

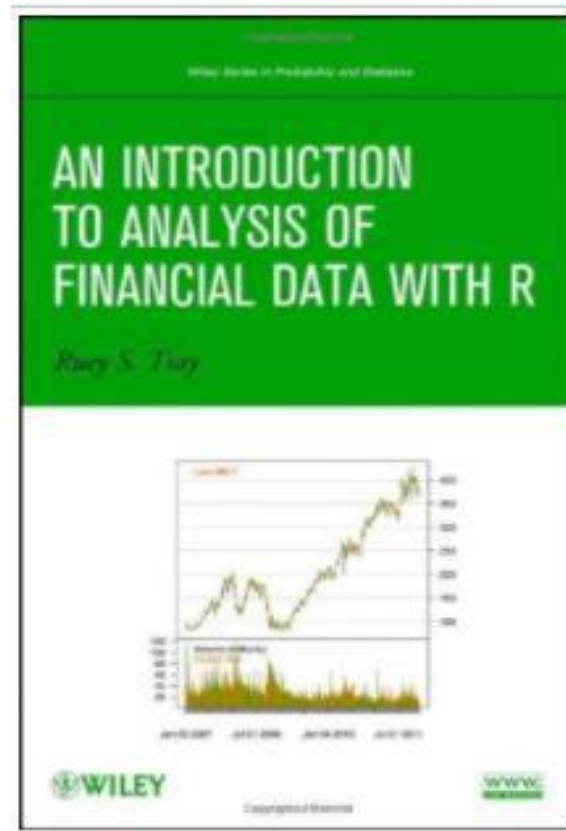
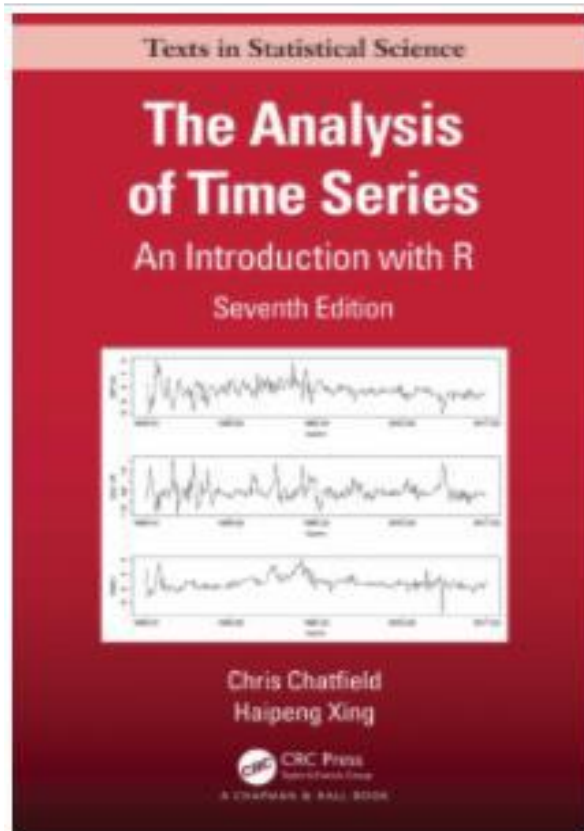
ESTATÍSTICA APLICADA II

Parte 3

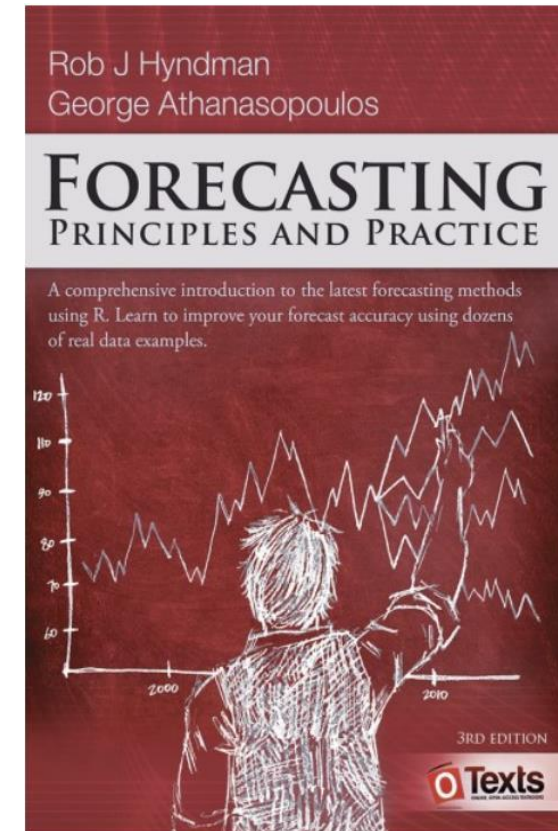
Prof. Arno P. Schmitz

Séries Temporais

Referências



<https://www.msperlin.com/adfeR/>



<https://otexts.com/fpp3/>

Estatística Aplicada II

Séries Temporais

- Conjunto de observações ordenadas no tempo.

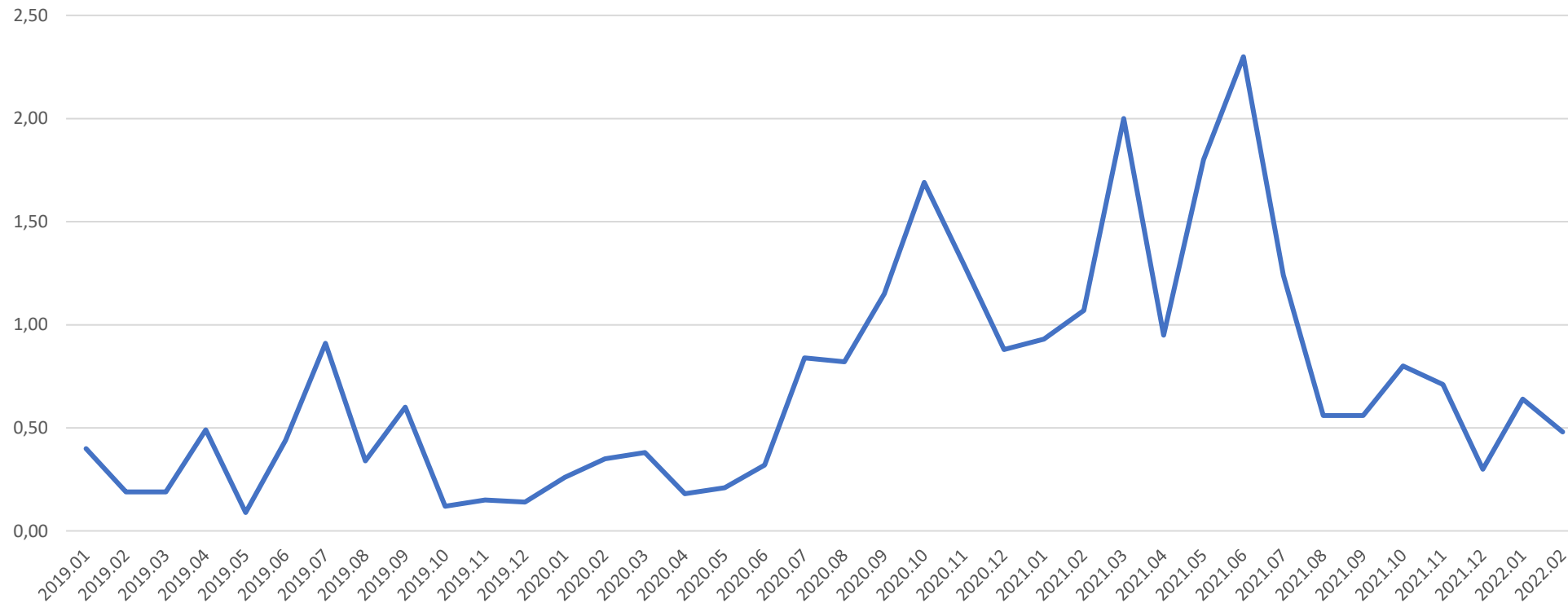
$$\{Y_t\}_{t=1}^n = \{Y_1, Y_2, \dots, Y_n\}$$

