

Universidade Federal de São Carlos
Centro de Ciências Exatas e de Tecnologia
Departamento de Estatística

Análise de Dados de Séries Temporais

Douglas de Paula Nestlehner

Novembro, 2021

Sumário

1	Introdução	2
2	Análise descritiva	3
2.1	Informações	3
2.2	Estacionariedade	4
2.2.1	Variabilidade	4
2.2.2	Verificando Tendência e Sazonalidade	5
2.3	Série diferenciada	7
2.4	Modelo Ajustado	7
2.4.1	Diagnostico do Modelo	8
2.4.2	Casos influentes	9
2.5	Previsões	10
2.5.1	SARIMA	10
2.5.2	Holt-Winters	11
2.5.3	Comparação	12
A	Código	13

Capítulo 1

Introdução

Esse relatório contém uma análise de série temporal, abordando os principais tópicos apresentados na disciplina de Séries Temporais 2021/2. Os dados utilizados na análise contém 240 observações, e não temos a informação de qual tempo essas observações correspondem (dias, semanas, meses, anos, etc.).

Com isso dividi a análise em 3 principais tópicos:

- Análise descritiva da série;
- Modelo ajustado;
- Previsões.

Nas próximas seções apresentarei os resultados obtidos e realizarei alguns comentários com o intuito de esclarecer/explicar as conclusões obtidas.

Capítulo 2

Análise descritiva

Quando analisamos uma série temporal a representação gráfica dos dados é fundamental, com a análise gráfica podemos revelar padrões de comportamento importantes. Nesse capítulo irei apresentar uma análise descritiva dos dados, com o intuito de verificar: tendência; padrões; alterações estruturais; sazonalidade; estacionariedade.

2.1 Informações

Temos disponível que a série a ser analisada contem 240 observações, não temos especificado o período de tempo (dias, semanas, meses, anos) portanto não temos conhecimento da periodicidade da série. Mas plotando o gráfico da série foi possível observar que o padrão sazonal se repetia a cada 12 observações, ou seja, os dados informados são mensais. Assim sendo, defini os dados como série temporal entre o período de janeiro de um ano 1 até dezembro de um ano 20. (não necessariamente esses dados se referem a esse período definido, mas para a condução do estudo achei interessante definir um período qualquer) e com frequência igual a 12, a Figura 2.1 mostra como a série foi definida.

```
x = ts(df2, start = c(1,1), frequency = 12)
```

Figura 2.1: Definindo a série

Depois de definir a série plotei o gráfico representado na Figura 2.2 com o intuito de ter uma visão geral sobre a série.

Observando então a Figura 2.2 já é possível realizar algumas observações interessantes: aparentemente ela tem sazonalidade, tem tendência, e a variância não é constante.

