Relatório do 2º Exercício
Análise e Predição de uma Série Temporal utilizando o Método Box & Jenkins
(Série Iguaçu – Foz do Areia)

Aluno: Maxwell Queiroz Francisco
Disciplina: Probabilidade e Processos Estocásticos
Professor: Mêuser Valença

Recife, 19 de dezembro de 2012

1. Introdução

De acordo Souza(1989), o modelo Box & Jenkins tem como objetivo principal realizar uma análise das séries temporais e por conseguinte realizar a previsão desta. Para isso, Box & Jenkins, criaram uma metodologia que permite que valores futuros de uma série possam ser previstos apenas tomando como referência valores atuais e valores passados. Tudo isso graças à correlação temporal entre os dados.

Uma série temporal é uma coleção de observações feitas sequencialmente ao longo do tempo. A característica mais importante deste tipo de dados é que as observações vizinhas são dependentes e estamos interessados em analisar e modelar esta dependência. (EHLERS, 2007)

De acordo com Ehlers(2007) todo o processo que tem como o objetivo efetuar a previsão de uma série pode ser decomposta algumas etapas, tais como:

a. **Identificação**

Abordagens como técnicas descritivas, técnicas de análise de gráficos e identificação de padrões são utilizadas nesta etapa. O objetivo de descrever algumas propriedades da série como o padrão de tendência, variação sazonal e observações discrepantes (*outliers*);

b. Estimativa

Seleção, comparação e adequação de modelos, estimação de parâmetros para uma melhor predição são realizadas nessa etapa, cujas ferramentas mais utilizadas são as funções de autocorrelação (ACF) e autocorrelação parcial (PACF);

c. Diagnóstico

Nesta etapa, é necessário verificar a adequação do modelo encontrado para com os dados utilizados. Para isso, pode-se aplicar alguns testes estatísticos, verificar se o conjunto de todas as correlações dos resíduos pode ser considerado estatisticamente igual a zero;

d. Previsão e análise dos resultados

Com os valores dos parâmetros definidos é então realizada a previsão com os modelos encontrados e realizado uma análise dos resultados, a métrica para avaliação destes resultados não é muito bem definida na literatura, mas é comum encontrar trabalhos que avaliam o desempenho através desvio absoluto médio (MAD), erro quadrático médio (MSE) e o erro médio percentual absoluto (MAPE);

2. Utilização do Método

Nesta seção será realizada a aplicação do método de Box & Jenkins para análise da série temporal proposta, Iguaçu – Foz do Areia, seguindo a metodologia indicada no trabalho de Ehlers[1] sobre Análise de Séries Temporais (notas de aula).

2.1 Análise Descritiva

A Figura 1 apresenta a série temporal de Iguaçu – Foz do Areia, com observações reais médias de vazão em m³/s, no período de observação de janeiro/1931 a dezembro de 1985 (75% inicias da série) e sua respectiva média anual (agregação temporal)

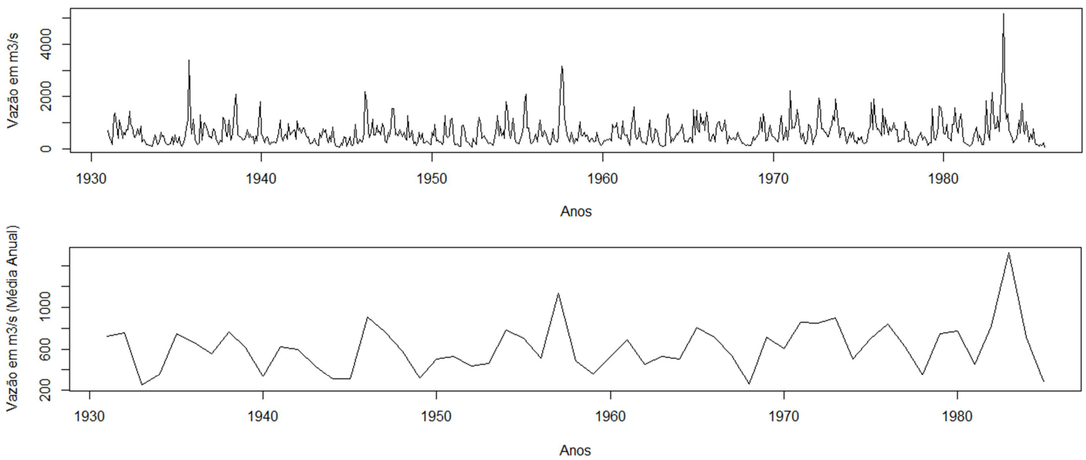

Figura 1 – Série original (cima) e média anual (baixo)

A Figura 2 mostra o resultado depois da aplicação de um filtro linear média móvel, onde este é utilizado para estimar a tendência da série.

