

Modelos SARIMA (Auto-Regressivo Integrado de Médias Móveis com Sazonalidade): parte 3

Prof. Caio Azevedo

Diagnóstico

- Voltemos a questão dos resíduos. Em geral, os modelos vistos no curso, podem ser escritos da seguinte forma (considerando em um amostra finita, ou seja, uma ST observada):

$$Y_t = h(\mathbf{Y}_{t-1}, \epsilon_{t-1}, \phi, \Phi, \theta, \Theta) + \epsilon_t$$

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

em que $\mathbf{Y}_{t-1} = (Y_1, Y_2, \dots, Y_{t-1})'$ e $\epsilon_{t-1} = (\epsilon_1, \epsilon_2, \dots, \epsilon_{t-1})'$.

- Assim, suponha os parâmetros sejam conhecidos. Sob a validade do modelo acima (considerando $\sigma = \sqrt{\sigma^2}$), temos que:

$$\epsilon_t = \frac{Y_t - h(\mathbf{Y}_{t-1}, \epsilon_{t-1}, \phi, \Phi, \theta, \Theta)}{\sigma} \sim NID(0, 1) \quad (1)$$

Diagnóstico

- Por outro lado pode-se considerar diversos (tipos de) **resíduos** (valores preditos dos erros) para os erros (ϵ_t).
- O resíduo apresentado **aqui** é bastante apropriado, como discutido. Contudo, podemos considerar um outro resíduo, mais simples de calcular.
- Com efeito, seja $\hat{h} = h(\mathbf{Y}_{t-1}, \epsilon_{t-1}, \hat{\phi}, \hat{\Phi}, \hat{\theta}, \hat{\Theta})$ um “preditor” de $h = h(\mathbf{Y}_{t-1}, \epsilon_{t-1}, \phi, \Phi, \theta, \Theta)$, por exemplo, obtido via máxima verossimilhança (MV), ou seja, $\hat{(\cdot)}$ denota o respectivo estimador de MV.

Diagnóstico

- Como os estimadores de MV são **consistentes**, então:

$$|\hat{h} - h| \xrightarrow[n \rightarrow \infty]{P} 0 \quad (2)$$

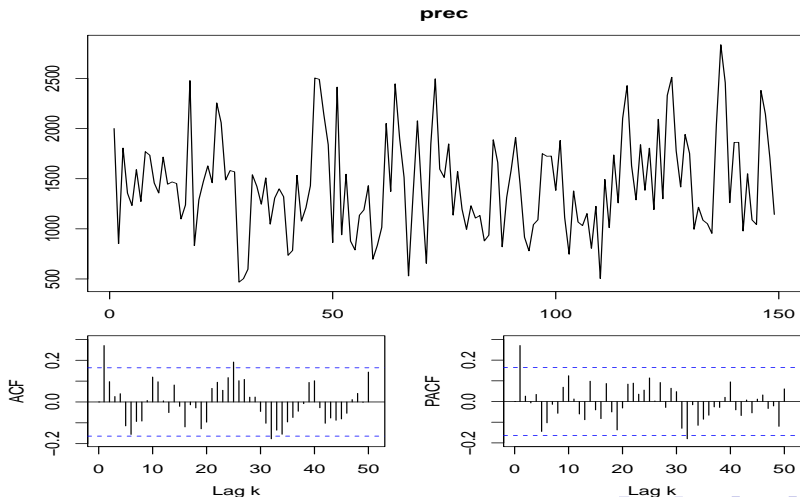
- Assim, o seguinte resíduo:

$$\hat{\epsilon}_t = \frac{Y_t - \hat{h}_t}{\hat{\sigma}}$$

tem distribuição aproximadamente normal, sob o bom ajuste do modelo e para n suficientemente grande.

- Com efeito, este é o resíduo utilizado pela função “sarima” do pacote “astsa”.
- Nos próximos slides revistaremos algumas ST anteriormente apresentadas (e eventualmente, analisadas).

Precipitação pluviométrica no município de Fortaleza/CE



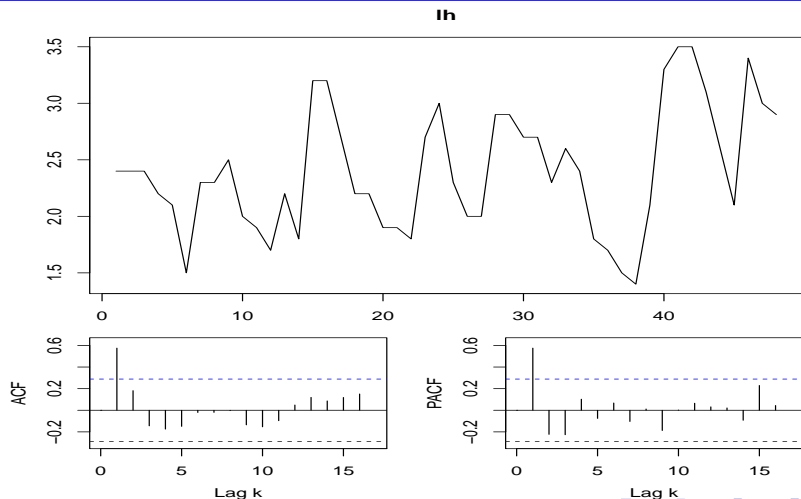
Precipitação pluviométrica no município de Fortaleza/CE

- Trata-se de uma série complicada de ser modelada, em particular, estando limitados ao conteúdo visto neste curso.
- De acordo com Morettin & Tolloi (2018), pag. 14, tal ST foi analisada por vários autores e, aparentemente, não apresenta periodicidades marcantes.
- Por outro lado, estudos de [Morettin et al \(1985\)](#) e [Harvey and Souza \(1987\)](#), indicam uma periodicidade de aproximadamente 13 anos.
- Portanto, não consideraremos a modelagem dessa ST.

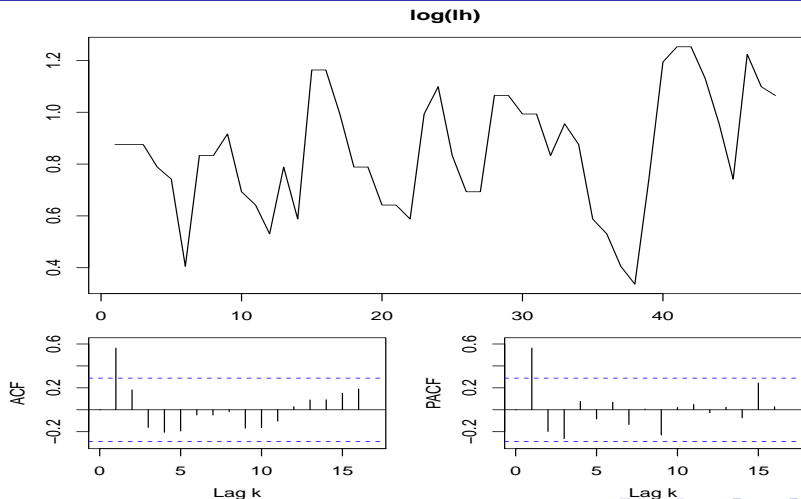
Dados sobre concentração do hormônio luteinizante (HL)

- Um modelo AR(1) fora utilizado ([aqui](#)).
 - Os resíduos apresentam assimetria positiva, embora, segundo os outros gráficos, o ajuste seja razoável.
 - A predição dos valores observados não foi muito boa, embora a predição para valores futuros, pareça razoável.
- Segundo [essa referência](#), os dados do hormônio luteinizante foram fornecidos com precisão de até uma casa decimal, indicando que há algum desvio entre o nível real de HL e as observações. O autor ajuste um modelo (além do escopo desta disciplina) o qual, aparentemente, resolve o problema.
- Entretanto, tentemos encontrar um modelo SARIMA que se ajuste (ainda melhor) aos dados.

Concentração do hormônio luteinizante

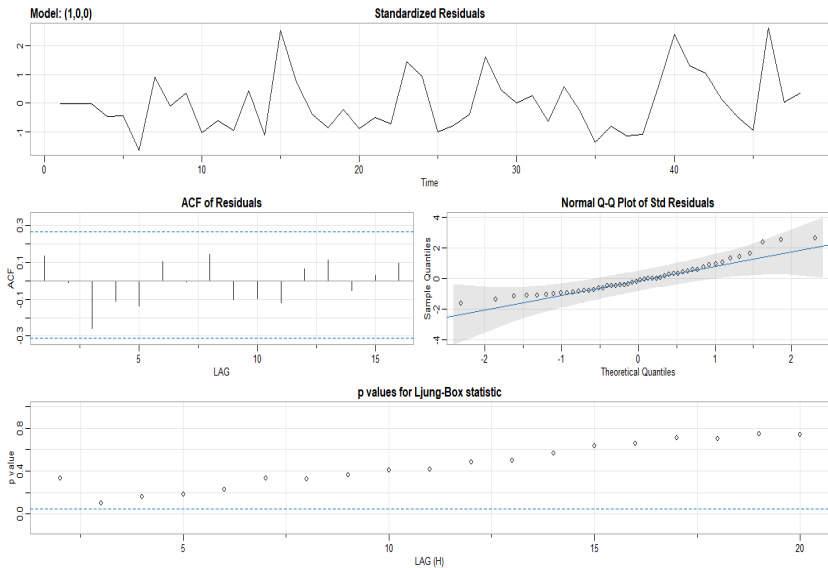


Log da concentração do hormônio luteinizante

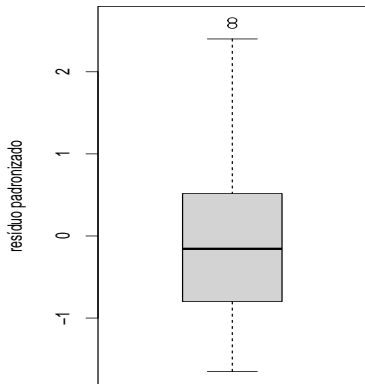
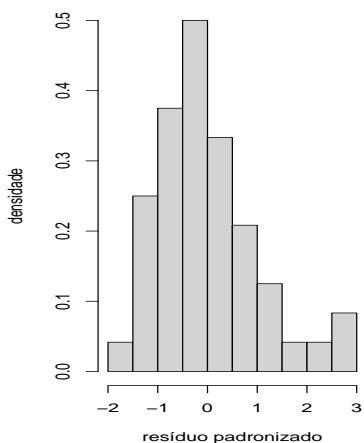


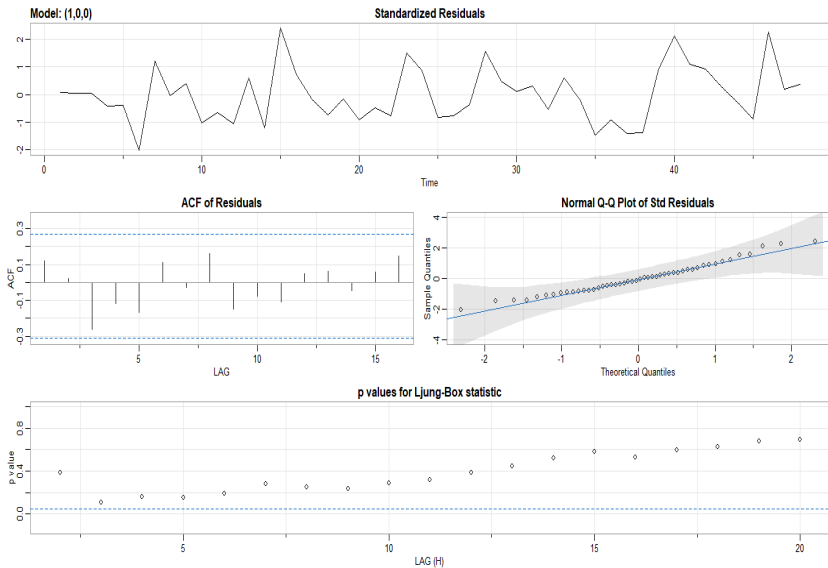
Dados sobre concentração do hormônio luteinizante

- Inicialmente, vamos ajustar um modelo AR(1) para o log (ln) da concentração.
- Em todos os slides a seguir, TSW: teste de [Shapiro-Wilks](#) para normalidade.

