

MAE0325 - Séries Temporais

Fernando Henrique Ferraz Pereira da Rosa

17 de dezembro de 2004

Prova Prática¹

a) Faça uma apresentação do seu conjunto de dados.

O conjunto de dados é composto de vazões mensais de uma hidrelétrica instalada no Rio Parnaíba, controlada pela Companhia Energética de Minas Gerais (CEMIG). O período das observações consideradas nessa análise é de janeiro de 1960 a dezembro de 1975 (192 observações).

Na Figura 1 temos o gráfico dessa série. De imediato suspeita-se a existência de um componente sazonal. Essa suspeita é confirmada pelo gráfico Boxplot mostrado na Figura 2, onde agrupa-se os valores mensalmente. Observa-se que a vazão é em geral menor nos meses de junho a outubro, e tende a ter seus maiores valores em janeiro, fevereiro e dezembro. Nota-se ainda que a série não aparenta ter tendência, e que alguns anos tiveram uma variabilidade maior.

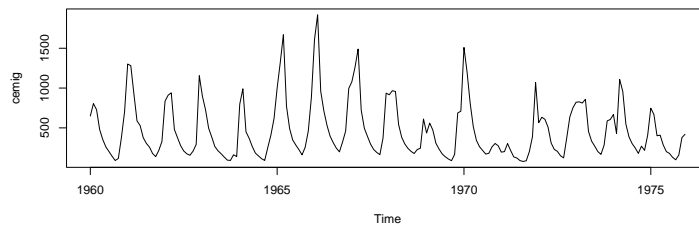


Figura 1: Gráfico da série de vazões, de jan/1960 a dez/1975.

b) Verifique se existe a necessidade de uma transformação logarítmica na série, com o objetivo de estabilizar a variância. Para responder a essa pergunta, você deverá dividir o seu conjunto de dados em seqüências de 12 observações consecutivas.

Com os comandos abaixo, no R, dividimos a série em grupos de 12 observações consecutivas e calculamos a média e o desvio padrão para cada grupo:

```
> y <- matrix(cemig, 12)
> medias <- apply(y, 2, mean)
> desvios <- apply(y, 2, sd)
```

¹Powered by L^AT_EX 2_ε, R 1.9.1 and Sweave

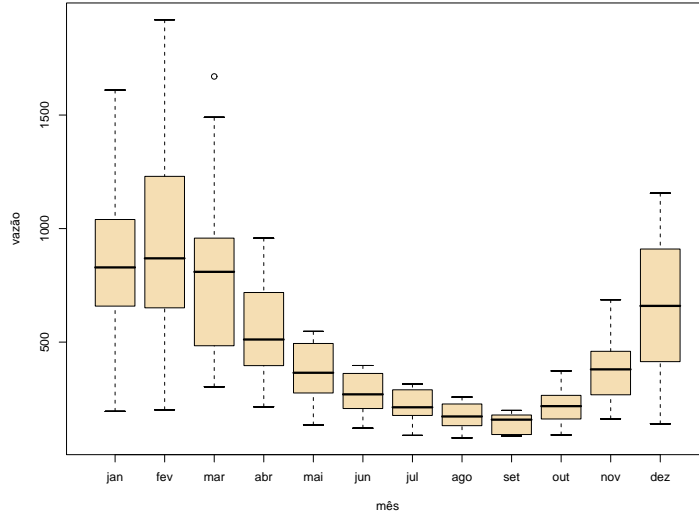


Figura 2: Gráfico Boxplot dos valores agrupados por mês.

Temos na Figura 3 o diagrama de dispersão dessas quantidades. Nota-se claramente uma associação linear positiva entre a média e o desvio-padrão, o que sugere que uma transformação logarítmica na série original é necessária para estabilizar a variância.

Prosseguiremos portanto com o restante das análises na série transformada:

$$Y_t = \ln(\text{cemig}), \quad t = 1, \dots, 192, \quad (1)$$

cujo gráfico está na Figura 4.

- c) Separe as 12 últimas observações e, com as demais, ajuste um modelo com componente sazonal estocástica (SARIMA) e apresente graficamente a série original e o modelo ajustado.

