

Análise de dados

Autor:

Adriana Eva Fernandes da Silva

Capítulo 1

Análise

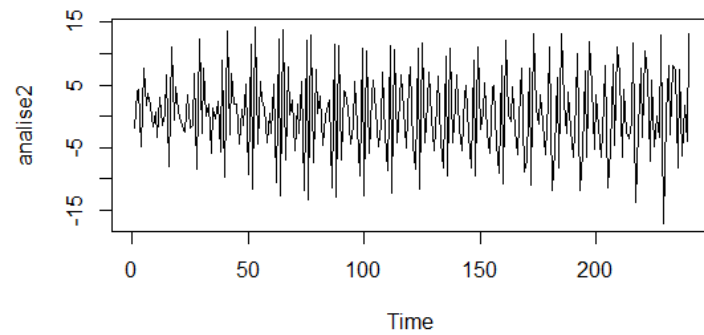


Figura 1.1: Gráfico do tsplot.

Percebe-se que a série não possui tendência mas há um padrão se repetindo o que pode ser uma sazonalidade. Primeiramente, vamos analisar a variabilidade das série.

1.1 Heterocedasticidade

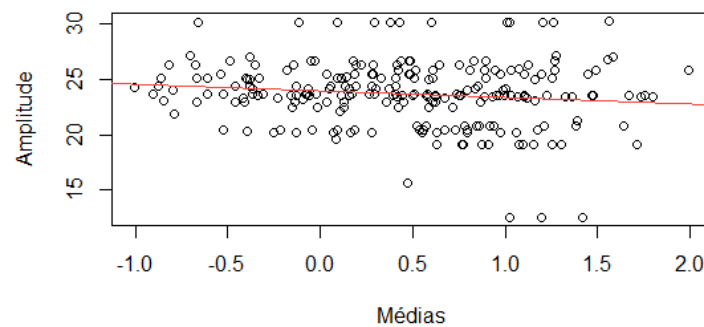


Figura 1.2: Gráfico de amplitude e média para a variabilidade da série.

A partir dos resultados obtidos no **R**, temos um R-quadrado de apenas 0.01244 e que os pontos no gráfico da figura 1.2, estão dispersos com uma reta insignificativa.

1.2 Tendência

Para ver se realmente a série tem tendência, vamos utilizar a função `adfTest`

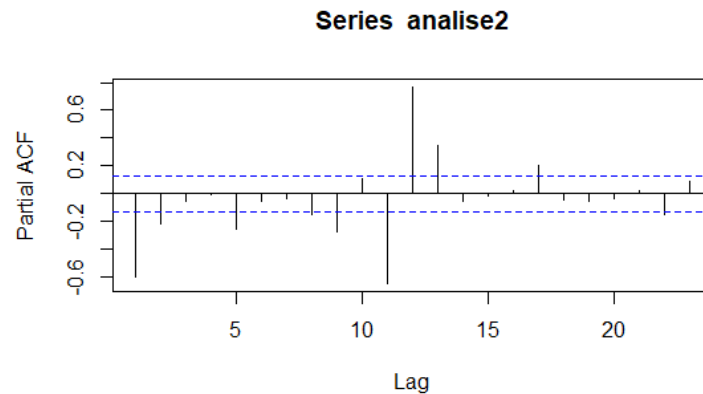


Figura 1.3: Autocorrelação parcial da série

Observando a figura 1.3 notamos que a o último lag que ficou acima do intervalo de confiança é o lag 17, então vamos utilizar ele para a função do Rstudio

Comandos e funções no R:

```
1 >adfTest( analise2 , lags=17)
2 #####SAIDA#####
3 Title :
4 Augmented Dickey–Fuller Test
5
6 Test Results :
7 PARAMETER:
8 Lag Order: 17
9 STATISTIC:
10 Dickey–Fuller : -1.6592
11 P VALUE:
12 0.09396
```

Listing 1.1: Código fonte em R

Após fazer o teste, como o p-valor resultou em 0,09396, maior que 0,05 então concluímos que há tendência na série. Vamos agora tratar dessa tendência.

