

Modelos ARIMA (Auto regressivos integrados de médias móveis: Parte 1

Prof. Caio Azevedo

Introdução

- Comumente em séries temporais observamos algum tipo de tendência estocástica.
- Por exemplo, considere um passeio aleatório dado por $Y_t = T_{t-1} + \epsilon_t$, em que $\epsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2)$. A seguir, encontra-se uma figura de uma série simulada segundo esse passeio aleatório.

Introdução

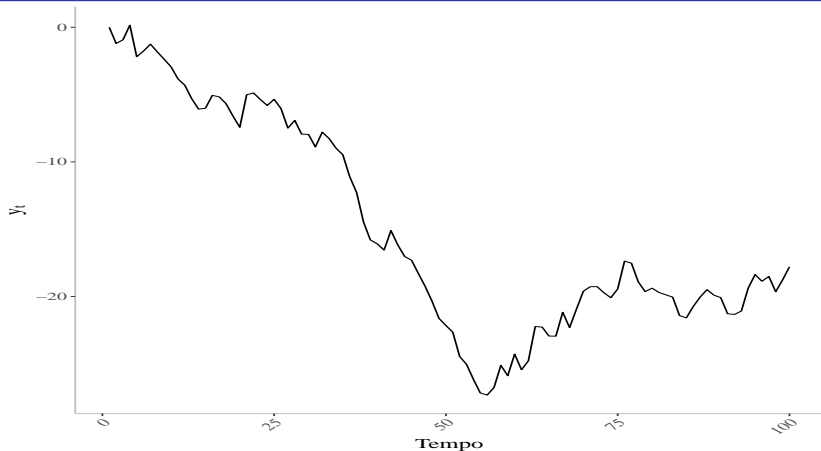


Figura: Série simulada para um passeio aleatório de tamanho 100 e $\sigma^2 = 1$.

Introdução

- Nota-se uma provável não estacionariedade, em que a média parece variar com o tempo. Ou seja, temos indícios de uma tendência nessa série.
- Na aula de [transformações de séries](#) discutimos alguns meios de retirarmos a tendência de uma série de forma a deixá-la estacionária. Aqui, nos focaremos no operador diferença de ordem d dado por:

$$\Delta_d Y_t = (1 - B)^d Y_t.$$

- Lembrando, se $d = 1$, então $\Delta_1 Y_t = Y_t - Y_{t-1}$, de $d = 2$ então $\Delta_2 Y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$, assim por diante.

Introdução

- Note que $(1 - B)^d$ está relacionado ao **operador defasagem**, ou seja:

$$(1 - B)Y_t = Y_t - BY_t = Y_t - Y_{t-1} = \Delta_1 Y_t$$

$$(1 - B)^2 Y_t = Y_t - 2BY_t + B^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2} = \Delta_2 Y_t$$

\vdots

$$(1 - B)^d Y_t = \left\{ \sum_{k=0}^d \binom{d}{k} 1^k (-B)^{d-k} \right\} Y_t = \left\{ \sum_{k=0}^d \binom{d}{k} (-B)^{d-k} \right\} Y_t$$

Introdução

- A ideia desse operador é que agora ao invés de estudarmos o comportamento da série em si, estudamos as distâncias entre as observações dessa séries, com a expectativa de que elas sejam estacionárias.
- Como exemplo, considere novamente a série simulada do passeio aleatório, agora observe essa série através de um gráfico de escadas.

Introdução

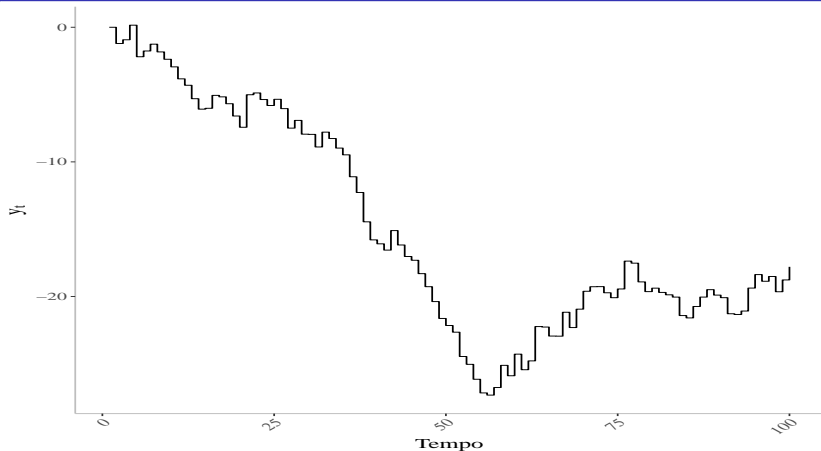


Figura: Gráfico de escada da série simulada para um passeio aleatório de tamanho 100 e $\sigma^2 = 1$.

Introdução

- Se aplicássemos a primeira diferença nessa série estaríamos analisando os saltos vistos no gráfico de escada, ou seja, a distância de uma observação para outra subsequente.
- Dessa forma, teríamos uma série sem tendência, com comportamento estacionário, assim como visto na figura a seguir.

Introdução

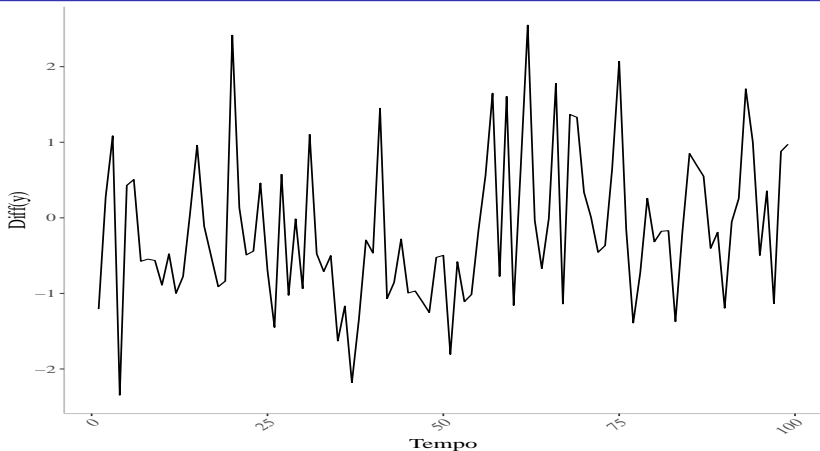


Figura: Primeira diferença da série simulada para um passeio aleatório de tamanho 100 e $\sigma^2 = 1$.

ARIMA

- Para séries com tendência estocástica, podemos fazer uso dos modelos ARMA integrados, também chamados de ARIMA (Auto-Regressivos Integrados de Médias Móveis), que consideram o operador diferença em um modelo ARMA.
- Formalmente temos que uma série (processo) $\{Y_t\}$ é ARIMA(p, d, q) se $(1 - B)^d Y_t$ é ARMA(p, q), i.e., se satisfaz a equação:

$$\phi(B)(1 - B)^d Y_t = \theta(B)\epsilon_t, \quad \epsilon_t \sim RB(0, \sigma^2),$$

em que $\phi(B)$ e $\theta(B)$ são os polinômios autoregressivos e de médias móveis, respectivamente, e $d \in \mathcal{N}^+$.

ARIMA - Propriedades

- $\{Y_t\}$ é estacionário apenas para o caso em que $d = 0$.
- Exceto quando $d = 0$ a média desse processo não pode ser determinada pela equação do processo.
- A seguir temos algumas séries simuladas de modelos ARIMA com comentários pertinentes.

ARIMA(1,1,0)

- Modelo

$$(1 - \phi B)(1 - B)Y_t = \epsilon_t$$

$$(1 - B - \phi B + \phi B^2)Y_t = \epsilon_t$$

$$Y_t - Y_{t-1} = \phi Y_{t-1} - \phi Y_{t-2} + \epsilon_t$$

$$Y_t = Y_{t-1} + \phi Y_{t-1} - \phi Y_{t-2} + \epsilon_t$$

$$\Delta_1 Y_t = \phi \Delta_1 Y_{t-1} + \epsilon_t$$

$$\epsilon_t \sim NID(0, \sigma^2)$$

- Escrever as equações dos outros modelos simulados.