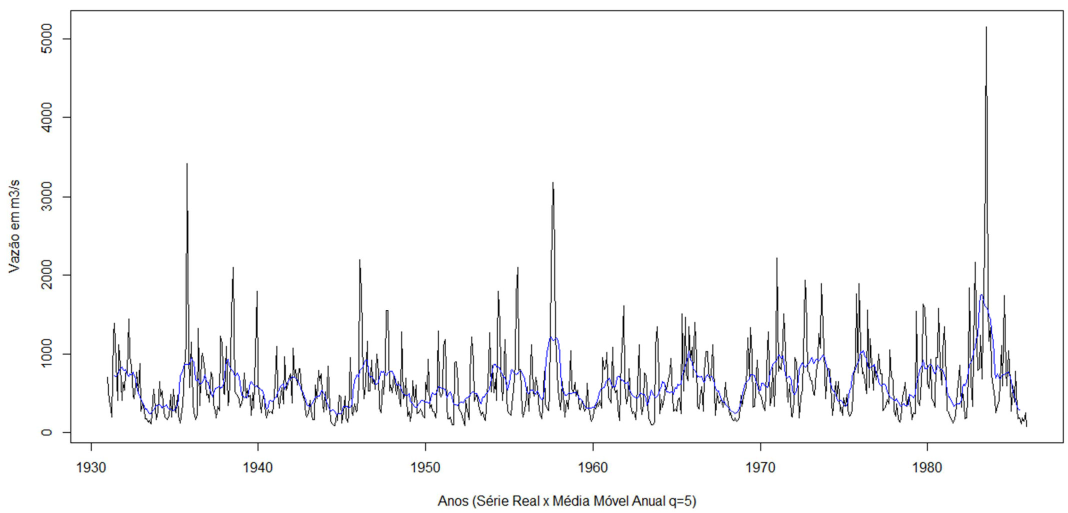

Figura 2 – Série Real x Média Móvel 1 ano (q=5)

[1] Análise de Séries Temporais, Ehlers, Acessado em 18/12/2012 através endereço eletrônico: http://www2.icmc.usp.br/~ehlers/stemp/

Onde, a linha na cor preta é a série original e a linha azul é a média móvel anual. Através da figura pôde-se analisar que a série não apresenta tendência, pois a linha azul gira em torno de uma mesma média por toda a amostra da série, sem indícios de crescimento ou decaimento constantes.

Na Figura 3 é apresentado o resultado da aplicação de uma ferramenta do R chamada "***decompose***", que estima sazonalidade e tendência via médias móveis. Para uma melhor visualização foi decomposto apenas os dez primeiros anos serie.

