

MODELOS ESTOCÁSTICOS DE PREVISÃO DA UMIDADE RELATIVA DO AR PARA BOA VISTA-RR

Data de submissão: 12/05/2023

Data de aceite: 03/07/2023

Alanderson Firmino de Lucas

<http://lattes.cnpq.br/0484302300145718>

Matheus Henrique Bezerra de Melo

<http://lattes.cnpq.br/2538146023601236>

Daniel Felipe da Silva

<http://lattes.cnpq.br/9510085354088329>

Thiago Gonçalves da Silva

<http://lattes.cnpq.br/9247213148151165>

Albert Smith Feitosa Suassuna Martins

<http://lattes.cnpq.br/1571415337290282>

Everton Felipe de Souza Pinheiro

<http://lattes.cnpq.br/3226955524305254>

Bárbara Araujo Silva

<http://lattes.cnpq.br/0059136452032385>

Paula Andressa Alves de Araújo

<http://lattes.cnpq.br/4857107569349842>

José Augusto Ferreira Neto

<http://lattes.cnpq.br/0401494432064481>

Israel Walter Hilário da Silva

<http://lattes.cnpq.br/5659345555242426>

RESUMO: Neste capítulo, foi apresentado o processo de determinação da ordem

do modelo ARMA para a série temporal da Umidade Relativa do Ar de Boa Vista/RR de 1961 a 2014. Como a série já era estacionária, não foi necessária a aplicação da primeira diferença simples para torná-la estacionária. A ordem do modelo ARMA foi determinada a partir dos gráficos da PACF e da ACF, resultando na ordem (1,0,2). A série temporal apresenta flutuações sazonais com picos no verão, extremos em 1997 devido ao El Niño e alguns outliers em anos com menor variabilidade da Umidade Relativa do Ar da região. Embora a série tenha variabilidade, ela apresenta comportamento constante em torno de sua média, o que torna a configuração original praticamente a mesma se aplicada quaisquer métodos de transformação na série.

PALAVRAS-CHAVE: Métodos, estocástica, previsão.

STOCHASTIC MODELS FOR PREDICTING RELATIVE HUMIDITY FOR BOA VISTA-RR

ABSTRACT: In this chapter, the process of determining the order of the ARMA model for the time series of Relative Air Humidity in Boa Vista/RR from 1961 to 2014 was presented. As the series was already stationary, it was

not necessary to apply the first simple difference to make it stationary. The order of the ARMA model was determined from the PACF and ACF graphs, resulting in the order (1,0,2). The time series shows seasonal fluctuations with peaks in the summer, extremes in 1997 due to El Niño, and some outliers in years with lower variability of the Relative Air Humidity of the region. Although the series has variability, it shows a constant behavior around its mean, which makes the original configuration practically the same if any transformation methods are applied to the series.

KEYWORDS: Methods, stochastic, forecast.

INTRODUÇÃO

Boa Vista é um município brasileiro e capital do estado de Roraima, Região Norte do país, sendo a capital mais setentrional (a norte) do Brasil e situada no Hemisfério Norte.

Neste estudo, apresentamos uma análise estocástica da cidade de Boa Vista-RR para reduzir o erro de projeção em séries cronológicas ao longo do tempo. Cada situação possui um conjunto único de problemas que requerem uma previsão específica, e a solução para um caso pode não ser necessariamente a melhor para outras situações. No entanto, certos princípios gerais são comuns à maioria dos problemas e devem ser incorporados em qualquer sistema de previsão.

Embora a mesma tenha variabilidade, esta também possui comportamento constante em torno de sua respectiva média, fazendo com que a configuração original desta seja praticamente a mesma se aplicada quaisquer métodos de transformação na série em questão, por exemplo: logarítimo natural da referida série. Onde é comprovado pelos testes de estacionariedade e ao plotar a série transformada.

Modelos Estocásticos de Previsão

Não é preciso efetivar uma transformação *Box Cox* na série original, em virtude da série em questão verificar-se homocedástica, ou seja, possuir variância constante, resultado este obtido após executar o teste de *Bartlett* de homogeneidade de variâncias, o qual fornece p-valor correspondente a 0.8529, conseqüentemente, é possível observar ao nível de significância de 5% que o p-valor é maior, assim, não se rejeita a hipótese nula de que há variâncias homogêneas.

