Aula 8: Estimação e testes de diagnóstico

Apresentação

Nesta aula apresentaremos os métodos de estimação de modelos ARMA, bem como testes para diagnosticar, a partir dos resíduos, se determinado modelo é ou não adequado para representar uma série.

Tratam-se da segunda e terceira etapas da metodologia proposta por Box e Jenkins, posterior à identificação e anterior à etapa de previsão.

Objetivos

- Estimar um modelo ARMA por máxima verossimilhança ou método dos momentos;
- Analisar os resíduos do modelo, verificando sua adequação para uma série temporal;
- Analisar uma série na prática, com todas as etapas necessárias para definir, estimar e efetuar o diagnóstico do modelo adequado para representar uma série real.

Estimação por máxima verossimilhança

O método geral para estimação de modelos de Box e Jenkins é o da máxima verossimilhança. Este método consiste em obter o valor dos parâmetros que torna o modelo mais "compatível" com a amostra observada. Esta função, no caso de uma amostra aleatória simples, em que as variáveis aleatórias são independentes e identicamente distribuídas, é dada por:

$$L\theta = \mathrm{f}\mathrm{Y}_1, \mathrm{Y}_2, ..., \mathrm{Y}_n = \prod_{i=1}^{n} \mathrm{f}\left(\mathrm{Y}_i\right),$$

Atenção! Para visualização completa da equação utilize a rolagem horizontal

em que $\boldsymbol{\theta}$ é o vetor de parâmetros de interesse.

Em linhas gerais, esta função representa a probabilidade de que a amostra observada $Y_1, Y_2, ..., Y_n$ seja selecionada, como função dos parâmetros desconhecidos, armazenados em θ. Ao determinarmos o valor de θ que maximiza L(θ), obtemos os valores dos parâmetros que tornam máxima a probabilidade de observar aquela amostra efetivamente observada. Trata-se, portanto, de uma abordagem de inquestionável elegância, sob o aspecto estatístico.

A aplicação deste método para modelos de séries temporais tem um caráter específico: Como não há independência, a função de verossimilhança precisa ser escrita em função das densidades condicionais (ou preditivas): $f\left(\mathrm{Y}_t \mid \mathrm{Y}_{t \to 1}\mathrm{Y}_{t \to 2}, \ \ldots\right)$. Isto porque, se as variáveis não são independentes,$f\left(\mathrm{Y}_1, \mathrm{Y}_2, ..., \mathrm{Y}_n\right) \ = \ \neq \ \prod_{i=1}^{n} \mathrm{f}\left(\mathrm{Y}_i\right)$, e sim:

$$f\left(Y_1, Y_2, \ldots, Y_n\right) = fY_n \mid Y_1, Y_2, \ldots, Y_{n-1} fY_{n-1} \mid Y_1, Y_2, \ldots, Y_{n-2} \ldots fY_2 \mid Y_1 fY_1$$

Atenção! Para visualização completa da equação utilize a rolagem horizontal

No caso de um modelo de séries temporais, no qual, exceto para o caso trivial do ruído branco, as variáveis apresentarão uma estrutura de dependência temporal, teremos:

$$L(\theta) = fY_n \mid Y_1, Y_2, \ldots, Y_{n-1} fY_{n-1} \mid Y_1, Y_2, \ldots, Y_{n-2} \ldots fY_2 \mid Y_1 fY_1,$$

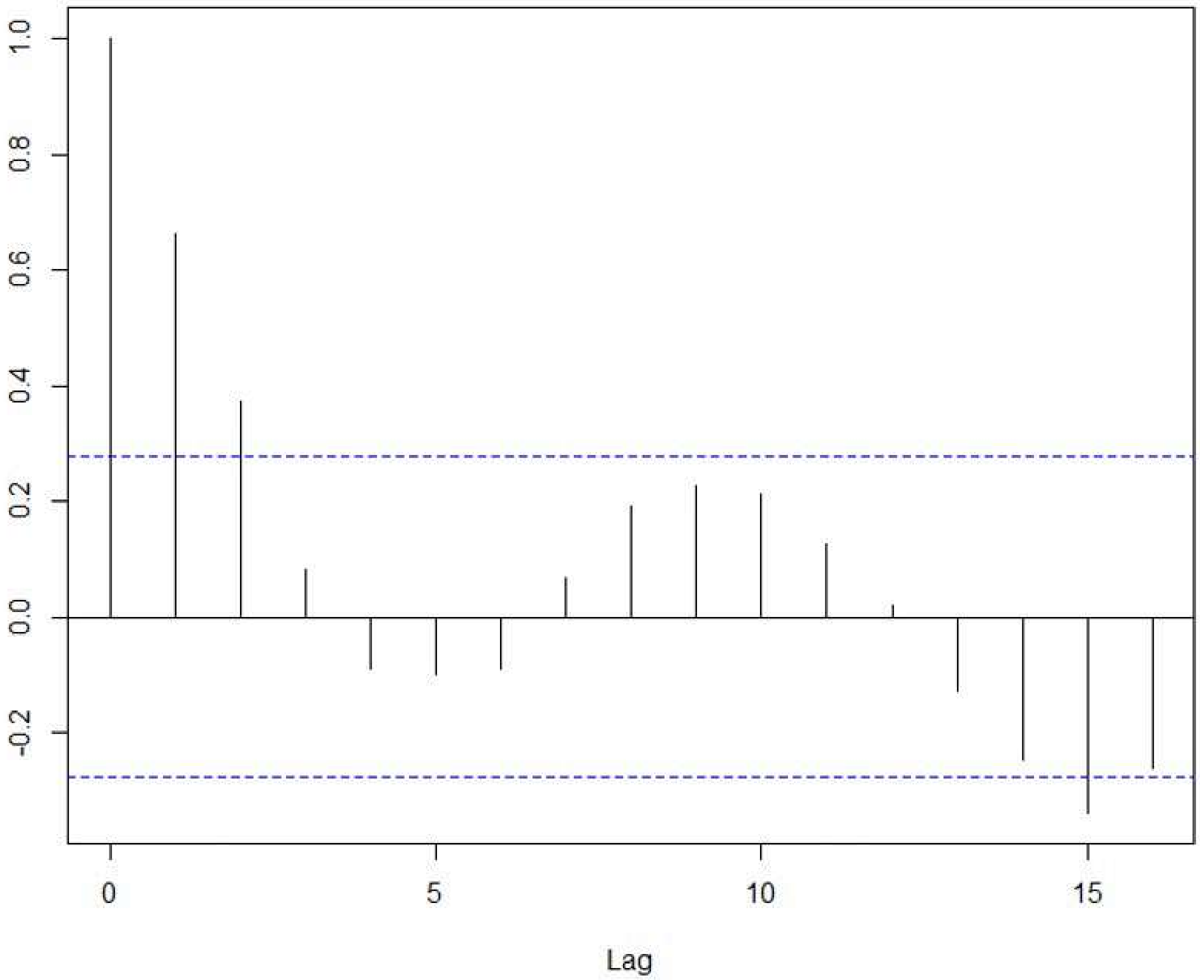
Atenção! Para visualização completa da equação utilize a rolagem horizontal