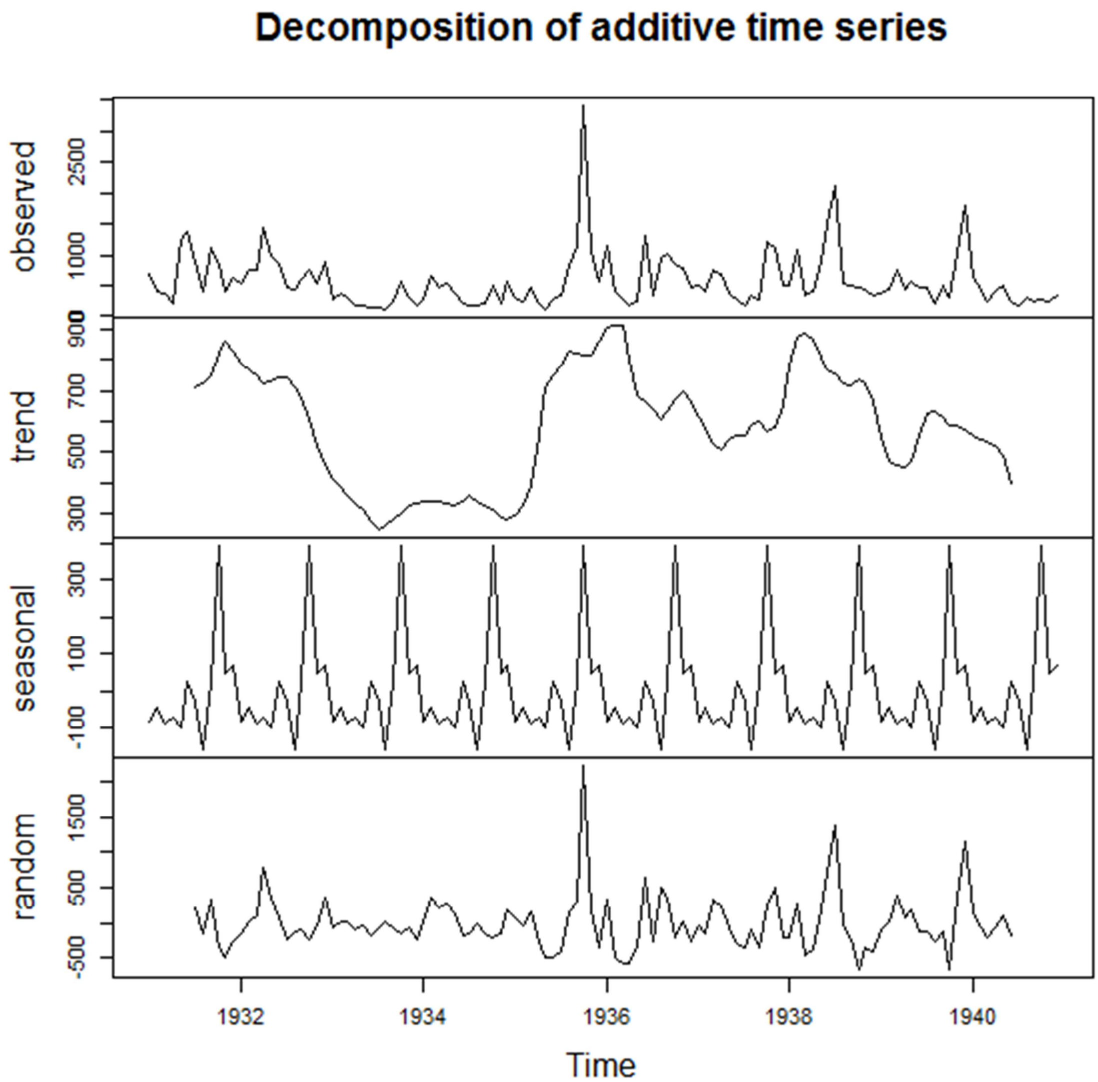

Figura 3 – Decomposição dos 10 primeiros anos da série

Pôde-se observar na linha descrita como "*seasonal*", uma sazonalidade em períodos de 12 meses, na linha "*trend*" verificou-se que não há tendência.

2.2 Estimativa de Modelos

Nesta etapa serão utilizadas as ferramentas de ACF e PACF para inferir um modelo adequado para série proposta e assim poder efetuar a previsão para o conjunto de teste.

Contudo, antes de iniciar a interpretação e cálculo desta estimativa, a literatura recomenda transformação logarítmica da série e testar sua estacionariedade, visto que o modelo Box & Jenkins é indicado apenas para séries estacionárias.

2.2.1 Estacionariedade

O método utilizado para verificar se uma determinada série é ou não estacionária, será o teste de Dickey-Fuller Aumentado (Dickey & Said, 1984), onde esse teste funcionará como um teste de hipótese que verifica por padrão, com um grau de confiança de 95%, se as séries utilizadas são ou não estacionárias.

A execução deste teste para a série proposta resultou em um *p-value* menor que 0.01, indicando a estacionariedade, visto que a hipótese nula (H0) foi rejeitada. Esse resultado serviu como um indicador que não é necessário realizar diferenciações na série o modelo Box & Jenkins.

2.2.2 Estimando MA(q)

A Figura 4 apresenta o correlograma (ACF) da série.