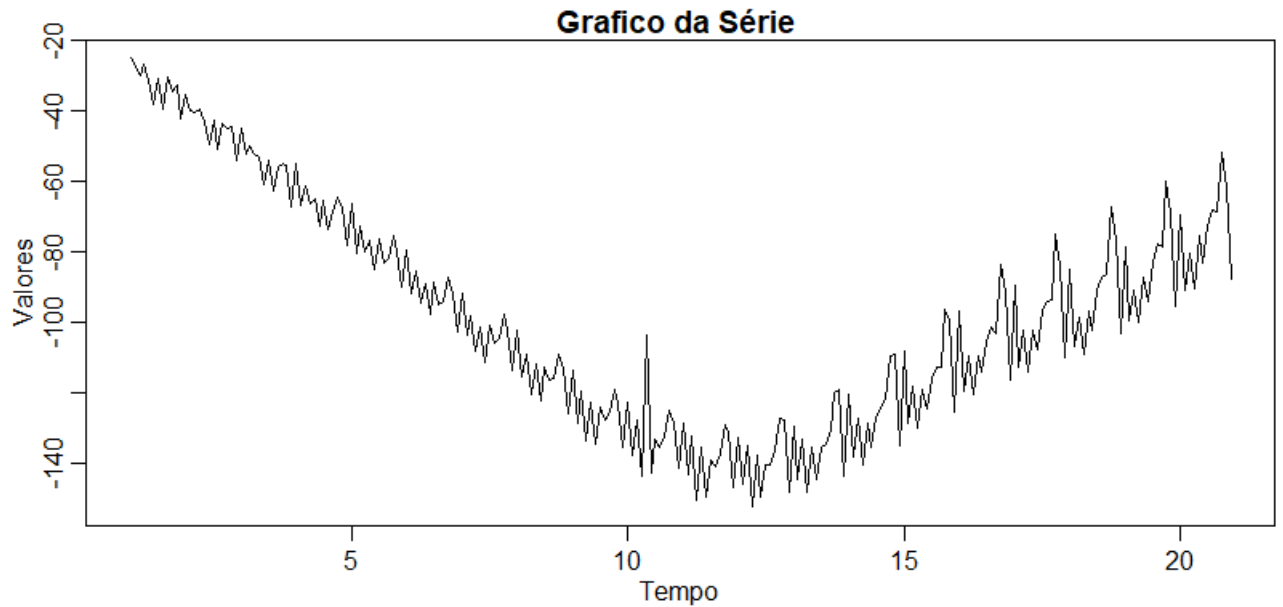


Figura 2.2: Gráfico da série temporal analisada.

Porém não é muito interessante tomar conclusões apenas com análise gráfica da série, para isso podemos realizar outras análises gráficas e testes que serão apresentados nas próximas seções, que vão nos permitir conclusões mais concretas.

2.2 Estacionariedade

Para análises futuras é interessante que a série analisada seja estacionária, ou seja, a série não aumenta ou diminui linearmente ou exponencialmente com o tempo (sem tendências) e não apresenta nenhum tipo de padrão de repetição (sem sazonalidade). Portanto, a seguir irei verificar se a série é estacionária ou não, caso a estacionariedade não for satisfeita, irei "transformá-la" para que seja.

2.2.1 Variabilidade

Com o intuito de verificar se a variância é constante ao longo do tempo, analisei o gráfico da amplitude média representado na Figura 2.3.

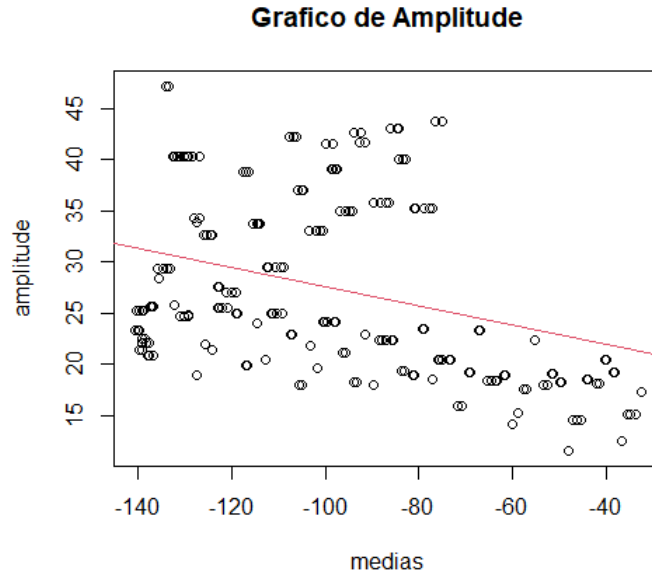


Figura 2.3: Gráfico de amplitude média, considerando um período igual a 12.

Observando que a variabilidade dos pontos decresce ao longo o tempo, e com um Adjusted R-squared = 0.09925, ao nível de significância $\alpha = 0.05$, conclui-se que a variabilidade não é constante.

2.2.2 Verificando Tendencia e Sazonalidade

Para poder verificar se a série tem tendencia e sazonalidade, usei o comando `decompose()` obtendo a Figura 2.4, optei por usar a decomposição multiplicativa, pois observando a série a variação sazonal aumenta ao longo tempo, e nesses caso o indicado é usar a decomposição multiplicativa.