em que θ é um vetor de parâmetros que contém $\theta_1, \theta_2, \ldots \quad \theta_p, \theta_1, \theta_2, \ldots \theta_p$ e σ^2. A maximização da função acima se dá por métodos numéricos.

Estimação pelo método dos momentos

É outra forma de obter estimativas dos parâmetros de modelos ARMA simples. Consiste em utilizar as expressões teóricas da FAC, que são funções dos coeficientes do modelo, substituir nestas fórmulas as estimativas das correlações, e resolver para os coeficientes.

No exemplo a seguir, vemos a FAC estimada para uma série temporal:


1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
5
10
15
Lag

FAC e FACP estimadas para a série do exemplo 1.3 - Fonte: Autor

Sabendo que a FACP estimada é truncada no *lag* 1 e a média do modelo é zero, identifique o modelo adequado e estime-o usando o método dos momentos.

Solução: Como a FAC decai exponencialmente e a FACP é truncada no *lag* 1 (ou seja, apresenta valor estatisticamente significante no *lag* 1 e valores não significantes para lags k = 2, 3, ...), identifica-se um AR(1). Como a média é zero, temos que $\theta_0 = 0,$ e, desta forma, a equação do modelo é: Yt = $\mathrm{Y}_t = \theta \mathrm{Y}_{\mathrm{t-1}} + \varepsilon$.

Por outro lado, sabe-se que a função de autocorrelação do modelo AR(1) é:

$$\rho_k = \phi^k, k = 0, \quad 1, \quad 2, \quad \ldots$$

Atenção! Para visualização completa da equação utilize a rolagem horizontal

Para k = 1, temos: $\rho_1 = \phi$.

Desta forma, a estimação pelo método dos momentos é trivial, bastando igualar:

$$\hat{\theta} = \hat{p}_1$$

Atenção! Para visualização completa da equação utilize a rolagem horizontal

Conclui-se assim que $\hat{\theta} = 0.63$, e o modelo final estimado é: $y_t = 063 y_{t1} + \varepsilon_{t.}$.

Vamos praticar?

O correlograma de uma série é truncado no lag 1, com $\hat{p}_1 = -0,4$. Identifique o modelo adequado para esta série e estime-o utilizando o método dos momentos.

Solução: Utilizando o procedimento de identificação apresentado na aula 7, vemos que se trata de um modelo $MA(1): Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$. Desta forma, temos que resolver a equação:

$$\rho_1 = \frac{-\theta}{(1+\theta^2)} = \quad -0.4$$

Para θ. Assim:

$$-\theta = -0.4 \quad 1 + \theta^2$$

$$\theta = 0.4 + 0.4\theta^2$$

$$0.4\theta^2 - \theta + 0.4 = 0$$

$$\theta = \frac{--1\sqrt{-1^2 - 40.40.4}}{20.4}$$

$$= \frac{1\sqrt{1-0.64}}{0.8} = \frac{1\sqrt{0.36}}{0.8} = \frac{1 \pm 0.6}{0.8}$$

Temos assim duas raízes:

$$\theta = \frac{1+0.6}{0,8} = 2e\theta = \frac{1-0.6}{0,8} = 0.5$$

Como θ=2 conduz a um modelo não inversível, temos que a solução é θ=0.5 e o modelo final estimado é $Y_t = \varepsilon_t - 0,5\varepsilon_{t-1}$.

Testes de significância estatística

Os testes de significância dos coeficientes de um modelo estimado baseiam-se na distribuição t de Student com n-(p+q) graus de liberdade. As hipóteses são:

- H_0 (hipótese nula): o coeficiente é não significante.
- H_1 (hipótese alternativa): o coeficiente é significante.

Atenção! Aqui existe uma videoaula, acesso pelo conteúdo online

Saiba mais

Este teste pode ser conduzido com base no p-valor calculado para a estimativa do parâmetro, fornecido pelo R para o conjunto de parâmetros do modelo estimado por meio do comando *coeftest*. O procedimento de teste é bem simples: se o p-valor é inferior a 0.05, rejeita-se a hipótese nula de que o coeficiente é não significante, concluindo-se pela significância do mesmo. Consideramos aqui o nível de significância mais usual de um teste, que é o de 0.05.

Testes de significância do modelo AR(2) identificado anteriormente, no R

library(lmtest) # carrega a biblioteca que contém a função coeftest a seguir

fit1 <- Arima(treino,order = c(2,0,0))

coeftest(fit1) # mostra a significância das estimativas

	Estimate	Std. Error	z value	Pr(>|z|)
ar1	0.295787	0.062258	4.7510	2.025e-06 ***
ar2	0.269755	0.062206	4.3365	1.448e-05 ***
intercept	0.090290	0.287114	0.3145	0.7532

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

fit2 <- Arima(treino,order = c(2,0,0),include.constant = F) # retira o θ_0 não significante

A saída da função "coeftest" é:

	Estimate	Std. Error	z value	Pr(>|z|)
ar1	0.318959	0.063842	4.9961	5.850e-07 ***
ar2	0.253023	0.063990	3.9541	7.682e-05 ***

A equação do modelo é: $Yt = 0.3189Y_{t-1} + 0.2530Y_{t-2}$. Porém, como mencionado na aula 7, este não necessariamente é o modelo final, uma vez que o procedimento de identificação tende a subidentificar o modelo, isto é, conduzir a valores de p e/ou q que menores do que os reais.