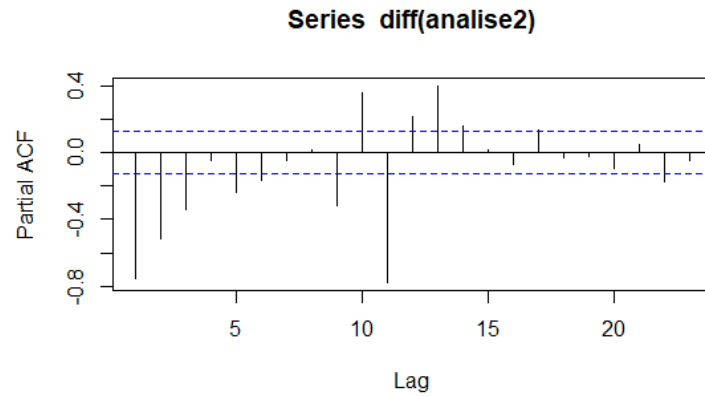


Figura 1.4: Gráfico com uma transformação para retirada da tendência

Analisando a 1.4 notamos que o maior lag significativo é no lag 1 então vamos refazer o teste para ver se conseguimos tratar da tendência.

Comandos e funções no R:

```

1 >adfTest(diff(analise2 ,lags=1))
2 #####SAIDA#####
3 Title :
4 Augmented Dickey–Fuller Test
5 Test Results:
6 PARAMETER:
7 Lag Order: 1
8 STATISTIC:
9 Dickey–Fuller : -25.9801
10 P VALUE:
11 0.01
12 Description:
13 Mon Nov 22 07:24:09 2021 by user: adria

```

Listing 1.2: Código fonte em R

Como o p-valor deu menor que 0,01 então conseguimos tratar da tendência.

1.3 Sazonalidade

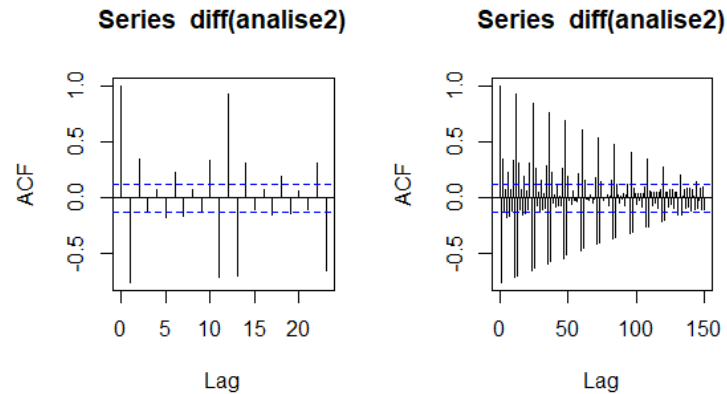


Figura 1.5: Gráficos sem tendência para a sazonalidade

Após tratar a tendência notamos uma sazonalidade não evidente na figura 1.5. Mas se colocarmos maiores lags, temos uma sazonalidade bem evidente.

Assim, foi feita uma diferença de período 12 (que é a frequência em que a sazonalidade ocorre) na série para tratarmos esse padrão de repetição.

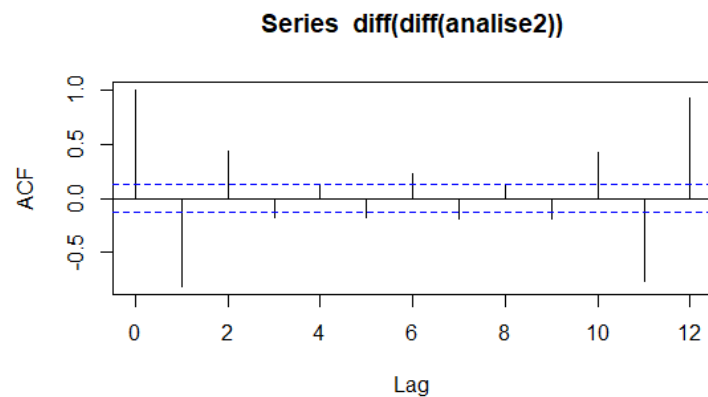


Figura 1.6: Gráfico com duas transformações

A figura 1.6 nos mostra a sazonalidade foi tratada ou seja, a série ficou estacionária e que os picos que ultrapassam o intervalo de confiança serão tratadas no ajuste do modelo,