ARIMA - Propriedades

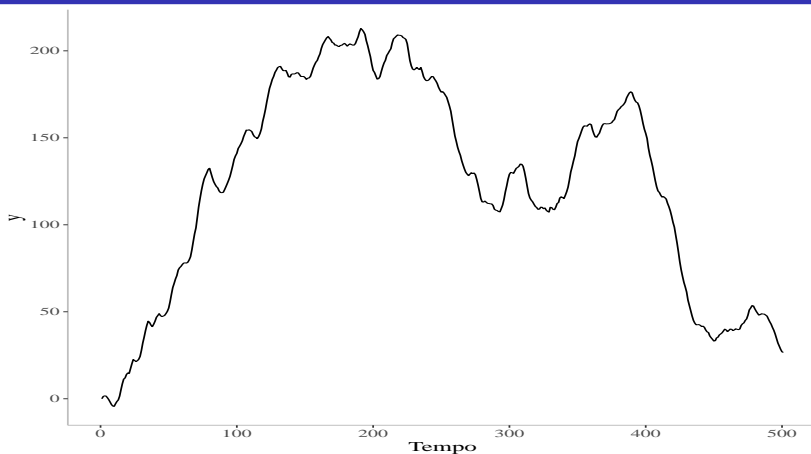


Figura: Série simulada para um modelo ARIMA(1, 1, 0) com $\phi = 0,9$.

ARIMA - Propriedades

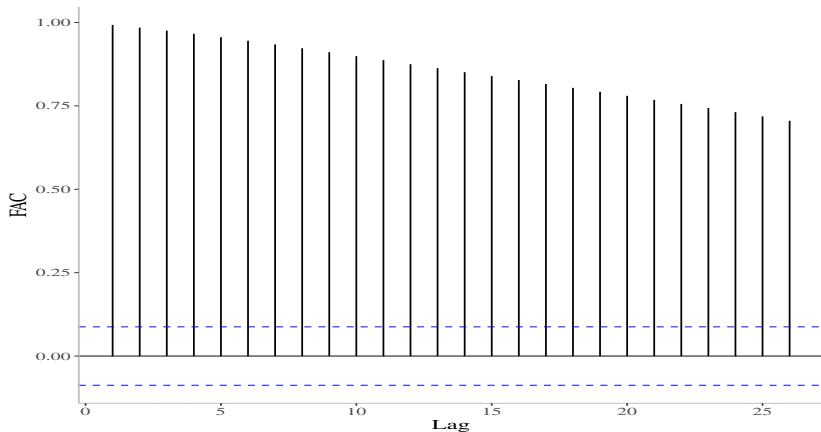


Figura: FAC da série simulada para um modelo ARIMA(1,1,0) com $\phi = 0,9$.

ARIMA - Propriedades

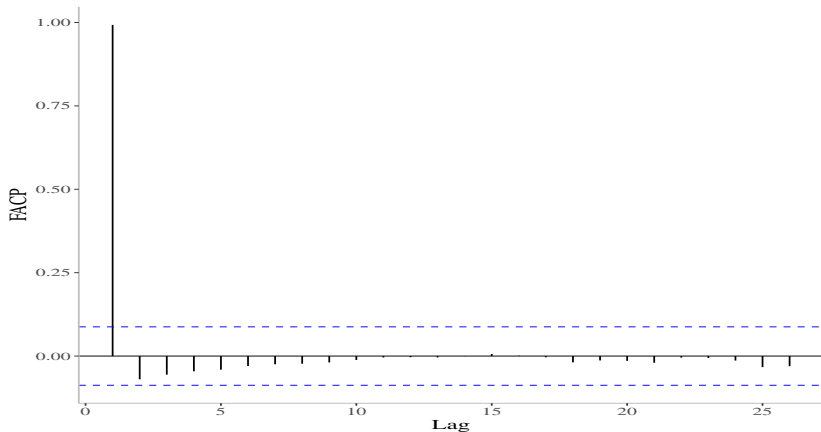


Figura: FACP da série simulada para um modelo ARIMA(1, 1, 0) com $\phi = 0,9$.

ARIMA - Propriedades

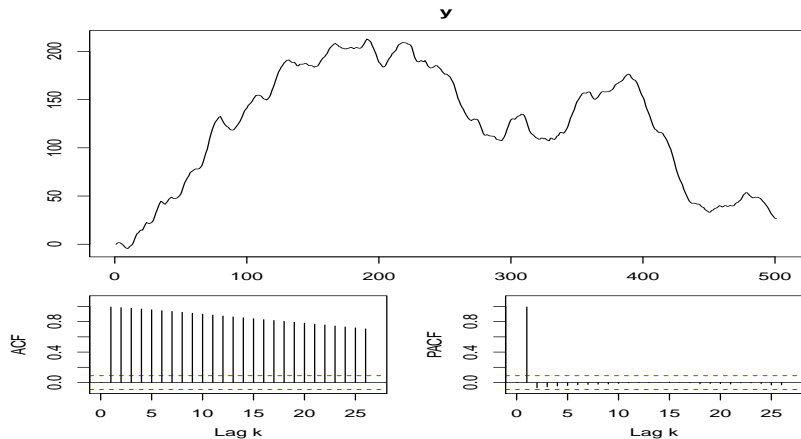


Figura: Gráficos da série simulada para um modelo ARIMA(1,1,0) com $\phi = 0,9$.

ARIMA - Propriedades

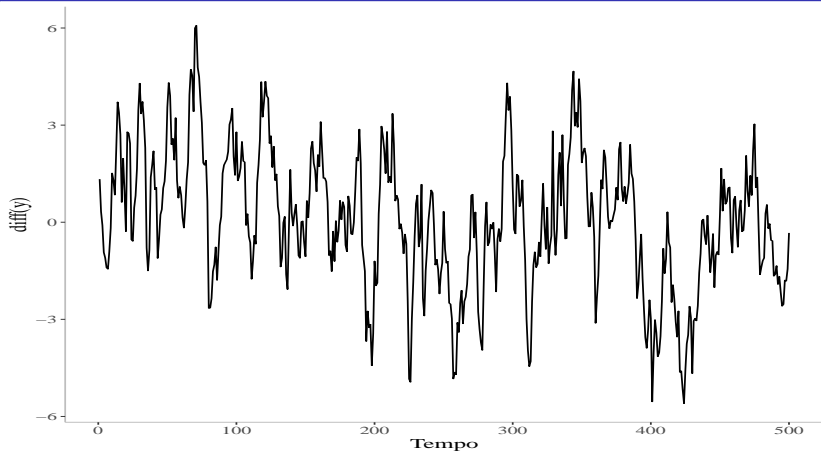


Figura: Primeira diferença da série simulada para um modelo ARIMA(1,1,0) com $\phi = 0,9$.

ARIMA - Propriedades

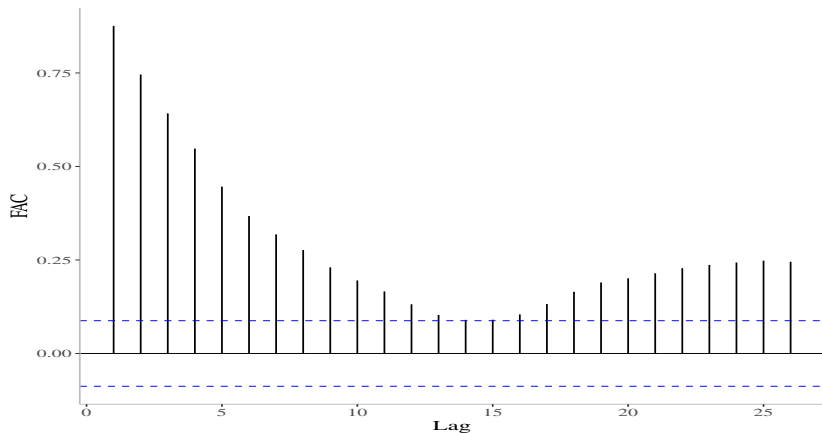


Figura: FAC da primeira diferença da série simulada para um modelo ARIMA(1, 1, 0) com $\phi = 0,9$.