ARMA

Um dos objetivos dos modelos de *Box & Jenkins* é tornar a série estacionária e para isso é necessário aplicar a primeira diferença simples com o objetivo de “eliminar” a tendência, por isso, é preciso por a série original no padrão de estacionária para ajustar o modelo ARIMA e/ou SARIMA.

Devido à série temporal original em estudo apresentar-se estacionária, como comprovado através dos testes citados, assim sendo, a ordem do modelo Integrado (I, d – diferenciação) é igual à zero, já que não é necessário aplicar a primeira diferenciação

na série arrolada no estudo, uma vez que a mesma é estacionária. Desta forma, foi determinada apenas a ordem do modelo Auto Regressivo (AR, p) e a ordem do modelo de Médias Móveis (MA, q) por meio do gráfico da PACF (Auto Correlação Parcial) e da ACF (Auto Correlação Total), respectivamente. Resultando no modelo estocástico de previsão denominado ARMA, derivado do ARIMA. No caso deste modelo ARMA, foi feita a PACF e a ACF na série original, visto que não foi realizada a primeira diferenciação simples na referida série devido esta se encontrar estacionária, como mostrado a seguinte:

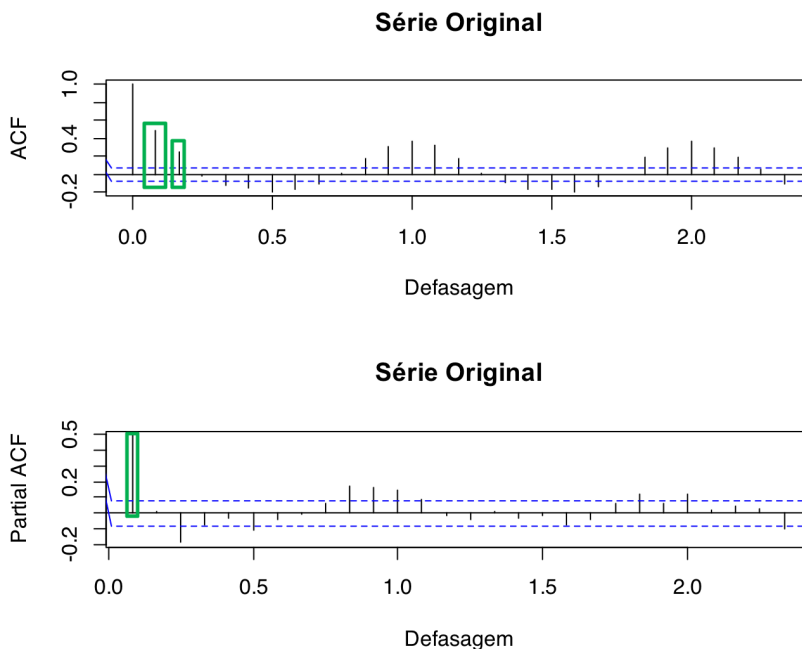


Figura 2 – ACF e PACF da série trabalhada original.

A partir dos gráficos acima se tem que o modelo é mascarado pela sazonalidade, ou seja, o modelo não está ajustado, assim, encontrando-se fora da faixa azul (Intervalo de Confiança) tanto da ACF quanto da PACF, não aleatório, dependente temporalmente e havendo significância na correlação com defasagem na sazonalidade, exceto com lag 0 que apresenta correlação significativa igual a 1 devido correlacionar a variável da série temporal gerada com ela mesma.

Desta forma, sabendo-se também que esta ordem é designada com o objetivo do modelo ser o mais parcimonioso possível, ou melhor, modelo explicado por menos parâmetros (menor ordem), então, para determinar a ordem p, d, q simples com lags sucessivos do modelo ARMA da série em questão foi feita a contagem dos lags mais significativos, isto é, com maior correlação na significância estatística (o que extrapola o intervalo de confiança), sendo que para o modelo auto regressivo (p , correlacionando com a própria variável com defasagem no tempo) visto através da PACF a ordem é 1,

pois é o único lag mais significativo logo após o primeiro lag, e para o modelo de médias móveis (q , correlacionando com o efeito aleatório) é definido aproximadamente no gráfico da ACF ordem 2, dado que também são os lags mais significativos após o lag 0 que tem correlação igual a 1 e também pelo fato da correlação decrescer exponencialmente, a medida que aumenta as defasagens no tempo, assim sendo, os pontos posteriores tem menor significância e, conseqüentemente, menor correlação. Logo, a ordem p , d , q do modelo ARIMA corresponde a $(1,0,2)$, como d (I, modelo Integrado) equivale a 0 o modelo é denominado ARMA.