1.4 Identificação de modelo

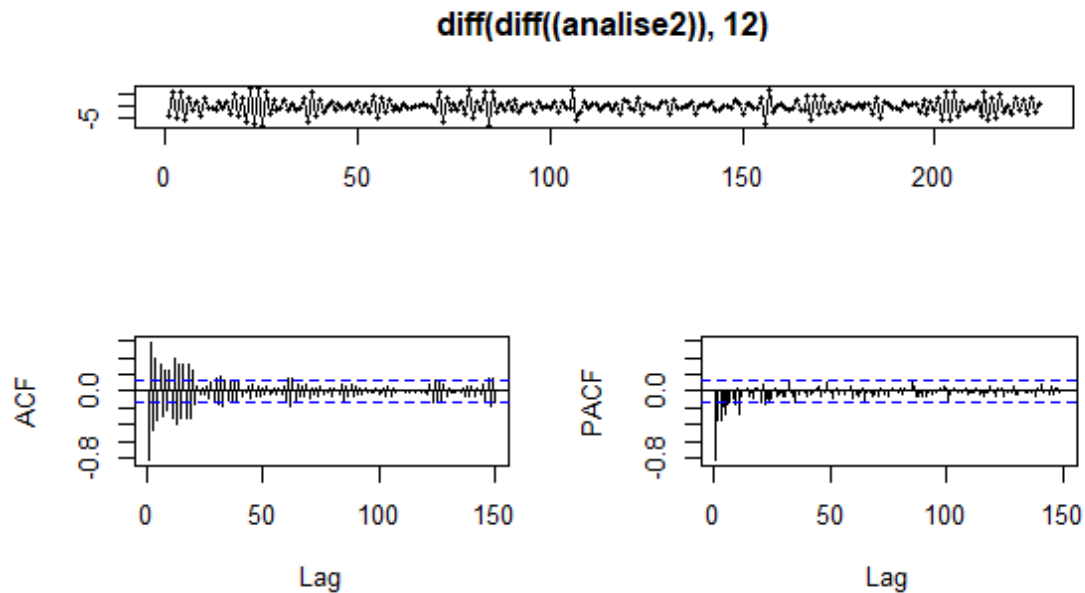


Figura 1.7: Gráficos da série sem sazonalidade e tendência, da função de autocorrelação e autocorrelação parcial

No gráfico da FAC percebemos que os valores para a função de autocorrelação decaem para zero, isso indica o modelo Médias Móveis (MA). Os valores da função de autocorrelação parcial decaem abruptamente, porém para a FACP isso indica um modelo AR. Percebemos também que temos que tratar dos lags significativos, vamos colocar um diferença sazonal de ordem 2 para conseguir tratar desses lags.

1.5 Estimação

Comandos e funções no R:

```
1 > sarima(analise2, 1, 1, 2, 1, 2, 1, 12) ### melhor modelo
2 #####SAIDA#####
3 Coefficients:
4           ar1           ma1           ma2           sar1           smal           constant
5      -0.7334    -0.7465    -0.2532     0.1944    -0.8671         0.0028
6 s.e.    0.0830     0.1179     0.1159     0.0901     0.0736     20.7469
7
8 sigma^2 estimated as 1.945:  log likelihood = -391.27,  aic = 796.54
```

Listing 1.3: Código fonte em R

O modelo **SARIMA(1,1,2)(0,1,1)**, nos retornou um único componente sazonal auto-regressivo estimado com um p-valor significativo, e parâmetro autorregressivo e médias móveis significativos já que, em módulo, as estimativas são maiores que os erros padrões.