ARIMA - Propriedades

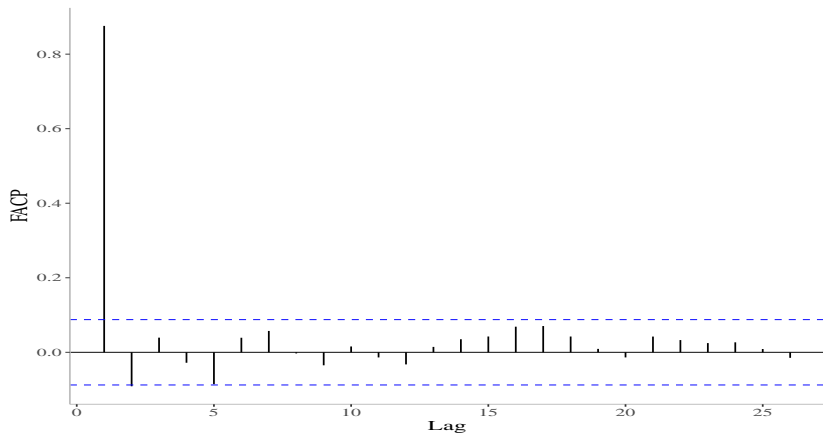


Figura: FACP diferença da série simulada para um modelo ARIMA(1,1,0) com $\phi = 0,9$.

ARIMA - Propriedades

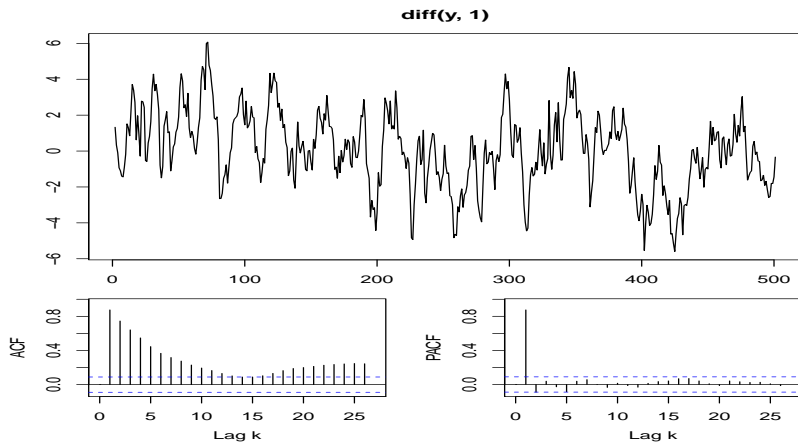


Figura: Gráficos da primeira diferença da série simulada para um modelo ARIMA(1, 1, 0) com $\phi = 0,9$.

ARIMA - Propriedades

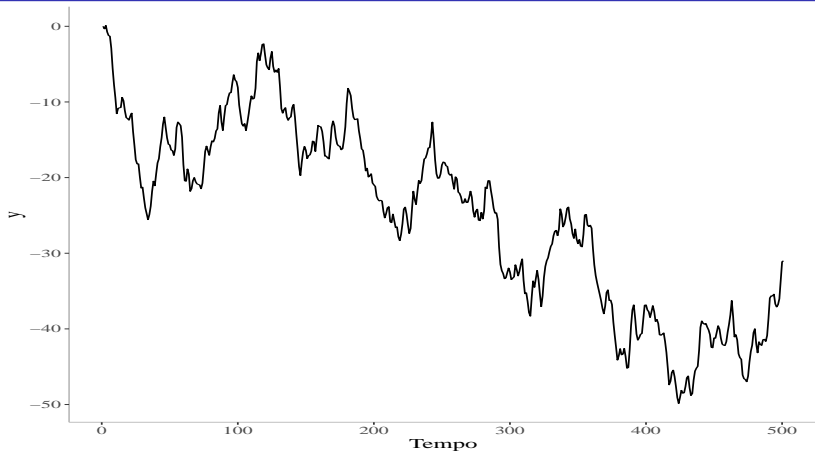


Figura: Série simulada para um modelo ARIMA(0, 1, 1) com $\theta = 0,8$.

ARIMA - Propriedades

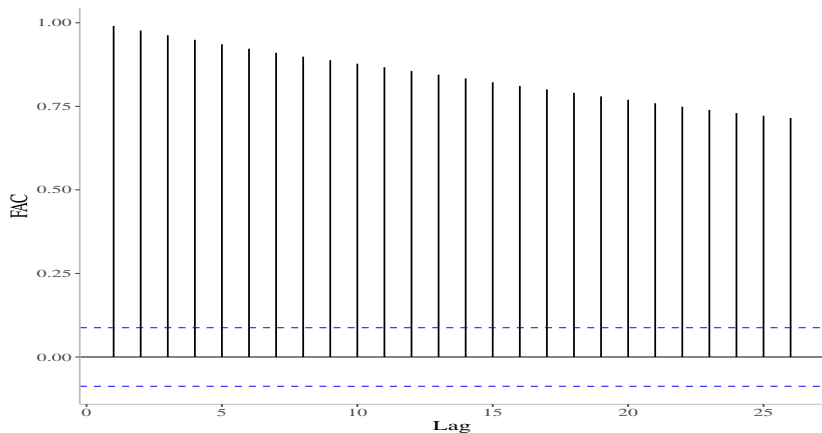


Figura: FAC da série simulada para um modelo ARIMA(0,1,1) com $\theta = 0,8$.

ARIMA - Propriedades

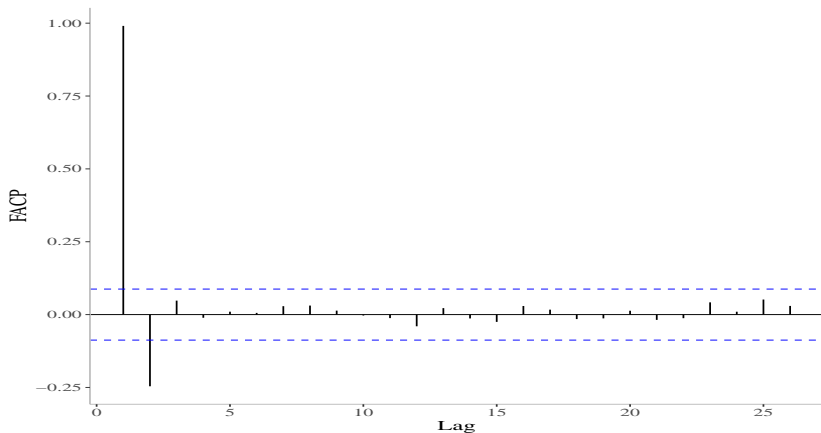


Figura: FACP da série simulada para um modelo ARIMA(0, 1, 1) com $\theta = 0,8$.

ARIMA - Propriedades

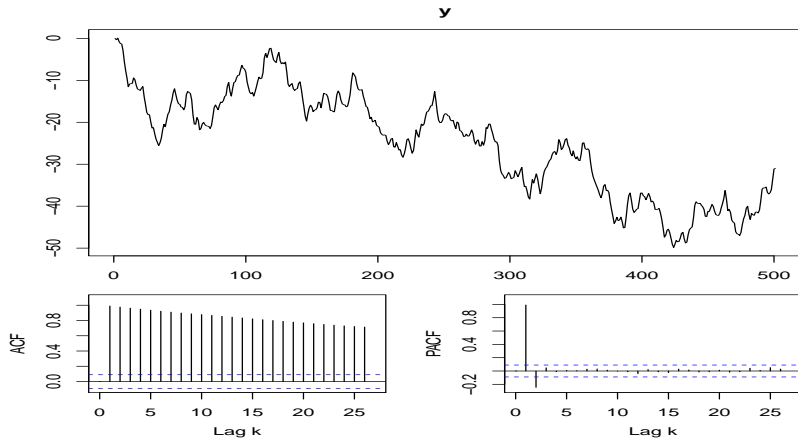


Figura: Gráficos da série simulada para um modelo ARIMA(0, 1, 1) com $\theta = 0,8$.