Testes de sobrefixação

Clique no botão acima.

O modelo identificado e estimado (no exemplo anterior, o AR(2)) é apenas um ponto de partida, do qual serão conduzidos testes de sobrefixação.

Tais testes consistem em estimar, além do modelo identificado, modelos com acréscimos de uma unidade em p, em q e em ambos, e avaliar a significância da estimativa estatística do(s) parâmetro(s) incidental(is), em geral ao nível de significância 0.05, em cada caso.

Por exemplo, se inicialmente foi identificado um AR(1), ou seja, p = 1 e q = 0, deve-se estimar, além deste modelo, um modelo com p = 2 e q = 0, ou seja, um AR(2), um modelo com p = 1 e q = 1, ou seja, um ARMA(1,1), e um modelo com p = 2 e q = 1, ou seja, um ARMA(2,1). Em outras palavras, devemos testar a significância das estimativas dos parâmetros ϕ_1 e/ou θ_1.

O passo seguinte é verificar, para cada modelo sobrefixado, se a estimativa do(s) parâmetro(s) introduzido(s) é significante. Ocorrerá então uma das seguintes situações (de 1 a 6) a seguir:

1 - Ambos resultam não significantes, inclusive quando especificadas conjuntamente. Neste caso, escolhe-se o modelo inicialmente identificado (no exemplo, AR(1)) como modelo final.

2 - Se apenas o ϕ_2 for significante, passamos a ter o modelo AR(2) como candidato a modelo final. Ele então deve ser comparado ao modelo inicial (veremos como na seção 8.5 e, caso saia vencedor, o procedimento de sobrefixação deve continuar a partir dele, isto é, deve-se testar um AR(3), um ARMA(2,1) e um ARMA(3,2) e repetir o procedimento, até que os parâmetros incidentais de todos os modelos sobrefixados em uma etapa sejam não significantes.

3 - Se apenas o θ_1 for significante, passamos a ter o modelo ARMA(1,1) como candidato a modelo final. Ele então deve ser comparado ao modelo inicial (veremos como na seção 8.5) e, caso saia vencedor, o procedimento de sobrefixação deve continuar, ou seja, devemos testar um ARMA(2,1), um ARMA(1,2) e um ARMA(2,2) e repetir o procedimento até que os parâmetros incidentais de todos os modelos sobrefixados em uma etapa sejam não significantes.

4 - Se ambos, ϕ_2 e θ_1, forem significantes apenas quando especificados isoladamente, temos um AR(2) e um ARMA(1,1) como candidatos a modelo final. Devemos então comparar esses modelos com o modelo inicial (veremos como na seção 8.5) e, a não ser que o modelo inicial saia vencedor, reiniciamos a partir do novo modelo nosso procedimento de identificação.

5 - Se ambos, ϕ_2 e θ_1, forem significantes apenas quando especificados conjuntamente, compara-se o ARMA(2,1) com o modelo inicial e, caso saia vencedor, reiniciamos a partir dele o procedimento de identificação, testando um ARMA(3,1), um ARMA(2,2) e um ARMA(3,2).

6 - Se ambos, ϕ_2 e θ_1, forem significantes, tanto isoladamente quanto conjuntamente, compara-se os três modelos correspondentes (AR(2), ARMA(1,1) e ARMA(2,1)) com o modelo AR(1) inicial (veremos como na seção 8.5) e, a não ser que o modelo inicial saia vencedor, reiniciamos a partir do novo modelo o processo de identificação.

Testes de sobrefixação para o modelo AR(2) no R

```
fit3<-Arima(Y,order = c(3,0,0), include.constant=F) # sobrefixando a parte AR


coeftest(fit3) # Os resultados são:

        Estimate    Std. Error   z value    Pr(>|z|)

ar1     0.290602    0.064607     4.4980   6.861e-06 ***
```

```
ar2    0.263847   0.065197    4.0469  5.189e-05 ***

ar3    0.020932   0.065035    0.3219    0.7476
```

Constata-se, assim, que o termo AR(3) não foi significante (p-valor = 0.7476).

```
fit4<-Arima(Y,order = (2,0,1), include.constant=F) # sobrefixando a parte MA

coeftest(fit4)
```

Os resultados dos comandos acima são:

```
        Estimate    Std. Error  z value    Pr(>|z|)

ar1      0.40245     0.25520    1.5770   0.11479

ar2      0.22654     0.12405    1.8262   0.06782 .

ma1  -0.11440     0.26349   -0.4342   0.66418
```

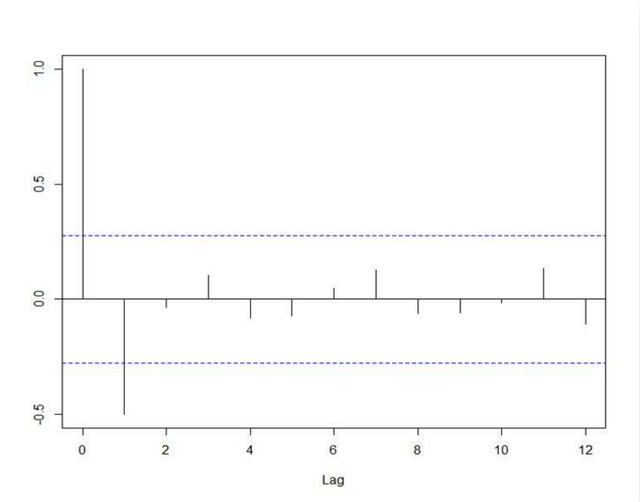
Constata-se, assim, que o termo MA(1) não foi significante (p-valor = 0.6642).

Quando ambos os parâmetros foram introduzidos simultaneamente, o modelo resultante (ARMA(3,1)) também resultou não significante. Desta forma, retornamos ao modelo previamente identificado e concluímos que o modelo final é o AR(2), armazenado em fit2.

Veja o exemplo que separamos para você.

Considere o procedimento de identificação a partir da FAC estimada, apresentada a seguir, e da FACP, que não é apresentada, mas sabe-se que é caracterizada por um decaimento exponencial. Aponte o modelo inicial e os testes de sobrefixação necessários.