1.6 Diagnóstico

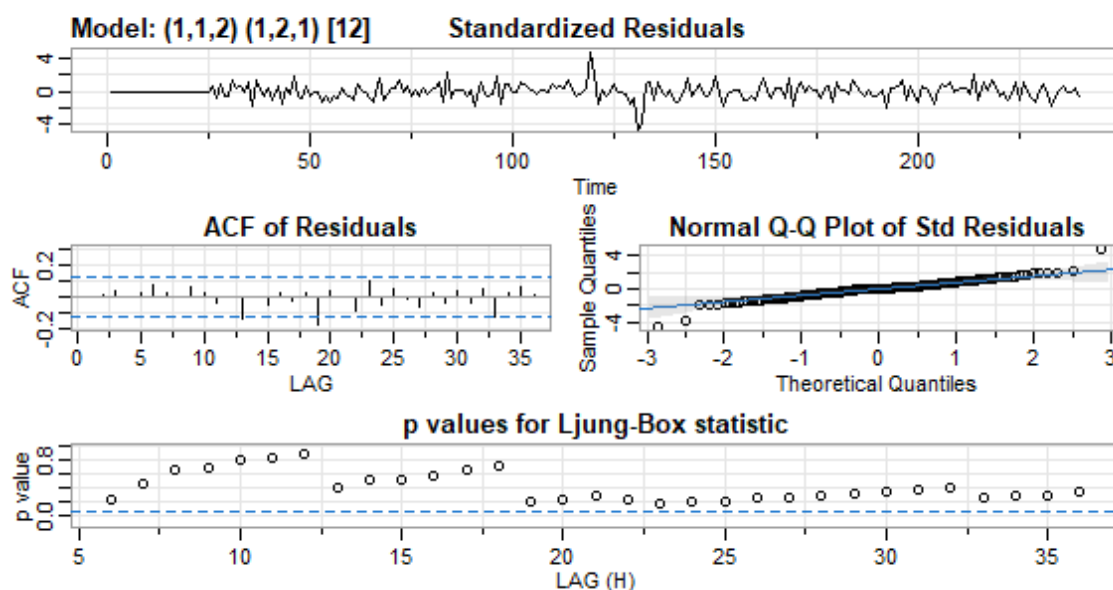


Figura 1.8: Gráfico de diagnóstico para os resíduos do modelo ajustado da série

Analisando a figura 1.9 no gráfico dos resíduos padronizados observamos que não há um padrão nos valores, ou seja, são valores aleatórios. No gráfico da função de autocorrelação dos resíduos podemos notar que, todos os valores para essa função estão dentro do intervalo de confiança, os resíduos tem distribuição aproximadamente Normal então podemos dizer que os resíduos são Ruído Branco, o que é confirmado pelo gráfico seguinte, os valores de p-valor dão maiores que 5% então não rejeitamos a hipótese nula de que o resíduo é Ruído Branco.

1.7 Detecção de outliers

Utilizando o software **R**, verificaremos se há outliers na série, e caso haja, fazer a incorporação deles no modelo.

Comandos e funções no R:

```
1 m1=arimax(analise2, order=c(1,1,2), seasonal=list(order=c(0,1,1),
2                                     period=12))
3 m1
4
5 Call:
6
7 m7=arimax(analise2, order=c(1,1,1), seasonal=list(order=c(1,2,1), period=12))
8 m7
9 Coefficients:
```

```

10 Coefficients :
11           ar1           ma1           sar1           sma1
12      -0.5610      -0.9997      0.1692      -0.8354
13 s.e.      0.0578      0.0240      0.0922      0.0709
14
15 sigma^2 estimated as 2:  log likelihood = -393.49,  aic = 794.97
16 > detectAO(m7)
17 [1] "No AO detected"
18 > detectAO(m7, robust=F)
19 [1] "No AO detected"
20 > detectIO(m7)
21           [,1]           [,2]           [,3]
22 ind      119.000000  131.000000  132.000000
23 lambda1      5.929996      -5.765493      -5.290666

```

Listing 1.4: Código fonte em R

Observando os resultados obtidos no **R**, temos dois outliers na série, etnã iremos incorporá-los:

Comandos e funções no R:

```

1 > m8=arimax(analise2, order=c(1,1,1), seasonal=list(order=c(1,2,1), period=12)
  , io=c(119,131,132))
2
3 > detectAO(m8)
4 [1] "No AO detected"
5 > detectAO(m8, robust=F)
6 [1] "No AO detected"
7 > detectIO(m8)
8 [1] "No IO detected"

