MODELO ARIMA – Exercícios de Revisão

1) A série temporal do arquivo de texto anexo representa a precipitação atmosférica anual (em mm) na cidade de Fortaleza entre 1851 e 1997.

a) Utilizando o método do Gráfico Nível-Dispersão, verifique se a série é estacionária na variância. Em caso negativo, qual seria a transformação adequada para se estabilizar sua variância? Divida, obrigatoriamente, os dados em 21 partes iguais e utilize $\alpha = 0{,}05$.

b) Após a possível transformação no item (a), a série resultante é estacionária na média? Em caso negativo, determine a quantidade de diferenças necessárias para que se obtenha uma série estacionaria na média. Justifique sua resposta utilizando os métodos e teoria expostos em aula.

2) A série temporal do arquivo de texto anexo representa a média anual da temperatura (º C) global terra-mar entre os anos de 1880 e 2014.

a) Utilizando o método do Gráfico Nível-Dispersão, verifique se a série é estacionária na variância. Em caso negativo, qual seria a transformação adequada para se estabilizar sua variância? Divida, obrigatoriamente, os dados em 15 partes iguais e utilize $\alpha = 0{,}05$.

b) Após a possível transformação no item (a), a série resultante é estacionária na média? Em caso negativo, determine a quantidade de diferenças necessárias para que se obtenha uma série estacionaria na média. Justifique sua resposta utilizando os métodos e teoria expostos em aula.

3) A série temporal do arquivo de texto anexo representa a precipitação atmosférica mensal (em mm) na cidade de Manaus entre 1961 e 2007.

a) Utilizando o método do Gráfico Nível-Dispersão, verifique se a série é estacionária na variância. Em caso negativo, qual seria a transformação adequada para se estabilizar sua variância? Divida, obrigatoriamente, os dados em 47 partes iguais e utilize $\alpha = 0{,}05$.

b) Após a possível transformação no item (a), a série resultante é estacionária na média? Em caso negativo, determine a quantidade de diferenças necessárias para que se obtenha uma série estacionaria na média. Justifique sua resposta utilizando os métodos e teoria expostos em aula.

4) A série temporal do arquivo de texto anexo representa o PIB per capita do Brasil entre 1901 e 2012 em milhares de US$ de 2013.

a) Utilizando o método do Gráfico Nível-Dispersão, verifique se a série é estacionária na variância. Em caso negativo, qual seria a transformação adequada para se estabilizar sua variância? Divida, obrigatoriamente, os dados em 16 partes iguais e utilize $\alpha = 0{,}05$.

b) Após a possível transformação no item (a), a série resultante é estacionária na média? Em caso negativo, determine a quantidade de diferenças necessárias para que se obtenha uma série estacionaria na média. Justifique sua resposta utilizando os métodos e teoria expostos em aula.

5) Considere novamente a série temporal estacionária do exercício 1.

a) Quais modelos AR(p) e MA(q) puros são apropriados para modelar a série? Justifique sua resposta utilizando os métodos e teoria expostos em aula.

b) De acordo com o critério do AICc, qual dos modelos do item (a) gera um melhor ajuste?

c) Utilizando o teste do logaritmo da razão de verossimilhança, teste a hipótese que um modelo ARMA(3,2) gera um ajuste melhor que o modelo escolhido no item (b), com $\alpha = 0{,}01$. Resolva o exercício utilizando o método do valor-p.

d) Ajuste aos dados o melhor modelo do item (c) e escreva sua equação no formato:

- $\hat{Y}_t = \hat{c} + \hat{\phi}_1 \cdot \hat{Y}_{t-1} + \cdots + \hat{\phi}_p \cdot \hat{Y}_{t-p} + \hat{\varepsilon}_t + \hat{\theta}_1 \cdot \hat{\varepsilon}_{t-1} + \cdots + \hat{\theta}_q \cdot \hat{\varepsilon}_{t-q}$

e) Por meio do teste de Ljung-Box, verifique se os resíduos do modelo ajustado no item (d) não são ruído branco, com $\alpha = 0{,}01$, $m = 20$ e o método do valor-p.

f) Com α = 0,01 e utilizando a correção de Bonferroni, verifique se algum resíduo padronizado é candidato a outlier. Na resolução, devem ser indicados no mínimo a região crítica, o menor e o maior resíduos padronizados. Caso haja *outliers*, identifique-os.
g) Teste a hipótese que os resíduos do modelo ajustado não seguem a distribuição normal. Utilizar α = 0,01 e o método do valor-p.