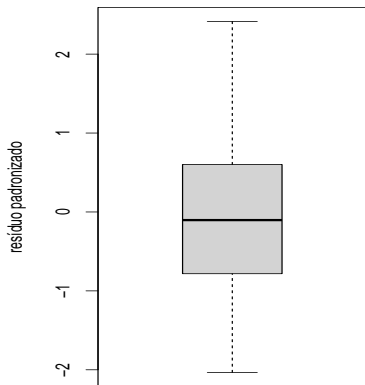
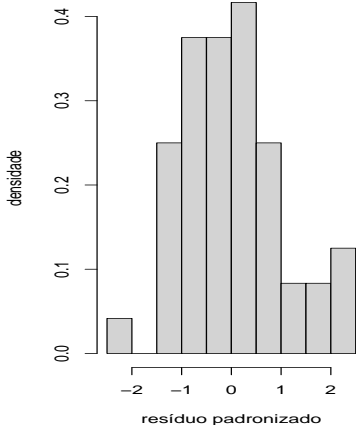


Histograma e Box-plot do RP: AR(1) - TSW ($p=0,0084$)





Histograma e Box-plot do RP: AR(1)-ln TSW ($p=0,3977$)



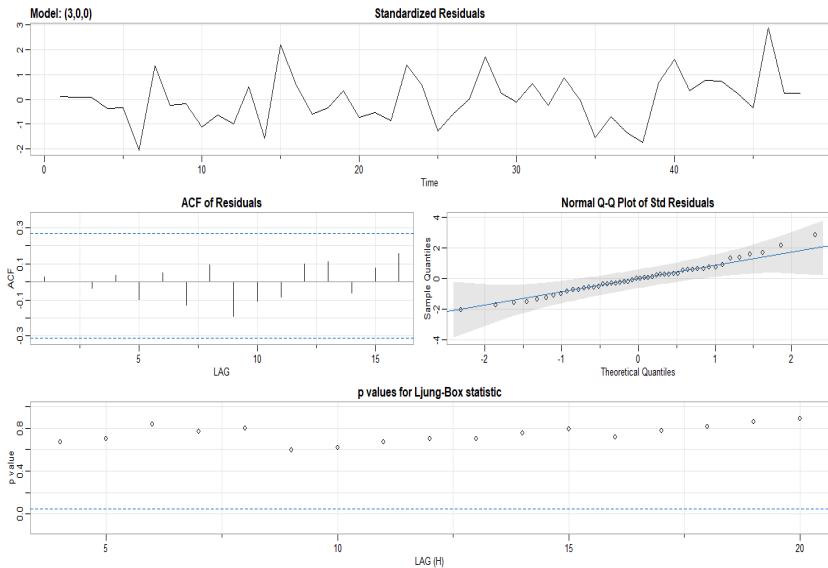
Critérios de Informação (CI)

Modelo	AIC	AICc	BIC
AR(1)	1,349	1,355	1,466
AR(2)	1,344	1,355	1,500
AR(3)	1,337	1,357	1,532
AR(1)-ln	-0,353	-0,347	-0,236
AR(2)-ln	-0,347	-0,336	-0,191
AR(3)-ln	-0,372	-0,353	-0,178

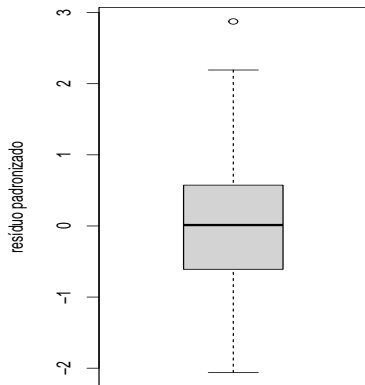
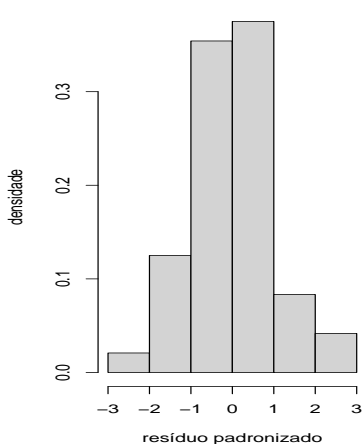
Acurácia da previsão de valores futuros

Modelo	EPM	EQM	EAM	EPAM
AR(1)	-396,866	117,000	10,601	396,866
AR(2)	-291,545	63,699	7,821	291,545
AR(3)	-202,889	33,231	5,429	202,889
AR(1)-ln	-154,353	2,272	1,485	154,353
AR(2)-ln	-136,221	1,774	1,311	136,221
AR(3)-ln	-111,768	1,218	1,072	111,768

Escolheremos o modelo AR(3)-ln, pois fora indicado por seis, das sete estatísticas (CI e CP-Critérios de Previsão). Em tempo, transformar não é, em geral, a melhor opção (com exceção do operador diferença).



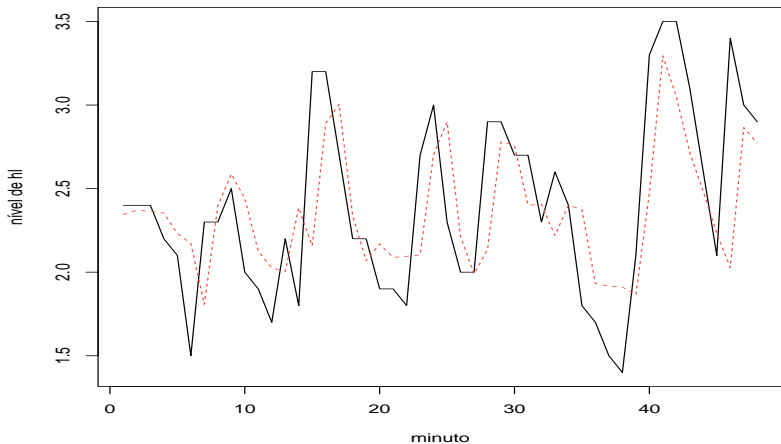
Histograma e Box-plot do RP: AR(3) - TSW ($p=0,6439$)



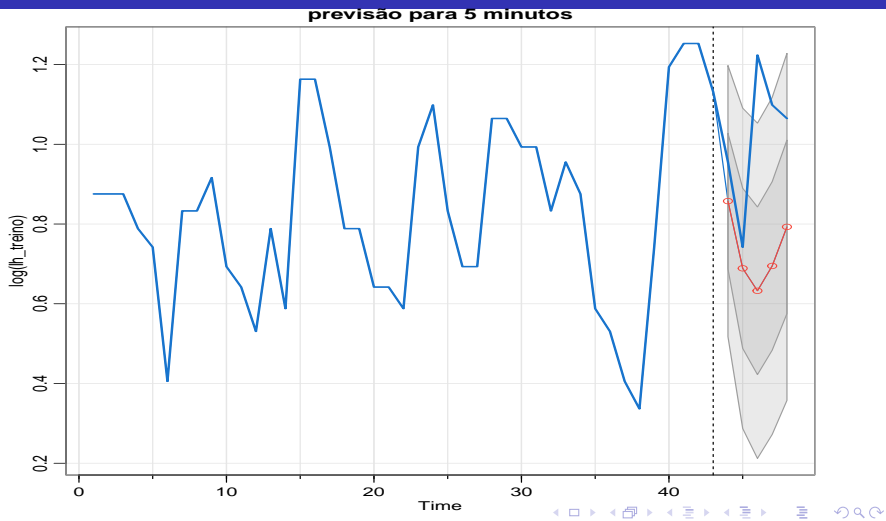
Estimativas do modelo AR(3)-In

Parâmetro	Estimativa	EP	IC(95%)	Estat-t	p-valor
ϕ_1	0,61	0,14	[0,34 ; 0,88]	4,46	0,0001
ϕ_2	-0,02	0,16	[-0,34 ; 0,30]	-0,13	0,8973
ϕ_3	-0,26	0,14	[-0,53 ; 0,02]	-1,83	0,0741
μ	0,85	0,04	[0,77 ; 0,92]	21,38	< 0,0001

Valores preditos para os dados observados



Valores preditos para dados futuros



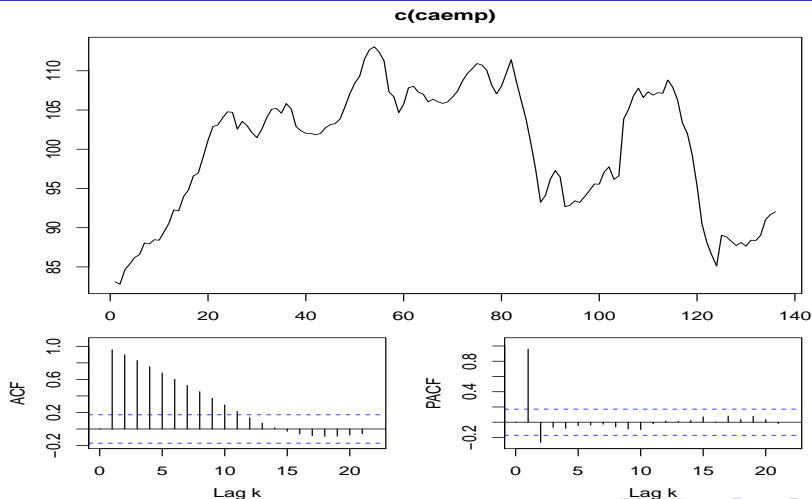
Comentários

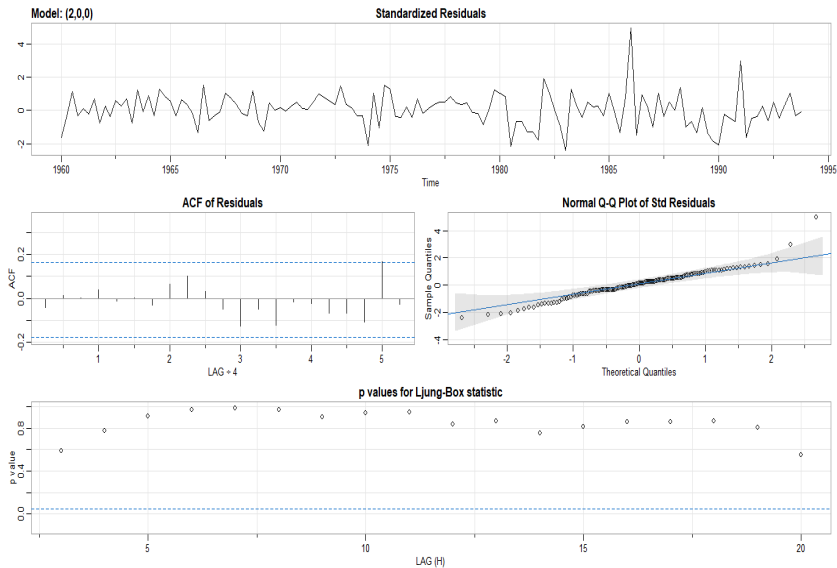
- O modelo, em termos de análise residual, se ajustou bem aos dados.
- No entanto, as previsões (valores observados e futuros) não foram satisfatórias. Provavelmente, isso se deve ao fato mencionado anteriormente.
- Em relação às estimativas, em princípio, não é problema ter parâmetros de ordem maior (ϕ_3) (marginalmente) significativos, e outros de ordem menor (ϕ_2) não significativos.
- O modelo ajustado indicou um processo estacionário ($z_1 = 1,363, z_2 = 2,108, z_3 = 1,363$).
- Uma alternativa é usar a modelagem apresentada [aqui](#), página 10.