Ajuste do Modelo ARMA

Em seguida, foi estimado o possível modelo ARMA da referida série com a ordem proposta a anterior. Observando se na ACF os resíduos com lag igual a 1 são estatisticamente iguais ou não a zero, indicando que os resíduos são correlacionados ou não e verificando também por meio da estatística do Teste de independência *Ljung-Box* do modelo que os valores- p são baixos ou elevados, além de mostrar os resíduos padronizados, através do *tsdiag*, cujo gráfico é exposto abaixo:

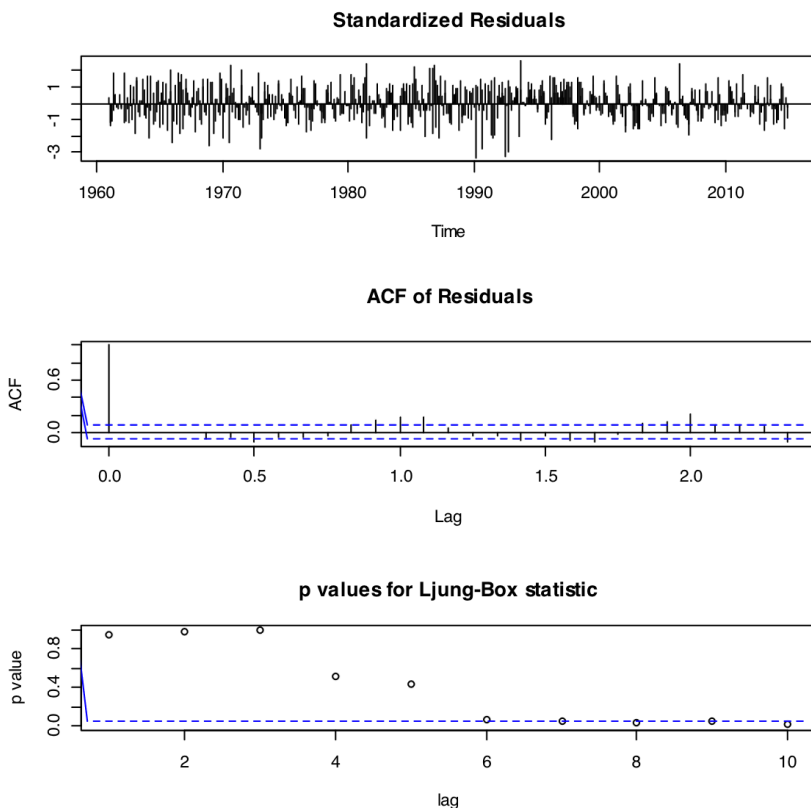


Figura 3 – Resíduos Padronizados, ACF dos resíduos e p-valor para Estatística *Ljung-Box* do modelo ARMA ajustado da série trabalhada original.

Averiguou-se que no modelo ARMA a ACF mostra que grande parte dos resíduos com lag igual a 1 não são estatisticamente iguais a zero, isto significa, são correlacionados por saírem do intervalo de confiança (significância na correlação) e a maioria dos valores da estatística de *Ljung-Box* são muito pequenos, deste modo, os erros são dependentes ao longo do tempo, e o modelo será descartado.

Portanto, o modelo ARMA é o modelo candidato para análise, desde que calcule a significância dos coeficientes do referido modelo.

AR1	MA1	MA2
4.048138e-02	2.537966e-03	1.453419e-06

Tabela 1– Coeficientes do modelo ARMA estimado da série original.

Nota-se que os coeficientes do modelo ARMA são todos significantes, visto que todos os coeficientes do modelo em estudo apresentam valor menor do que o nível de significância de 0,05.

Previsão do Modelo ARMA

Feitas todas as análises anteriores e não sendo violada nenhuma das suposições é possível fazer previsão de um ano para 2015 baseados no modelo ARMA.