Começamos olhando o gráfico das auto-correlações e auto-correlações parciais estimadas da série transformada, sem as 12 últimas observações, na Figura 5. A fac amostral indica o comportamento sazonal da série, e a facp se comporta de maneira similar.

Consideremos agora a fac e facp amostrais das diferenças de ordem 1 (Δ), 12 (Δ_{12}) e 1 e 12 ($\Delta_1\Delta_{12}$), na Figura 6. Nota-se que o uso de diferença simples não melhora significativamente as fac e facp, optamos então por usar apenas uma diferença sazonal, ou seja $D = 1$ e $d = 0$.

Concentrando-nos no gráfico das fac e facp de Δ_{12} , vemos que um termo de médias móveis sazonal de ordem 1 pode ser apropriado, pois a correlação parcial de ordem 12 é significativa. Observemos ainda que a fac tem um comportamento de decrescimento exponencial, e que a facp, fora os lags sazonais, tem somente $\phi_{11} \neq 0$. Isso sugere um componente autoregressivo

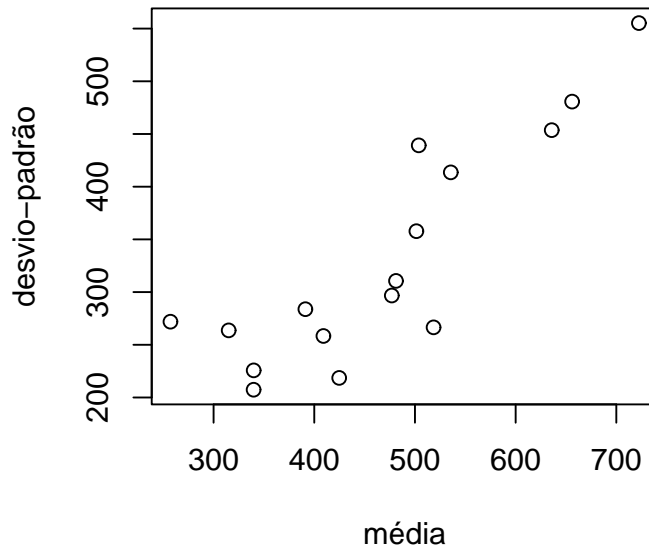


Figura 3: Gráfico de médias por desvio-padrão para grupos de 12 observações da série.

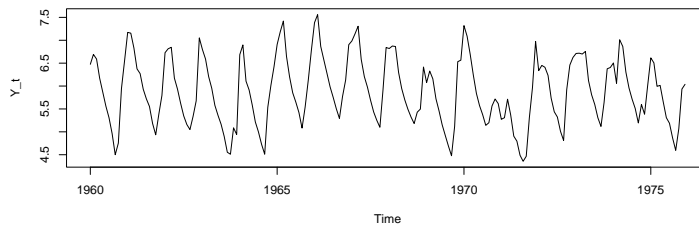


Figura 4: Gráfico da série transformada de acordo com a equação (1).

de ordem 1 na parte não-sazonal do modelo. Optamos então por um modelo preliminar dado por:

$$\text{SARIMA}(1, 0, 0) \times (0, 1, 1)_{12} \quad (2)$$

Ajustamos esse modelo no R com o comando:

```
> modelo1 <- arima(y, order = c(1, 0, 0), seasonal = list(order = c(0,
+ 1, 1), period = 12))
```

Obtendo as estimativas:

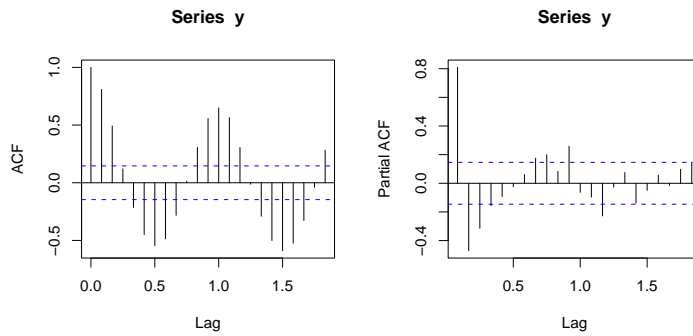


Figura 5: Gráfico da fac e facp para Y_t .