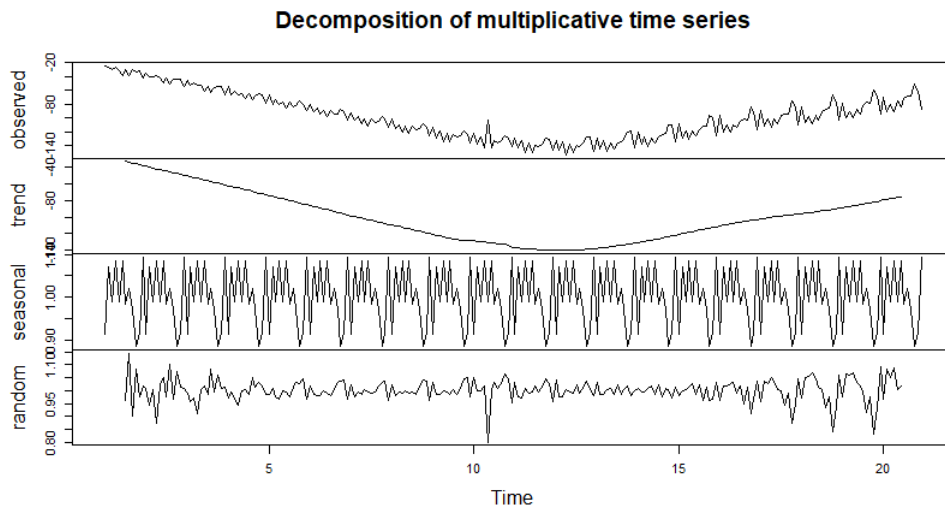


Figura 2.4: Decomposição da série em tendencia - sazonalidade - aleatoriedade

Sendo assim, pode-se observar que aparentemente a série tem tendencia e sazonalidade, indicando então a não estacionariedade. Para poder confirmar esse argumento plotei o ACF e PACF:

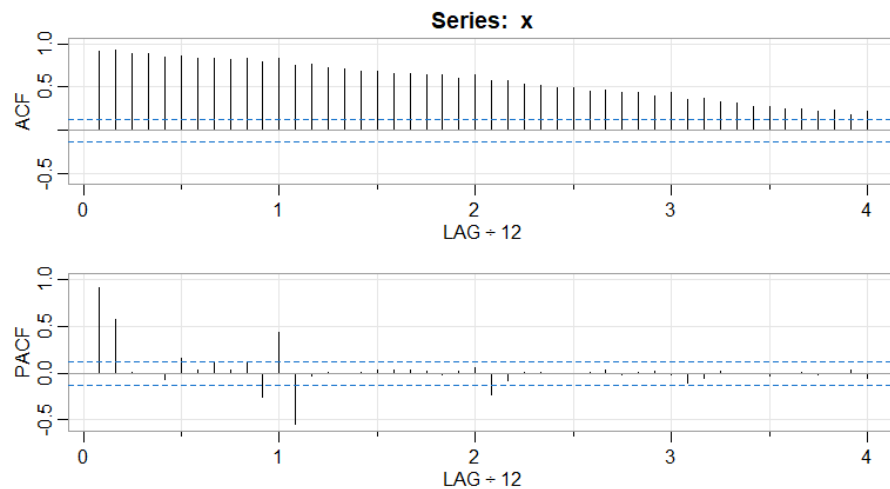


Figura 2.5: ACF e PACF da série

No ACF observa-se que o decaimento das observações (os traços) é lenta e não chega a 0, o que indica a presença de tendencia. Também realizei o teste de Cox-Stuart para pode verificar se realmente a série tem tendencia:

Teste de Cox-Stuart

$$\text{valor-p} = 0.01338$$

Concluindo, que ao nível de significância $\alpha = 0.05$, rejeita H_0 , ou seja, a série tem tendencia.

Também realizei o teste de Dickey-Fuller aumentado `adf.test()`, que é um teste de raiz unitária usado para checar estacionariedade de uma série, utilizei Lag Order = 12, pois observando o PACF verifica-se que o ultimo valor significativo esta no Lag 1 (igual a 12).

Teste de Dickey-Fuller

$$\text{valor-p} = 0.9718$$

Sendo assim, ao nivel de significancia $\alpha = 0.05$ não rejeita-se H_0 , ou seja, a série não é estacionaria.

Obs: Realizei esse teste apenas no intuito de confirmar/reforçar as afirmações anteriores, pois a não estacionariedade já foi provada quando verificado: tendencia, sazonalidade e varaibilidade não constante.

2.3 Série diferenciada

Como a estacionariedade não foi satisfeita na série original, calculei a série diferenciada pelo comando `diff(serie)`. Realizei os testes novamente e a estacionariedade ainda não foi aceita (foi verificado sazonalidade pelo gráfico do `acf`) e no teste de Dickey Fuller $p - valor > \alpha$.