```

Listing 1.5: Código fonte em R

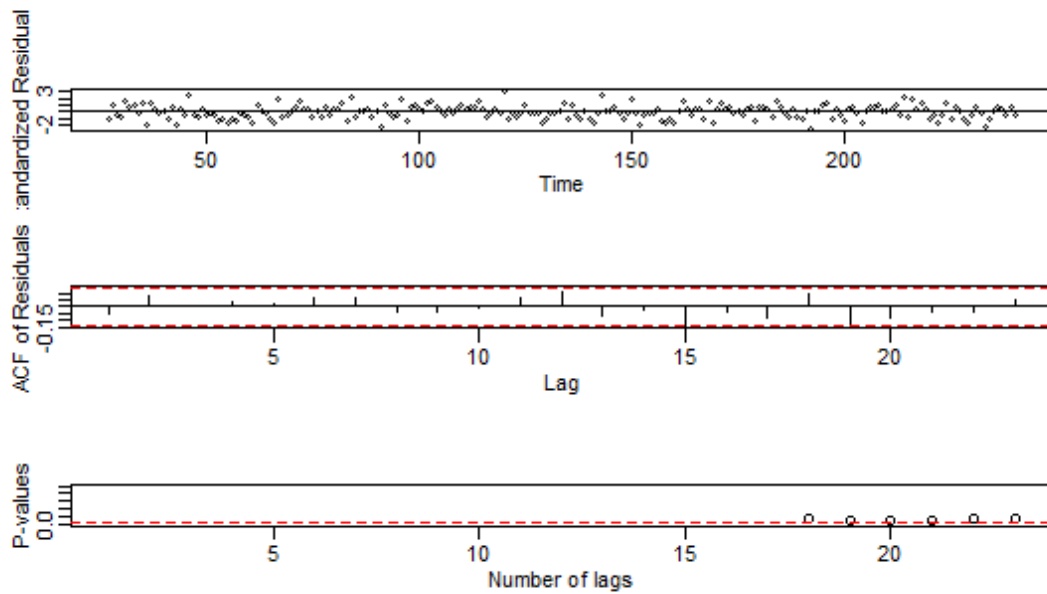


Figura 1.9: Gráfico de diagnóstico para os resíduos do modelo ajustado da série

Sendo assim, temos o melhor modelo ajustado $\text{ARIMA}(1,1,1)(1,2,1)[12]$ com os outliers incorporados

1.8 Previsão pelo método de alisamento exponencial

Para o presente trabalho iremos utilizar o alisamento exponencial de Holt-Winters com sazonalidade multiplicativa

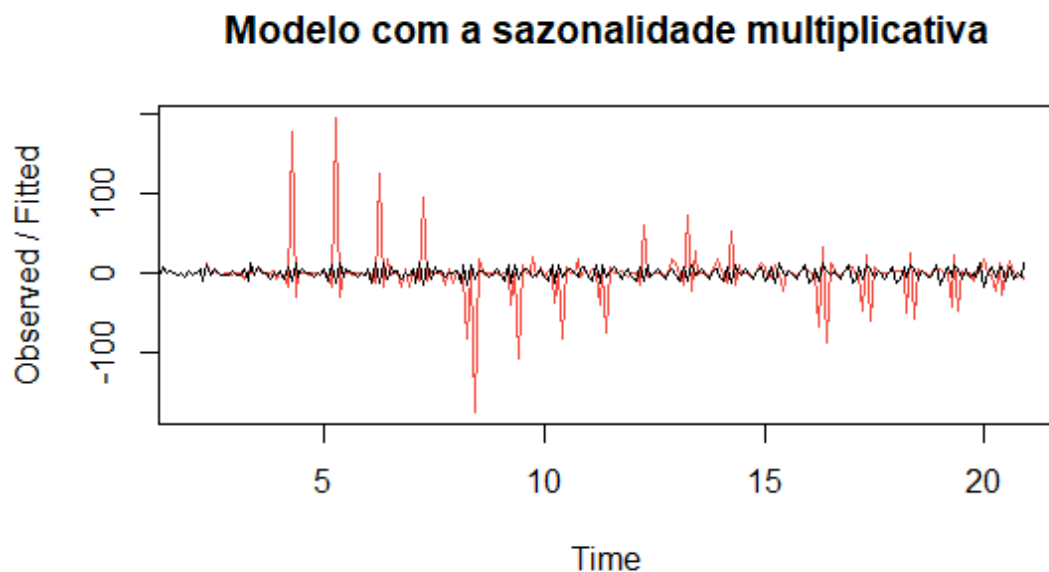


Figura 1.10: Gráfico para a previsão do modelo com a sazonalidade multiplicativa.

Agora que já sabemos o método de alisamento exponencial de Holt-Winters mais adequado para a série, iremos fazer previsões. Para isso o R calculou os melhores valores para os coeficientes de alfa, beta, gama, pois eles minimizam os erros dos valores previstos.