Aplicando o *predict* para a previsão do modelo, calculando o intervalo de confiança (IC) com nível de confiança de 95% e criando um *dataframe* das previsões com seus respectivos IC, tem-se que:

	Point Forecast
JAN 2015	72.42288
FEB 2015	73.61124
MAR 2015	75.40857
APR 2015	75.74983
MAY 2015	75.81462
JUN 2015	75.82693
JUL 2015	75.82926
AUG 2015	75.82970
SEP 2015	75.82979
OCT 2015	75.82980
NOV 2015	75.82981
DEC 2015	75.82981

Tabela 2 – Previsão do Modelo ARMA para 2015.

A previsão foi realizada para todo o ano de 2015, a fim de que seja avaliado o desempenho do Modelo ARMA. Para melhor análise do gráfico da previsão a posteriori e como forma de comparação com dados reais é tida a visualização dos valores da previsão, como é exibido na Tabela 4.

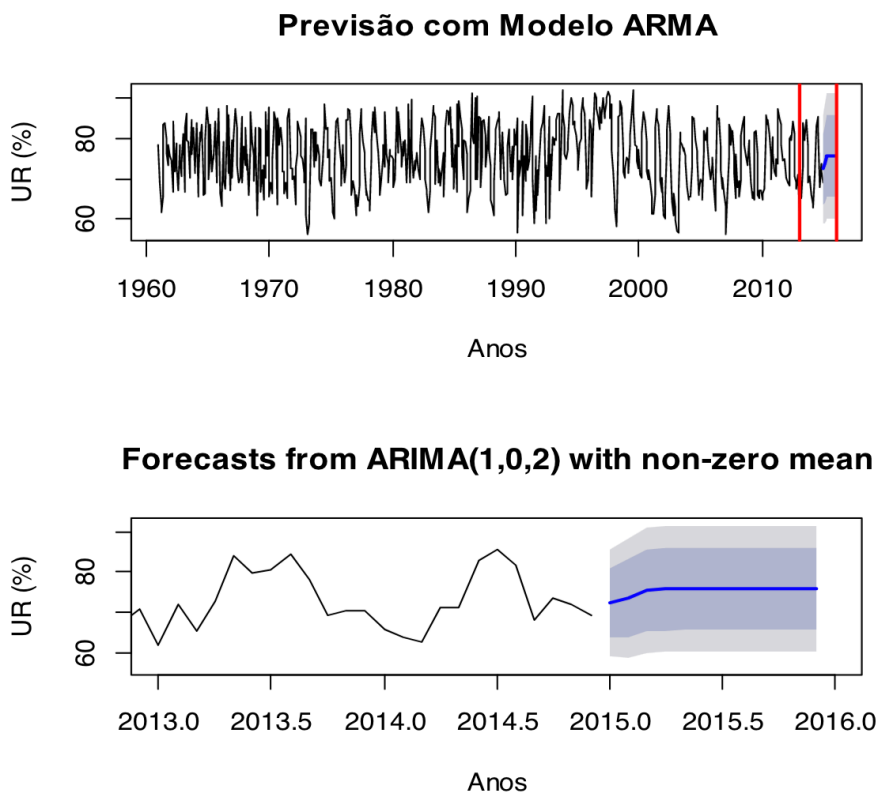


Figura 4 – Gráfico da previsão do Modelo ARMA da série de Umidade Relativa de Boa Vista-RR para 2015.

É notado que este modelo de previsão subestima a umidade relativa do ar da capital Roraimense nos meses correspondentes ao período chuvoso e superestima nos demais meses do ano.

SARIMA

Embora a série temporal original debatida seja estacionária, ainda é possível visualizar a sazonalidade na mesma, desta forma, como o modelo de previsão SARIMA é ajustado com sazonalidade é necessário aplicar a primeira diferenciação sazonal para tornar a série mais aleatória e, assim, definir a ordem do presente modelo. Além do que, foi feita a PACF e a ACF da série com a primeira diferenciação sazonal, como configurado abaixo:

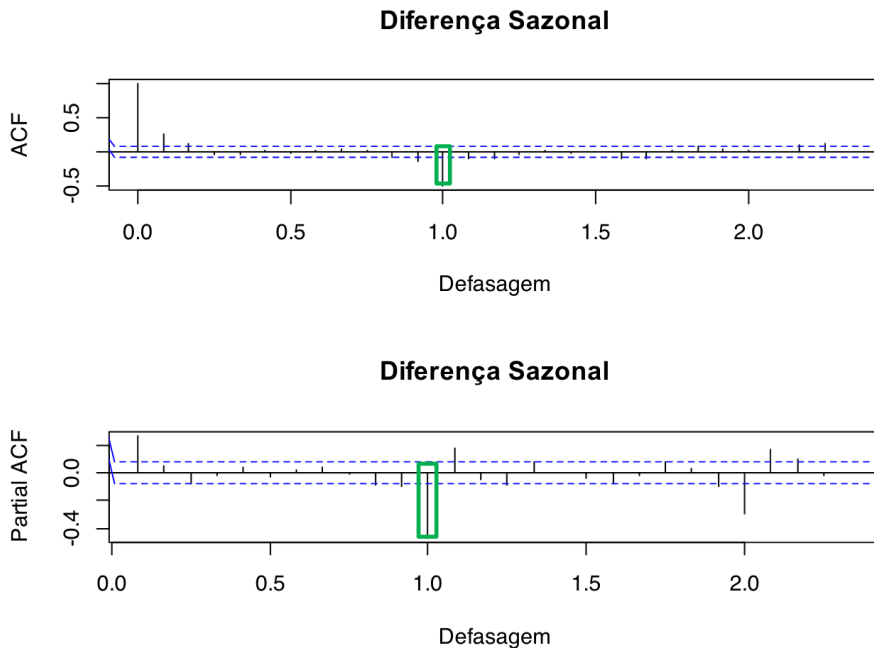


Figura 5 – ACF e PACF da série com diferenciação sazonal.

Com base na Figura 23 tem que o modelo encontra-se exterior a faixa azul (Intervalo de Confiança) no gráfico da ACF e da PACF, de modo que este modelo é dependente por toda extensão de tempo e possui significância na correlação com defasagem sazonal.

Para definir a ordem P, D, Q com lags sazonais do modelo SARMA da série em questão foi obtida a contagem dos lags mais significativos, sendo que para o modelo auto regressivo (P) a ordem é 1, pois o lag sazonal 12 é o mais significativo da defasagem sazonal de 1 a 12, e para o modelo de médias móveis (Q) é definido próximo a ordem 1, dado que também é o lag sazonal 12 mais significativo da sazonalidade de lag 1 a 12 e após o lag 0 que tem correlação igual a 1. Então, a ordem P, D, Q do modelo SARIMA corresponde a (1,1,1), sabendo-se que D (I, modelo Integrado) é correlato a 1, em razão de ter aplicado somente a primeira diferenciação sazonal.

É primordial verificar se a série após a diferenciação sazonal tem padrão de estacionária, para isso é testado esta condição a partir dos tres testes de estacionariedade na mesma série, cujos testes são os de ADF, PP e o KPSS, se teve os resultados a seguir do p-valor: 0.01, 0.01 e 0.1. Assim, pode-se declarar que logo após os testes se tem que a série com diferenciação sazonal é estacionária, uma vez que o p-valor é menor do que o nível de significância (0,05) no teste de ADF e PP, rejeitando a hipótese nula de que a série não é estacionária e o p-valor é maior do que o nível de significância de 5% no teste de KPSS, não rejeitando H_0 , onde esta hipótese diz que a série é estacionária, como comprovado na figura a diante:

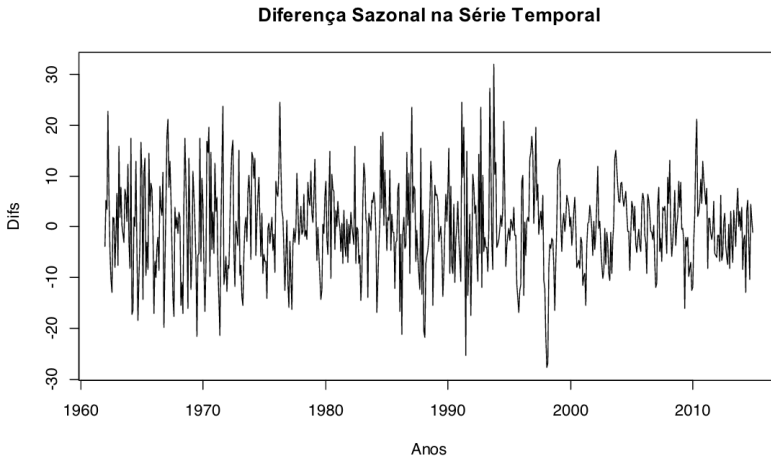


Figura 6 – Série com diferenciação sazonal.