Assim sendo, calculei a série diferenciada duas vezes `diff(diff(serie))`, obtendo o resultado representado na Figura 2.6

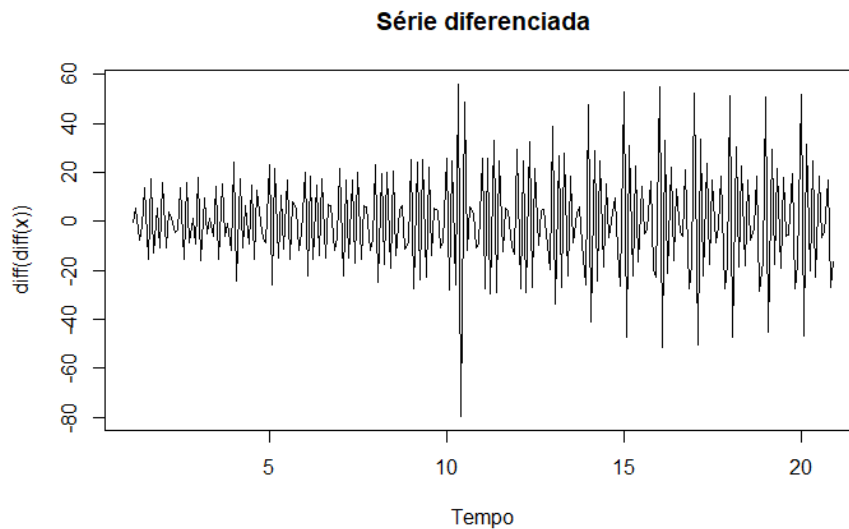


Figura 2.6: Série diferenciada duas vezes

Realizando o teste de Dickey-Fuller obtive:

Teste de Dickey-Fuller

$$\text{valor-p} < 0.01$$

Concluindo que ao nível de significância $\alpha = 0.05$ rejeita-se H_0 , ou seja, a série é estacionária.

Assim sendo, para encontrar estacionariedade foi preciso calcular a série diferenciada duas vezes.

2.4 Modelo Ajustado

Como foi verificado na análise descritiva que a série tem tendência e sazonalidade, o modelo a ser ajustado, trata-se de um modelo SARIMA(p, d, q)(P, D, Q)[12], e que para verificar estacionariedade foi necessário aplicar uma diferenciação para retirar tendência e outra para a sazonalidade, então $p = 1$ e $D = 1$.

Para determinar os demais parâmetros do modelo, analisei os gráficos ACF e PACF da série diferenciada uma e duas vezes, mas estava levando muito tempo para identificar e testar os possíveis parâmetros, então optei por usar a função "auto.arima" disponível no R, que gerou o seguinte modelo:

$$M1 = SARIMA(0, 1, 2)(0, 1, 1)[12]$$

Para poder verificar se o modelo foi bem ajustado analisei os gráfico representados na Figura 2.7