FAC estimada para a série temporal



Fonte: Autor

Atenção! Aqui existe uma videoaula, acesso pelo conteúdo online

Solução: Como a FAC é truncada no *lag* 2, e dada a característica da FACP, identifica-se um MA(2). Como p = 0 e q = 2, os primeiros testes de sobrefixação deverão ser conduzidos para os modelos com p = 1 e q = 2, ou seja, um ARMA(1,2), com p = 0 e q = 3, ou seja, um MA(3), e com p = 1 e q = 3, ou seja, um ARMA(1,3).

Uma questão em aberto é: Como comparar os modelos no processo de sobrefixação? A princípio, bastaria olhar para a soma dos quadrados dos resíduos (escolhendo o menor). O modelo vencedor continua na disputa. Mas a comparação entre modelos com diferentes números de parâmetros não é tão simples, e será objeto de estudo a seguir.

Critérios de informação

O procedimento de identificação apresentado anteriormente foi mostrado de forma simplificada. Na prática, ao sobrefixar um AR(1) para um ARMA(1,1), mesmo que o parâmetro θ_1 seja significante, não necessariamente se conclui que o ARMA(1,1) é superior. É fato que um modelo com mais parâmetros (no caso, o ARMA(1,1)) tende a fornecer um melhor ajuste.

Comentário

Desta forma, se compararmos estes modelos via SQR ou função de verossimilhança, é certo que o ARMA(1,1) sairá vencedor. Porém, é fato também que um bom modelo deve ser parcimonioso, ou seja, explicar bem o comportamento de uma série utilizando o menor número de parâmetros possíveis. Desta forma, somos conduzidos a utilizar critérios de informação, que servem para comparar modelos com número de parâmetros diferentes.

Os critérios de informação usuais são:

$\mathrm{AIC} = \ln\hat{\sigma}^2 + p + q\frac{2}{T}$

Critério de Schwarz ou Bayesiano: $BIC \quad ou \quad SBC \quad = \quad \ln\hat{\sigma}^2 + p + q\frac{\ln T}{T}$.

em que $\hat{\sigma}^2 = \frac{\sum_{t=1}^{T}\hat{\varepsilon}_t^2}{T}$ é a média dos quadrados dos resíduos.

Saiba mais

Em linhas gerais, eles computam uma função crescente da soma de quadrados dos resíduos e do número de parâmetros. Assim, queremos que o valor resultante seja o menor possível. A segunda parcela penaliza pelo número de parâmetros. Nota-se que esta penalização é mais severa no BIC, para T $\geq$ 8 (pois ln(8) > 2). O BIC é, portanto, mais parcimonioso do que o AIC.

Veja o exemplo que separamos para você.

Considere que no primeiro exemplo que apresentamos nessa aula, o parâmetro incidental na parte AR tenha sido significativo, ao passo que o da parte MA não. Quando especificados isoladamente, em um ARMA(1,3), o modelo resultante foi não significativo. Desta forma, teríamos os modelos MA(2) e ARMA(1,2) como possíveis candidatos, antes de continuar o processo. Suponha agora que esses modelos tenham apresentado os seguintes valores para os critérios de informação:

	AIC	BIC
MA(2)	568	493
ARMA(1,2)	615	577

Como ambos os critérios apontam para o MA(2), concluímos que este é o modelo final. Note que, neste caso, não é necessário continuar o processo, uma vez que os modelos sobrefixados correspondentes ao MA(2) já haviam sido testados (entre eles, o próprio ARMA(1,2)).

Uma vez definido o modelo, é necessário efetuar os testes de diagnóstico. Isto porque, se o modelo não for "aprovado" na análise de resíduos, deve-se abortar o processo e retornar ao início da modelagem. É essencial que os resíduos do modelo sejam "bem comportados".

Obtendo os Critérios de informação para o modelo AR(2)

summary(fit2)

Os resultados são apresentados a seguir.

sigma^2 estimated as 3.807:
log likelihood=-475.09
AIC=956.19 BIC=966.48

Observe que os critérios de informação só têm sentido quando há comparação de modelos. Os valores de AIC e BIC acima, portanto, não possuem interpretação em termos absolutos.

Diagnósticos

Uma vez identificado e estimado um modelo, devemos verificar se os resíduos apresentam as propriedades esperadas para eles. Ou seja, se eles se comportam como um ruído branco.

Lembrando as propriedades do ruído branco:

Atenção! Aqui existe uma videoaula, acesso pelo conteúdo online

$$E\left(\varepsilon_t\right) = 0, \quad \forall t V\left(\varepsilon_t\right) = \sigma^2, \quad \forall t Corr\left(\varepsilon_i, \varepsilon_j\right) = 0, \quad \forall t \neq j\,.$$

Atenção! Para visualização completa da equação utilize a rolagem horizontal

O passo mais importante é detectar a possibilidade de autocorrelação. Se os resíduos apresentarem algum padrão de dependência temporal, o modelo é invalidado e, neste caso, deve-se buscar uma alternativa a partir do procedimento de identificação (ou seja, voltamos à etapa 1).

Desta forma, deve-se verificar se os resíduos são descorrelacionados como um ruído branco, ou seja, se:

$$\rho_1 = \rho_2 = \ \dots \ = 0_l$$

Atenção! Para visualização completa da equação utilize a rolagem horizontal

sendo ρ_j a autocorrelação de *lag* j dos resíduos. Uma possibilidade é verificar se o IC para ρ_j:

$$IC_{100(1-\alpha)\%}\left(\rho_j\right) = \rho_j \pm z_{\frac{\alpha}{2}}\sqrt{V}\rho_j,$$

Atenção! Para visualização completa da equação utilize a rolagem horizontal

sendo $\sqrt{\widetilde{V}}\hat{\rho}_j \cong \frac{1}{T}$, contém o zero, para cada j = 1,2...

Comentário

Um problema com esta abordagem é que a (multi)colinearidade pode levar a conclusões erradas a respeito da significância dos testes individuais, que têm baixo poder.

A alternativa recomendada é o teste de Ljung-Box, que investiga a significância conjunta das autocorrelações estimadas.

Denotando por ρ_j a autocorrelação de *lag* j da série de resíduos, as hipóteses do teste são:

H_0: $\rho_1 = \rho_2 = \rho_3 \ldots = \rho_K = 0$.

H_1: ao menos um ρ_j (j = 1, 2, ..., K) ρ 0.