Dados sobre índice de emprego no Canadá

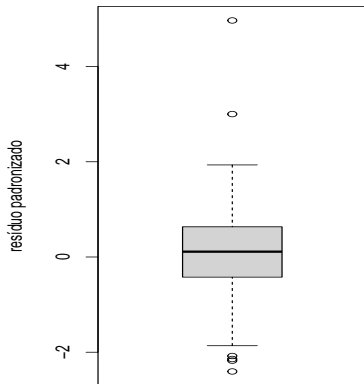
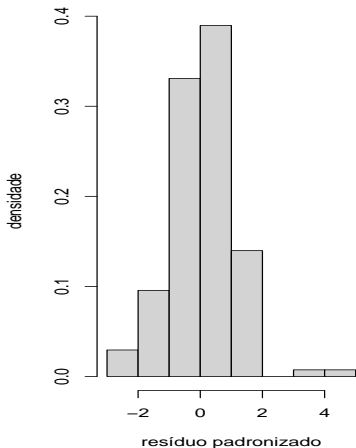
- Os dados já foram analisados ([aqui](#)).
- O modelo AR(2) mostrou-se o melhor entre os modelos AR(i), $i=1,2,3$.
- Compararemos, novamente, os 3 modelos.
- OBS: o ajuste de modelos ARMA(p,q), $p, q \geq 1$ levou a erros-padrão assintóticos, negativos.

Dados sobre índice de emprego no Canadá



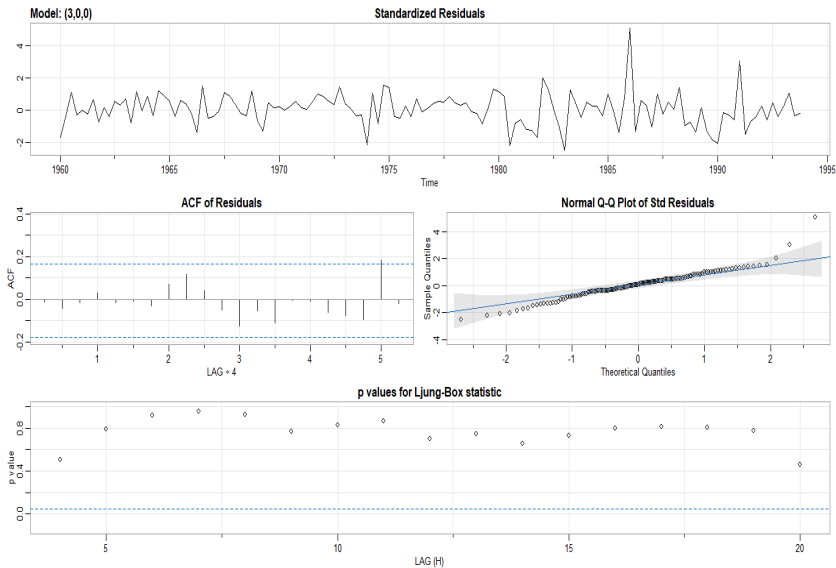


Histograma e Box-plot do RP: AR(2) - TSW ($p < 0,0001$)

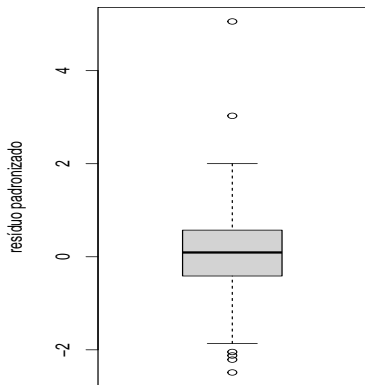
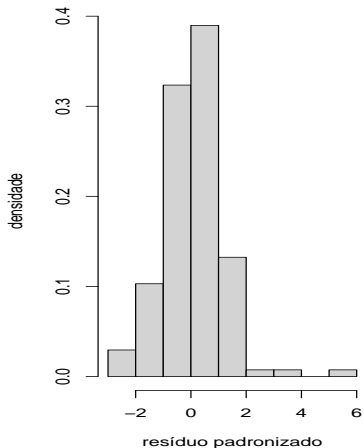


Comentários

- Apesar dos resíduos indicarem que os erros são RB, há valores discrepantes.
- Com efeito, o gráfico QQ com envelopes, o histograma e o box-plot indicam que os resíduos têm uma distribuição de caudas pesadas (possivelmente assimétrica).
- Com efeito, o ajuste de um modelo AR(2), supondo que $\epsilon_t \stackrel{iid}{\sim} t(0, \sigma^2, \nu)$ (erros), indica $\tilde{\nu} = 6,44$ (o que sugere caudas mais pesadas do que as da $N(0, \sigma^2)$, para os erros) ([link](#)).
- Assim, apesar do modelo não ter se ajustado bem, continuaremos a comparação entre os modelos AR(i) ($i = 1, 2, 3$), considerando normalidade para os erros.



Histograma e Box-plot do RP: AR(3) - TSW ($p < 0,0001$)



Critérios de Informação

Modelo	AIC	AICc	BIC
AR1	3,867	3,868	3,931
AR2	3,629	3,631	3,715
AR3	3,639	3,642	3,746

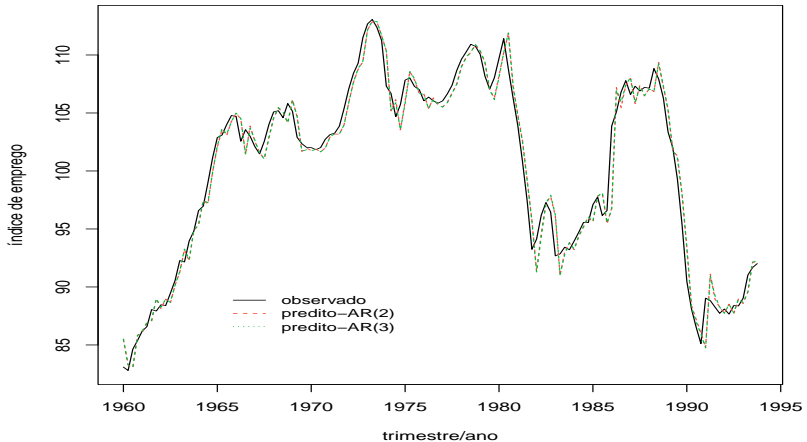
melhor modelo: AR(2)

Cr terios de Previs o

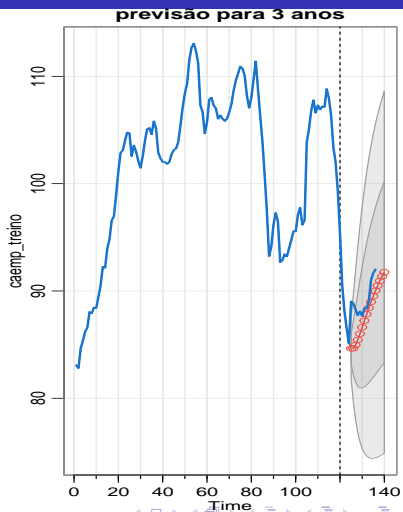
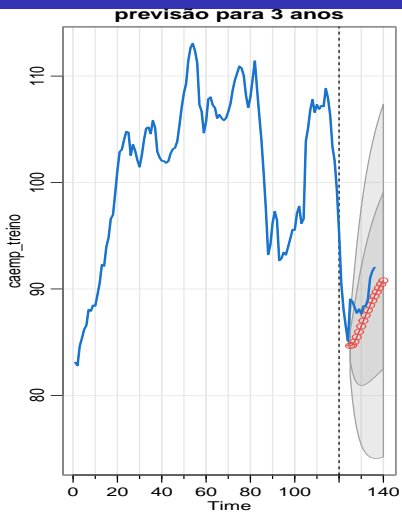
Modelo	EPM	EQM	EAM	EPAM
AR(1)	4,011	14,632	3,597	4,011
AR(2)	2,663	6,881	2,379	2,663
AR(3)	2,409	6,085	2,149	2,409

melhor modelo: AR(3)

Valores preditos (observados)



Valores preditos (futuros)



Estimativas

Parâmetro	Estimate	EP	IC(95%)	Estat. t	p-valor
ϕ_1	1,45	0,07	[1,30 ;1,60]	19,36	< 0,0001
ϕ_1	-0,48	0,08	[-0,63 ; -0,33]	-6,25	< 0,0001
μ	97,50	4,39	[88,89 ; 106,11]	22,19	< 0,0001
ϕ_1	1,42	0,09	[1,25 ;1,59]	16,67	< 0,0001
ϕ_2	-0,38	0,15	[-0,66 ; -0,09]	-2,59	0,0106
ϕ_3	-0,07	0,09	[-0,24 ; 0,10]	-0,79	0,4083
μ	97,81	4,10	[89,78 ; 105,85]	23,86	< 0,0001

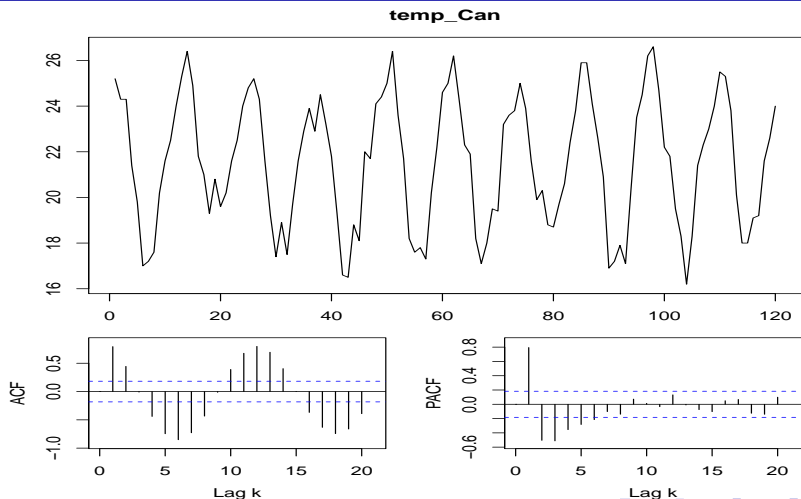
Comentários

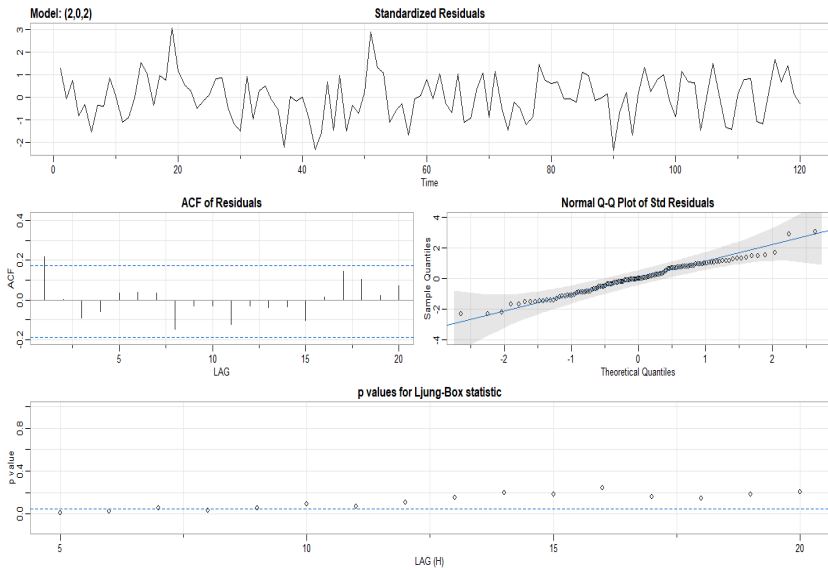
- Optaremos pelo modelo AR(3), apesar dos Critérios de Informação terem indicado o AR(2) e do parâmetro ϕ_3 ter sido não significativo. Note os critérios de previsão indicaram o modelo AR(3). Provavelmente, esse “conflito” se deve a não normalidade dos resíduos.
- Modelos ARMA(p,q), como mencionado anteriormente, apresentaram problemas no ajuste (erros-padrão assintóticos, negativos).
- Sugestão: comparar modelos na classe AR(p), com os equivalente com erros t de Student assimétricos.