ARIMA - Propriedades

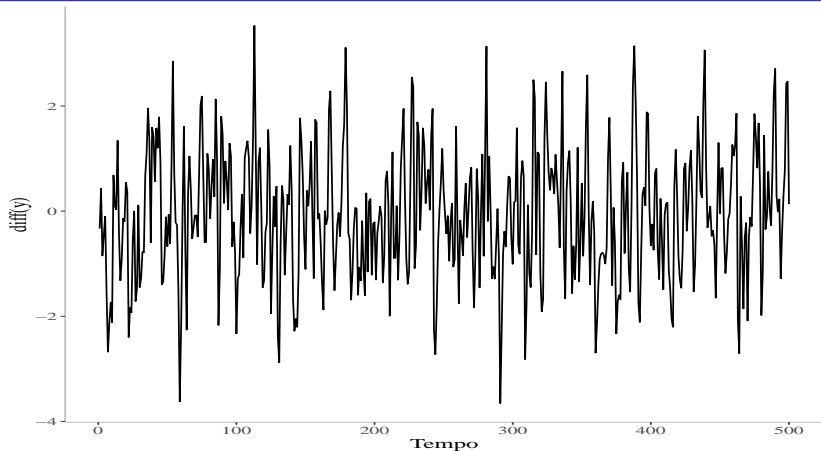


Figura: Primeira diferença da série simulada para um modelo ARIMA(0,1,1) com $\theta = 0,8$.

ARIMA - Propriedades

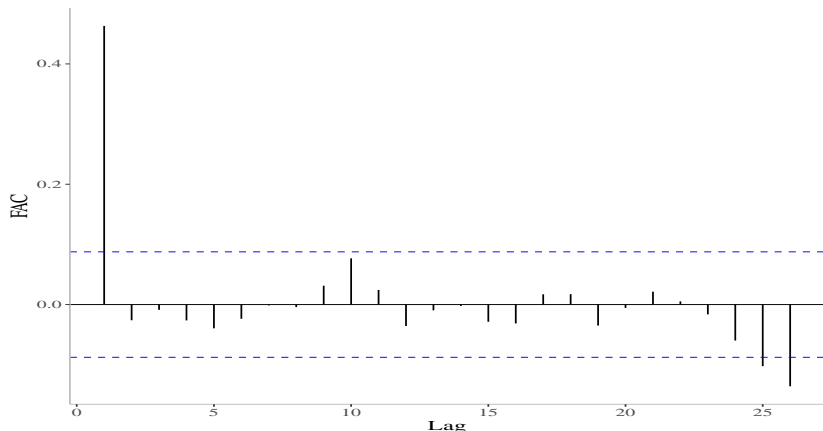


Figura: FAC da primeira diferença da série simulada para um modelo ARIMA(0, 1, 1) com $\theta = 0,8$.

ARIMA - Propriedades

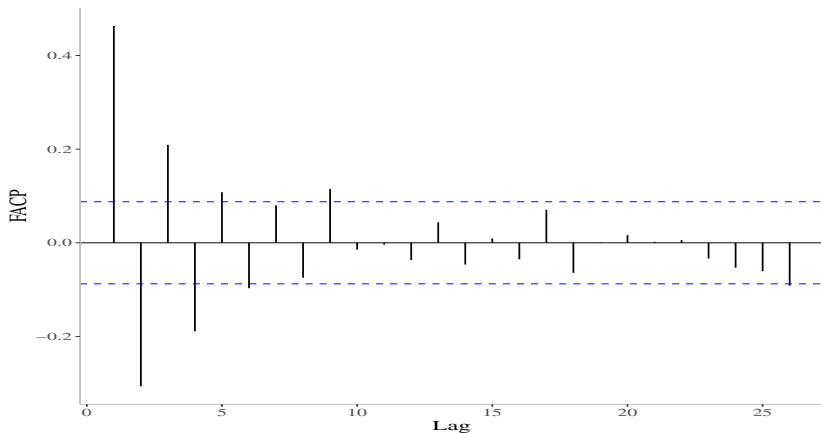


Figura: FACP diferença da série simulada para um modelo ARIMA(0,1,1) com $\theta = 0,8$.

ARIMA - Propriedades

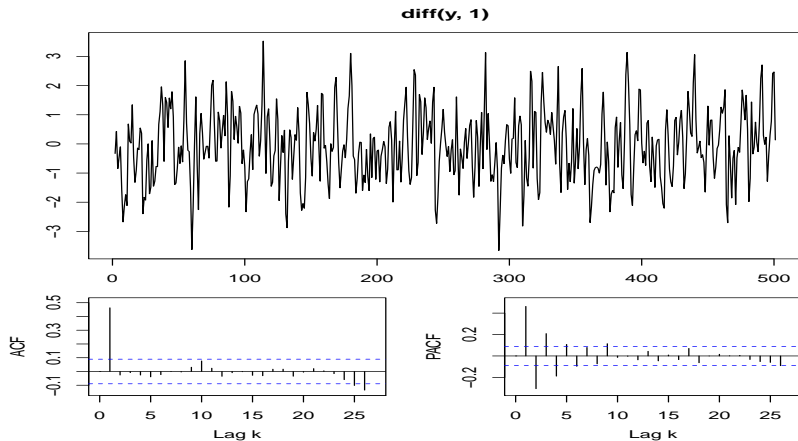


Figura: Gráficos da primeira diferença da série simulada para um modelo ARIMA(0, 1, 1) com $\theta = 0,8$.

ARIMA - Propriedades

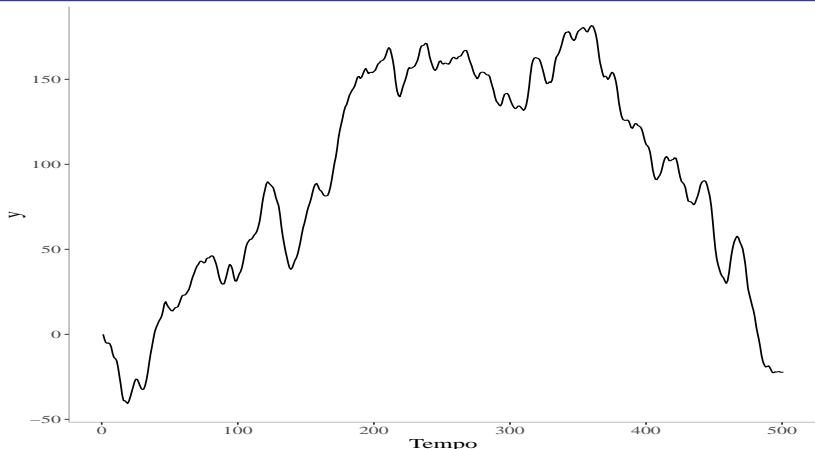


Figura: Série simulada para um modelo ARIMA(1, 1, 1) com $\phi = 0,8$ e $\theta = 0,8$.

ARIMA - Propriedades

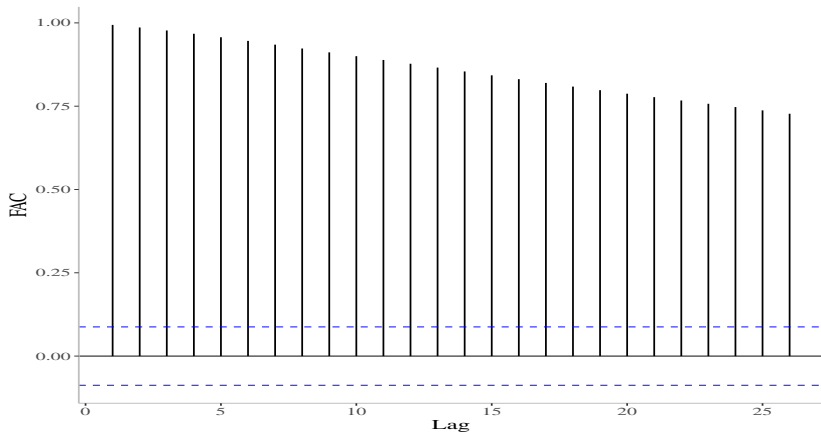


Figura: FAC da série simulada para um modelo ARIMA(1, 1, 1) com $\phi = 0,8$ e $\theta = 0,8$.