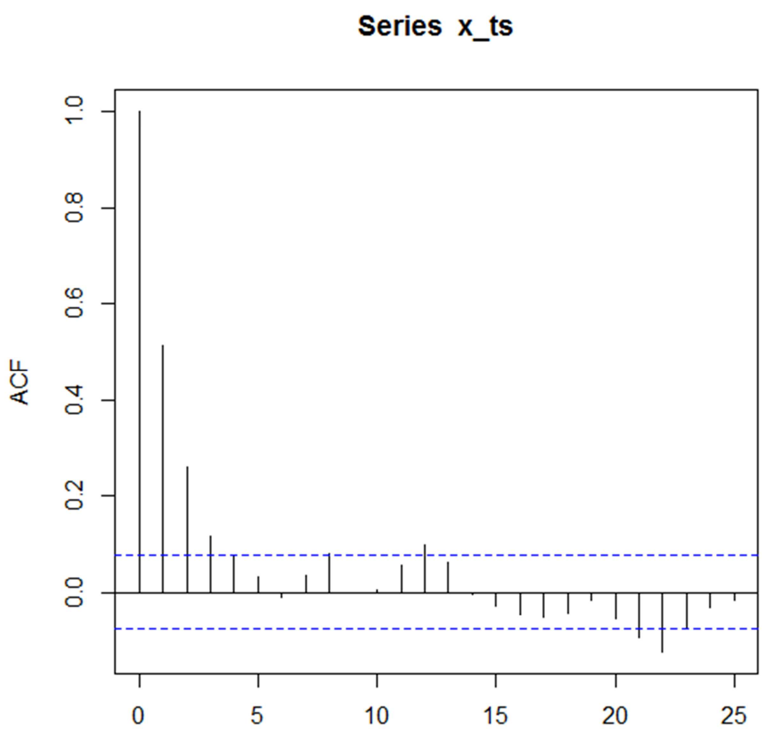

Figura 4 – Correlograma da série (ACF)

Para se estimar o coeficiente “q”, foi contado o número de defasagens fora do intervalo do de confiança, para que fosse testado no modelo ARIMA a parcimônia de cada combinação. Visualmente os valores de “q” indicados foram 3,6,9 e 12. Contudo, observou-se a repetição de “picos” nas defasagens com lag 12 e 24, isto pode indicar que um modelo MA(Q) sazonal possa resultar melhoria.

2.2.3 Estimando AR(p)

A Figura 5 apresenta o correlograma parcial (PACF) da série.

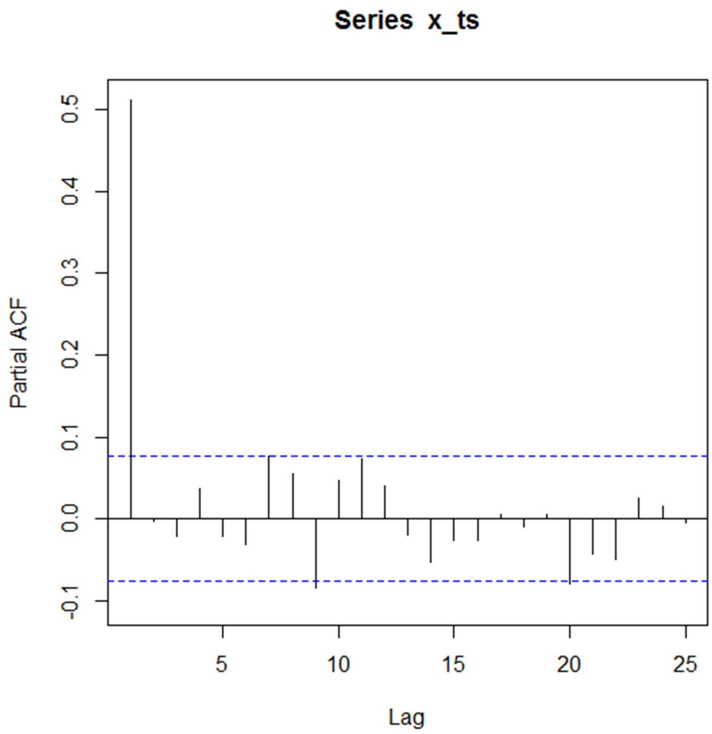

Figura 5 – Correlograma parcial da série (PACF)

Foi inferido o valor de p, de maneira análoga ao coeficiente apresentado anteriormente, sendo que com a função PACF. Observou-se pico fora do intervalo de confiança apenas na primeira defasagem, levando a crer que um valor ideal para o modelo AR(p) seja 1 (hum);

2.3 Diagnóstico

Para verificação da parcimônia entre modelos sugeridos na etapa anterior, foi utilizado o valor de AIC (*Akaike Information Criteria*) como indicador de parcimônia. A Tabela 1 mostra as diferentes este valor para diversas combinações de coeficientes.