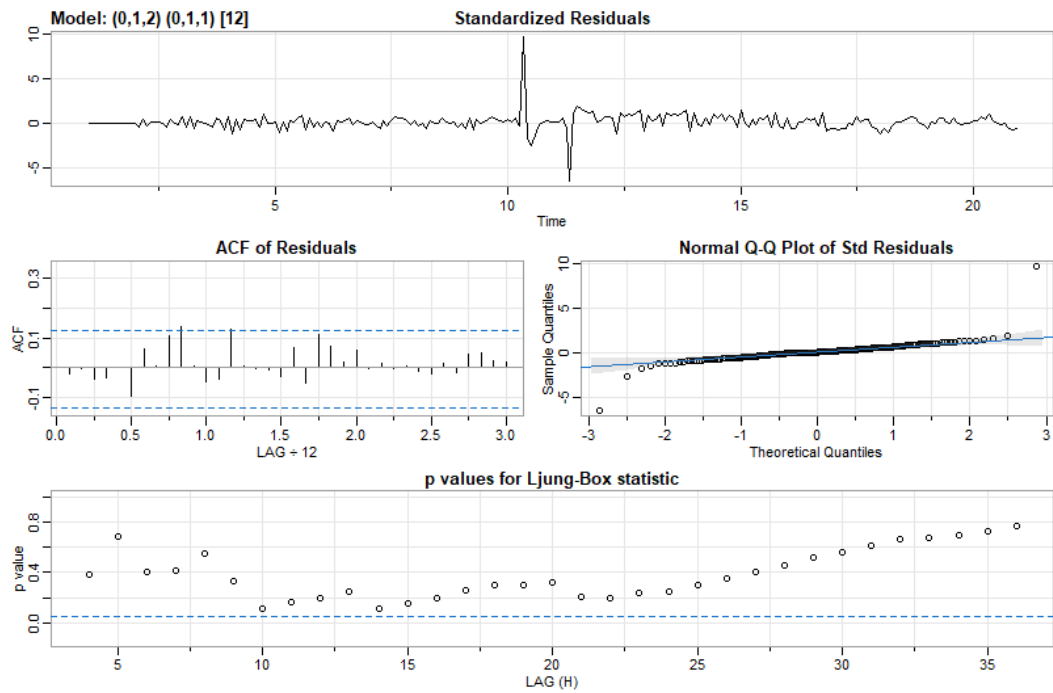


Figura 2.7: Gráfico dos resíduos padronizados, gráfico da função de autocorrelação (resíduos), gráfico Normal - QQ (resíduos) e gráfico de Ljung-Box para o modelo.

Apesar de ser possível observar alguns valores discrepantes nos resíduos (aproximadamente no tempo 11 e 13) é possível dizer que o modelo foi bem ajustado, pois as suposições sobre os resíduos são satisfeitas, no gráfico do ACF as correlações são bem pequenas, e os p-valores da estatística de Ljung-Box estão "comportados", indicando ruído branco.

2.4.1 Diagnostico do Modelo

Para verificar se os parametros do modelo são significativos usei o comando sarima() que retornou o resultado representado na Figura 2.8

```

> arimax(x, order = c(0,1,2), seasonal = list(order = c(0,1,1), period = 12))

Call:
arimax(x = x, order = c(0, 1, 2), seasonal = list(order = c(0, 1, 1), period = 12))

Coefficients:
          ma1          ma2          sma1
      -0.7759    0.0935   -0.2515
s.e.    0.0656    0.0621    0.0618

sigma^2 estimated as 9:  log likelihood = -572.24,  aic = 1150.48

```

Figura 2.8: Diagnostico do modelo.

Como as estimativas dos parâmetros em modulo são maiores que os seus respectivos erros padrões, conclui-se que os parâmetros são significativos.

2.4.2 Casos influentes

Afim de verificar possíveis casos que estivessem interferindo no modelo, verifiquei a existencia de observações atípicas aditivas AO (Additive Outlier) e atípicas de inovação IO (Innovation Outlier), obtendo os resultados representados na Figura 2.9

```

> detectAO(M1)
      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]      [,8]
ind    112.000000 113.000000 126.000000 127.000000 128.000000 129.000000 138.000000 150.0000
lambda2  4.244597  5.733057  5.299416  5.405492  4.693052  3.921883  3.957492  3.7235
> detectIO(M1)
      [,1]      [,2]      [,3]
ind    113.0000 115.000000 125.000000
lambda1 14.4265 -3.849271 -9.644976

```

Figura 2.9: Casos influentes

Como não se tem mais informações sobre a base de dados, optei por manter os casos outliers no modelo.

2.5 Previsões

Em seguida, utilizando o modelo obtido realizei previsões para os próximos 15 meses, utilizando o método de previsão pelo modelo SARIMA e o método de Holt-Winters.

2.5.1 SARIMA

O método de previsão pelo modelo SARIMA obteve os seguintes resultados apresentados na Figura 2.10.

```
> sarima.for(x,15,0,1,2,0,1,1,12)
$pred
      Jan      Feb      Mar      Apr      May      Jun      Jul      Aug      Sep
21 -61.34763 -82.66945 -72.85793 -82.57102 -68.24566 -75.82572 -64.97283 -60.51976 -60.78673
22 -53.23176 -74.51245 -64.70093
      Oct      Nov      Dec
21 -43.49776 -53.25671 -79.55909
22

$se
      Jan      Feb      Mar      Apr      May      Jun      Jul      Aug      Sep      Oct
21 3.000037 3.074434 3.218697 3.356765 3.489374 3.617125 3.740516 3.859964 3.975824 4.088403
22 5.362828 5.556963 5.801324
      Nov      Dec
21 4.197964 4.304738
22
```

Figura 2.10: Predição de 15 meses pelo método SARIMA

Obtendo também esses resultados na forma gráfica, representado na Figura 2.11.