ARIMA - Propriedades

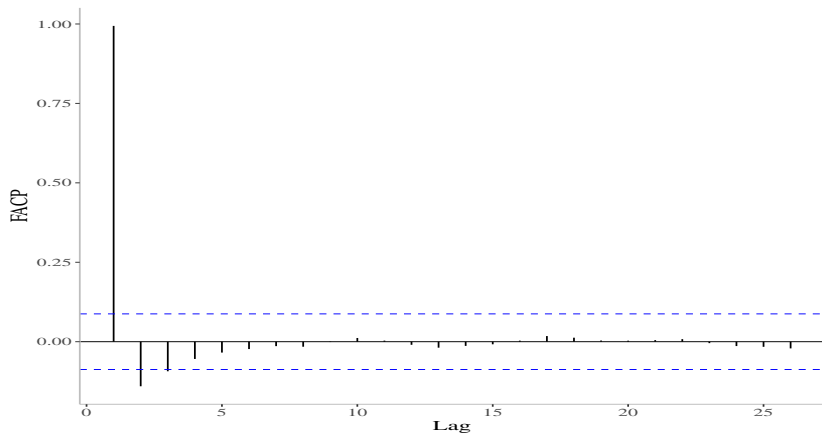


Figura: FACF da série simulada para um modelo ARIMA(1,1,1) com $\phi = 0,8$ e $\theta = 0,8$.

ARIMA - Propriedades

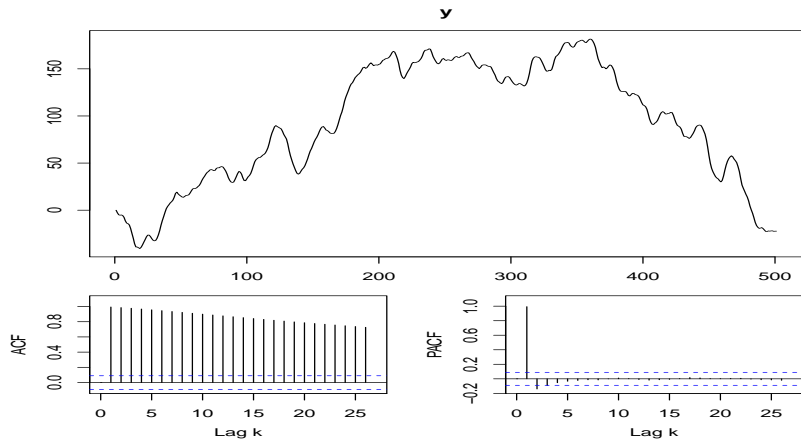


Figura: Gráficos da série simulada para um modelo ARIMA(1,1,1) com $\phi = 0,8$ e $\theta = 0,8$.

ARIMA - Propriedades

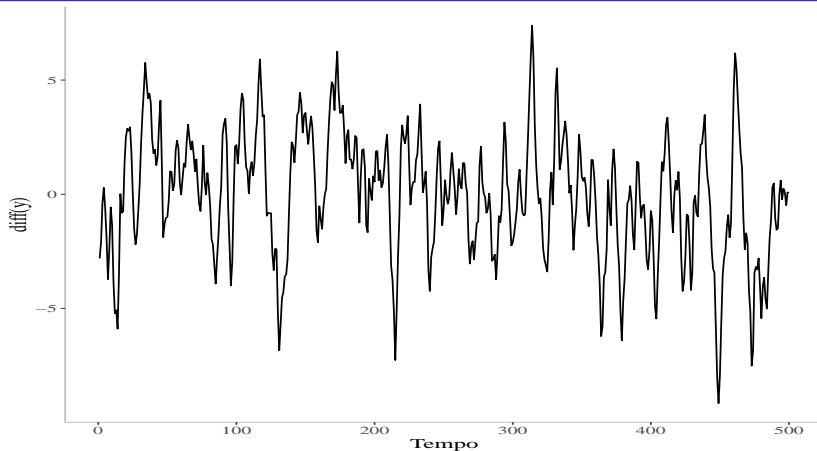


Figura: Primeira diferença da série simulada para um modelo ARIMA(1,1,1) com $\phi = 0,8$ e $\theta = 0,8$.

ARIMA - Propriedades

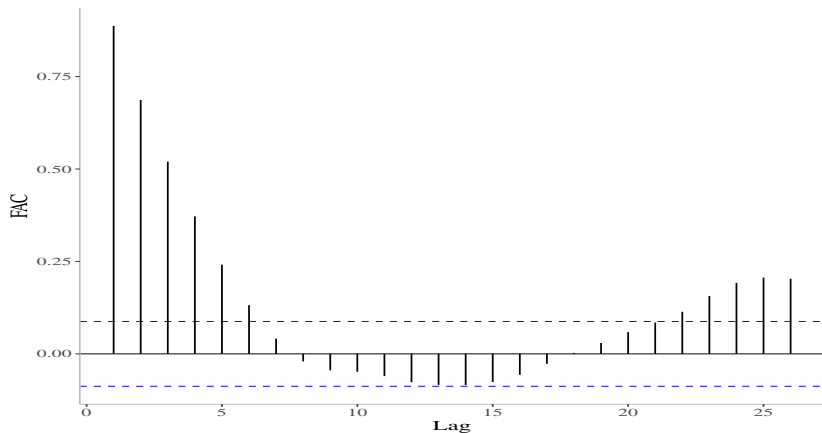


Figura: FAC da primeira diferença da série simulada para um modelo ARIMA(1, 1, 1) com $\phi = 0,8$ e $\theta = 0,8$.

ARIMA - Propriedades

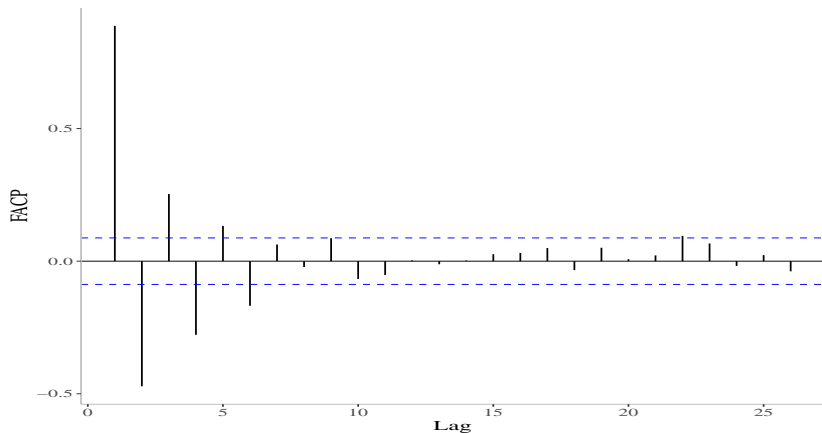


Figura: FACP diferença da série simulada para um modelo ARIMA(1,1,1) com $\phi = 0,8$ e $\theta = 0,8$.