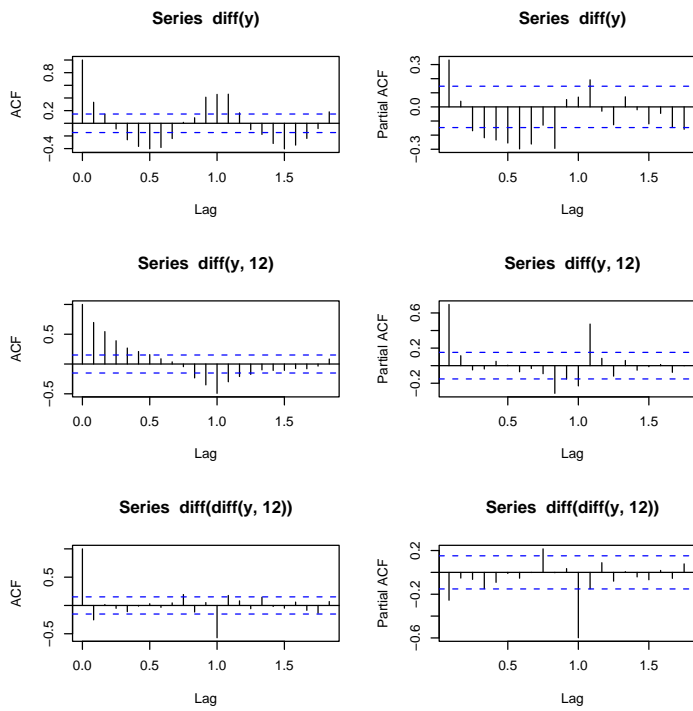


Figura 6: Gráfico da fac e facp para Δ , Δ_{12} e $\Delta\Delta_{12}$.

$$\begin{aligned}
 \hat{\phi}_1 &= 0.7376 & , & & \hat{\sigma}(\hat{\phi}_1) &= 0.0523 \\
 \hat{\Theta}_1 &= -0.8021 & , & & \hat{\sigma}(\hat{\Theta}_1) &= 0.0616 \\
 \hat{\sigma}^2 &= 0.097 & \text{ e } & & AIC &= 103.93
 \end{aligned}$$

Logo ambos os coeficientes são significantes. O valor estimado da variância também é razoável, considerando que a variância dos dados transformados é

quase 6 vezes maior: 0.5488. Na Figura 7 temos os gráficos de diagnóstico dos resíduos do modelo 2 proposto. Observa-se que a fac não tem nenhuma correlação não nula, e que os p-valores do teste de independência dos resíduos estão sempre acima de 40%. Concluí-se que o modelo está bem ajustado, com os resíduos aparentando ser ruído branco.

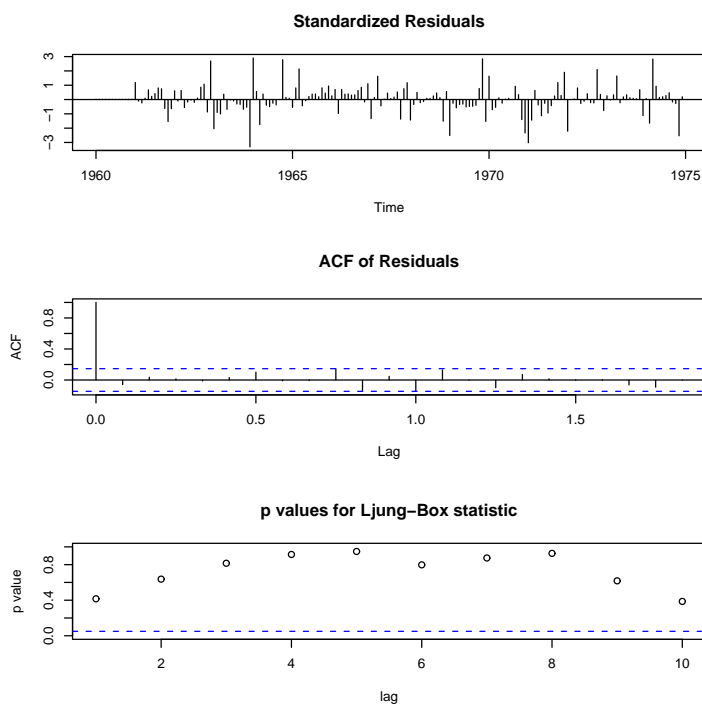


Figura 7: Gráfico de diagnósticos do modelo 2.