Atenção! Para visualização completa da equação utilize a rolagem horizontal

em que K é um valor suficientemente grande. Para séries não sazonais, k entre 10 e 20 costuma ser suficiente. Para séries sazonais, é importante considerar 2 ou 3 *lags* sazonais. Se estamos falando de uma série mensal com sazonalidade anual, K entre 30 e 40 é apropriado.

Estatística do teste de Ljung-Box:

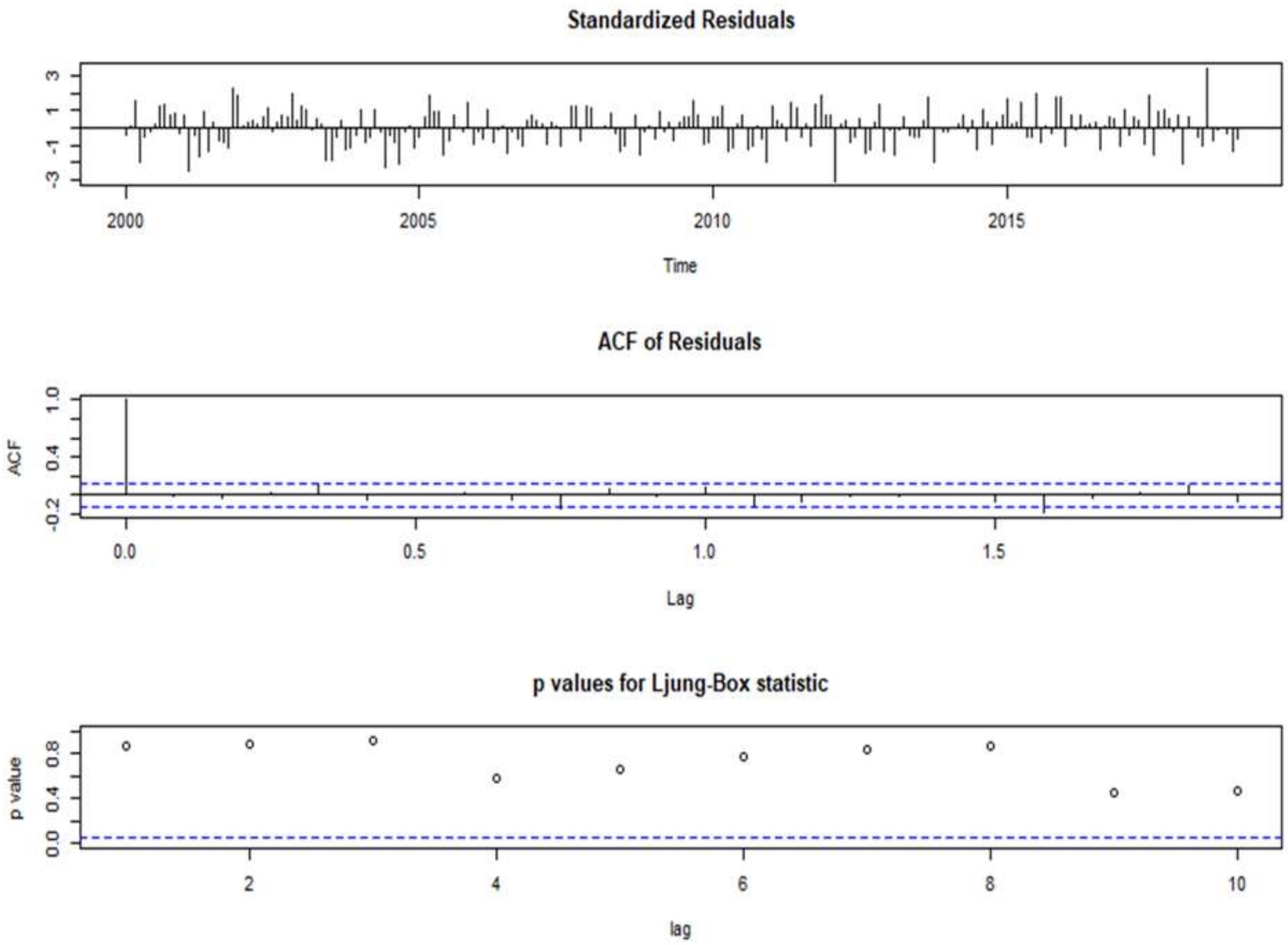
$$0 = nn + 2\sum_{j=1}^{K} \frac{\hat{\rho}_j^2}{(n -)}$$

Atenção! Para visualização completa da equação utilize a rolagem horizontal

Sob H_0, $Q \sim X^2_{K-(p+q)}$. P"or exemplo, para o AR(p): $Q \sim X^2_{K-p}$

Teste de Ljung-Box no R, para o modelo AR(2)

tsdiag(fit) # O resultado é mostrado na figura a seguir.