Comandos e funções no R:

```
1 > m <- HoltWinters(serie , seasonal = "mult")
2 Holt-Winters exponential smoothing with trend and multiplicative seasonal
  component .
3
4 Call:
5 HoltWinters(x = serie , seasonal = "mult")
6
7 Smoothing parameters:
8   alpha: 0.3
9   beta  : 0.1
10  gamma: 0.1
```

Listing 1.6: Código fonte em R

Comandos e funções no R:

```
1 > predict(m, 12)
2           Jan           Feb           Mar           Apr           May           Jun
3 21  0.4363264  0.1549393 -0.2442984  2.2033290 -2.0484473  4.4486165
4           Jul           Aug           Sep           Oct           Nov           Dec
5 21  0.4015429 -1.6935900 -0.3541495 -0.7461742  0.4485310 29.7000840
```

Listing 1.7: Código fonte em R

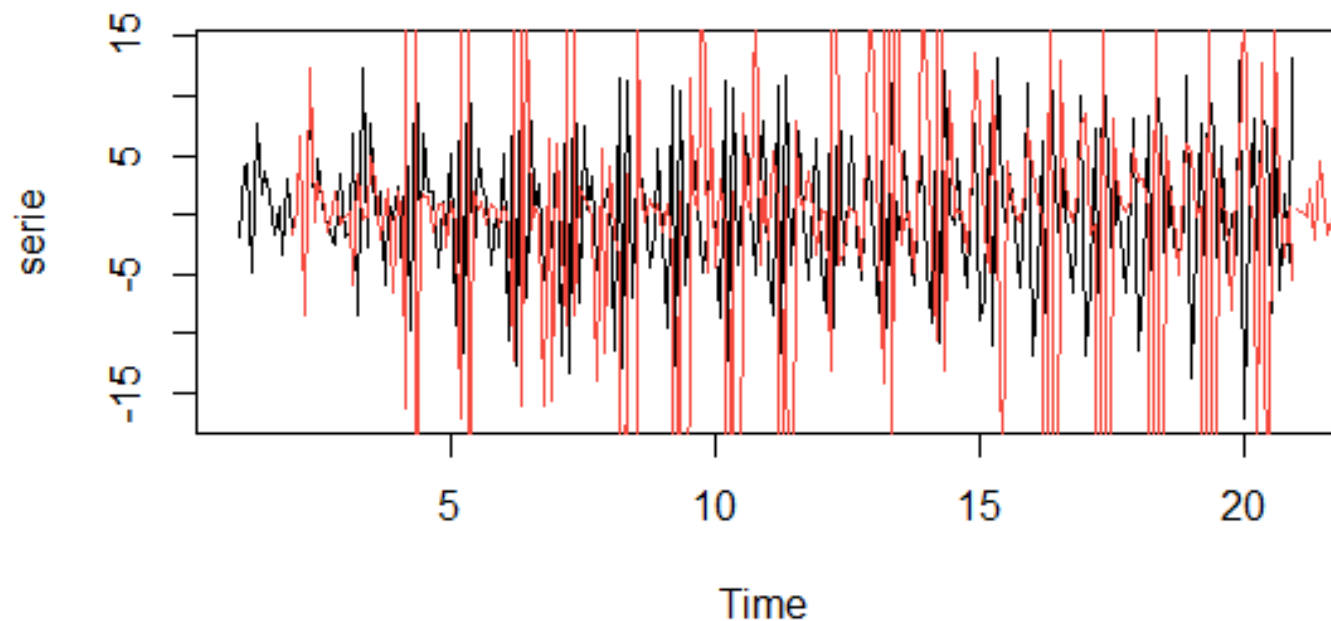


Figura 1.11: Gráfico da previsão segundo o método de Holt Winters com sazonalidade multiplicativa

Analisando o gráfico da figura 1.11 e as funções em R, a previsão é a série, fique com menos picos ao passar do tempo. Neste caso, a série vai se tornando menos sazonal. Onde os valores negativos e os valores positivos não tem tanta amplitude entre si.

1.9 Previsão pelo método ARIMA

Para o ajuste de modelo foi obtido um modelo **SARIMA(1,1,2)(0,1,1)[12]**. Para fazer a previsão o ARIMA utiliza dados passados para prever valores futuros, usando principalmente a autocorrelação e médias móveis.