ARIMA - Propriedades

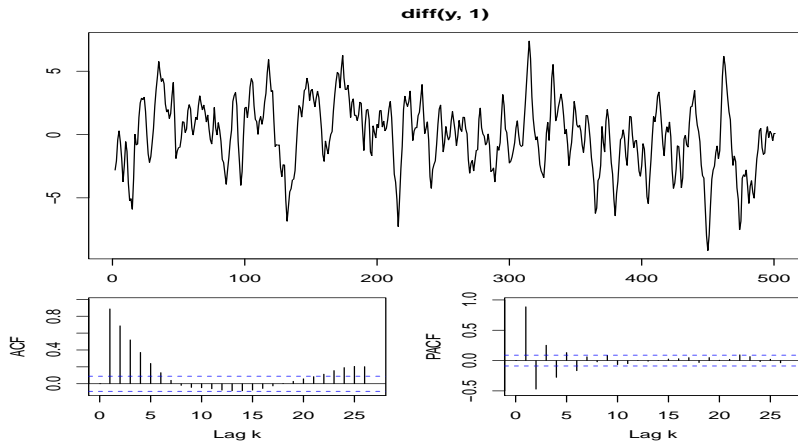


Figura: Gráficos da primeira diferença da série simulada para um modelo ARIMA(1, 1, 1) com $\phi = 0,8$ e $\theta = 0,8$.

ARIMA - Propriedades

- Note que a FAC e FACP pode não estar bem definidas para **processos não estacionários**.
- Contudo, a FAC e FACP amostrais são importantes ferramentas de identificação da necessidade de uso de um modelo ARIMA apropriado.

ARIMA - Propriedades

- Dito isso, em geral, observa-se que modelos ARIMA são apropriados quando:
 - 1 A série apresentar tendência e sua diferenciação parecer estacionária (isso pode indicar que processo estacionário pode ser apropriado para uma dado valor de d).
 - 2 Na FAC, as autocorrelações amostrais apresentarem valores grandes e decaimento lento.
 - 3 Na FACP, a primeira autocorrelação parcial amostral for bem próxima de 1 e as restantes próximas de 0.

ARIMA - Propriedades

- Muitas das propriedades para modelos ARIMA podem ser derivadas através da série diferenciada $X_t = (1 - B)^d Y_t$.
- Em termos de análise exploratória podemos tentar identificar as ordens p e q através da FAC e FACP da série X_t .
- Para previsões podemos considerar a previsão m passos a frente de X_t , dada por $X_n(m)$, obter as previsões via métodos já vistos para **modelos ARMA** e assim recuperar $Y_n(m)$ através de $X_n(m)$. Por exemplo, considerando $d = 1$, temos que $Y_n(m) = X_n(m) - X_n(m - 1)$.

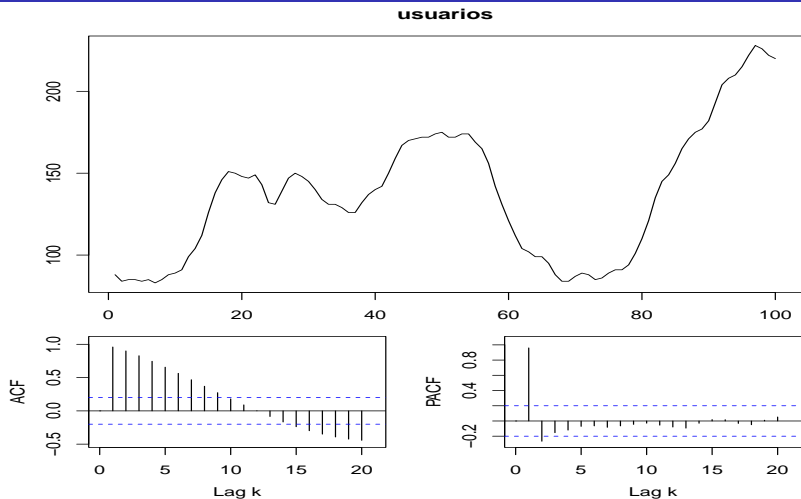
ARIMA - Propriedades

- A **estimação** pode ser feita usando os mesmos métodos para o modelo ARMA, ou seja, obtendo estimativas iniciais via método de mínimos quadrados condicionais para se obter, de modo iterativo, estimativas de máxima verossimilhança.
- A **análise residual** também é feita considerando-se a série Y_t e, assim, podemos usar todo o ferramental visto como análise gráfica, FAC, FACP, teste de Ljung-Box, etc.

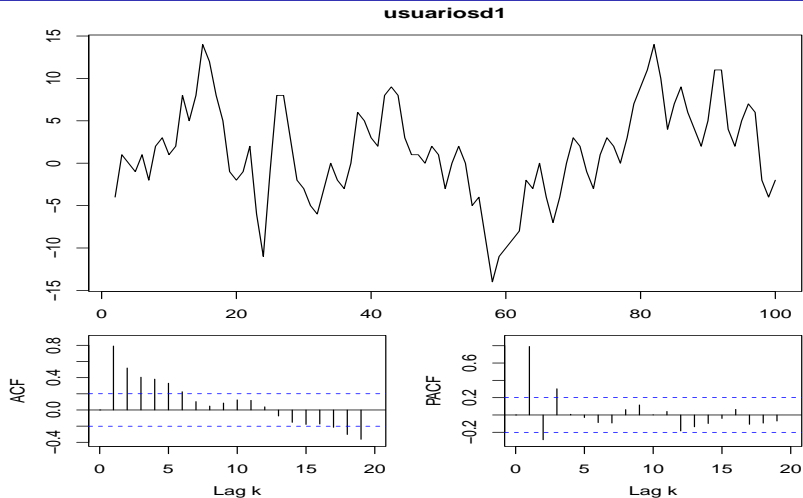
ARIMA - Exemplo

- Considere a série `WWWusage` do pacote `datasets` que trata do número de usuários conectados à Internet por meio de um servidor a cada minuto.
- A seguir, apresenta-se gráficos de análise exploratória para esses dados. Nota-se que a série possui tendência estocástica e que os gráficos indicam que o uso de um modelo $ARIMA(3, 1, 0)$ pode ser apropriado.

Gráficos de ST, FAC e FACP da ST



Gráficos de ST, FAC e FACP da ST (Δ_1)



ARIMA - Exemplo

- Modelo inicial:

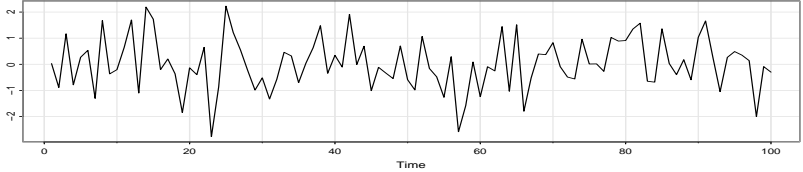
$$\Delta_1 Y_t = c + \phi_1 \Delta_1 Y_{t-1} + \phi_2 \Delta_1 Y_{t-2} + \phi_3 \Delta_1 Y_{t-3} + \epsilon_t$$
$$\epsilon_t \sim NID(0, \sigma^2).$$

- Exercício: escrever os outros modelos.
- Os resíduos que o pacote apresenta nos gráficos, aparentemente, não são os mesmos que ele solta via comando

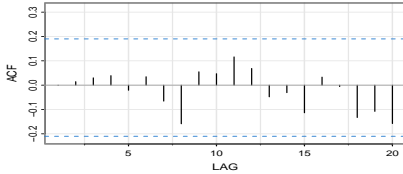
```
fitARIMA310<-sarima(usuarios,p=3,d=1,q=0)
fitARIMA310$fit$residuals
```

Model: (3,1,0)