Na Figura 8 temos o gráfico da série original e da prevista (tracejada) pelo modelo 2. Observamos que em geral o ajuste foi bom, entretanto os picos de maior variabilidade não foram previstos.

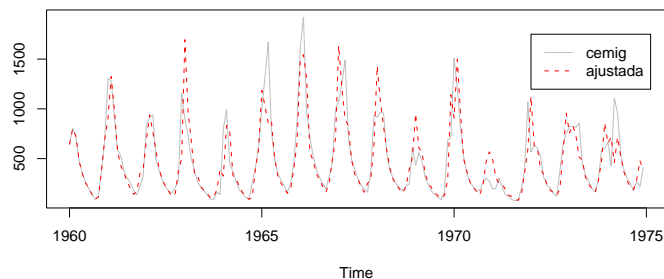


Figura 8: Gráfico da série das vazões e modelo 2 ajustado.

- d) Faça o ajustamento utilizando a metodologia Sazonalidade - Determinística - Método de Regressão (Cap. 3) apresentando graficamente a série original e o modelo ajustado.

Através dessa metodologia, procuramos modelar a série transformada através de componentes determinísticos, supondo que as observações são independentes entre si. Ajustamos um modelo do tipo:

$$Y_{ij} = T_{ij} + S_j + a_{ij}, \quad i = 1, \dots, 180, \quad j = 1, \dots, 12$$

onde o índice i representa o período, o índice j o mês, e a_{ij} é um ruído branco. T é um efeito de tendência polinomial de grau m , e S é um efeito sazonal determinístico com valores fixos para cada mês α_j :

$$T_t = \sum_{j=0}^m \beta_j t^j \text{ e } S_t = \sum_{j=1}^{12} \alpha_j D_{jt},$$

onde:

$$D_{jt} = \begin{cases} 1 & , \text{ se o período } t \text{ corresponde ao mês } j, \\ -1 & , \text{ se o período } t \text{ corresponde ao mês } 12, \\ 0 & , \text{ caso contrário, } j = 1, \dots, 12. \end{cases}$$

Em particular, tomaremos $m = 1$, ou seja, um efeito de tendência linear simples. Ajustemos então o modelo de regressão linear múltipla correspondente a:

$$Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^{12} \alpha_j D_{jt} + a_t, \quad (3)$$

com os comandos abaixo obtemos esse ajuste no R:

```
> t <- 1:length(y)
> modelo.det <- lm(y ~ t + mes, contrasts = list(mes = contr.sum))
```

Na Tabela 1 temos a Tabela de Análise de Variância para as componentes do modelo 3, onde confirmamos que o efeito de tendência é não significativo e o de sazonalidade mensal é.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
t	1	0.12	0.12	0.64	0.4246
mes	11	66.59	6.05	32.06	0.0000
Residuals	167	31.53	0.19		

Tabela 1: Análise de variância para as componentes do modelo

Na Figura 9 temos o gráfico da série de vazões e o modelo 3 ajustado, que considera todos os efeitos determinísticos.

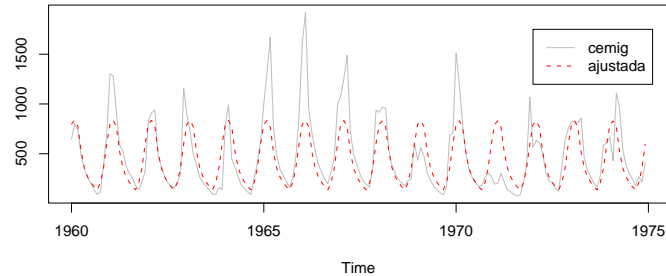


Figura 9: Gráfico da série das vazões e modelo 3 ajustado (sazonalidade determinística).

e) Qual metodologia foi mais adequada?

Comparando as Figuras 9 e 8, fica claro que o ajuste por SARIMA é bem mais adequado nesse caso. O ajuste determinístico ignora toda a estrutura de correlação dos erros e não consegue prever nenhum dos picos da série, em contraste com o modelo SARIMA.

f) Faça previsões até 12 passos a frente, utilizando os dois métodos acima e apresente graficamente a série original e as previsões.