6) A série temporal do arquivo de texto anexo representa a taxa de variação anual (em porcentagem) do consumo residencial de energia elétrica no Brasil entre 1962 e 2010.

a) Quais modelos AR(p) e MA(q) puros são apropriados para modelar a série? Justifique sua resposta utilizando os métodos e teoria expostos em aula.
b) De acordo com o critério do BIC, qual dos modelos do item (a) gera um melhor ajuste?
c) Com α = 0,05 e utilizando o teste do logaritmo da razão de verossimilhança, teste a hipótese que um modelo MA(2) se ajusta melhor à série do que um modelo AR(1). Resolva o exercício utilizando o método do valor-p.
d) Ajuste aos dados o modelo AR(1) e escreva suas equações no formato:
   - $\hat{Y}_t = \hat{c} + \hat{\phi}_1 \cdot \hat{Y}_{t-1} + \cdots + \hat{\phi}_p \cdot \hat{Y}_{t-p} + \hat{\varepsilon}_t + \hat{\theta}_1 \cdot \hat{\varepsilon}_{t-1} + \cdots + \hat{\theta}_q \cdot \hat{\varepsilon}_{t-q}$
e) Por meio do teste de Ljung-Box, verifique se os resíduos do modelo AR(1) ajustado não são ruído branco, com α = 0,05. Utilizar m = 10 e o método do valor-p.
f) Com α = 0,05 e utilizando a correção de Bonferroni, verifique se algum resíduo padronizado do modelo AR(1) ajustado é candidato a outlier. Na resolução, devem ser indicados no mínimo a região crítica, o menor e o maior resíduos padronizados. Caso haja *outliers*, identifique-os.
g) Excluindo o *outlier* detectado no item (f), teste a hipótese que os resíduos do modelo AR(1) ajustado não seguem a distribuição normal. Utilizar α = 0,05 e o método do valor-p.

7) Considere novamente a série temporal não estacionária na média do exercício 2. Aplique na mesma a 1º diferença. Seja X_t a série após esta transformação, isto é, $X_t = \Delta Y_t$:

a) Utilizando o teste do logaritmo da razão de verossimilhança, teste a hipótese que o modelo ARMA(0,2) gera um ajuste melhor que o modelo ARMA(0,1), com α = 0,05. Resolva o exercício utilizando o método do valor-p.
b) Para a série X_t, ajuste o melhor modelo do item (a) e escreva sua equação no formato:
   - $\Delta^d \hat{Y}_t = \hat{c} + \hat{\phi}_1 \cdot \Delta^d \hat{Y}_{t-1} + \cdots + \hat{\phi}_p \cdot \Delta^d \hat{Y}_{t-p} + \hat{\varepsilon}_t + \hat{\theta}_1 \cdot \hat{\varepsilon}_{t-1} + \cdots + \hat{\theta}_q \cdot \hat{\varepsilon}_{t-q}$
c) Por meio do teste de Ljung-Box, verifique se os resíduos do modelo ajustado no item (b) não são ruído branco, com α = 0,05, m = 10 e método do valor-p.
d) Verifique se algum resíduo padronizado é candidato a *outlier*, com α = 0,05 e utilizando a correção de Bonferroni. Na resolução, devem ser indicados no mínimo a região crítica, o menor e o maior resíduos padronizados. Caso haja *outliers*, identifique-os.
e) Teste a hipótese que os resíduos do modelo ajustado no item (b) não seguem a distribuição normal, com α = 0,05.

8) A série temporal não estacionária (na média e na variância) do arquivo de texto anexo representa o número de passageiros embarcados em voos regulares domésticos e internacionais ente 1927 e 2010. Aplique à mesma o logaritmo natural (para estabilizar a variância) e, ao resultado, a 1ª diferença (para estabilizar a média), isto é, $Y_t = \ln Z_t \Rightarrow X_t = \Delta Y_t$.

a) De acordo com o critério do AICc, qual modelo se ajusta melhor à série X_t: ARMA(1,1) ou ARMA(1,2)?
b) Para a série X_t, ajuste o melhor modelo do item (a) e escreva sua equação no formato:

- $\Delta^d Y_t = c + \phi_1 \cdot \Delta^d Y_{t-1} + \cdots + \phi_p \cdot \Delta^d Y_{t-p} + \varepsilon_t + \theta_1 \cdot \varepsilon_{t-1} + \cdots + \theta_q \cdot \varepsilon_{t-q}$

c) Por meio do teste de Ljung-Box, verifique se os resíduos do modelo ajustado no item (b) não são ruído branco, com α = 0,01, m = 15 e o método do valor-p.
d) Verifique se algum resíduo padronizado é candidato a *outlier*, com α = 0,10 e utilizando a correção de Bonferroni. Na resolução, devem ser indicados no mínimo a região crítica, o menor e o maior resíduos padronizados. Caso haja *outliers*, identifique-os.
e) Excluindo o *outlier* detectado no item (d), teste a hipótese que os resíduos do modelo ajustado no item (b) não seguem a distribuição normal, com α = 0,01.