Exemplos:

- Índice B3 diário;
- Retorno das ações da Petrobrás mensal;
- Índices de Inflação mensal;
- Taxa de Câmbio Real/US\$ diário.

Elementos das Séries Temporais

INCC-M - (% a.m.) INCCMG12



Taxa

Medida

INCC

Fato

Mensal

Unidade de
Medida

Séries Temporais

Tipos de Séries Temporais

Univariadas e Multivariadas:

- **Univariadas:** apenas uma variável conectada ao tempo;
- **Multivariadas:** duas ou mais variáveis conectadas ao tempo.

Objetivos do Estudo de Séries de Tempo

- Investigação do mecanismo que gera a série temporal;
- Descrever o comportamento da série;
- Identificar periodicidades relevantes nos dados;
- Fazer previsões de valores futuros da série.

Séries Temporais

Classificação das Séries Temporais

- **Determinística:** quando pode ser descrita por uma função matemática para estabelecer exatamente os valores futuros da série.
- **Estocástica:** quando os valores futuros da série somente podem ser estabelecidos em termos probabilísticos, pois o modelo compõe-se também de um termo aleatório (erro estocástico).

Métodos

SIMPLES = Naive, Mean, Drift;

CLÁSSICOS = Decomposição, Suavização exponencial, ARIMA, SARIMA;

OUTRAS = Regressão, Redes neurais.

Séries Temporais

ESTATÍSTICAS DE ERRO DAS PREVISÕES

ME: Mean Error – Média da diferença entre o realizado e o previsto

$$erro_t = Y_t - \hat{Y}_t$$

$$ME = \frac{\sum_{t=1}^h erro_t}{h}$$

MAE: Mean Absolute Error – Média da diferença absoluta entre realizado e previsto

$$MAE = \frac{\sum_{t=1}^h |erro_t|}{h}$$

RMSE : Root Mean Square Error – Desvio padrão total da amostra da diferença entre o previsto e o realizado

$$RMSE = \sqrt{\frac{\sum_{t=1}^h (erro_t)^2}{h}}$$

Séries Temporais

ESTATÍSTICAS DE ERRO DAS PREVISÕES

MPE: Mean Percentage Error – Diferença percentual do erro

$$MPE = \frac{\sum_{t=1}^h \frac{\text{erro}_t}{Y_t}}{h} \cdot 100$$

MAPE: Mean Absolute Percentage Error – Diferença absoluta percentual do erro

$$MAPE = \frac{\sum_{t=1}^h \frac{|\text{erro}_t|}{Y_t}}{h} \cdot 100$$

TIC: Theil Inequality Coefficient (Theil's U) – Grau de ajuste da previsão; quanto menor, melhor; Próximo de zero é melhor.

$$\text{Theil's } U = \sqrt{\frac{\sum_{t=1}^h \left(\frac{\hat{Y}_{t+1} - Y_{t+1}}{Y_t} \right)^2}{\sum_{t=1}^h \left(\frac{Y_{t+1} - Y_t}{Y_t} \right)^2}}$$

Séries Temporais

MÉTODOS CLÁSSICOS

Decomposição: Projeções por decomposição da série

Suavização exponencial: Método de amortecimento (suavização), ideal para tendências e inclui variação sazonal (ETS)

Aditivo – para variação sazonal constante;

Multiplicativo – para variação sazonal significativa na série.

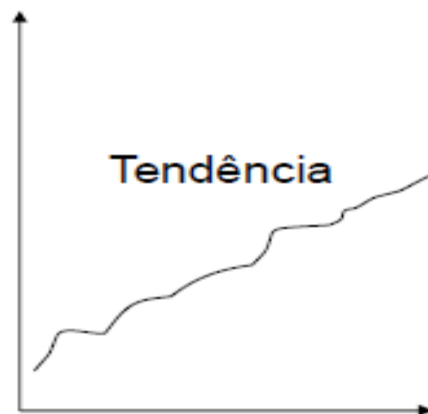
ARIMA: Modelo autorregressivo (AR) – as observações atuais tem relação com passado (curto prazo); integração (I) – Existe uma tendência persistente no tempo e; Média Móvel (MA) –. Uma determinada observação atual pode ser explicada pelos resíduos (parte não explicada da previsão) do passado.

SARIMA: Modelo ARIMA com Sazonalidade.

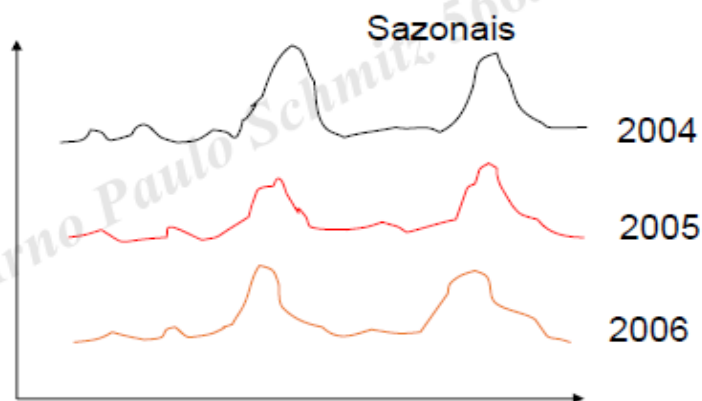
Séries Temporais

COMPONENTES DE UMA SÉRIE TEMPORAL

TENDÊNCIA: Movimento oculto nos dados, seguindo uma direção – crescente, decrescente ou estacionária. Séries sem tendência são chamadas de séries estacionárias.



