018). *Princípios e Práticas de Previsão*. Universidade Monash, Austrália. <https://otexts.org/fpp2/>
- [9] Levenberg, Kenneth (1944). Um Método para a Solução de Certos Problemas Não Lineares em Mínimos Quadrados. *Quarterly of Applied Mathematics*, Vol. 2, pp. 164-168.
- [10] Marquardt, Donald (1963). Um Algoritmo para Estimativa de Mínimos Quadrados de Parâmetros Não Lineares. *SIAM Journal on Applied Mathematics*, Vol. 11, No. 2, pp. 431-441. doi:10.1137/0111030
- [11] Rahardja, D. (2005). Combinações de Gráficos X versus Gráficos X/MR: Casos IID e Casos Não IID. *Engenharia da Qualidade*, v. 17, nº 2, pp. 189–196.
- [12] Rahardja, D., Yang, Y. e Zhang, Z. (2016). Uma revisão abrangente dos dados binários independentes ou emparelhados de duas amostras – com ou sem efeitos de estrato. *Journal of Modern Applied Statistical Methods*, vol. 15, No. 2, pp. 215-223.
- [13] Rahardja, D. (2017). Uma Revisão dos Testes de Múltiplas Amostras para o Tipo de Dados Contínuos. *Journal of Modern Applied Statistical Methods*, vol. 16, No. 1, pp. 127-136.
- [14] Rahardja, D., e Wu, H. (2018). Revisão metodológica estatística para dados de tempo para evento. *Jornal de Estatística e Sistemas de Gestão*, vol. 21, No. 1, pp. 189-199.
- [15] SAS Institute Inc. (1995). *SAS Manual Software SAS/ETS: Time Series Forecasting System*, Versão 6, Primeira Edição, pp. 225–235. Cary, Carolina do Norte: SAS Institute Inc.
- [16] SAS Institute Inc. (2013). *Guia de Procedimentos do Base SAS® 9.4: Procedimentos Estatísticos*, Segunda Edição. Cary, Carolina do Norte: SAS Institute Inc.

- [17] SAS Institute Inc. (2015). SAS® Forecast Studio 14.1: Guia do usuário.
Cary, Carolina do Norte: SAS Institute Inc.
- [18] SAS Institute Inc. (2017). Guia do usuário do SAS/ETS® 14.3. Cary, NC:
SAS Institute Inc.
- [19] Sypsas, PT (1989). Identificando Padrões em Dados de Séries Temporais
Múltiplas. *Journal of Information and Optimization Sciences* Vol.10, No. 3, pp.
471–494.
- [20] Winters, PR (1960). Previsão de vendas por médias móveis exponencialmente
ponderadas. *Management Science*, 6, 324-342. [https://doi.
org/10.1287/mnsc.6.3.324](https://doi.org/10.1287/mnsc.6.3.324)
- [21] Zhou, M. (2017). Consistência do Estimador Kaplan-Meier quando os dados
são distribuídos de forma independente e não idêntica. Relatório *Técnico*,
Departamento de Estatística, Universidade de Kentucky, Lexington, KY
40506-0027, propriedade do arquivo criada em 11 de setembro de 2017, pp. 1–15. [www.
ms.uky.edu/~mai/sta709/Tech351-1.pdf](http://www.ms.uky.edu/~mai/sta709/Tech351-1.pdf)

Recebido em abril de 2019