Modelo	p	d	q	P	D	Q	AIC
ARMA	1	-	6	-	-	-	1081.7
ARMA	1	-	9	-	-	-	1075.3
ARMA	1	-	12	-	-	-	1078.8
SARMA	**1**	**-**	**9**	**0**	**-**	**1**	**1067.2**
SARMA	1	-	9	0	-	2	1068.9

Tabela 1 – Valores de AIC

2.3 Previsão e Análise dos Resultados

A etapa anterior indicou um menor valor de AIC para o modelo SARMA(1,0,9)(0,0,1), ou seja, que este modelo se mostrou mais parcimonioso que os demais. Contudo, foi realizado um teste também para o modelo ARMA (1,9);

Através desta informação, foi iniciado o processo de previsão para o conjunto de teste (25% da amostra) para os dois modelos propostos e em seguida, com os vetores dos meses previstos e vetor dos meses observados, foi possível calcular o MAPE (Erro Percentual Médio Absoluto) para verificar o desempenho do modelo.

A Figura 6 mostra duas séries sobrepostas, onde a linha preta se refere a série observada (valor esperado) e a linha em azul é a série prevista para o modelo SARMA(1,0,9)(0,0,1);

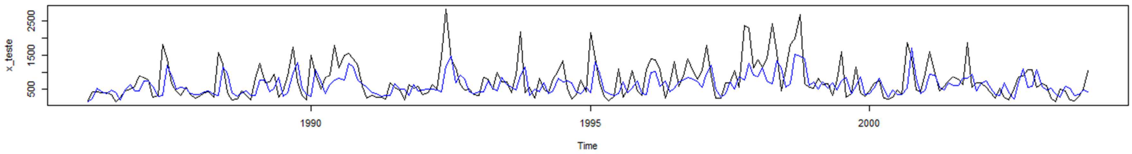

Figura 6 – Série Real x Série Prevista – Modelo SARMA(1,0,9)(0,0,1);

A Figura 7 mostra duas séries sobrepostass, onde a linha preta se refere a série observada (valor esperado) e a linha em azul é a série prevista para o modelo ARMA(1,9);

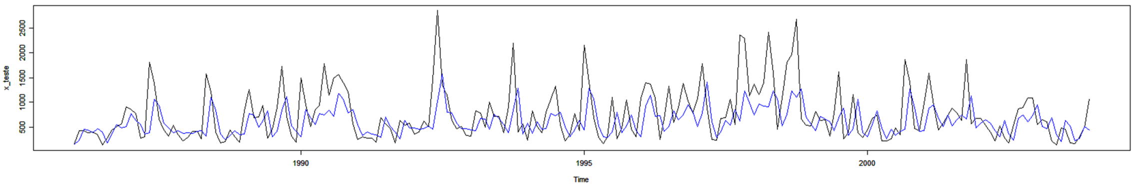

Figura 7 – Série Real x Série Prevista – Modelo ARMA(1,9)

A Tabela 2 apresenta os resultados calculados de MAPE para os modelos propostos. Observou-se que em termos de valor de MAPE o modelo SARMA obteve uma discreta melhoria, contudo seu custo computacional e tempo foi bem acentuado em relação ao modelo ARMA.

Modelo	p	d	q	P	D	Q	AIC	MAPE
ARMA	1	-	9	-	-	-	1075.3	48.736
SARMA	**1**	**-**	**9**	**0**	**-**	**1**	**1067.2**	**48.431**

Tabela 2 – Comparações MAPE entre Modelos

Referências

SOUZA, Reinaldo Castro. *Modelos Estruturais para Previsão de Séries Temporais : Abordagens Clássica e Bayesiana*. 17° Colóquio Brasileiro de Matemática. Rio de Janeiro, 1989.

S. E. Said and D. A. Dickey (1984): *Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order*. Biometrika 71, 599–607.

Ljung, G. M. and Box, G. E. P. (1978), *On a measure of lack of fit in time series models*. Biometrika 65, 553564.

Ehlers, Ricardo S. (2007), *Análise de Séries Temporais*