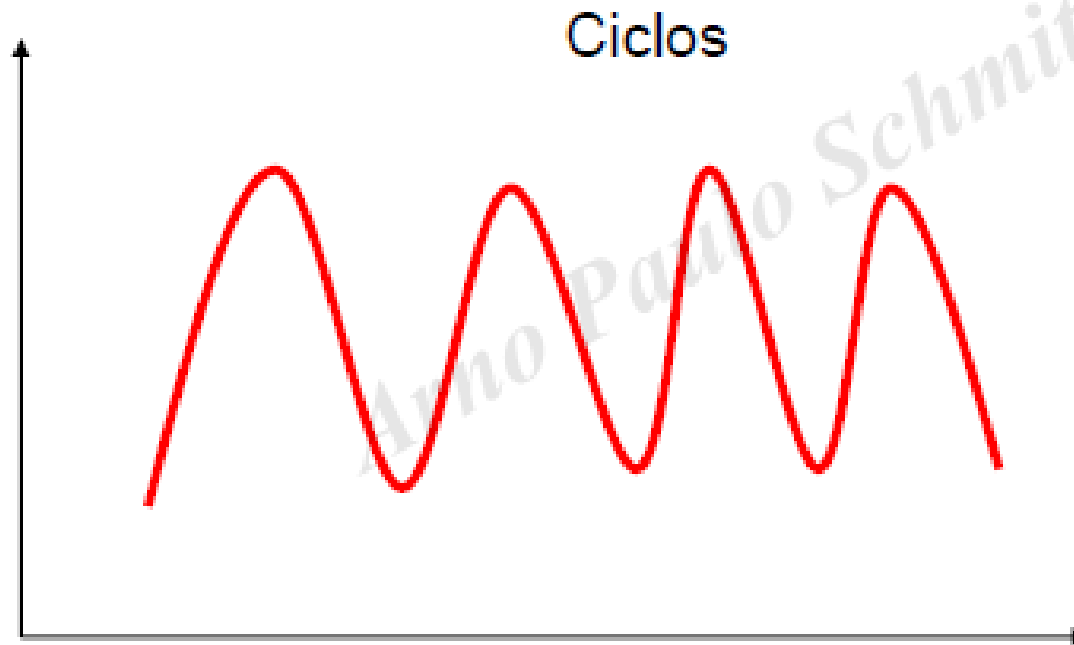
SAZONALIDADE: Flutuações regulares dentro de um período completo de tempo (dia, semana, mês, etc.) – Representam um tipo de padrão que se repete (picos e depressões).



Séries Temporais

COMPONENTES DE UMA SÉRIE TEMPORAL

Ciclos: Flutuações de longo prazo nos dados e são similares aos fatores sazonais. Padrão que se repete com regularidade mas sem período fixo.



Séries Temporais

Série com tendência, sazonalidade e variações cíclicas

- Muitas séries apresentam além de uma tendência, variações cíclicas e sazonais;
- Estas variações ocorrem devido ao clima, fatores econômicos, tempo, etc;
- Se esses componentes podem ser identificados, a sua consideração pode ajudar a melhorar as previsões.

Métodos de decomposição:

i) Aditivo ($Y = T + C + S + E$)

Y = valor da série no instante t;

T = componente de tendência para o instante t;

C = componente cíclica para o instante t;

S = componente sazonal para o instante t;

E = componente (de erro) aleatório para o instante t;

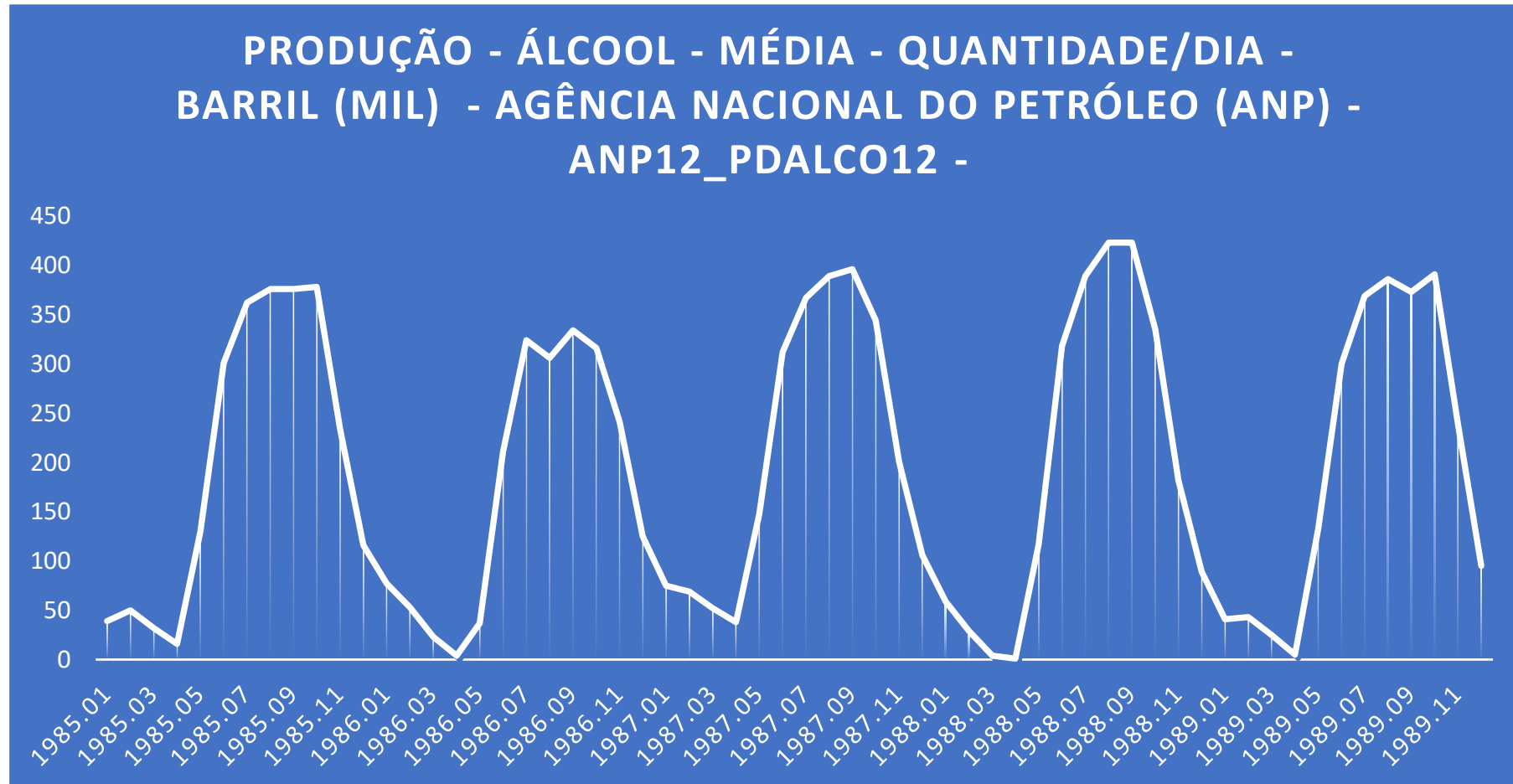
i) Multiplicativo ($Y = T * C * S * E$);

ii) Damped (amortecimento da tendência em modelos com nenhuma sazonalidade; sazonalidade aditiva ou; sazonalidade multiplicativa)