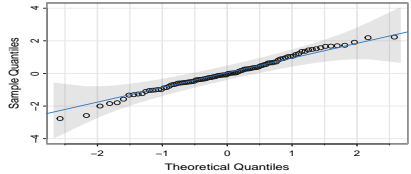
Standardized Residuals



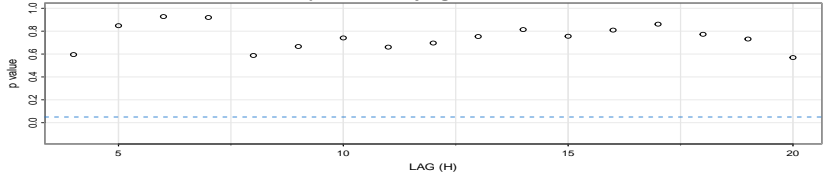
ACF of Residuals

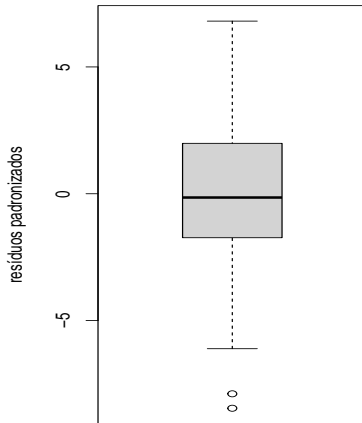
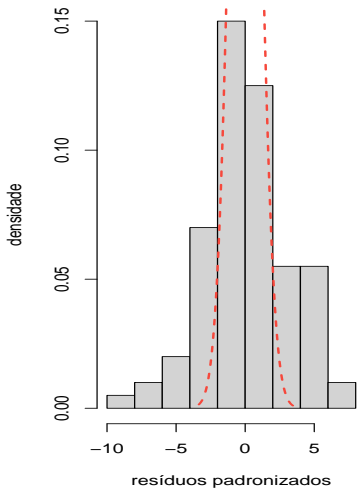


Normal Q-Q Plot of Std Residuals



p values for Ljung-Box statistic





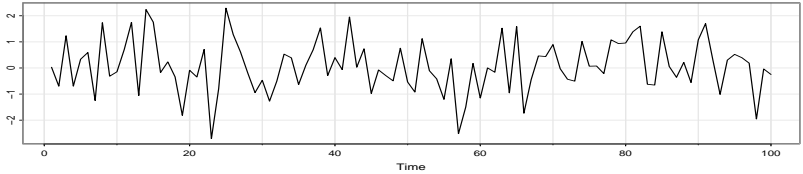
ARIMA - Exemplo

Par.	Est.	Ep	IC(95%)	Estat. t	p-valor
ϕ_1	1,15	0,10	[0,96;1,33]	12,02	<0,0001
ϕ_2	-0,66	0,14	[-0,92;-0,39]	-4,88	<0,0001
ϕ_3	0,33	0,09	[0,15;0,52]	3,53	<0,0001
c	0,98	1,65	[-2,25;4,21]	0,59	0,5540

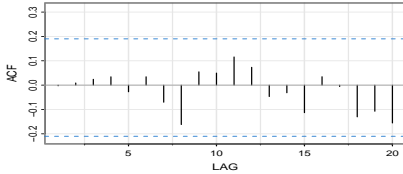
A constante (c) não é significativa. Reajustaremos este modelo sem esse parâmetro.

Model: (3,1,0)

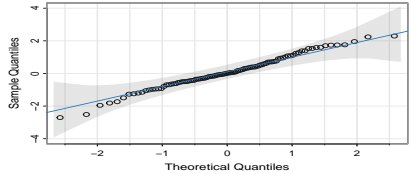
Standardized Residuals



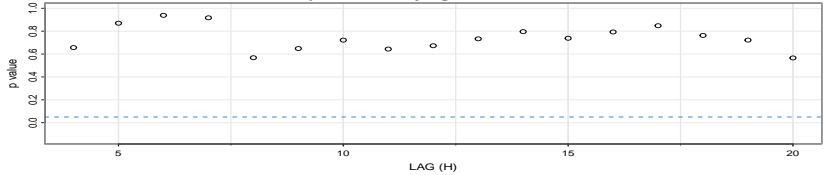
ACF of Residuals

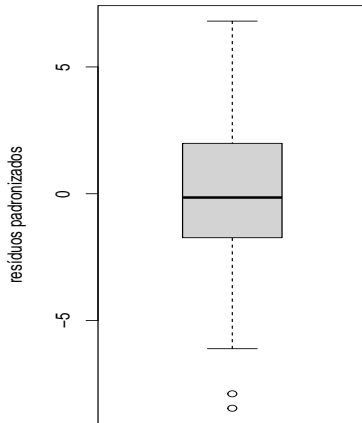
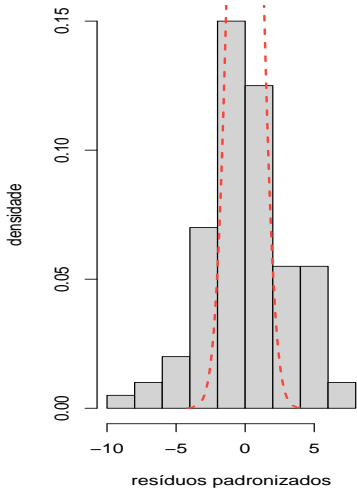


Normal Q-Q Plot of Std Residuals



p values for Ljung-Box statistic

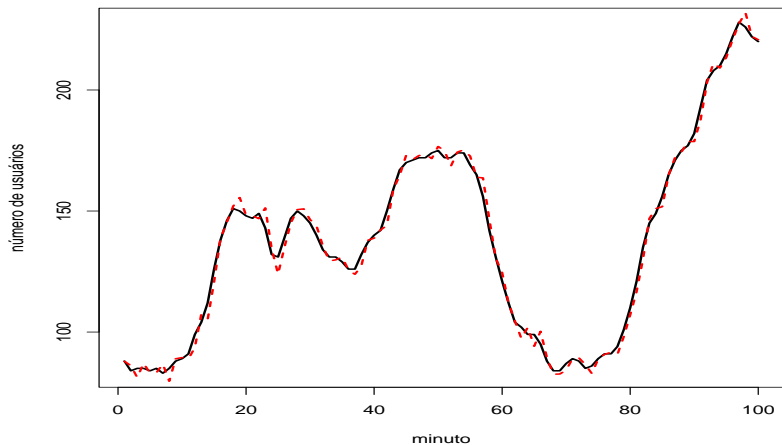




ARIMA - Exemplo

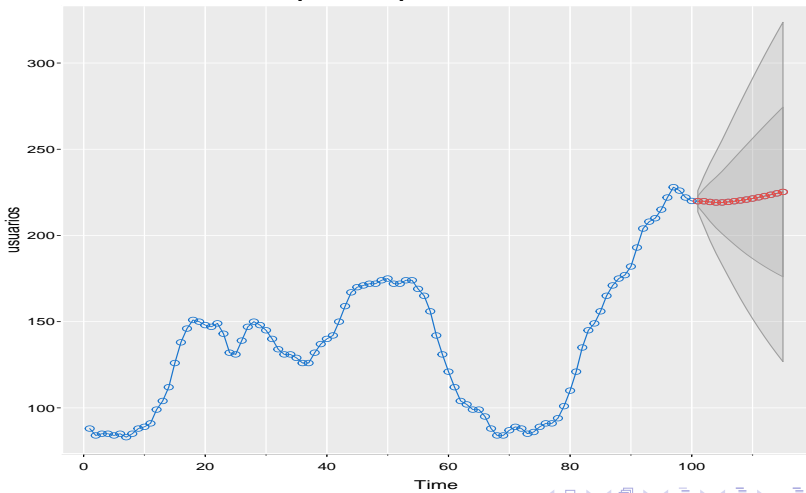
Par.	Est.	Ep	IC(95%)	Estat. t	p-valor
ϕ_1	1,15	0,10	[0,97 ; 1,34]	12,12	<0,0001
ϕ_2	-0,66	0,14	[-0,93 ; -0,40]	-4,89	<0,0001
ϕ_3	0,34	0,09	[0,16 ; 0,53]	3,62	<0,0001