Ajuste do Modelo SARIMA

Neste momento foi ajustado o modelo SARIMA da série original com a ordem designada anteriormente. Observando os resíduos padronizados, a ACF dos resíduos e o comportamento do p-valor do teste *Ljung-Box* de independência dos erros por meio da função *tsdiag*, onde é gerado o gráfico inferiormente.

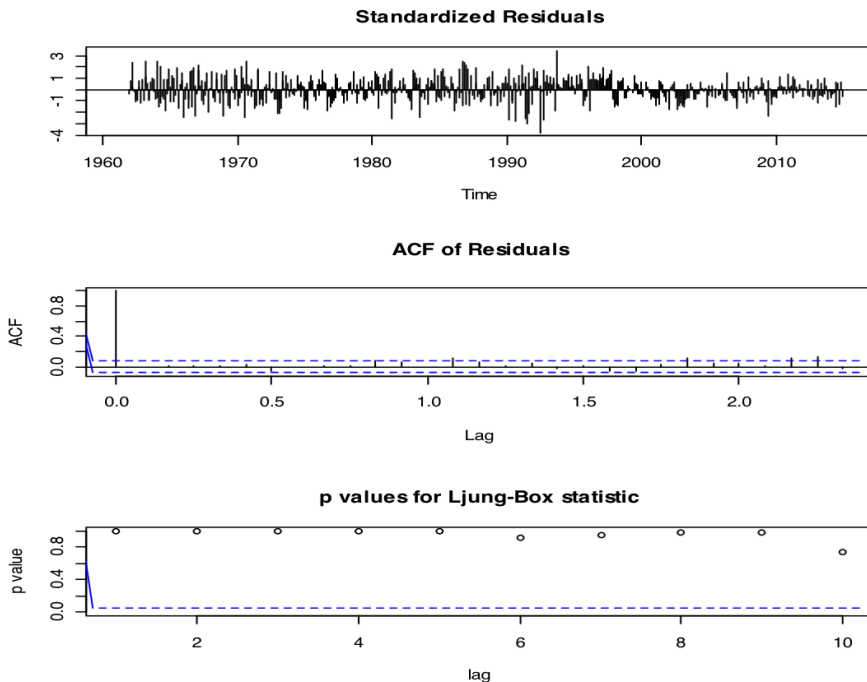


Figura 7 – Resíduos Padronizados, ACF dos resíduos e p-valor do teste de *Ljung-Box* do modelo SARIMA ajustado da série trabalhada original.

Constatou-se que a ACF dos resíduos do modelo SARIMA demonstra que os resíduos com lag igual a 1 não são estatisticamente iguais a zero, isto é, são correlacionados por não ultrapassarem o intervalo de confiança e os valores do teste de *Ljung-Box* são elevados, por conseguinte, os erros são independentes ao longo do tempo, onde será melhor especificado na análise dos resíduos, posteriormente.

O modelo SARIMA é candidato para análise, sendo que é preciso calcular a significância dos coeficientes do modelo investigado, como apresentado na Tabela 5:

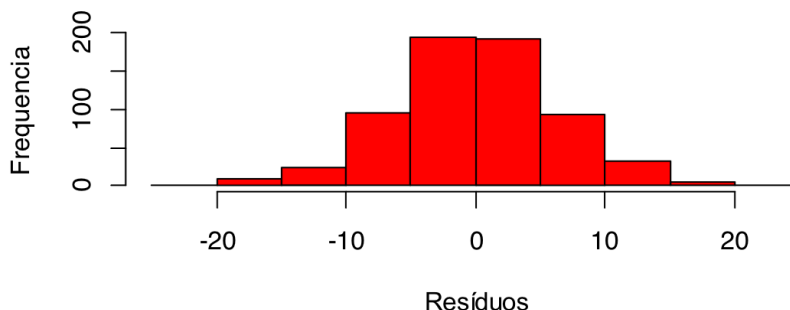
AR1	MA1	MA2	SAR1	SMA1
2.115215e-01	1.282862e-01	2.261688e-03	3.558653e-01	2.638393e-140

Tabela 3 – Coeficientes do modelo SARIMA estimado da série original.