Séries Temporais

Decomposição - Modelo Aditivo

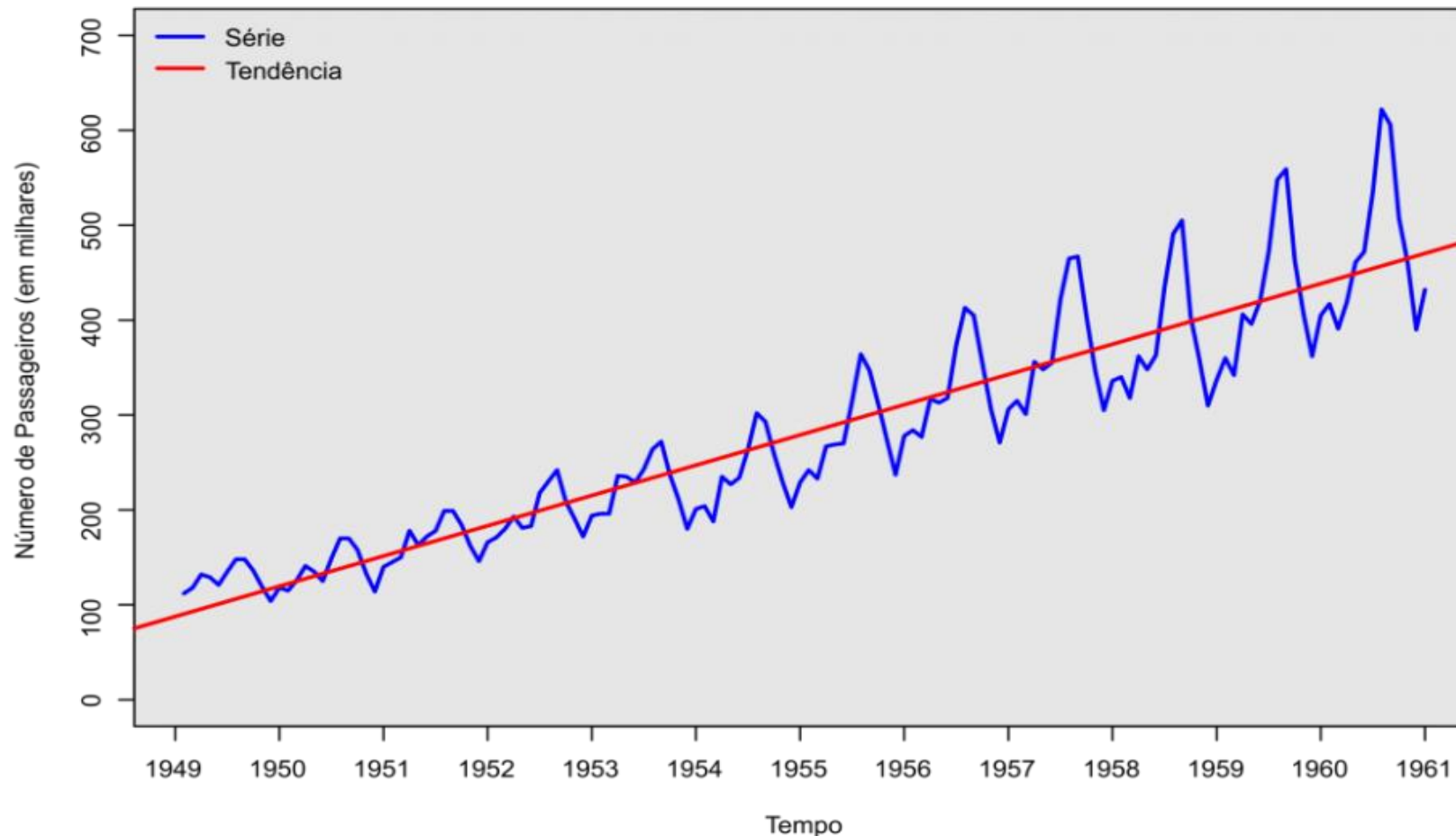
- O **modelo aditivo** geralmente é considerado mais adequado para séries em que as flutuações sazonais permanecem aproximadamente do mesmo tamanho ao longo da série.



Séries Temporais

Decomposição – Modelo Multiplicativo

- O modelo multiplicativo é normalmente aplicado a séries em que o tamanho dos efeitos sazonais aumentam.