Previsão: preto - observado; vermelho - predito

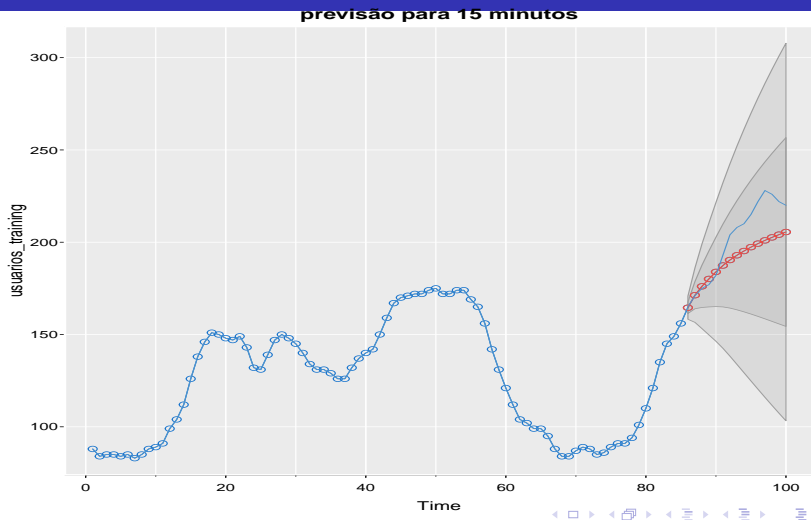


Previsão: dados futuros

previsão para 15 minutos



Previsão: dados futuros observados

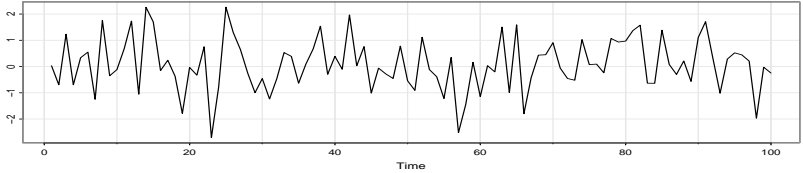


ARIMA - Exemplo

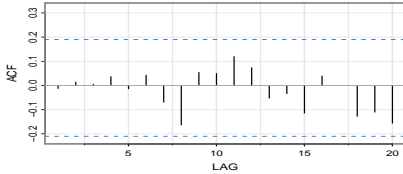
- Em termos de análise residual, o modelo se ajustou muito bem, embora o gráfico de Q-Q indique que o ajuste possa (deveria) ser melhorado.
- A previsão para os dados observados foi muito boa e para os dados futuros, satisfatória. Entretanto, esta última previsão deve ser melhorada.
- O modelo é estacionário e causal pois $\phi(z) = 1 - 1,1513z + 0,6612z^2 - 0,3407z^3 = 0$ tem uma solução real $z \approx 1,1823$ e duas soluções imaginárias $z \approx 0,3792 - 1,5293i$ e $z \approx 0,3792 + 1,5293i$ que estão fora do círculo unitário pois $|1,1823| > 1$ e $|0,3792 - 1,5293i| = |0,3792 + 1,5293i| = \sqrt{0,3792^2 + 1,5293^2} \approx 1,5756 > 1$.

Model: (4,1,0)

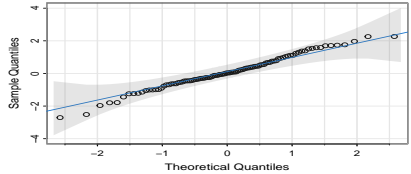
Standardized Residuals



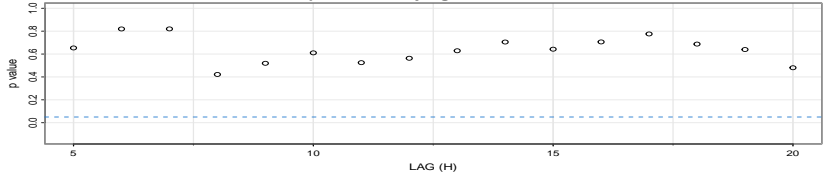
ACF of Residuals



Normal Q-Q Plot of Std Residuals

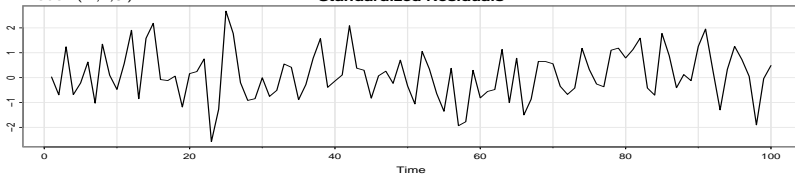


p values for Ljung-Box statistic

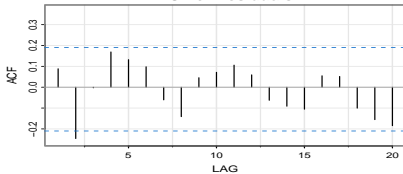


Model: (2,1,0)

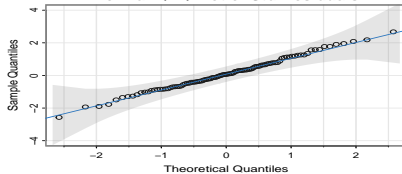
Standardized Residuals



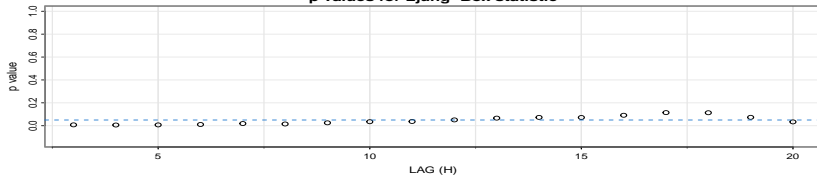
ACF of Residuals



Normal Q-Q Plot of Std Residuals

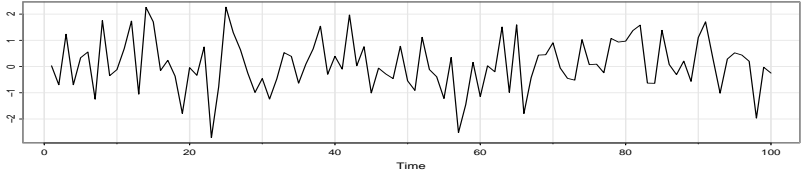


p values for Ljung-Box statistic

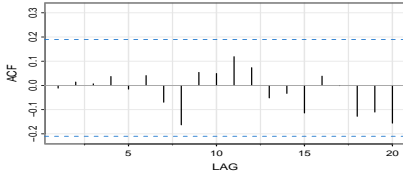


Model: (3,1,1)

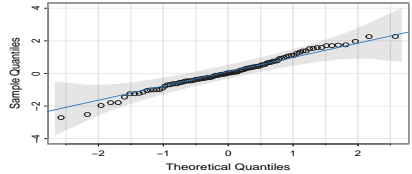
Standardized Residuals



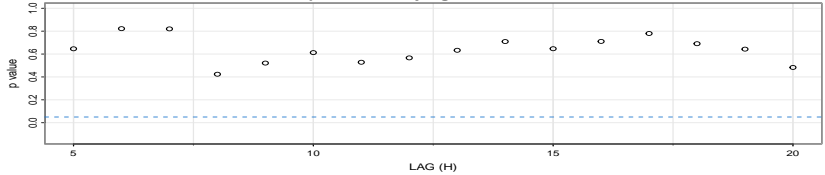
ACF of Residuals



Normal Q-Q Plot of Std Residuals



p values for Ljung-Box statistic



Critérios de Informação (modelos com $c = 0$)

Modelo	AIC	AICc	BIC
ARIMA(3,1,0)	5,172	5,174	5,277
ARIMA(4,1,0)	5,191	5,196	5,322
ARIMA(2,1,0)	5,275	5,276	5,353
ARIMA(3,1,1)	5,191	5,196	5,322

Critérios de previsão: valores observados (modelos com $c = 0$)

Modelo	EQM	EAM	EPM	EPAM
ARIMA(3,1,0)	9,270	2,367	0,275	1,891
ARIMA(4,1,0)	9,263	2,369	0,273	1,893
ARIMA(2,1,)	10,523	2,580	0,295	2,047
ARMA(3,1,1)	9,264	2,369	0,273	1,892

Critérios de previsão: valores futuros observados (modelos com $c = 0$)

Modelo	EQM	EAM	EPM	EPAM
ARIMA(3,1,0)	220,939	11,945	5,058	5,538
ARIMA(4,1,0)	380,385	15,533	7,104	7,176
ARIMA(2,1,)	893,216	25,216	11,800	11,800
ARMA(3,1,1)	311,635	13,993	6,302	6,460