Dados de temperaturas mensais ($^{\circ}\text{C}$) da cidade de Cananéia/SP: 1976 a 1985

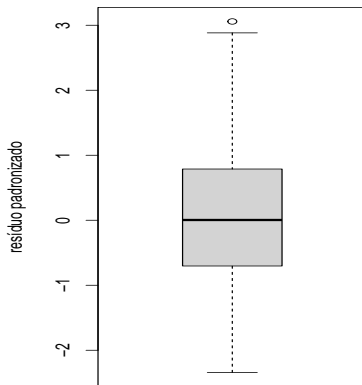
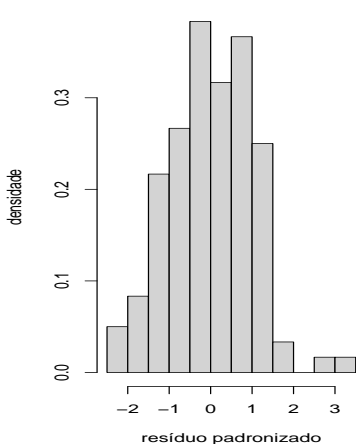
- Os dados já foram analisados ([aqui](#)).
- Dos modelos comparados ($\text{ARMA}(p,q)$), o modelo $\text{ARMA}(2,2)$ foi escolhido. No entanto, não apresentou um bom ajuste, sob os aspectos residual e preditivo.
- Com efeito, percebe-se na ST (próxima figura) uma sazonalidade não passível de ser capturada por modelos $\text{ARMA}(p,q)$, apesar de haver estacionariedade.
- Vamos considerar modelos $\text{SARMA}(p, q)(P, Q)_s$.

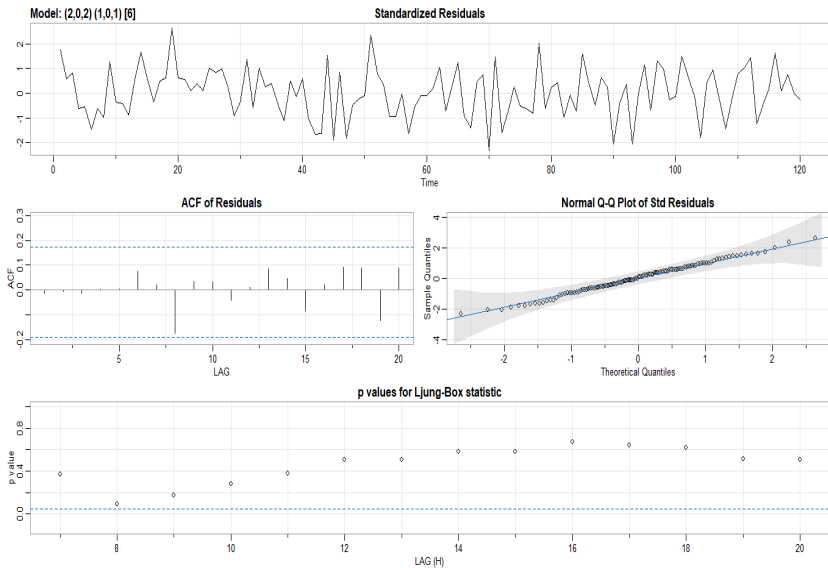
Dados sobre temperatura (ST,FAC,FACP)



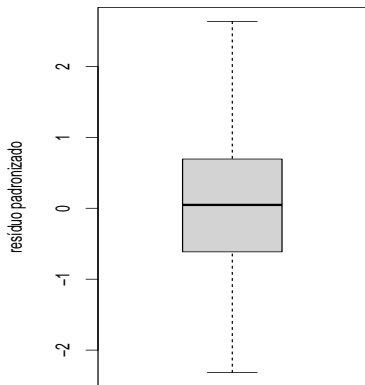
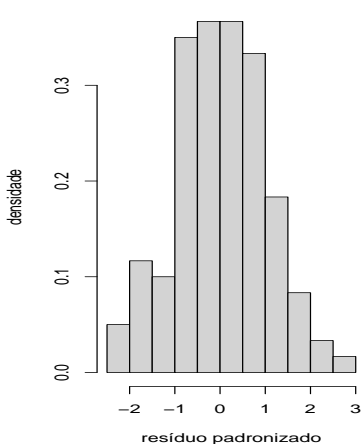


Hist. e BP do RP: ARMA(2,2) - TSW ($p = 0,1692$)





$SARMA(2, 0, 2)(1, 0, 1)_{s=6}$ - TSW ($p=0,8981$)



Critérios de Informação

Modelo	AIC	AICc	BIC
ARMA(1,1)	3,799	3,800	3,892
ARMA(2,1)	3,122	3,125	3,238
ARMA(1,2)	3,694	3,697	3,810
ARMA(2,2)	2,890	2,894	3,029
<i>SARMA</i> (2, 1)(1, 1) _{s=6}	2,929	2,936	3,092
<i>SARMA</i> (2, 2)(1, 1) _{s=6}	2,903	2,911	3,088

Os modelos *SARMA*(1, 1)(1, 1)_{s=6} e *SARMA*(1, 2)(1, 1)_{s=6} apresentaram problemas na convergência do processo iterativo de estimação.

Critérios de Previsão

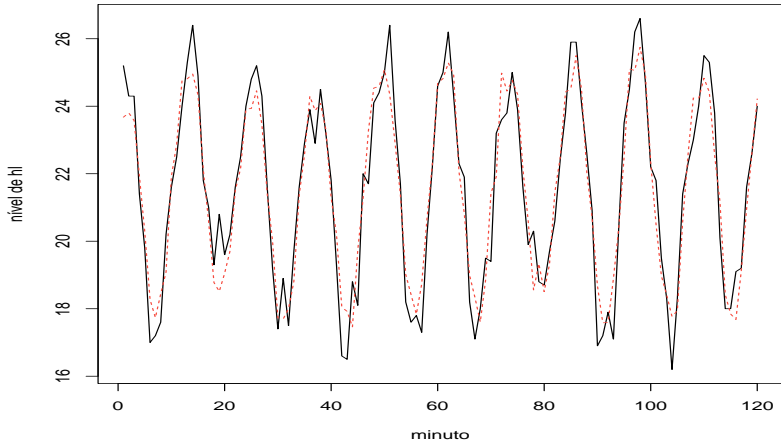
Modelo	EPM	EQM	EAM	EPAM
ARMA(1,1)	-5,029	6,388	2,227	10,989
ARMA(2,1)	-5,153	6,079	2,147	10,646
ARMA(1,2)	-0,061	2,064	1,342	6,319
<i>SARMA</i> (2, 1)(1, 1) _{s=6}	1,751	0,715	0,698	3,211
<i>SARMA</i> (2, 2)(1, 1) _{s=6}	1,787	0,726	0,701	3,224

O modelo ARMA(2,2) apresentou problemas no ajuste, para a amostra treino. Como os modelos SARIMA apresentaram os melhores ajustes mas, a diferença em termos de previsão é módica, escolheremos o modelo *SARMA*(2, 2)(1, 1)_{s=6}.

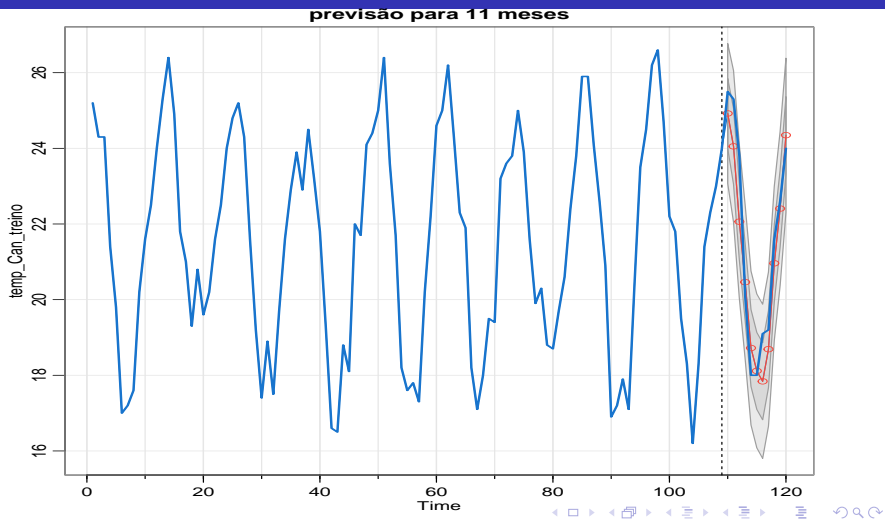
Estimativas do modelo selecionado

Parâmetro	Est.	EP	IC(95%)	Estat. t	p-valor
ϕ_1	1,46	0,17	[1,13 ; 1,78]	8,71	< 0,0001
ϕ_2	-0,53	0,16	[-0,85 ; -0,21]	-3,26	0,0015
θ_1	-1,10	0,20	[-1,48 ; -0,71]	-5,57	< 0,0001
θ_2	0,10	0,20	[-0,29 ; 0,48]	0,50	0,6179
Φ_1	-1,00	<0,01	[<-1,00 ; >-1,00]	-4297,92	< 0,0001
Θ_1	0,98	0,04	[0,91 ; 1,06]	24,95	< 0,0001
μ	21,51	0,02	[21,46 ; 21,56]	869,80	< 0,0001

Valores preditos (observados)



Valores preditos (futuros)