Pelo método SARIMA, obtemos as previsões através do comando `predict`, no R:

```
> predict(modelo1, 12)
```

```
$pred
```

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
1975	6.200	6.247	6.444	6.215	5.788	5.529	5.327	5.106
	Sep	Oct	Nov	Dec				
1975	4.937	5.477	5.928	6.367				

```
$se
```

	Jan	Feb	Mar	Apr	May	Jun	Jul
1975	0.3116	0.3871	0.4226	0.4407	0.4502	0.4553	0.4581
	Aug	Sep	Oct	Nov	Dec		
1975	0.4596	0.4604	0.4608	0.4610	0.4612		

Entretanto essas previsões são para a variável transformada Y_t . Se quisermos as previsões para a variável na escala original, temos que aplicar uma transformação inversa. O problema é que ao fazermos:

$$\hat{Z}_t(h) = e^{\hat{Y}_t(h)},$$

estamos obtendo uma previsão viesada. A previsão ótima nesse caso, de acordo com a seção 9.6 do livro texto, é:

$$e^{\hat{Y}_t(h) + \frac{1}{2}V_y(h)}.$$

Levando em conta essa correção, temos as previsões na escala original, pelo modelo 2 (SARIMA) dadas na Tabela 2. Na mesma Tabela, temos também a previsão para os valores através do modelo determinístico 3, e os valores reais observados.

	original	sarima	det
jan/1975	747.00	517.32	791.19
fev/1975	669.00	556.48	828.16
mar/1975	401.00	687.86	767.52
abr/1975	409.00	550.96	519.87
mai/1975	285.00	361.07	356.65
jun/1975	202.00	279.33	267.77
jul/1975	179.00	228.62	213.69
ago/1975	130.00	183.31	168.76
set/1975	99.00	154.92	136.97
out/1975	159.00	266.01	207.11
nov/1975	378.00	417.42	356.51
dez/1975	419.00	647.34	595.48

Tabela 2: Previsões segundo os dois modelos e série original.

Na Figura 10 temos o gráfico só dos últimos 12 meses da série original, com as previsões pelos dois modelos.

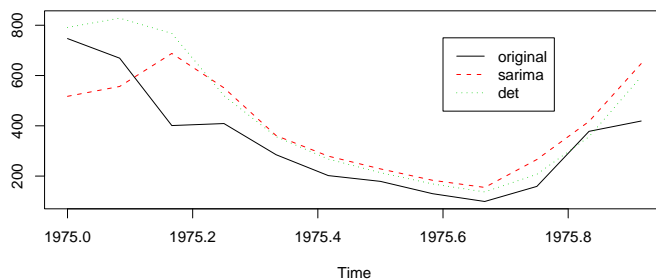


Figura 10: Previsões e série original do ano de 1975 para a série de vazões.

Por fim na Figura 11 temos a série completa e os valores previstos. Observamos que os dois métodos obtiveram resultados similares para a previsão desse período em particular, mas recomenda-se o uso do modelo SARIMA, entre outras razões, pela possibilidade de se ajustar os valores preditos de acordo com novas observações.

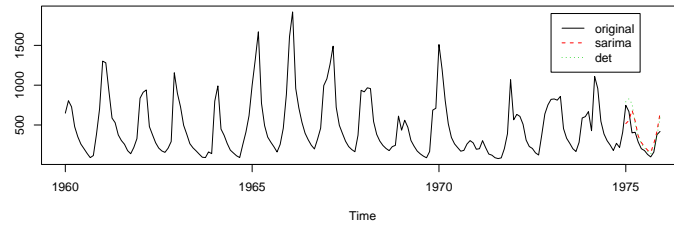


Figura 11: Previsões e série original, período completo.

Sobre

A versão eletrônica desse arquivo pode ser obtida em <http://www.feferraz.net>

Copyright (c) 1999–2005 Fernando Henrique Ferraz Pereira da Rosa.
É dada permissão para copiar, distribuir e/ou modificar este documento sob os termos da Licença de Documentação Livre GNU (GFDL), versão 1.2, publicada pela Free Software Foundation;
Uma cópia da licença em está inclusa na seção intitulada "Sobre / Licença de Uso".