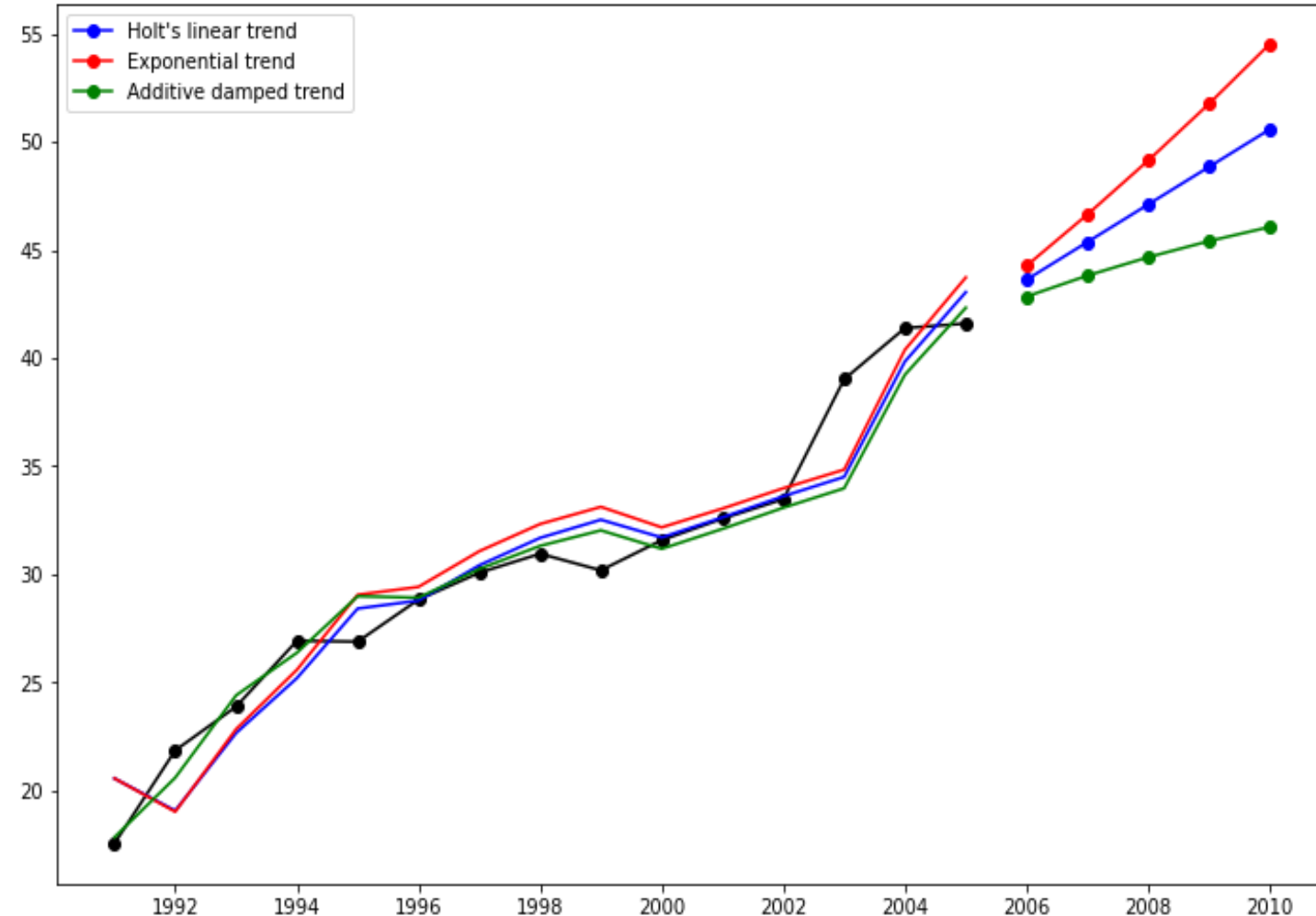


Fonte: <https://operdata.com.br/blog/caracteristicas-das-series-temporais/>

Séries Temporais

Decomposição – Damped (amortecimento da tendência)

- Os modelos com Damped tem a função de amortecer (se necessário) a tendência para melhorar as previsões.



Séries Temporais

Modelos de Suavização Exponencial

- **Suavização Exponencial Simples (SES)** - Série temporal que não apresenta tendência e nem sazonalidade ;
- **Suavização Exponencial de Holt (SEH)** - Série temporal com Tendência mas sem sazonalidade;
- **Suavização Exponencial de Holt-Winters (SEHW)** - Série temporal com Tendência e Sazonalidade.

Séries Temporais

Modelo de Suavização Exponencial Simples (SES)

- Aplica pesos maiores às observações mais recentes captando melhor as mudanças de comportamento.
- Os valores preditos são iguais ao último valor exponencial suavizado.

$$\{Y_t\}_{t=1}^n = \{Y_1, Y_2, \dots, Y_n\}$$

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha) \hat{Y}_t$$

$$0 \leq \alpha \leq 1 \quad , \quad t = 1, 2, 3, \dots \quad , \quad \hat{Y}_1 = Y_1$$

α = Constante de suavização (smoothing)

Séries Temporais

Modelo de Suavização Exponencial Simples (SES) - Previsão

Intervalo de Confiança: $\hat{Y}_t(h) = \hat{Y} \pm Z_{\frac{\alpha}{2}}.SE(h)$

Desvio padrão da estimativa: $SE = \sqrt{\frac{1}{n-1} \sum_{t=1}^{n-1} (\hat{Y}_t - Y_t)^2}$

Desvio padrão da previsão: $SE_h = \sqrt{1 + (h - 1) \cdot \alpha^2}$

Séries Temporais

Modelo de Suavização Exponencial de Holt (SEH)

- Para séries temporais com tendência linear e sem sazonalidade.

$$\{Y_t\}_{t=1}^n = \{Y_1, Y_2, \dots, Y_n\}$$

$$\hat{\mu}_t = \alpha Y_t + (1 - \alpha)(\hat{\mu}_{t-1} + \hat{T}_{t-1})$$

$$\hat{T}_t = \beta(\hat{\mu}_t - \hat{\mu}_{t-1}) + (1 - \beta)(\hat{T}_{t-1})$$

$$\hat{Y}_{t+1} = \hat{\mu}_t + \hat{T}_t$$

$$\hat{\mu}_1 = Y_1 ; \quad \hat{T}_1 = 0 \quad ; \quad 0 \leq \alpha \leq 1 \quad ; \quad 0 \leq \beta \leq 1 \quad ; \quad t = 1, 2, \dots, n$$

Séries Temporais

Modelo de Suavização Exponencial de Holt (SEH) - Previsões

$$\text{Intervalo de Confiança: } \hat{Y}_t(h) = \hat{Y} \pm Z_{\frac{\alpha}{2}}.SE(h)$$

$$\text{Desvio padrão da estimativa: } SE = \sqrt{\frac{1}{n-1} \sum_{t=1}^{n-1} (\hat{Y}_t - Y_t)^2}$$

$$\text{Desvio padrão da previsão: } SE_h = \sqrt{1 + k \cdot \alpha^2}$$

$$k = \sum_{i=2}^h (1 + \beta \cdot (i - 1))^2$$

Séries Temporais

Modelo de Suavização Exponencial de Holt-Winters (SEHW Aditivo)

- Para séries temporais com sazonalidade constante (c = período sazonal) e com tendência

$$\{Y_t\}_{t=1}^n = \{Y_1, Y_2, \dots, Y_n\}$$

$$\hat{Y}_t = \hat{L}_{t-1} + \hat{T}_{t-1} + \hat{S}_{t-c}$$

$$\hat{L}_t = \alpha(\hat{Y}_t - \hat{S}_{t-c}) + (1 - \alpha)(\hat{L}_{t-1} + \hat{T}_{t-1})$$

$$\hat{T}_t = \beta(\hat{L}_t - \hat{L}_{t-1}) + (1 - \beta)(\hat{T}_{t-1})$$

$$\hat{S}_t = \gamma(\hat{Y}_t - \hat{L}_{t-1}) + (1 - \gamma) \cdot \hat{S}_{t-c}$$

$$\hat{Y}_{t+h} = \hat{L}_t + h\hat{T}_t + \hat{S}_{t+h-ch'}$$

$$h' = \text{INT} \left(\frac{(h-1)}{c} \right) + 1$$

$$0 \leq \alpha \leq 1 \quad ; \quad 0 \leq \beta \leq 1 \quad ; \quad 0 \leq \gamma \leq 1 - \alpha \quad ; \quad t = 1, 2, \dots, n$$

Séries Temporais

Modelo de Suavização Exponencial de Holt-Winters (SEHW Aditivo) - Predições

$$\text{Intervalo de Confiança: } \hat{Y}_t(h) = \hat{Y} \pm Z_{\frac{\alpha}{2}}.SE(h)$$

$$\text{Desvio padrão da estimativa: } SE = \sqrt{\frac{1}{n-1} \sum_{t=1}^{n-1} (\hat{Y}_t - Y_t)^2}$$

$$\text{Desvio padrão da previsão: } SE_h = SE \cdot \sqrt{1 + \sum_{i=1}^{h-1} \varphi_i^2}$$

$$\varphi_i = \begin{cases} \alpha (1 + \beta_i), & \text{para } c \neq i - 1 \\ \alpha(1 + \beta_i) + \gamma(1 - \alpha), & \text{para } c = i - 1 \end{cases}$$

Séries Temporais

Modelo de Suavização Exponencial de Holt-Winters (SEHW Multiplicativo)