Teste de Ljung-Box para os resíduos do modelo estimado. Fonte: Autor

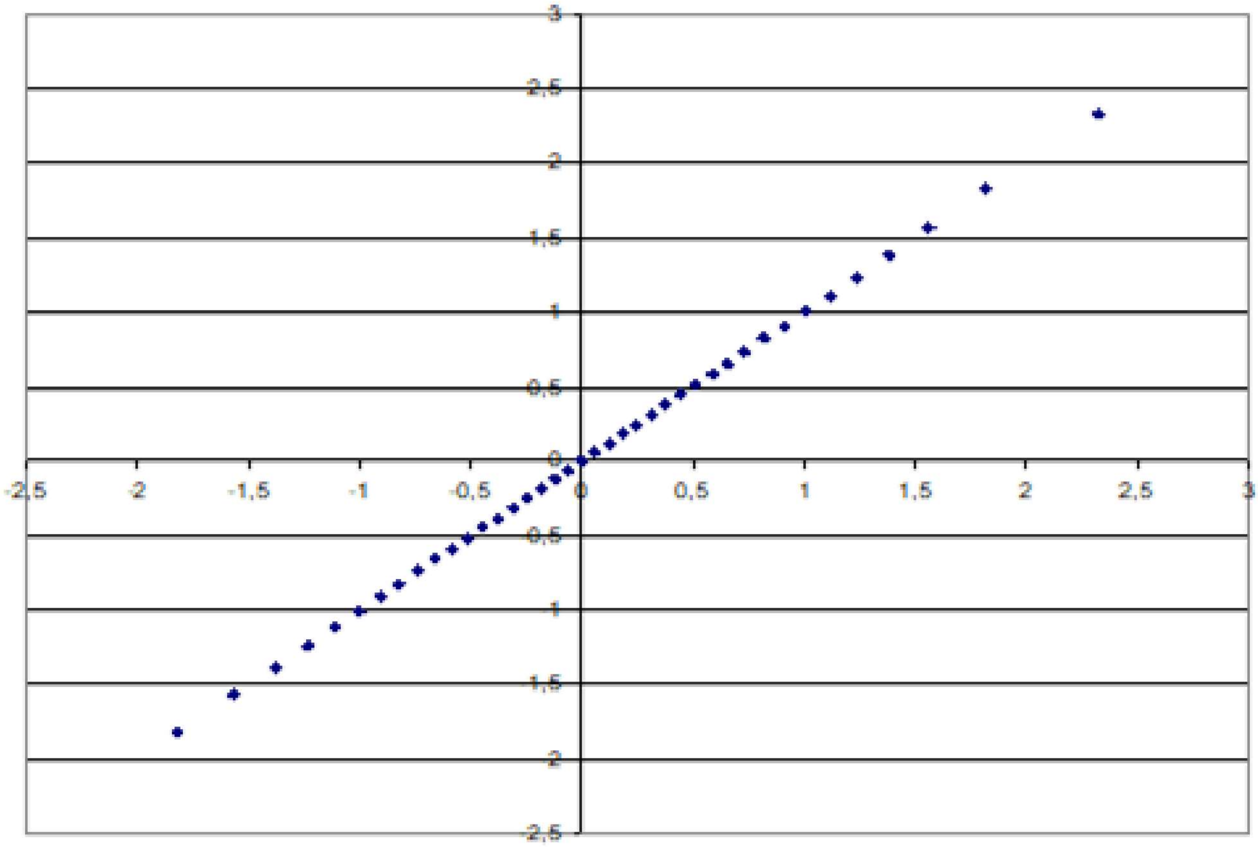
Saiba mais

A linha pontilhada azul no terceiro gráfico representa o limite de 0.05 para o p-valor. Todos os p-valores são maiores do que 0,05, e assim a hipótese de erros seguindo um ruído branco não é rejeitada.

Também importante é a detecção de desvios da normalidade, os quais podem ser detectados mediante um gráfico dos resíduos denominado QQ-plot. Trata-se de um gráfico dos percentis da distribuição dos resíduos padronizados e da distribuição Normal padrão.

O formato “ideal” de um QQ-Plot é o de uma reta de 45º, como apresentado a seguir

:



QQ-PLot “Ideal” Fonte: Autor

Também pode ser usado um teste de Normalidade. O mais usual para resíduos de modelos de séries temporais é o teste de Jarque-Bera, cujas hipóteses são:

- H_0: erros Normais;
- H_1: erros não Normais.

A estatística do teste de JB é:

$$JB = \frac{n}{6}a^2 + \frac{(k-3)^2}{4}.$$

Atenção! Para visualização completa da equação utilize a rolagem horizontal

Em que a e k são os estimadores dos coeficientes de assimetria e de curtose: $a = \frac{\sum_{i=1}^{n}\left(\hat{\varepsilon}_i - \hat{\varepsilon}\right)^3}{nS^3}$ $k = \frac{\sum_{i=1}^{n}\left(\hat{\varepsilon}_i - \hat{\varepsilon}\right)^4}{nS^4}$

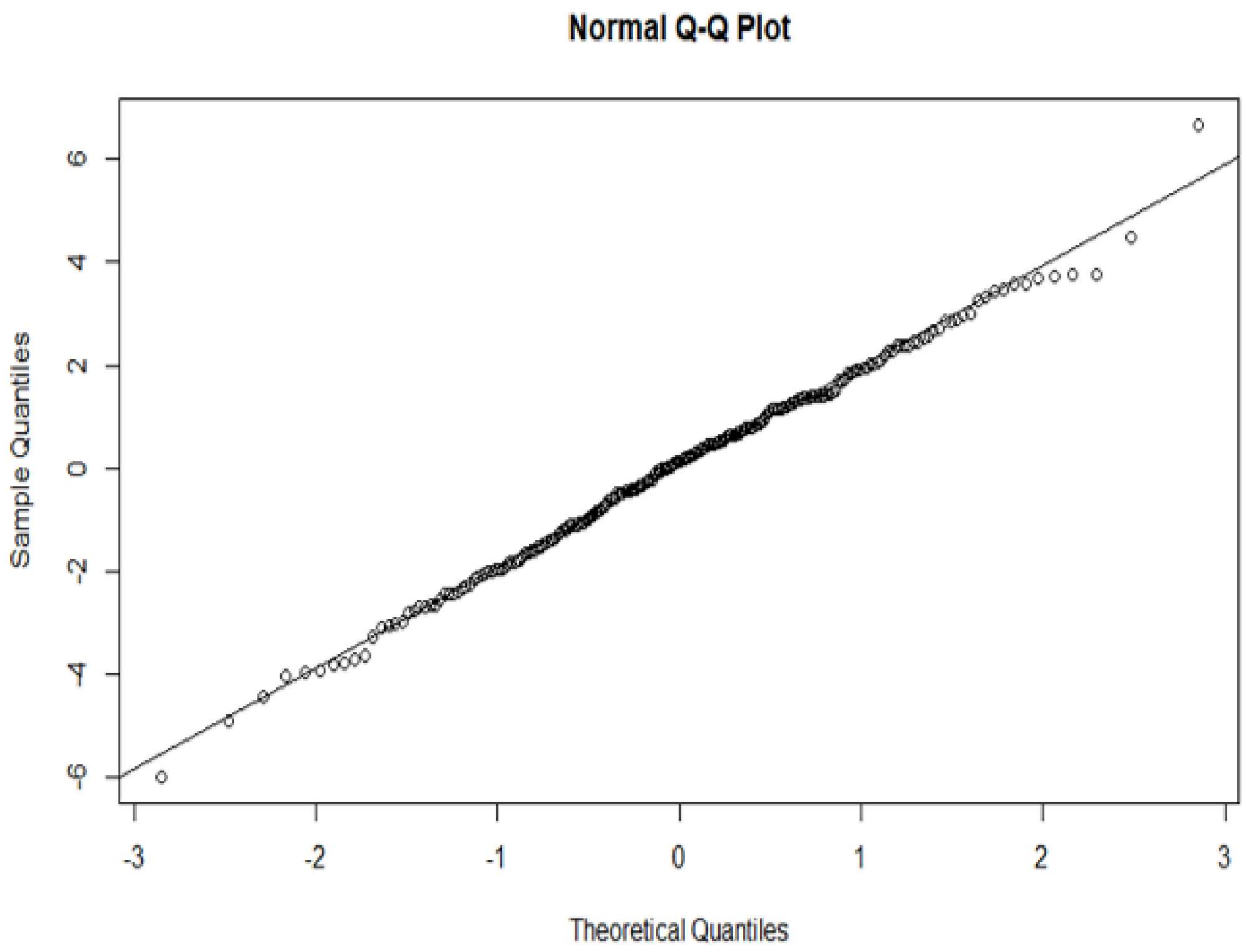
Sob H_0, JB segue distribuição qui-quadrado com 2 graus de liberdade. Como no caso dos testes de significância, é usual tomar a decisão baseada no p-valor. Assim, se o p-valor do teste for inferior a 0.05, rejeita-se a hipótese de Normalidade. Nota-se que, tanto no teste de normalidade quanto no teste de ausência de autocorrelação, o que se quer é obter p-valores elevados, maiores do que 0.05.