9) Utilizando a semente aleatória "254216", simule uma série temporal de 10.000 termos a partir do seguinte processo estocástico MA(3):

- $Z_t = \varepsilon_t + 0{,}8 \cdot \varepsilon_{t-1} + 0{,}5 \cdot \varepsilon_{t-2} + 0{,}2 \cdot \varepsilon_{t-3}$; $\sigma^2 = 1$

a) Determine as auto correlações amostrais até a defasagem 5.
b) Utilizando o resultado do item (a), construa um intervalo de confiança 95% aproximado para as auto correlações do processo MA(3) dado até a defasagem 5.
c) Sabendo que as auto correlações teóricas de um processo MA(3) são dadas pelas fórmulas abaixo, verifique se os intervalos de confiança do item (b) contém os valores populacionais verdadeiros.

- $\rho_1 = \dfrac{\theta_1 + \theta_1 \cdot \theta_2 + \theta_2 \cdot \theta_3}{1 + \theta_1^2 + \theta_2^2 + \theta_3^2}$; $\rho_2 = \dfrac{\theta_2 + \theta_1 \cdot \theta_3}{1 + \theta_1^2 + \theta_2^2 + \theta_3^2}$; $\rho_3 = \dfrac{\theta_3}{1 + \theta_1^2 + \theta_2^2 + \theta_3^2}$; $\rho_\tau = 0$, $\tau \geq 4$

10) Classifique os processos estocásticos abaixo e determine quais são estacionários. **Justifique** as respostas

a) $Z_t = \varepsilon_t + 0{,}8 \cdot \varepsilon_{t-1} - 0{,}6 \cdot \varepsilon_{t-2} + 0{,}4 \cdot \varepsilon_{t-3} - 0{,}2 \cdot \varepsilon_{t-4}$
b) $Z_t = 0{,}8 \cdot Z_{t-1} - 0{,}8 \cdot Z_{t-2} + 0{,}8 \cdot Z_{t-3} + \varepsilon_t$
c) $Z_t = -0{,}8 \cdot Z_{t-1} + 0{,}6 \cdot Z_{t-2} - 0{,}4 \cdot Z_{t-3} + 0{,}2 \cdot Z_{t-4} + \varepsilon_t$
d) $Z_t = 0{,}5 \cdot Z_{t-1} - 0{,}9 \cdot Z_{t-2} + \varepsilon_t - 0{,}7 \cdot \varepsilon_{t-1} - 0{,}7 \cdot \varepsilon_{t-2}$
e) $Z_t = -0{,}5 \cdot Z_{t-1} + 0{,}9 \cdot Z_{t-2} + \varepsilon_t - 0{,}7 \cdot \varepsilon_{t-1} - 0{,}7 \cdot \varepsilon_{t-2} - 0{,}7 \cdot \varepsilon_{t-3}$
f) $\Delta Z_t = -0{,}8 \cdot \Delta Z_{t-1} - 0{,}5 \cdot \Delta Z_{t-2} + \varepsilon_t$
g) $\Delta^3 Z_t = 0{,}5 \cdot \Delta^3 Z_{t-1} + \varepsilon_t + 0{,}5 \cdot \varepsilon_{t-1} - 0{,}5 \cdot \varepsilon_{t-2}$

11) Uma série com tendência determinística pode ser reduzida a uma série estacionária por meio da sua estimação por mínimos quadrados e posterior subtração da série. Na sequência, os resíduos da regressão são modelados como uma processo ARMA(p,q). Considere a série temporal com tendência determinística do arquivo de texto anexo.

a) Estime a tendência determinística por meio de uma regressão linear simples.
b) Seja Y os resíduos da regressão do item (a). Utilizando o BIC, determine qual modelo se ajusta melhor à série Y_t: ARMA(2,1) ou ARMA(3,1)
c) Para a série Y, ajuste o melhor modelo do item (c) e escreva sua equação no formato:

- $\hat{Y}_t = \hat{c} + \hat{\phi}_1 \cdot \hat{Y}_{t-1} + \cdots + \hat{\phi}_p \cdot \hat{Y}_{t-p} + \hat{\varepsilon}_t + \hat{\theta}_1 \cdot \hat{\varepsilon}_{t-1} + \cdots + \hat{\theta}_q \cdot \hat{\varepsilon}_{t-q}$

Resultados Numéricos

1)
a) Valor-p = 0,0958
b) d = 0

2)
a) Valor-p = 0,414
b) d = 1

3)
a) Valor-p = 0,0133. Transformação raiz quadrada: $1-\hat{\beta}_1 = 1 - 0{,}3474 = 0{,}6526 \cong 0{,}5$
b) d = 0