- Para séries temporais com sazonalidade (de)crescente (c = período sazonal) e com tendência

$$\{Y_t\}_{t=1}^n = \{Y_1, Y_2, \dots, Y_n\}$$

$$\hat{Y}_t = \hat{L}_{t-1} + \hat{T}_{t-1} \cdot \hat{S}_{t-c}$$

$$\hat{L}_t = \alpha(\hat{Y}_t / \hat{S}_{t-c}) + (1 - \alpha) \cdot (L_{t-1} + T_{t-1})$$

$$\hat{T}_t = \beta(\hat{L}_t - \hat{L}_{t-1}) + (1 - \beta)(T_{t-1})$$

$$\hat{S}_t = \gamma(\hat{Y}_t / \hat{L}_{t-1}) + (1 - \gamma) \cdot S_{t-c}$$

$$\hat{Y}_{t+h} = (\hat{L}_t + \hat{T}_t) \cdot \hat{S}_{t+h-ch'}$$

$$h' = INT\left(\frac{(h-1)}{c}\right) + 1$$

$$0 \leq \alpha \leq 1 \quad ; \quad 0 \leq \beta \leq 1 \quad ; \quad 0 \leq \gamma \leq 1 - \alpha \quad ; \quad t = 1, 2, \dots, n$$

Séries Temporais

Modelo de Suavização Exponencial de Holt-Winters (SEHW Multiplicativo) - Predições

$$\text{Intervalo de Confiança: } \hat{Y}_t(h) = \hat{Y} \pm Z_{\frac{\alpha}{2}}.SE(h)$$

$$\text{Desvio padrão da estimativa: } SE = \sqrt{\frac{1}{n-1} \sum_{t=1}^{n-1} (\hat{Y}_t - Y_t)^2}$$

$$\text{Desvio padrão da previsão: } SE_h = SE \cdot \sqrt{1 + \sum_{i=1}^{h-1} \varphi_i^2}$$

$$\varphi_i = \begin{cases} \alpha (1 + \beta_i), & \text{para } c \neq i - 1 \\ \alpha(1 + \beta_i) + \gamma(1 - \alpha), & \text{para } c = i - 1 \end{cases}$$

Séries Temporais

Modelo de Suavização Exponencial - Damped

Tendência Aditiva com Damping:

- Sazonalidade nenhuma: $\hat{Y}_{t+h} = L_t + \phi_h T_t$
- Sazonalidade aditiva: $\hat{Y}_{t+h} = L_t + \phi_h T_t + S_{t-m+h}$
- Sazonalidade multiplicativa: $\hat{Y}_{t+h} = L_t + \phi_h T_t \cdot S_{t-m+h}$

Tendência Multiplicativa com Damping:

- Sazonalidade nenhuma: $\hat{Y}_{t+h} = L_t \cdot T_t^\phi \cdot S_{t-m+h}$
- Sazonalidade aditiva: $\hat{Y}_{t+h} = L_t \cdot T_t^\phi + S_{t-m+h}$
- Sazonalidade multiplicativa: $\hat{Y}_{t+h} = L_t \cdot T_t^\phi \cdot S_{t-m+h}$

Séries Temporais

Modelo Generalizado ETS (Error, Trend, Season)

- Modela **por experimentação e escolhe** o melhor modelo de acordo com o comportamento da série temporal.
- ➔ Modelos “sem” tendência e “sem” sazonalidade;
- ➔ Modelos “com” tendência (dumped – aditivo ou multiplicativo) e “sem” sazonalidade;
- ➔ Modelos “sem” tendência e “com” sazonalidade (aditiva ou multiplicativa);
- ➔ Modelos “com” tendência (dumped – aditivo ou multiplicativo) e “com” sazonalidade (aditiva ou multiplicativa);

Séries Temporais

Modelos ARIMA (Box – Jenkins)

“Os dados falam por si”

Características:

- Robusto: pode ser aplicado em praticamente qualquer tipo de série temporal;
- Funciona melhor com dados estáveis, com poucos *outliers* (pode-se remover)
- **Requer dados estacionários (média zero,**
 - ➔ Pode ser transformada usando diferenciação: remove tendências
 - ➔ Diferenciação: subtrai a observação atual da anterior
 - ➔ Diferenciação pode ser feita 1x: diferenciação de primeira ordem
 - ➔ Diferenciação 2x: diferenciação de segunda ordem (mais raro)

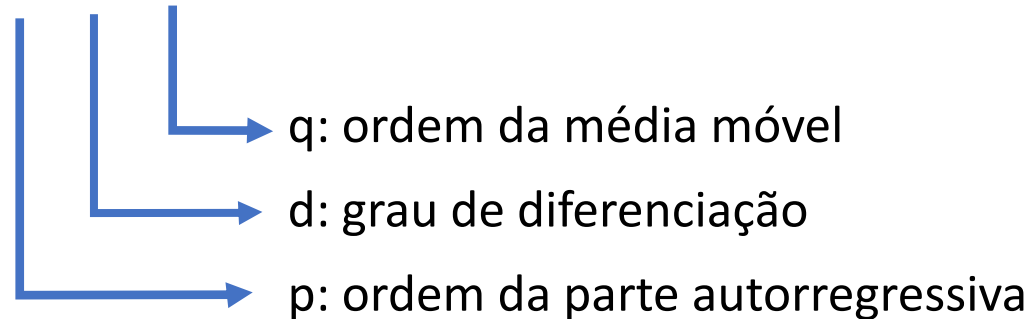
Séries Temporais

Tipos de Modelos – ARIMA (não Sazonal)

- Modelos auto-regressivos (AR): avalia a relação entre os períodos (*lags/defasagens* - p) – extrai a influência;
- Integrado (I): Aplicado à diferenciação, quando necessário (d);
- Modelos médias móveis (MA): avalia erros entre períodos e extrai esses erros (q);
- Modelos auto-regressivos e de médias móveis (ARMA)
- Modelos auto-regressivos integrados e de médias móveis (ARIMA)

AR I MA

(p , d , q)



Séries Temporais

Tipos de Modelos – ARIMA (não Sazonal)

- **Modelos Auto-Regressivos (AR):** Os valores correntes de uma série Y_t dependem apenas de seus valores passados e dos erros aleatórios.

$$\text{Modelo AR(1)} \rightarrow Y_t = \phi_1 Y_{t-1} + \varepsilon_t$$

- **Modelos de Média Móvel (MA):** Os valores correntes de uma série Y_t dependem apenas dos valores dos erros aleatórios.

$$\text{Modelo MA(1)} \rightarrow Y_t = \theta \varepsilon_{t-1} + \varepsilon_t$$

- **Modelos Auto-Regressivos e de Média Móvel:** Os valores correntes de uma série Y_t dependem de seus valores passados e dos erros aleatórios.

$$\text{Modelo ARMA(1,1)} \rightarrow Y_t = \phi_1 Y_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t$$

Séries Temporais

ESTATÍSTICAS DE SÉRIES TEMPORAIS

ACF: k-Order Autocorrelation Function – Função de autocorrelação de ordem “k” - esta é a correlação entre os valores das séries que estão separados por k intervalos (lags).

$$ACF_k = \frac{cov(Y_t, Y_{t+k})}{variância(Y_t)}$$

* A ACF determina a ordem “q” do MA(q)

PACF: Função de Autocorrelação parcial - Na defasagem k , esta é a correlação entre os valores das séries que estão separados por k intervalos (lags), contabilizando os valores dos intervalos entre eles.