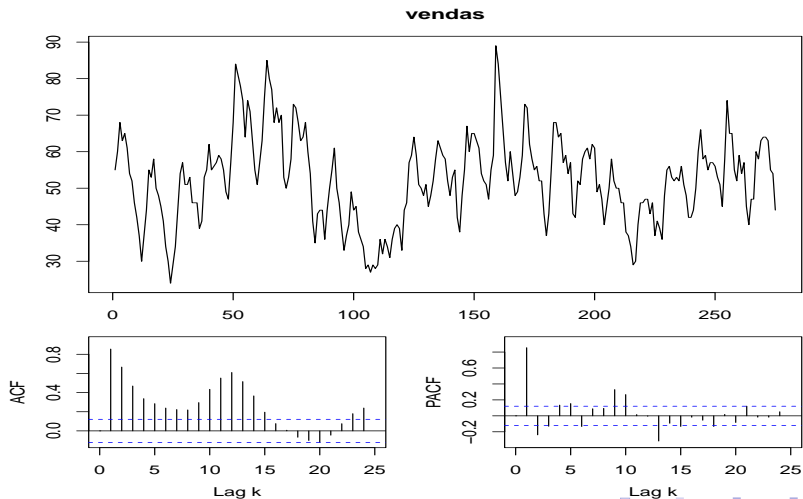
Comentários

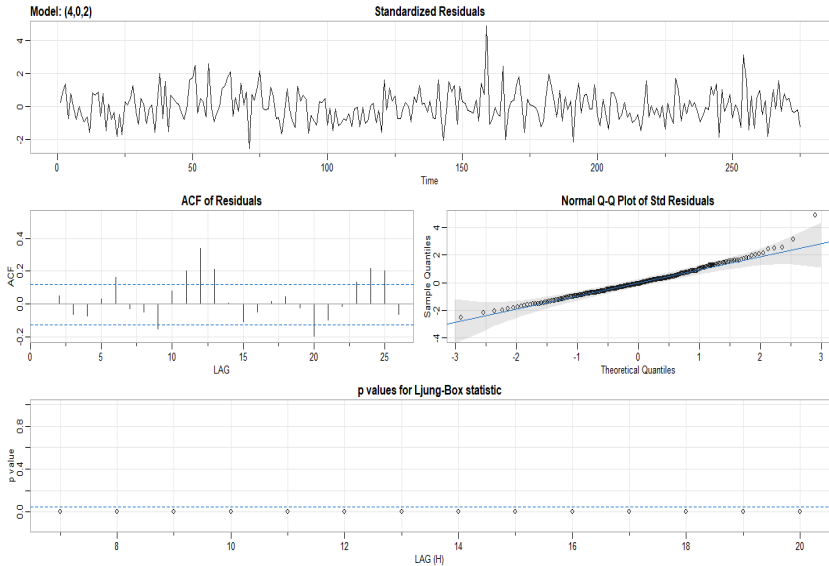
- O modelo escolhido mostrou um ajuste bastante satisfatório, tanto do ponto de vista residual quanto de predição (observados e futuros).
- Não se vê necessidade de tentar melhorar, ao menos dentro do contexto do curso.
- Características das raízes dos polinômios (estacionário, causal e invertível - regular e sazonal):
 - $\phi(B)$: $z_1 = 1,306$; $z_2 = 1,449$; $\theta(B)$: $z_1 = 1,000$; $z_2 = 10,235$.
 - $\Phi(B)$: $z_1 > 1,000$; $z_2 > 1,000$; $z_3 > 1,000$; $z_4 > 1,000$; $z_5 > 1,000$; $z_6 > 1,000$.
 - $\Theta(B)$: $z_1 > 1,002$; $z_2 > 1,002$; $z_3 > 1,002$; $z_4 > 1,002$; $z_5 > 1,002$; $z_6 > 1,002$;

Dados de vendas mensais de casas

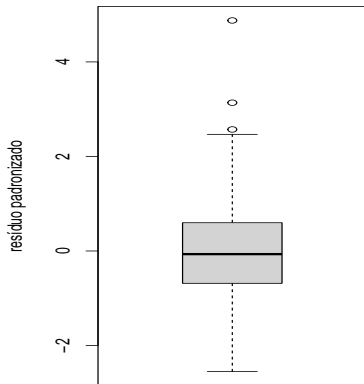
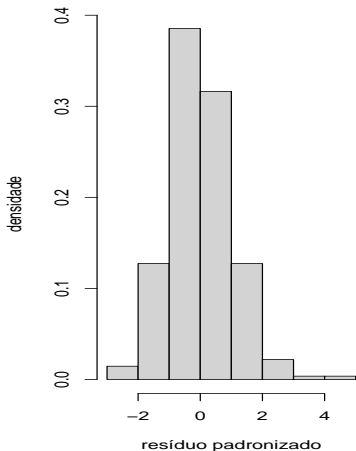
- Analisados, anteriormente, [aqui](#).
- De todos os modelos considerados ($ARMA(p,q)$), o selecionado fora o $ARMA(4,2)$.
- Entretanto, a análise residual indicou um mal ajuste (erros não RB, presença de ACF nos resíduos e não normalidade destes).
- Por outro lado, as previsões (daos observados e futuros) foram razoáveis.
- Além disso (veja gráfico a seguir) parece existir uma sazonalidade multiplicativa, embora a ST pareça ser estacionária. Compare este padrão com o desta [figura](#) (slide 38)

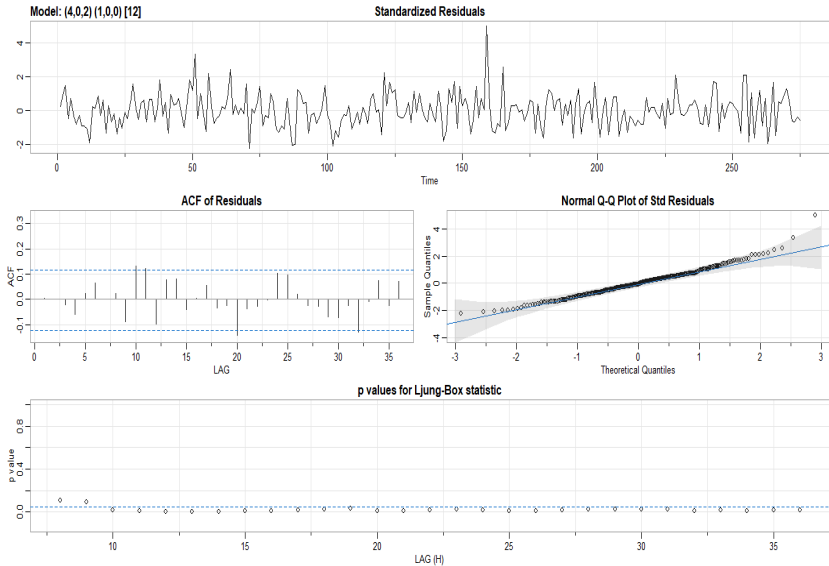
Dados sobre vendas de casa (ST,FAC,FACP)



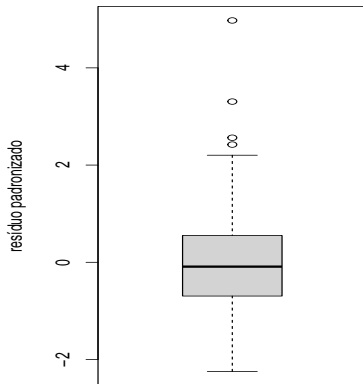
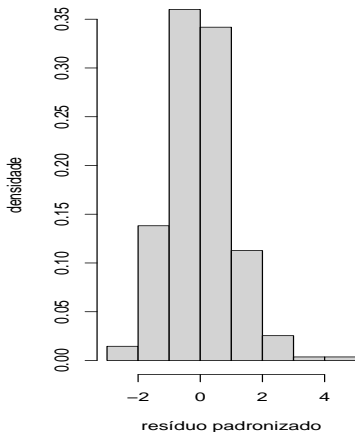


ARMA(4,2) - TSW ($p=0,1692$)





SARMA(4, 2)(1, 0)_{s=12} - TSW (p=0,1692)

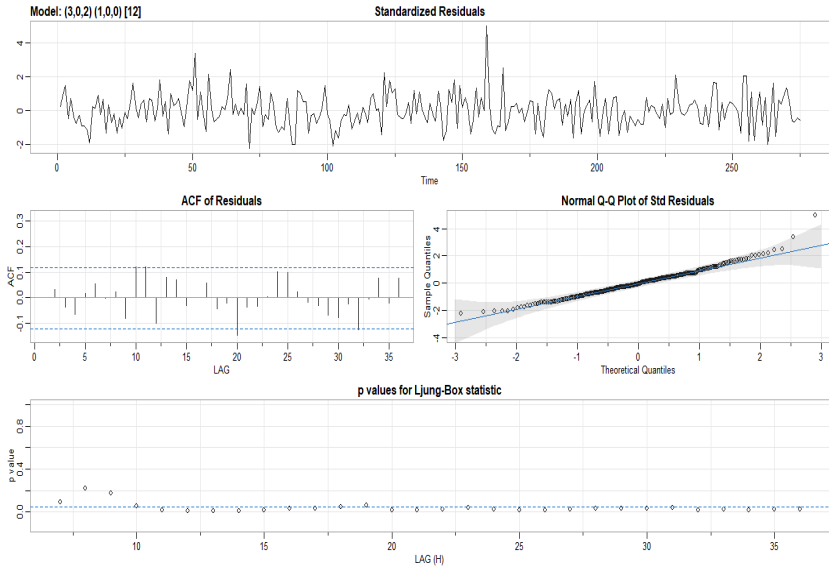


Critérios de Informação

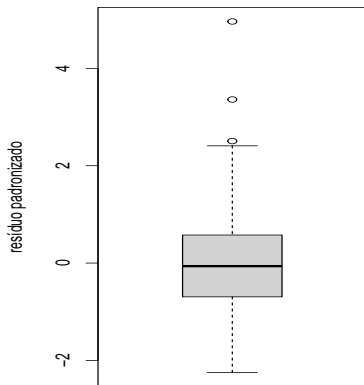
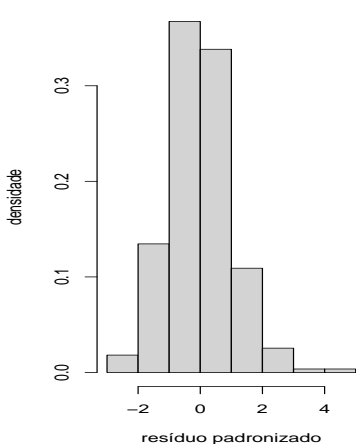
Modelo	AIC	AICc	BIC
<i>SARMA</i> (1, 1)(1, 0) _{s=12}	6,151	6,151	6,217
<i>SARMA</i> (1, 2)(1, 0) _{s=12}	6,156	6,157	6,235
<i>SARMA</i> (1, 3)(1, 0) _{s=12}	6,163	6,165	6,256
<i>SARMA</i> (1, 4)(1, 0) _{s=12}	6,157	6,159	6,262
<i>SARMA</i> (2, 1)(1, 0) _{s=12}	6,157	6,158	6,236
<i>SARMA</i> (2, 2)(1, 0) _{s=12}	6,163	6,165	6,256
<i>SARMA</i> (2, 3)(1, 0) _{s=12}	6,157	6,158	6,262
<i>SARMA</i> (2, 4)(1, 0) _{s=12}	6,163	6,165	6,282
<i>SARMA</i> (3, 1)(1, 0) _{s=12}	6,153	6,154	6,245
<i>SARMA</i> (3, 2)(1, 0) _{s=12}	6,160	6,161	6,265
<i>SARMA</i> (3, 3)(1, 0) _{s=12}	6,164	6,166	6,282
<i>SARMA</i> (3, 4)(1, 0) _{s=12}	6,138	6,140	6,269
<i>SARMA</i> (4, 1)(1, 0) _{s=12}	6,158	6,160	6,263
<i>SARMA</i> (4, 2)(1, 0) _{s=12}	6,165	6,167	6,283
<i>SARMA</i> (4, 3)(1, 0) _{s=12}	6,107	6,109	6,238
<i>SARMA</i> (4, 4)(1, 0) _{s=12}	6,143	6,146	6,288