Compreende-se que a maioria dos coeficientes do modelo SARIMA não são significantes, visto que mais da metade dos coeficientes do modelo averiguado possuem valor maior do que o nível de significância de 5%.

Análise dos Resíduos

Histograma dos Resíduos



Normal Q-Q Plot

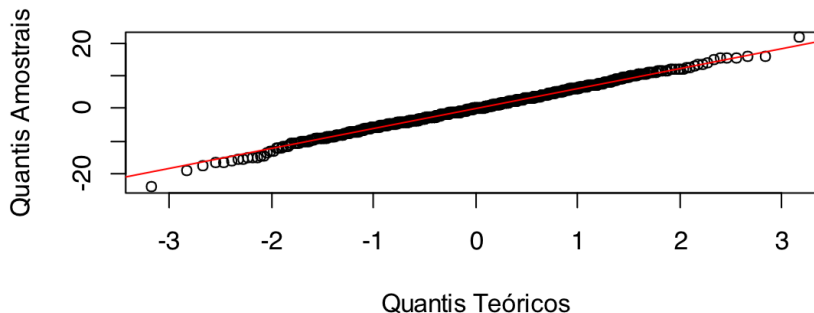


Figura 8 – Histograma e Normal Q-Q Plot dos resíduos do Modelo SARIMA.

A Figura 26 evidencia graficamente que os resíduos seguem uma distribuição normal por consequência da simetria da distribuição dos erros, como revelado no Histograma, além disso, pelo motivo de uma reta se ajustar bem a todos os erros, porém se irá aplicar o teste de *Shapiro-Wilk* para confirmar. Onde se verifica ao nível de significância de 0,05 que não se rejeita a hipótese nula de que os erros seguem uma distribuição normal, pois o p-valor é maior do que o nível de significância citado ($0.5127 > 0,05$), quer dizer, os resíduos vêm de uma distribuição normal, comprovando a análise gráfica.

Desta forma, a média dos erros é similar à zero (-0.05) e a variância dos resíduos do modelo desenvolvido não é homogênea, portanto, logo depois de submeter ao teste de *Bartlett* resultou-se no p-valor de $2.2e-16$, com isso, repara-se que ao nível de significância de 5% o p-valor é menor, isto posto, se rejeita a hipótese nula de que há variâncias homogêneas, desse jeito, não possui variância constante.

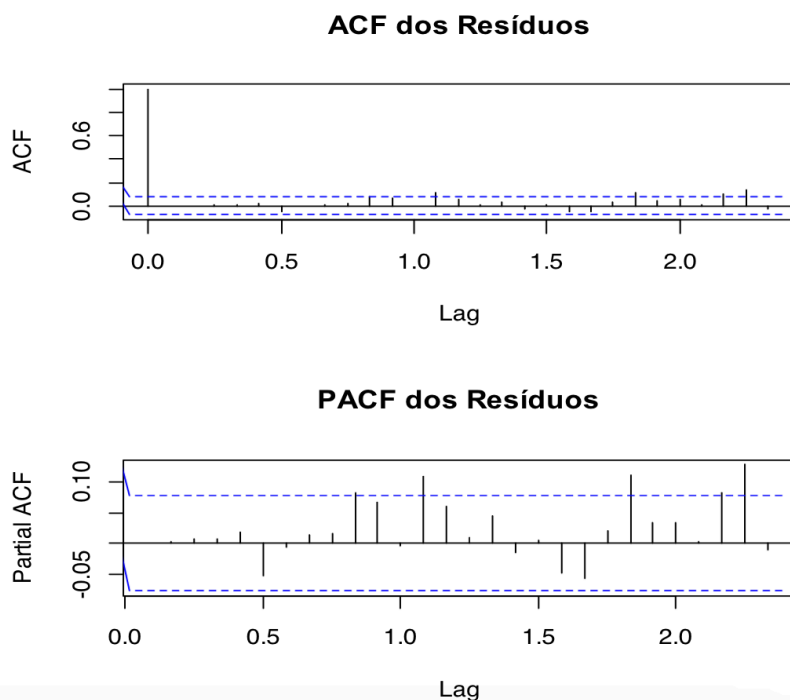


Figura 9 – ACF e PACF dos resíduos do Modelo SARIMA.