- A PACF determina a defasagem “p” do AR(p);

$$\hat{\phi}_{kk} = \frac{Y_k - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} Y_j}{1 - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} Y_j}$$

Séries Temporais

ESTATÍSTICAS DE SÉRIES TEMPORAIS

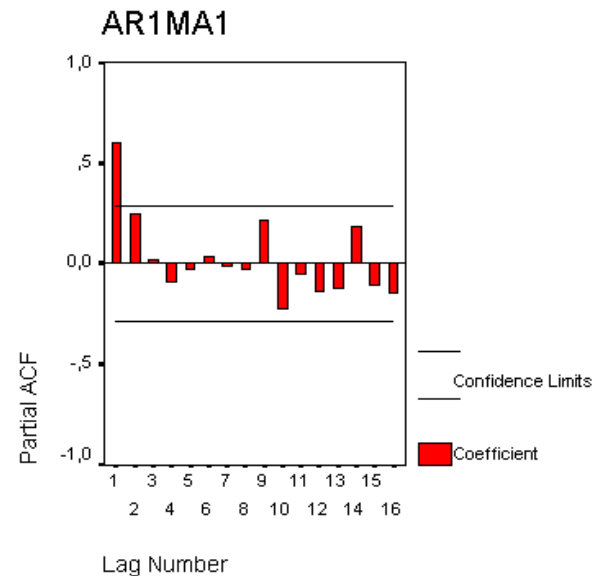
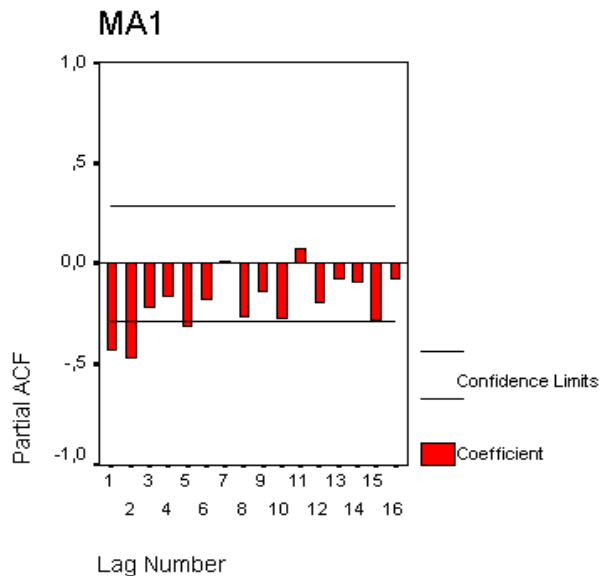
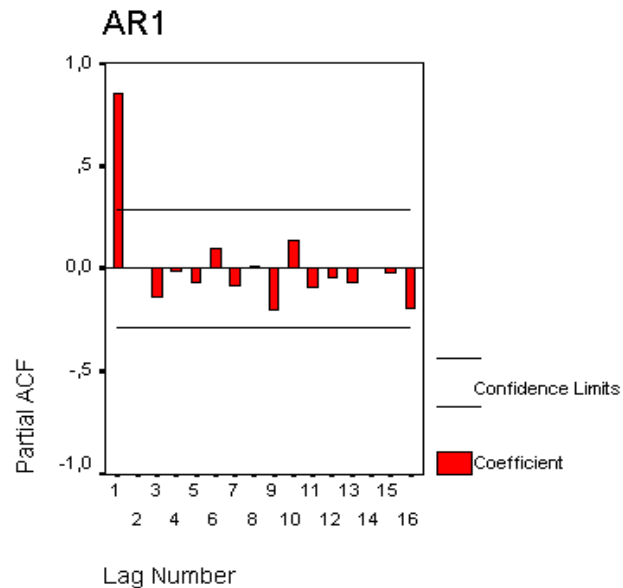
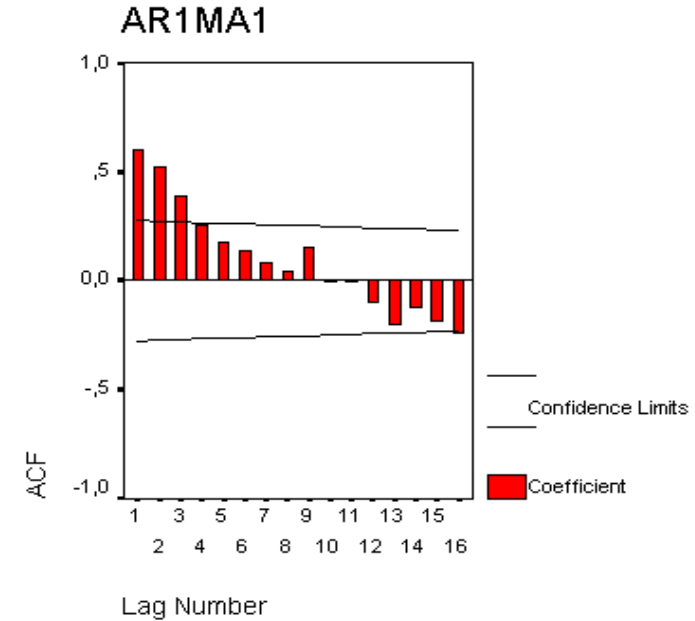
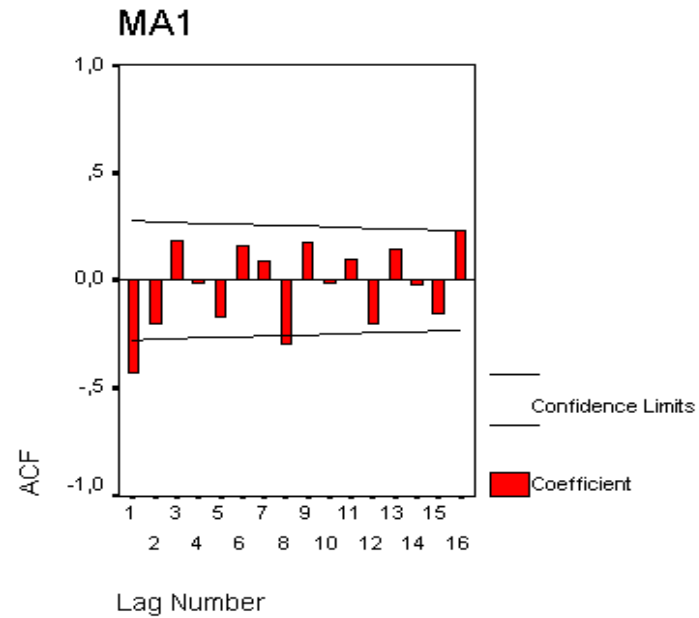
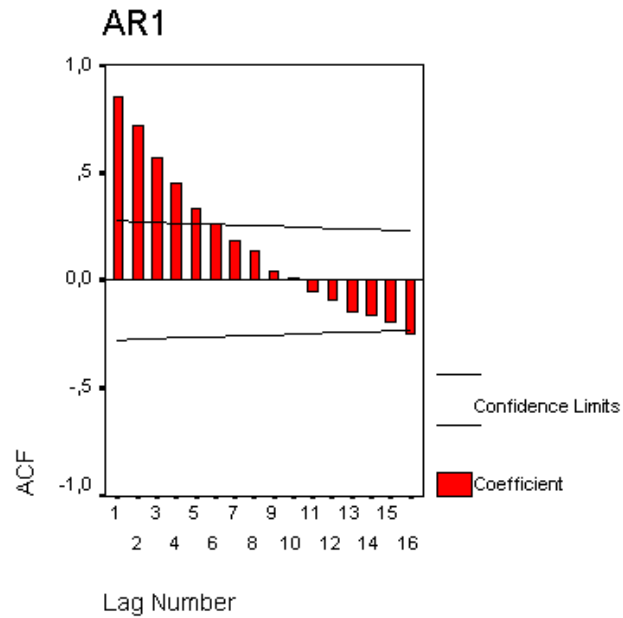
FAC (ACF) – Função de Autocorrelação