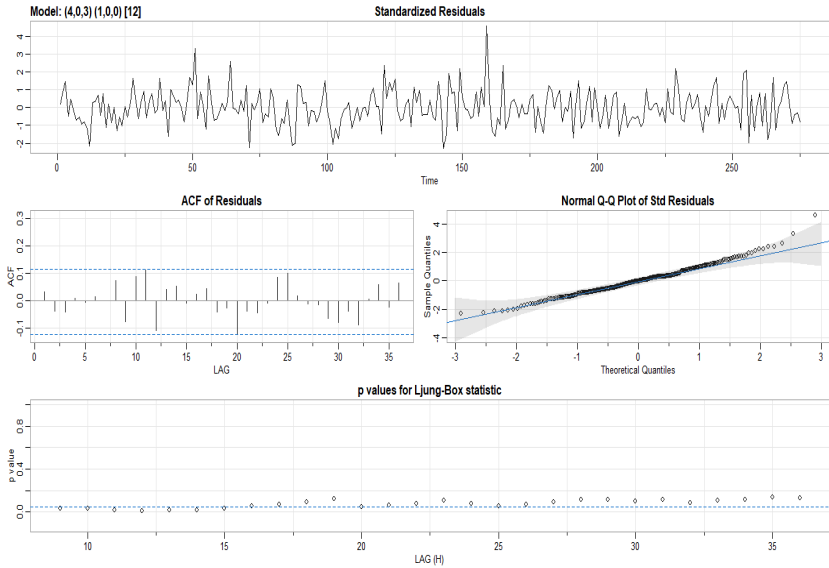
Cr terios de Previs o

Modelo	EPM	EQM	EAM	EPAM
<i>SARMA</i> (1, 1)(1, 0) _{s=12}	-0,552	128,163	9,986	19,912
<i>SARMA</i> (1, 2)(1, 0) _{s=12}	4,173	84,528	7,344	13,083
<i>SARMA</i> (1, 3)(1, 0) _{s=12}	3,932	91,705	7,593	13,380
<i>SARMA</i> (1, 4)(1, 0) _{s=12}	4,593	124,480	9,895	17,750
<i>SARMA</i> (2, 1)(1, 0) _{s=12}	-0,705	136,082	10,535	20,842
<i>SARMA</i> (2, 2)(1, 0) _{s=12}	4,234	77,533	6,988	12,456
<i>SARMA</i> (2, 3)(1, 0) _{s=12}	4,003	83,163	7,144	12,626
<i>SARMA</i> (2, 4)(1, 0) _{s=12}	4,879	102,408	8,792	15,779
<i>SARMA</i> (3, 1)(1, 0) _{s=12}	-0,294	112,146	8,992	17,831
<i>SARMA</i> (3, 2)(1, 0) _{s=12}	3,572	67,070	7,254	13,048
<i>SARMA</i> (3, 3)(1, 0) _{s=12}	4,179	74,086	7,542	13,433
<i>SARMA</i> (3, 4)(1, 0) _{s=12}	4,212	119,498	9,629	17,419
<i>SARMA</i> (4, 1)(1, 0) _{s=12}	-1,067	128,853	9,941	19,928
<i>SARMA</i> (4, 2)(1, 0) _{s=12}	3,844	85,215	7,425	13,275
<i>SARMA</i> (4, 3)(1, 0) _{s=12}	2,999	92,140	7,706	13,663
<i>SARMA</i> (4, 4)(1, 0) _{s=12}	3,671	122,527	9,922	17,928

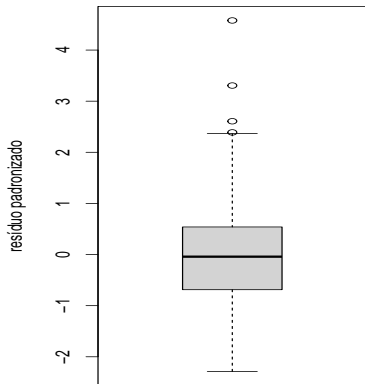
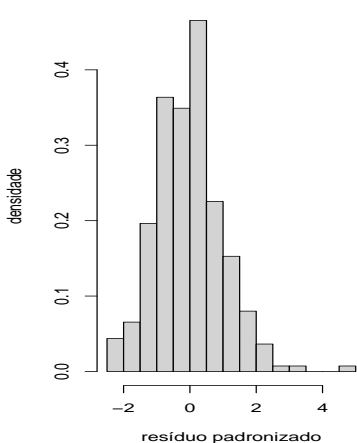


SARMA(3, 2)(1, 0)_{s=12} - TSW (p=0,0002)

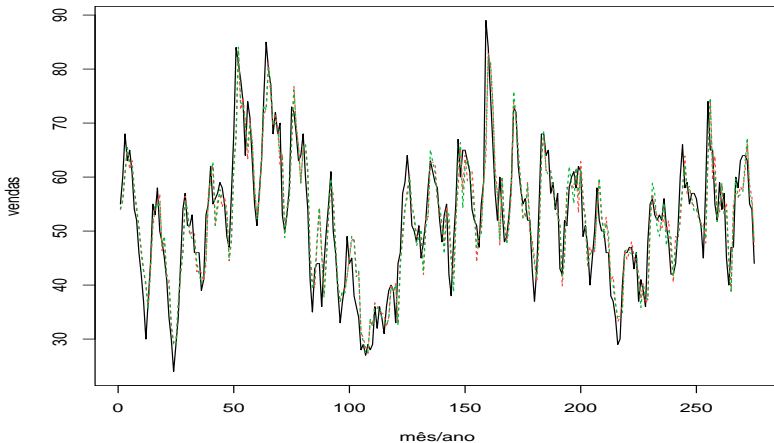




SARMA(4, 3)(1, 0)_{s=12} - TSW (p=0,0011)



Valores preditos para os dados observados



Valores preditos para dados futuros



Prâmetro	Est.	EP	IC(95%)	Estat. t	p-valor
ϕ_1	0,15	0,51	[-0,84 ; 1,14]	0,30	0,7642
ϕ_2	0,82	0,07	[0,67 ; 0,96]	11,10	<0,0001
ϕ_3	-0,23	0,40	[-1,02 ; 0,55]	-0,59	0,5564
θ_1	0,76	0,51	[-0,24 ; 1,75]	1,48	0,1397
θ_2	-0,23	0,50	[-1,22 ; 0,75]	-0,46	0,6428
Φ_1	0,55	0,05	[0,45 ; 0,66]	10,27	<0,0001
μ	52,67	3,66	[45,50 ; 59,84]	14,40	<0,0001

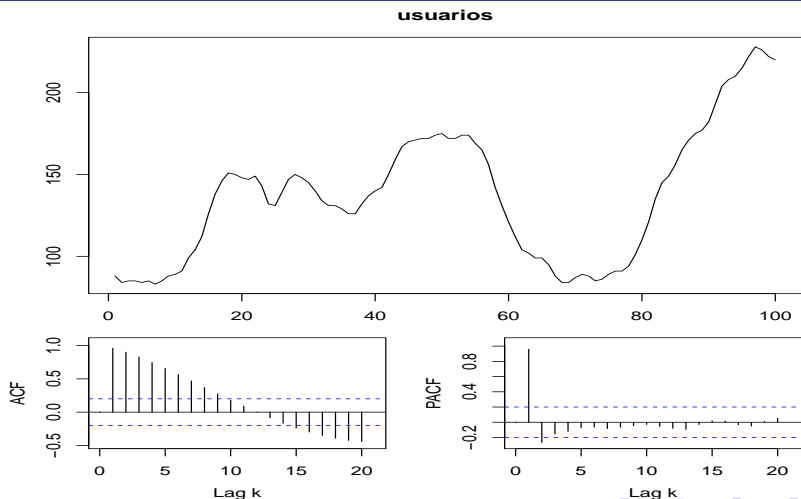
Comentários

- O modelo escolhido é o $SARMA(3, 2)(1, 0)_{s=12}$ (melhor previsão, bem destacada com relação aos demais modelos).
- Do ponto de vista de análise residual, o modelo não se ajustou bem (não normalidade e comportamento não compatível com RB dos erros).
- Embora a previsão para os dados razoáveis tenha sido satisfatória, para os dados futuros, não o foi.
- Sugestão: utilizara estrutura $SARMA(p, q)(P, D)_s$ com erros t de Student assimétricos.

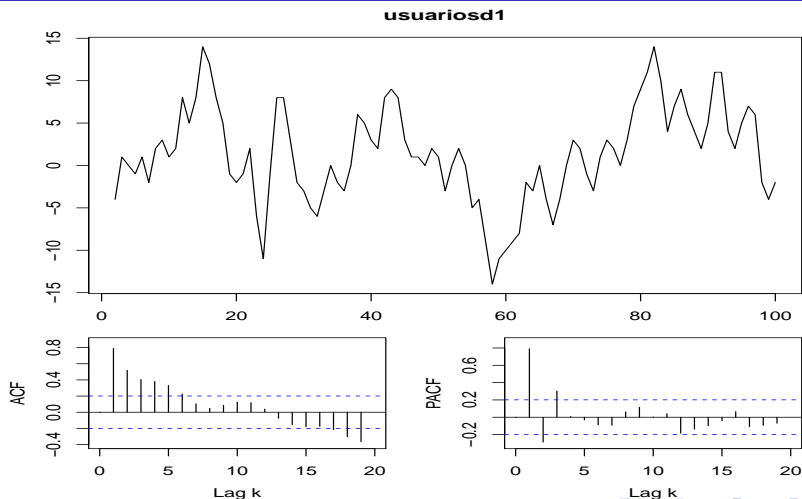
Dados do número de usuários conectados à Internet por meio de um servidor a cada minuto

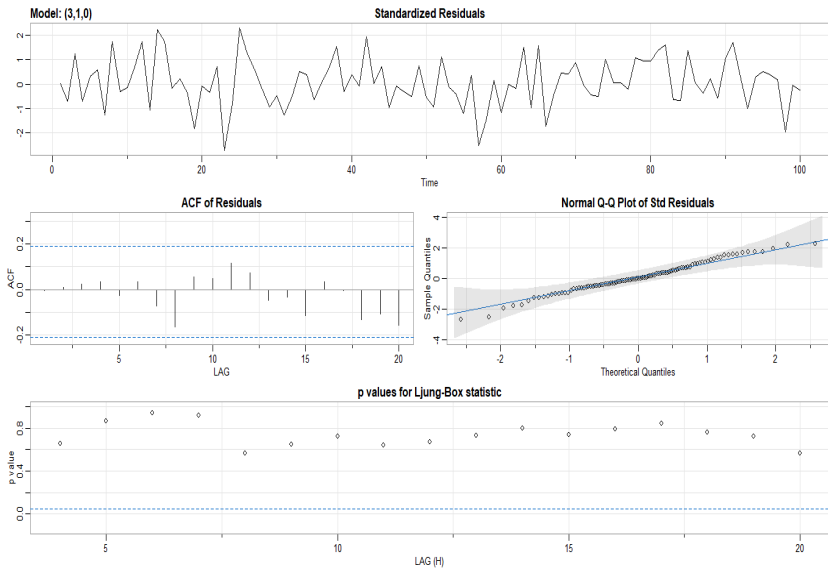
- Anteriormente analisados aqui.
- Série não estacionária com um padrão semelhante à um processo $ARIMA(p,d,q)$.
- Anteriormente, o modelo $ARIMA(3,1,0)$ (sem constante) fora selecionado, apresentando um bom ajuste (residual + previsão).
- Vamos comparar (mais) modelos (sem constante) nessa classe, usando critérios adicionais.