QQ-Plot e Teste de Normalidade para os resíduos

Para obter os resíduos do modelo, basta usar o símbolo $ após o fit2, que é o objeto que contém o modelo final definido para a série.

- qqnorm(fit2$residuals) # o QQ-plot em si
- qqline(fit2$residuals) # linha de 45º

O resultado é apresentado a seguir.


Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
-6
-4
-2
0
2
4
6
-3
-2
-1
0
1
2
3

QQ-Plot para os resíduos do modelo definido para a série. Fonte: Autor

Os comandos para a aplicação do Teste de Jarque-Bera são apresentados a seguir:

- library(tseries) # contém a função do teste
- jarque.bera.test(fit$residuals)

O resultado é apresentado a seguir.

Resultados:
X-squared = 0.43329, df = 2, p-value = 0.8052

Atenção! Para visualização completa da equação utilize a rolagem horizontal

O p-valor é maior do que 0.05, e assim não se rejeita a hipótese de que os erros do modelo sigam distribuição Normal a este nível de significância.

Atividade

A investigação do número de alunos que ingressam mensalmente no ensino superior segue um modelo AR(1). A FAC amostral dos dados analisados apresenta os seguintes valores para os cinco primeiros *lags*:

k	1	2	3	4	5
ρ_k	0.50	0.25	0.125	0.0625	0.03125

Atenção! Para visualização completa da tabela utilize a rolagem horizontal

A estimativa do parâmetro ϕ pelo método dos momentos é:

a) 0.05
b) 0.25
c) 0.4
d) 0.5
e) 0.75

Um modelo MA(1) foi identificado para uma série temporal. Os testes de sobrefixação iniciais são conduzidos com base na estimação dos modelos:

a) AR(1) e MA(1)

b) AR(1) e MA(2)

c) MA(2) e ARMA(1,1)

d) MA(2) e ARMA(1,2)

e) AR(1) e ARMA(1,1)

Uma das etapas mais importantes da análise de uma série temporal é a análise de resíduos, em particular o teste de Ljung-Box, cuja estatística é:

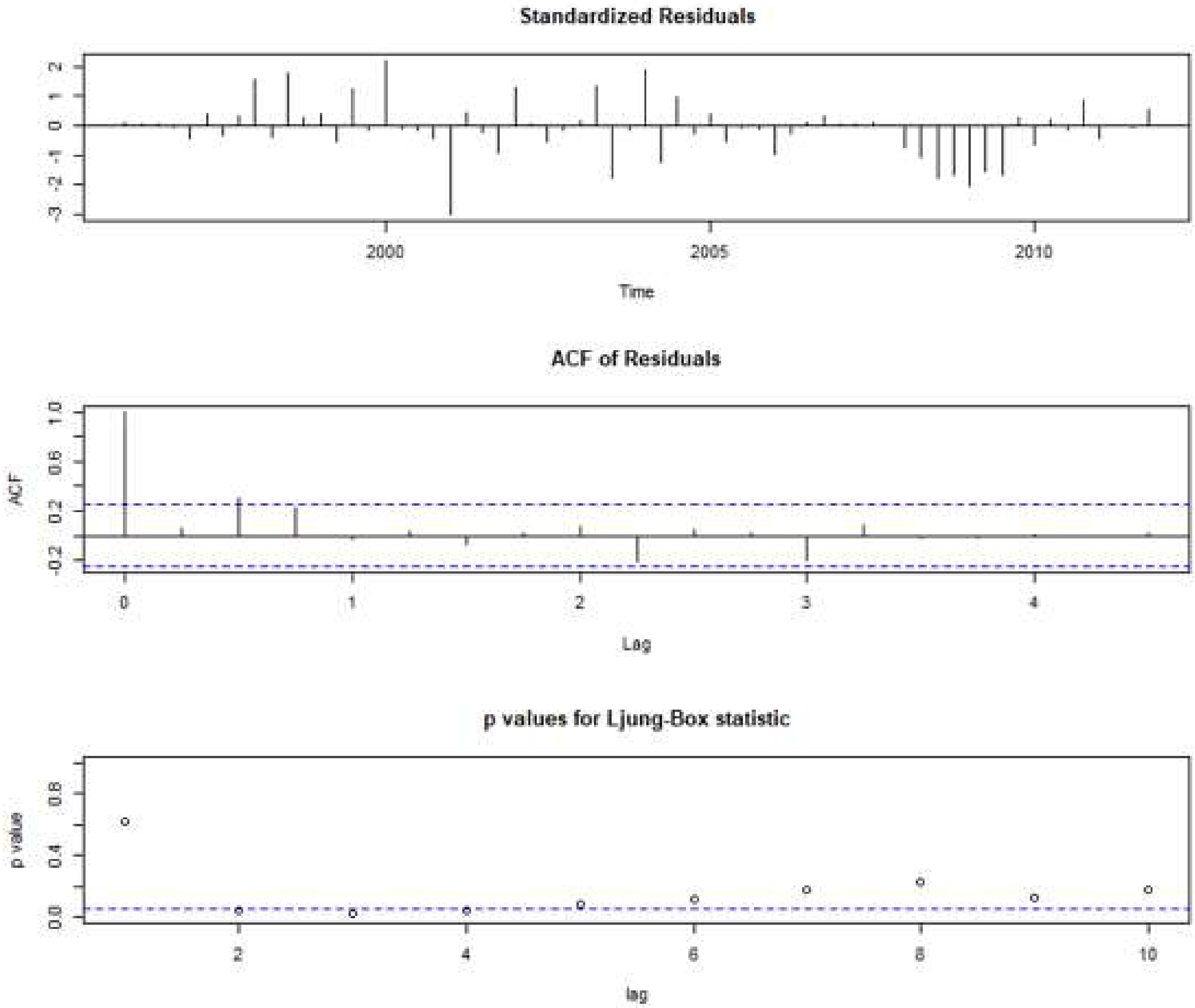
$$Q = nn + 2\sum_{j=1}^{K} \frac{\rho_j^2}{(n - j)}$$