Baseando-se na Figura 27 é correto afirmar que a maioria dos resíduos encontra-se dentro da faixa azul (Intervalo de Confiança) na ACF e na PACF, embora haja lags que ultrapassem o intervalo de confiança, estes atingem uma correlação muito pequena, sendo estes irrelevantes, conseqüentemente, não havendo significância na correlação com defasagem igual a 1 na sazonalidade, que por sua vez são estatisticamente iguais a zero, isto quer dizer que os erros não são correlacionados e o teste de *Ljung-Box* concede

p-valor proporcional a 0.2081, então, é maior do que o nível de significância de 5% e há evidências suficientes para não rejeitar a hipótese nula que os erros são independentes e identicamente distribuídos, logo, os mesmos são independentes temporalmente.

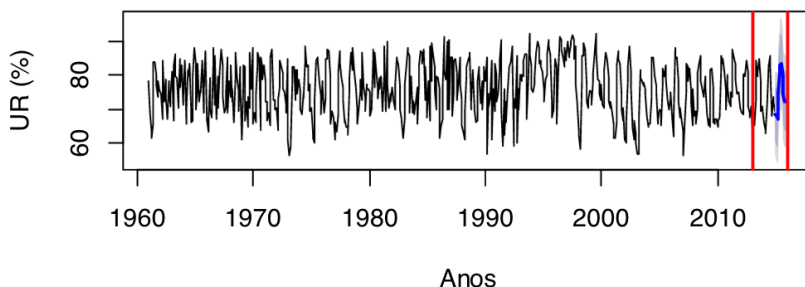
Previsão do Modelo SARIMA

	Point Forecast
JAN 2015	68.45529
FEB 2015	67.78420
MAR 2015	67.25075
APR 2015	74.94270
MAY 2015	81.19612
JUN 2015	82.86798
JUL 2015	83.46412
AUG 2015	80.23754
SEP 2015	74.39992
OCT 2015	72.20200
NOV 2015	72.58995
DEC 2015	72.15019

Tabela 4 – Previsão do Modelo SARIMA para 2015.

A previsão também foi feita para todo o ano de 2015, para que seja validado o desempenho do Modelo SARIMA. Os valores da previsão mensal para 2015 são vistos na Tabela 6.

Previsão com Modelo SARIMA



Forecasts from ARIMA(1,0,2)(1,1,1)[12]

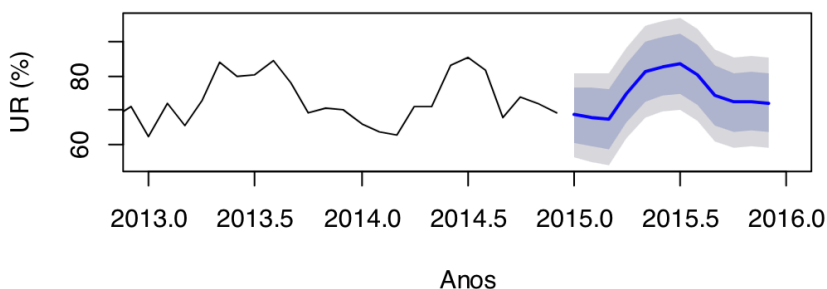


Figura 10 – Gráfico da previsão para 2015 do Modelo SARIMA da série de Umidade Relativa de Boa Vista-RR.

Neste modelo de previsão subestima a umidade relativa de Boa Vista-RR na maior parte do ano, exceto nos meses de Maio, Junho, Julho e Agosto que equivalem aos meses da quadra chuvosa na cidade, onde neste mesmo período superestima ou estima proporcional ao dado observado em determinado mês.

	ME	RMSE	MAE	MPE	MAPE	MASE	AIC
ARMA	-0.0033	6.529988	5.305245	-0.7870	7.140206	0.8661656	4334.9
SARIMA	-0.0552	6.25744	4.92734	-0.7682	6.63293	0.804467	4183.9

Tabela 7 – Medidas de Diagnóstico dos Modelos ARMA e SARIMA.

Tendo em vista a análise dos resíduos e os coeficientes de cada modelo e a Tabela 7 pode-se atestar que o melhor modelo de previsão é o SARIMA, embora não tenha significância em todos os coeficientes do modelo, é o modelo que está dentro de quase todos os pressupostos de normalidade, apenas não possui variância constante, este também apresenta menor AIC e se aproxima mais do erro igual a 0 em quase todas as medidas de diagnóstico, exceto no Erro Médio.