Dados sobre número de usuários (ST,FAC,FACP)

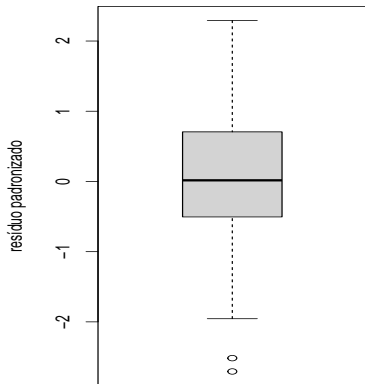
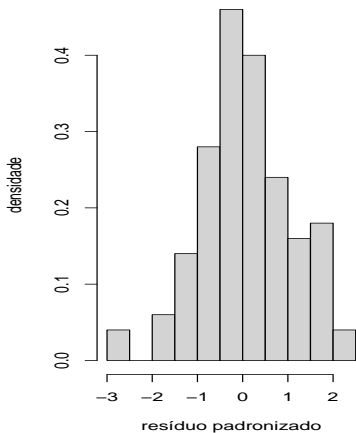


Dados sobre número de usuários (Δ_1) (ST,FAC,FACP)





ARIMA(3,1,0) - TSW ($p=0,5951$)



Critérios de Informação

Modelo	AIC	AICc	BIC
ARIMA(1,1,0)	5,346	5,346	5,398
ARIMA(2,1,0)	5,275	5,276	5,353
ARIMA(3,1,0)	5,172	5,174	5,277
ARIMA(4,1,0)	5,191	5,196	5,322
ARIMA(1,1,1)	5,195	5,196	5,274
ARIMA(2,1,1)	5,215	5,218	5,320
ARIMA(3,1,1)	5,191	5,196	5,322
ARIMA(4,1,1)	5,212	5,218	5,369

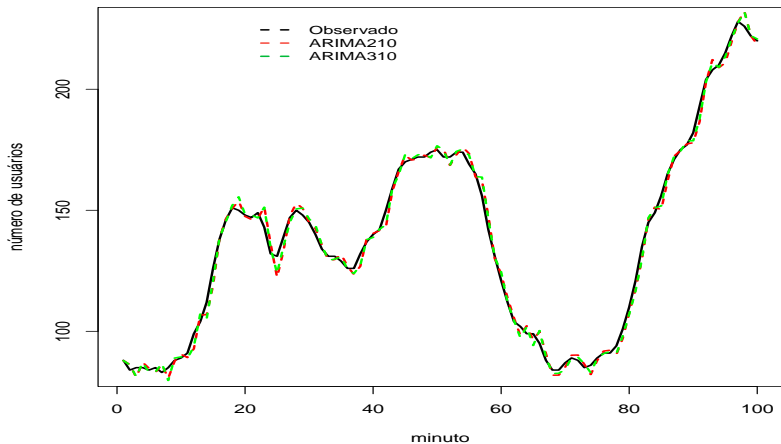
modelo escolhido: ARIMA(3,1,0)

Cr terios de Previs o

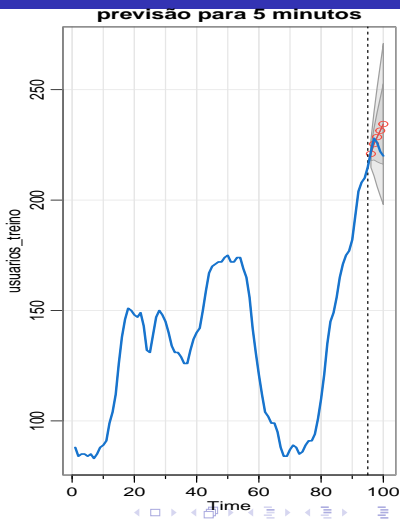
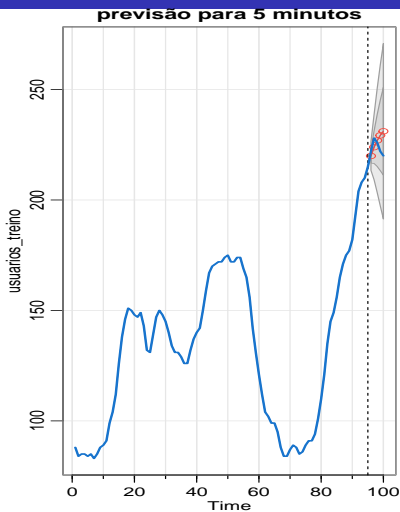
Modelo	EPM	EQM	EAM	EPAM
ARIMA(1,1,0)	3,261	61,063	7,322	3,261
ARIMA(2,1,0)	2,146	32,227	4,837	2,146
ARIMA(3,1,0)	2,377	34,830	5,344	2,377
ARIMA(4,1,0)	2,561	39,680	5,755	2,561
ARIMA(1,1,1)	2,556	41,076	5,749	2,556
ARIMA(2,1,1)	2,593	42,022	5,834	2,593
ARIMA(3,1,1)	2,531	38,869	5,688	2,531
ARIMA(4,1,1)	2,493	37,862	5,604	2,493

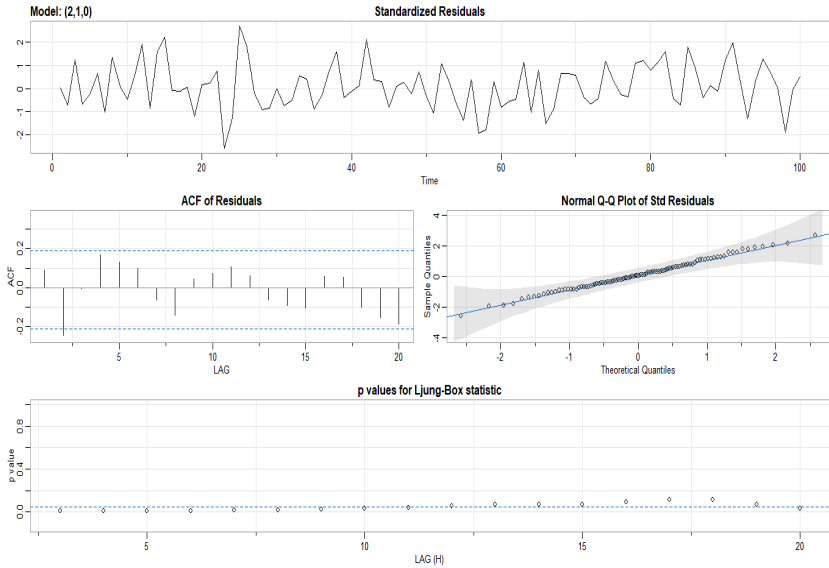
modelo escolhido: ARIMA(2,1,0)

Valores preditos para os dados observados

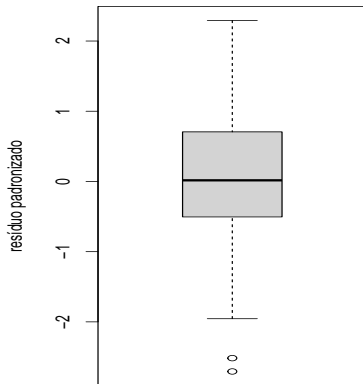
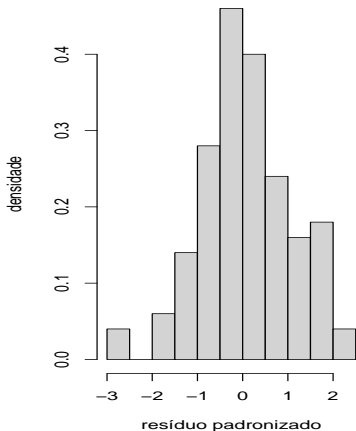


Valores preditos para dados futuros





ARIMA(2,1,0) - TSW ($p=0,9872$)



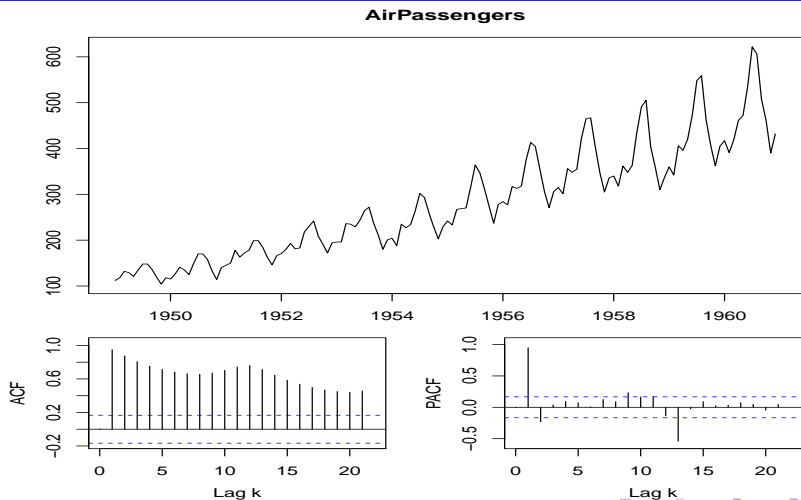
Comentários

- O modelo ARIMA(2,1,0) não se ajustou bem aos dados, devido à existência de ACF's significativas para os resíduos, apesar dos outros gráficos indicarem um bom ajuste.
- Como os CI indicaram o modelo ARIMA(3,1,0) e o modelo ARIMA(2,1,0) não se ajustou bem aos dados, pelo item anterior, apesar dele ter apresentado melhor previsão para dados futuros, escolheremos o modelo ARIMA(3,1,0).
- Para melhorar a previsão para dados futuros, podemos considerar uma extensão do modelo ARIMA(3,1,0) com mudança de regima.

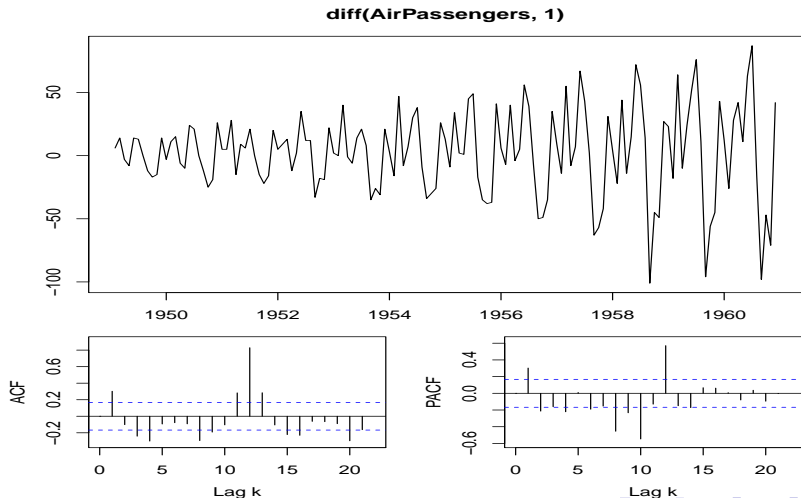
Dados sobre número de passageiros

- Voltando para a série dos passageiros, vamos analisar novamente o comportamento dela.
- [Aqui](#) utilizamos o modelo AR(1), o qual não se ajustou bem aos dados.
- Primeiramente nota-se que essa série parece ter uma variância não constante no tempo, logo, uma transformação logarítmica é útil de forma a estabilizar essa variância.