4)
a) Valor-p = 0,000011. Transformação logarítmica: $1-\hat{\beta}_1 = 1 - 0{,}9369 = 0{,}0631 \cong 0$
b) d = 1

5)
a) MA(1) ou AR(1).

b) AICc[AR(1)] = 2233,941; AICc[MA(1)] = 2235,211

c) Valor-p = 0,1449914

d) $\hat{Y}_t = 1025{,}164 + 0{,}2909 \cdot \hat{Y}_{t-1} + \hat{\varepsilon}_t$

e) Valor-p = 0,5016

f) Menor rp: –2,1447607535; maior rp = 2,5828418565; região de aceitação: [–3,983082; 3,983082]

g) Valor-p = 0,04404

6)
a) MA(2) ou AR(1).

b) BIC[AR(1)] = 276,9739; BIC[MA(2)] = 277,4463

c) Valor-p: 0,064432

d) AR(1): $\hat{Y}_t = 3{,}3179 + 0{,}4928 \cdot \hat{Y}_{t-1} + \hat{\varepsilon}_t$.

e) Valor-p: 0,414.

f) Menor rp: – 5,049279703; maior rp = 1,831359987; região de aceitação: [–3,284839; 3,284839]. O resíduo correspondente ao ano de 2000 (isto é, o 39º) é um *outlier*.

g) Valor-p = 0,3107

7)
a) Valor-p: 0,020350

b) $\Delta\hat{Y}_t = 0,0064 + \hat{\varepsilon}_t - 0,4531 \cdot \hat{\varepsilon}_{t-1} - 0,1872 \cdot \hat{\varepsilon}_{t-2}$

c) Valor-p = 0,1325

d) Menor rp: –2,81402961; maior rp: 2,25314820; região de aceitação: [–3,55839; 3,55839]

e) Valor-p = 0,1806

8)
a) AICc[ARMA(1,1)] = –42,49555; AICc[ARMA(1,2)] = –40.34757

b) $\Delta\hat{Y}_t = 0,9542 \cdot \Delta\hat{Y}_{t-1} + \hat{\varepsilon}_t - 0,7379 \cdot \hat{\varepsilon}_{t-1}$

c) Valor-p = 0,8345

d) Menor rp: –2,92904691; maior rp: 4,16997200; região de aceitação:[–3,237737; 3,237737]. O primeiro resíduo padronizado é um *outlier*.

e) Valor-p = 0,005609

9)
a) r_1 = 0,676; r_2 = 0,343; r_3 = 0,106; r_4 = –0,001 e r_5 = –0,005

b)
$\rho_1 = 0,676 \pm 2\times0,01000 \quad \Rightarrow \quad 0,6560 \le \rho_1 \le 0,6960$
$\rho_2 = 0,343 \pm 2\times0,01383 \quad \Rightarrow \quad 0,3153 \le \rho_2 \le 0,3707$
$\rho_3 = 0,106 \pm 2\times0,01466 \quad \Rightarrow \quad 0,0767 \le \rho_3 \le 0,1353$
$\rho_4 = -0,001 \pm 2\times0,01474 \quad \Rightarrow \quad -0,0304 \le \rho_4 \le 0,0285$
$\rho_5 = -0,001 \pm 2\times0,01474 \quad \Rightarrow \quad -0,0304 \le \rho_5 \le 0,0285$

c)
ρ_1 = 1,3/1,93 (≅0,6736); ρ_2 = 0,66/1,93 (≅0,3420); ρ_3 = 0,2/1,93 (≅0,1036); $\rho_4 = \rho_5$ = 0. Todos os intervalos de confiança do item (b) contêm os valores populacionais verdadeiros.

10)
a) MA(4) ou ARMA(0,4). Estacionário.

b) AR(3) ou ARMA(3,0). Estacionário.

c) AR(4) ou ARMA(4,0). Não estacionário.

d) ARMA(2,2). Estacionário.

e) ARMA(2,3). Não estacionário.

f) ARIMA(2,1,0). Não estacionário.

g) ARIMA(1,3,2). Não estacionário.

11)
a) $\hat{T}_t = 20,08 + 0,50 \cdot t$

b) BIC[ARMA(2,1)] = 1025,656, BIC[ARMA(3,1)] = 1029,979

c) $\hat{Y}_t = 0,9060 \cdot \hat{Y}_{t-1} - 0,5920 \cdot \hat{Y}_{t-2} + \hat{\varepsilon}_t - 0,7344 \cdot \hat{\varepsilon}_{t-1}$