- Apresenta as autocorrelações em uma série temporal;
- As linhas horizontais mostram o intervalo de confiança;
- A 1ª autocorrelação é igual a 1. Cada traço do gráfico mostra uma defasagem e uma correlação (autocorrelação).

FACP (PACF) – Função de Autocorrelação Parcial

- Mede a autocorrelação não entre lags, mas entre diferentes intervalos de lags.

Séries Temporais



Séries Temporais

Características das Funções de Autocorrelação

	Padrão típico da FAC	Padrão típico da FACP
AR(p)	Decai exponencialmente para zero ou com padrão de onda senoidal amortecida, ou ambos.	Valores significativos, ou seja, não nulos, até a defasagem p
MA(q)	Valores significativos, ou seja, não nulos, até a defasagem q	Decai exponencialmente para zero.
ARMA(p,q)	Decai exponencialmente para zero.	Decai exponencialmente para zero.

Séries Temporais

Tipos de Modelos – ARIMA (não Sazonal)

- **Modelos Integrados** → a questão da “Estacionariedade”

[...] um **processo estocástico é estacionário** se suas média e variância forem constantes ao longo do tempo e o valor da covariância entre dois períodos de tempo depender apenas da distância ou defasagem entre os dois períodos, e não do período de tempo efetivo em que a covariância é calculada (ENDERS, 2003).

$$\text{Modelo ARIMA}(1,1,1) \rightarrow (Y_t - Y_{t-1}) = \phi_1(Y_{t-1} - Y_{t-2}) + \theta \varepsilon_{t-1} + \varepsilon_t$$

Teste de Raiz Unitária para detectar “Estacionariedade” – Teste de Dickey-Fuller

H0 : a série não é estacionária (apresenta raiz unitária)

HA : a série é estacionária (sem raiz unitária)

Séries Temporais

Tipos de Modelos – ARIMA (não Sazonal)

- $p = 1$, significa que uma determinada observação pode ser explicada pela observação prévia + erro
- $p = 2$, significa que uma determinada observação pode ser explicada por duas observações prévias + erro
- $d = 0$, significa que não é aplicada diferenciação
- $d = 1$, significa que será aplicada diferenciação de primeira ordem
- $d = 2$, significa que será aplicada diferenciação de segunda ordem
- $q = 1$, significa que uma determinada observação pode ser explicada pelo erro da observação prévia
- $q = 2$, significa que uma determinada observação pode ser explicada pelo erro de duas observações prévias.

Exemplos:

- AR(1) ou ARIMA(1,0,0) – Apenas elemento auto-regressivo de 1ª ordem
- AR(2) ou ARIMA(2,0,0) – Apenas elemento auto-regressivo de 2ª ordem
- MA(1) ou ARIMA(0,0,1) – Apenas média móvel
- ARMA(1,1) ou ARIMA(1,0,1) – Auto-regressão e média móvel de 1ª ordem

Séries Temporais

Modelos ARIMA – Como definir os valores de **p**, **d**, e **q**

- **p - ordem da parte autorregressiva:** PACF (Função de Autocorrelação Parcial) – Verifica a autocorrelação entre as observações mais atuais e observações prévias e seus erros.
- **d - grau de diferenciação:** Teste de Estacionariedade (p. ex. teste de Dickey-Fuller aumentado/Teste de Raiz Unitária) – Verifica a tendência.
- **q - ordem da média móvel:** ACF (Função de Autocorrelação) – Verifica a autocorrelação entre as observações mais atuais e os resíduos prévios.

Detalhes:

- **Série Estacionária:** média e variância são constantes no tempo e o valor da covariância entre dois períodos de tempo depender apenas da distância ou defasagem entre os dois períodos, e não do período de tempo efetivo em que a covariância é calculada.

Formas detectar estacionariedade:

- ➔ Gráfico ACF da série (processo não estacionário apresenta lento decaimento da sua função de autocorrelação. Série com tendência é o motivo mais comum para não estacionariedade).
- ➔ Teste de Estacionariedade Aumentado de Dickey Fuller (Teste de Raiz Unitária).

Séries Temporais

Metodologia de Box-Jenkins

- **Identificação** → descobrir os valores apropriados de p e q . Para perceber em que períodos de defasagem existe correlação com a variável dependente ou de correlação entre as observações com k períodos de defasagem;
- **Estimação** → estimar os parâmetros dos termos auto-regressivo e de média móvel incluídos no modelo;
- **Checagem** → Testar se o modelo escolhido sobre os resíduos estimados e verificar se são ruídos brancos (completamente aleatórios e iid - normalidade); se são, podemos aceitar o ajuste específico; se não são, devemos começar tudo novamente.
- Existem formas automáticas de **identificação e estimativa**, a **checagem** deve ser feita por testes.

Séries Temporais

Função de Estimação do Modelo ARIMA

- Pode-se pensar num modelo ARIMA como uma função de regressão populacional para Y_t em que há apenas 2 tipos de “variáveis explicativas”:
 - (1) Valores passados de $Y_t \rightarrow$ parte “autorregressiva”.
 - (2) Valores presentes e passados do distúrbio normal - u_t (ou “inovação”) \rightarrow parte de “médias móveis”.

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + u_t - \theta_1 u_{t-1} - \dots - \theta_q u_{t-q}$$

- Hiperparâmetro p : a defasagem máxima de Y_t presente na equação.
- Hiperparâmetro q : a defasagem máxima de u_t presente na equação.
- Hiperparâmetro d : ordem de integração, se o processo for não-estacionário

➔ Para modelos com sazonalidade utiliza-se o modelo SARIMA

Estatística Aplicada II

FIM DA PARTE 3 !!!

OBRIGADO