Atenção! Para visualização completa da equação utilize a rolagem horizontal

em que K é um número arbitrário

.

Ao analisar os resíduos de um modelo de série temporal, deparamo-nos com os seguintes gráficos:



Assinale a alternativa correta.

a) Como praticamente todos os valores de autocorrelação estão dentro do intervalo de confiança, o modelo é válido, independentemente do número K de termos na estatística Ljung-Box.

b) Como a maior parte dos p-valores está acima da linha de 0.05, então o modelo é válido, independentemente do número K de termos computados na estatística Ljung-Box.

c) Como a maior parte dos p-valores está acima da linha de 0.05, então o modelo não é válido, independentemente do número K de termos computados na estatística de Ljung-Box.

d) O modelo não é válido, caso sejam considerados no máximo K = 4 termos na estatística de Ljung-Box.

e) O modelo somente é válido se considerarmos no máximo K = 4 termos na estatística de Ljung-Box.

Notas

Título modal [1]

Lorem Ipsum é simplesmente uma simulação de texto da indústria tipográfica e de impressos. Lorem Ipsum é simplesmente uma simulação de texto da indústria tipográfica e de impressos. Lorem Ipsum é simplesmente uma simulação de texto da indústria tipográfica e de impressos.

Título modal [1]

Lorem Ipsum é simplesmente uma simulação de texto da indústria tipográfica e de impressos. Lorem Ipsum é simplesmente uma simulação de texto da indústria tipográfica e de impressos. Lorem Ipsum é simplesmente uma simulação de texto da indústria tipográfica e de impressos.

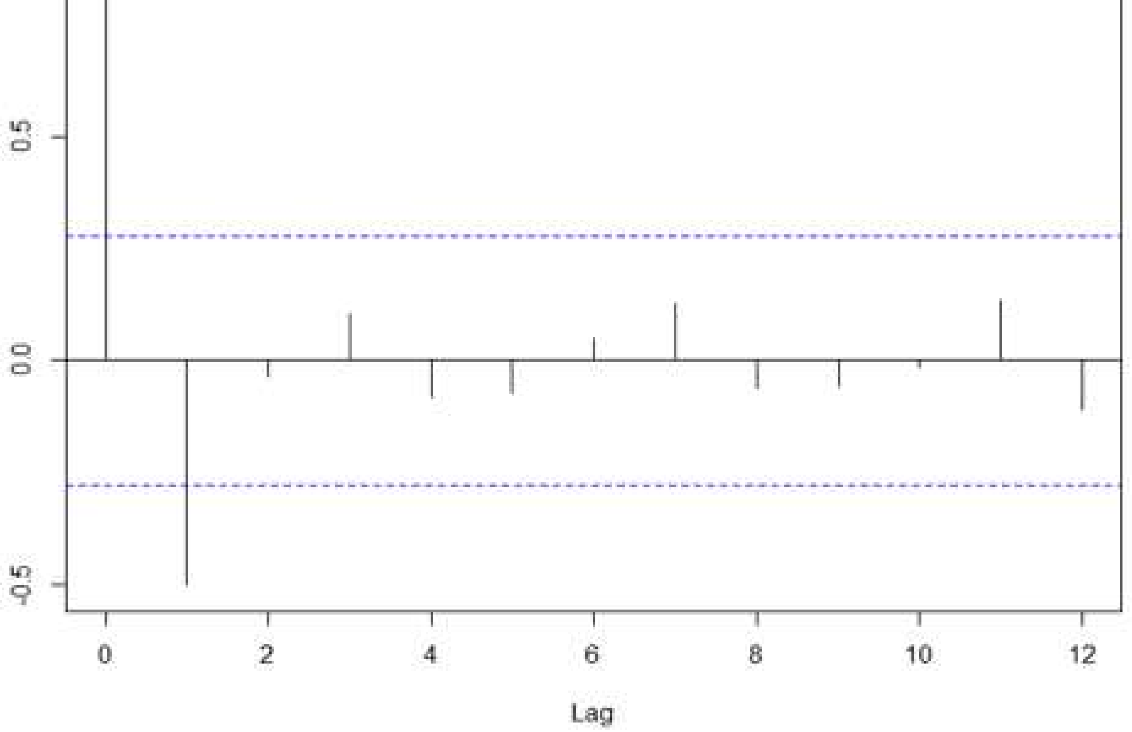
Referências

BOX, G.E.P.; JENKINS, G.M. **Time series analysis**: Forecasting and control. San Francisco: Holden-Day, Inc., 1976.

MORETTIN, P.A.; TOLOI, C.M.C. **Análise de Séries Temporais**. São Paulo: Edgard Blucher, 2004.

Próxima aula
Explore mais

- Série temporal a partir de um modelo estimado.
- Gere modelos de séries temporais artificialmente (usando, por exemplo, o Excel), e execute todo o processo de análise de uma série temporal de acordo com a sequência da figura a seguir.



- Compare o modelo final com o modelo teórico que você gerou. O que se pode concluir?