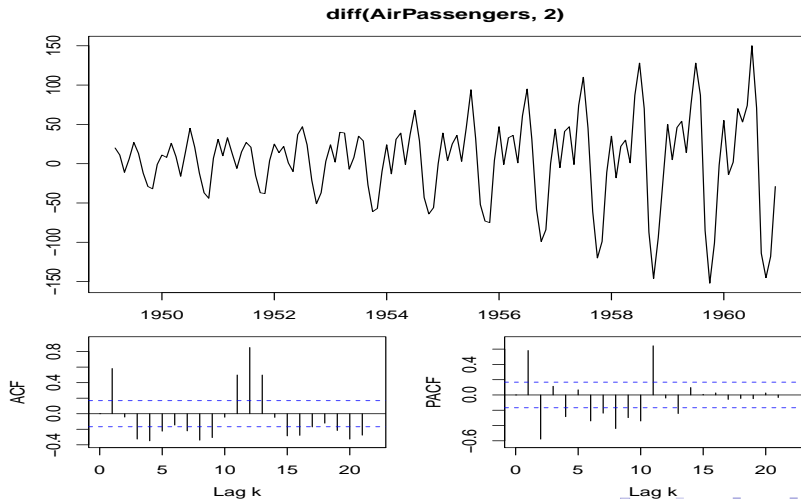
Gráficos ST, FAC, FACP



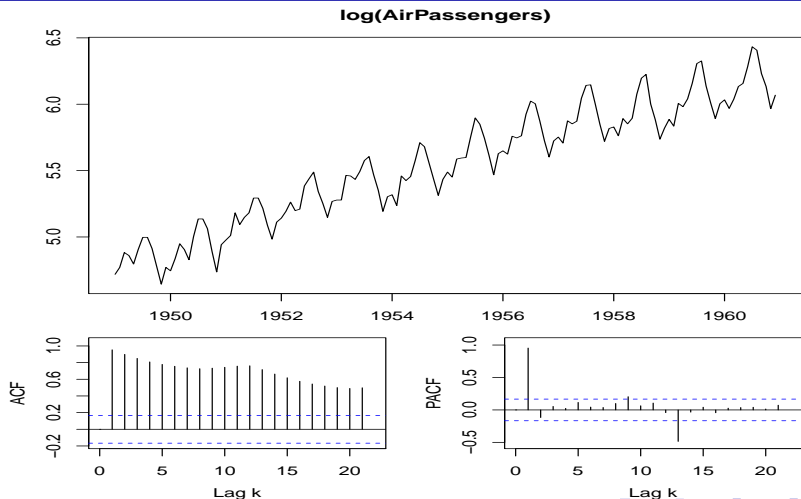
Gráficos ST, FAC, FACP (Δ_1)



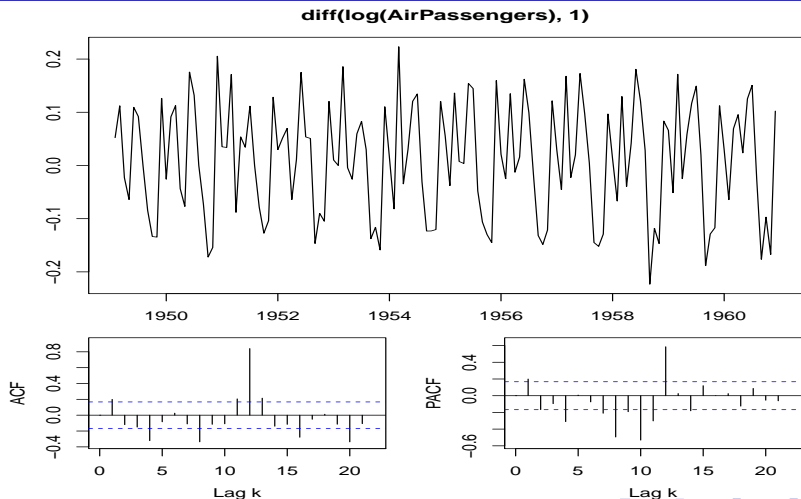
Gráficos ST, FAC, FACP(Δ_2)



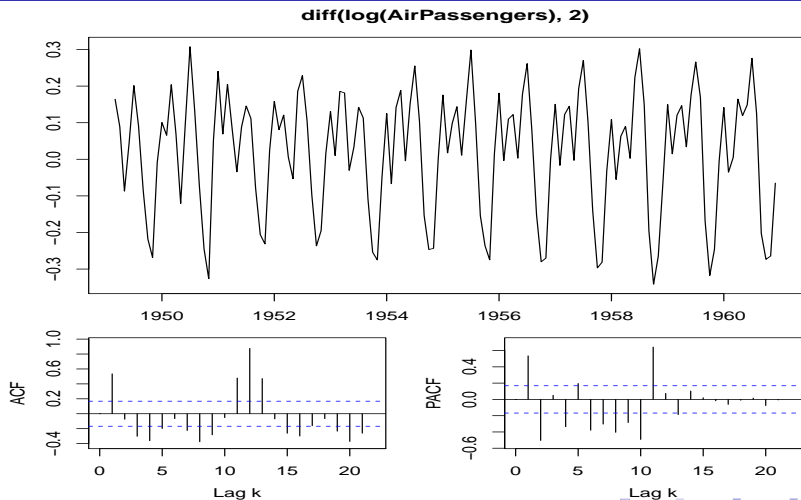
Gráficos ST, FAC, FACP-In



Gráficos ST, FAC, FACP-In (Δ_1)



Gráficos ST, FAC, FACP-In (Δ_2)

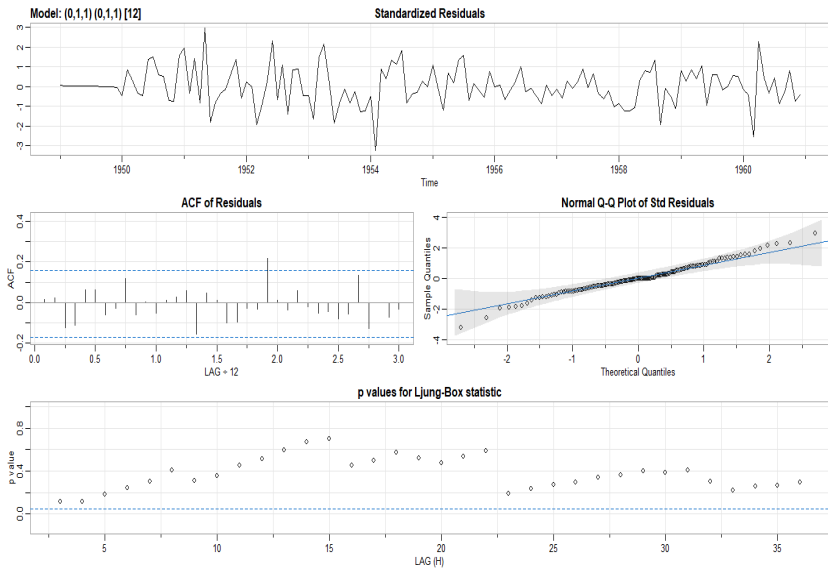


Comentários

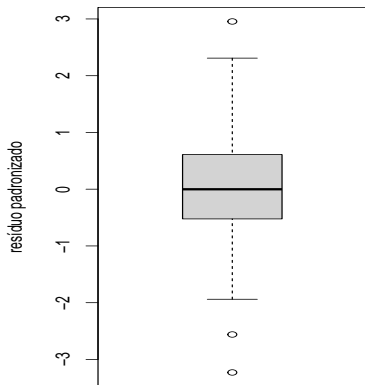
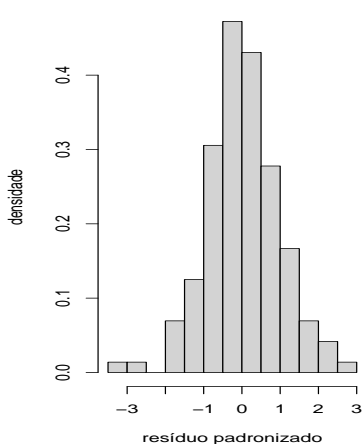
- Percebe-se através da FAC que as autocorrelações amostrais de 12 em 12 meses são diferentes de zero e altas, ou seja, temos indícios de que um mês específico em um ano tem alta correlação com os outros mesmos meses dos outros anos, ou seja, todos os Janeiros, Fevereiros, etc.
- Portanto, seria útil termos um modelo que tanto incorpora-se a correlação mês a mês (período 1) quanto ano a ano (período 12), para isso o modelo ARIMA sazonal (SARIMA) é útil.

Modelagem SARIMA

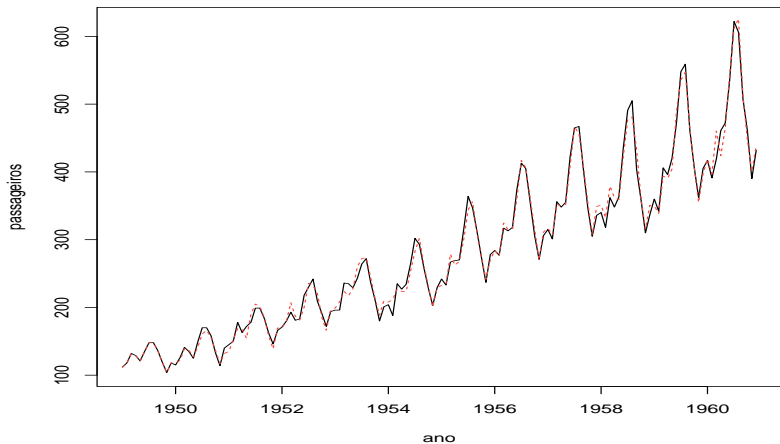
- Na literatura esse banco de dados é bem modelado pelo modelo $SARIMA(0, 1, 1)(0, 1, 1)_{12}$, inclusive esse modelo leva o nome de AIR-LINE devido a ter sido usado primeiramente nessa aplicação.
- Aqui, retiramos as 24 últimas observações de forma a avaliarmos a previsão para esse modelo.



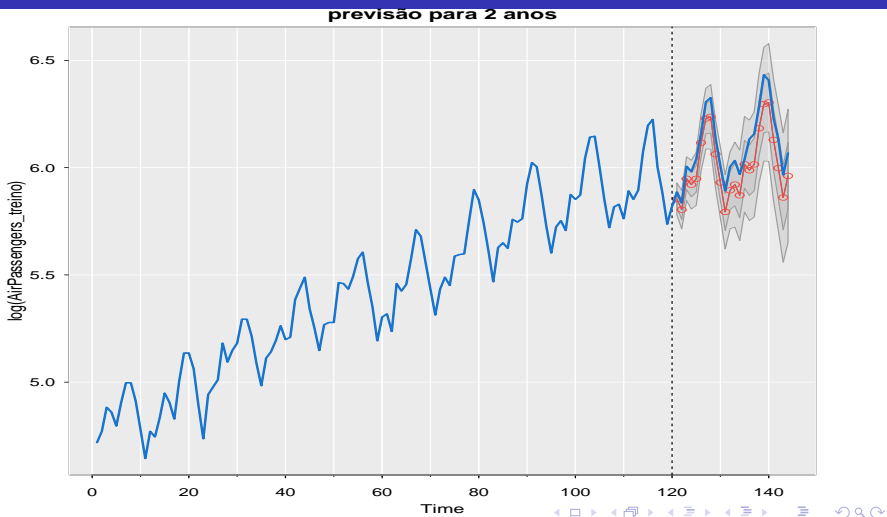
$SARIMA(0, 1, 1)(0, 1, 1)_{s=12}$ - TSW ($p=0,1674$)



Valores preditos para os dados observados



Valores preditos para dados futuros



Modelagem SARIMA

- O modelo se ajustou de forma satisfatória Porém, o ajuste pode ser melhorado, particularmente no que tange à diminuir as ACF's e, assim, conseguir um comportamento, por parte dos resíduos, mais próximo de um RB.
- Em relação à previsão:
 - Valores observados: previsão satisfatória.
 - Valores futuros: sub-previstos.
- Sugestão: considerar modelos da família $SARIMA(p, d, q)(P, D, Q)_s$, assumindo uma distribuição positiva (assimétrica, heterocedástico). Também pode ser apropriado considerar modelos de volatilidade estocástica.