Comandos e funções no R:

```
1 > sarima.for(analise2,12,1,1,2,0,1,1,12)###melhor modelo
2 $pred
3 Time Series:
4 Start = 241
5 End = 252
6 Frequency = 1
7 [1] -17.9191876  0.7318043  8.0233224 -2.9720701  7.3442365
   7.6174847
```

```

8 [7] -8.4409039  7.7559403 -6.5618095  2.0409987 -4.0614659
9      13.1098923
10 $se
11 Time Series:
12 Start = 241
13 End = 252
14 Frequency = 1
15 [1] 1.340289 1.475367 1.551423 1.592630 1.617604 1.631092 1.639700
16      1.644156
      [9] 1.647201 1.648650 1.649755 1.650208

```

Listing 1.8: Código fonte em R

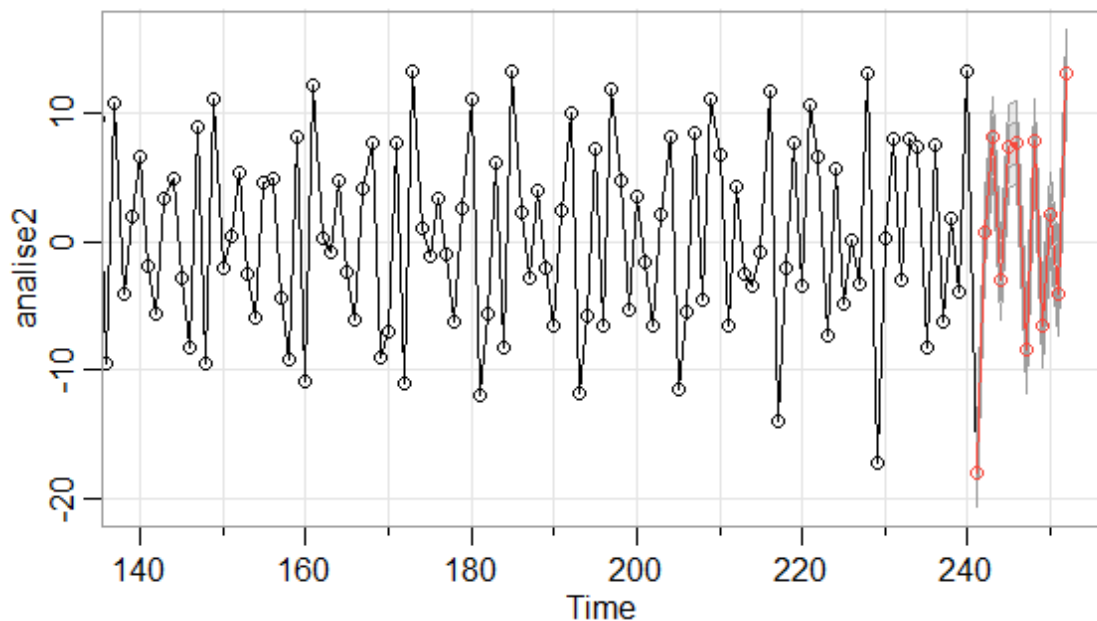


Figura 1.12: Gráfico da série original e previsão segundo o modelo SARIMA(1,1,2)(0,1,1)[12].

Analisando o gráfico da figura 1.12 e os códigos em Rstudio, a previsão é que a série siga o mesmo formato da série original pois a linha em vermelha está bem próxima do intervalo de previsão.

1.10 Comparação entre as previsões

Para conseguirmos comparar as previsões feitas pelo método Holt-Winters e pelo método Sarima, iremos fazer a previsão para uma parte da série correspondente ao ano anterior da série, para ser possível comparar com o valor real desse período e assim conseguir obter a acurácia de ambos.

Comandos e funções no R:

```

1 ##Acuracia metodo Holt-Winters
2 > accuracy(prev1, serie[229:240])
3           ME      RMSE      MAE      MPE      MAPE
4 Test set -2.110559  9.160658  7.617051  86.51195 107.5741
5
6 ##Acuracia metodo Sarima
7           ME      RMSE      MAE      MPE      MAPE
8 Test set -0.08731167  0.77848  0.7415082 -33.10864  37.18534

```

Listing 1.9: Código fonte em R

Ao comparar as medidas de acurácia ME, RMSE, MAE, MPE e MAPE de ambos os modelos, podemos observar que o melhor modelo de previsão para essa série é o SARIMA que possui todos valores menores para as medidas.