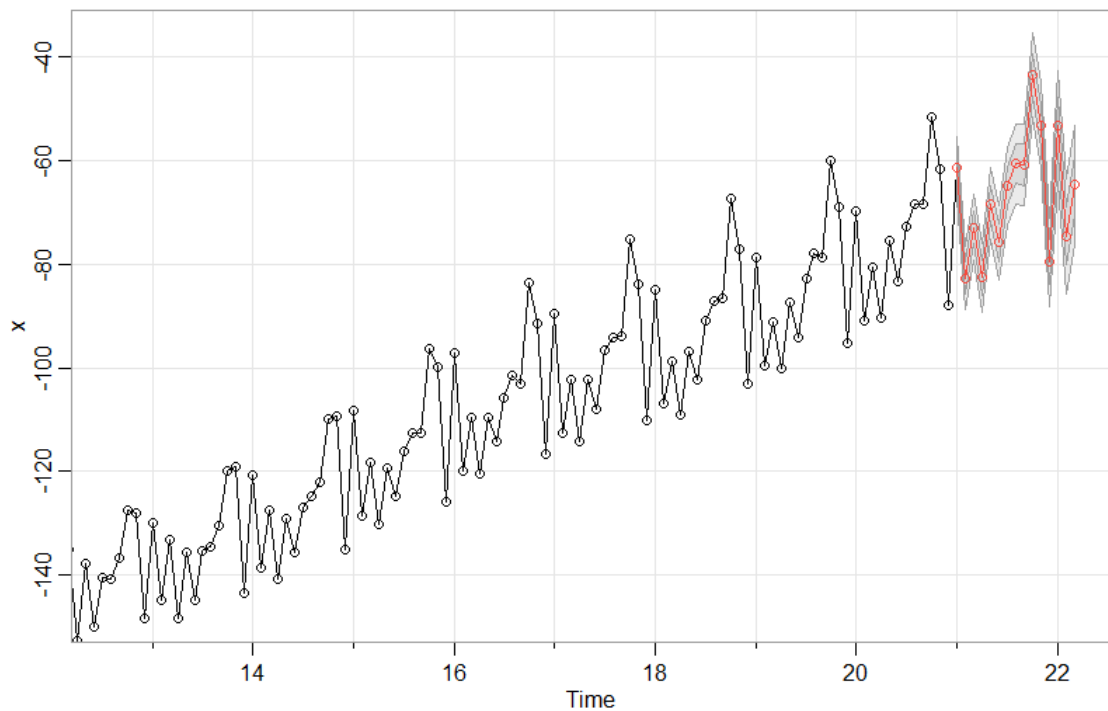


Figura 2.11: Predição de 15 meses pelo método SARIMA

Em que a linha vermelha representa os valores preditos, e as linhas cinzas o intervalo de confiança.

2.5.2 Holt-Winters

Utilizando o método de Holt-Winter obtive os resultados representados na Figura 2.12.

```
> HW3.pred <- predict(HW3, 15, prediction.interval = TRUE, level=0.95)
> HW3.pred
```

	fit	upr	lwr
Jan 21	-61.83226	-56.70177	-66.96276
Feb 21	-79.61970	-74.24805	-84.99135
Mar 21	-71.16353	-65.67813	-76.64893
Apr 21	-79.13023	-73.29928	-84.96117
May 21	-67.08841	-61.23918	-72.93763
Jun 21	-73.33455	-67.03858	-79.63053
Jul 21	-64.04017	-57.73340	-70.34693
Aug 21	-60.27165	-53.79231	-66.75100
Sep 21	-60.48250	-53.65024	-67.31476
Oct 21	-46.06477	-39.64403	-52.48551
Nov 21	-54.30682	-47.03318	-61.58047
Dec 21	-75.93922	-64.53478	-87.34366
Jan 22	-53.99690	-43.77441	-64.21938
Feb 22	-69.42263	-57.41303	-81.43223
Mar 22	-61.95114	-50.13853	-73.76375

Figura 2.12: Predição de 15 meses pelo método Holt-Winter

Também plotei os resultados obtido Figura 2.13

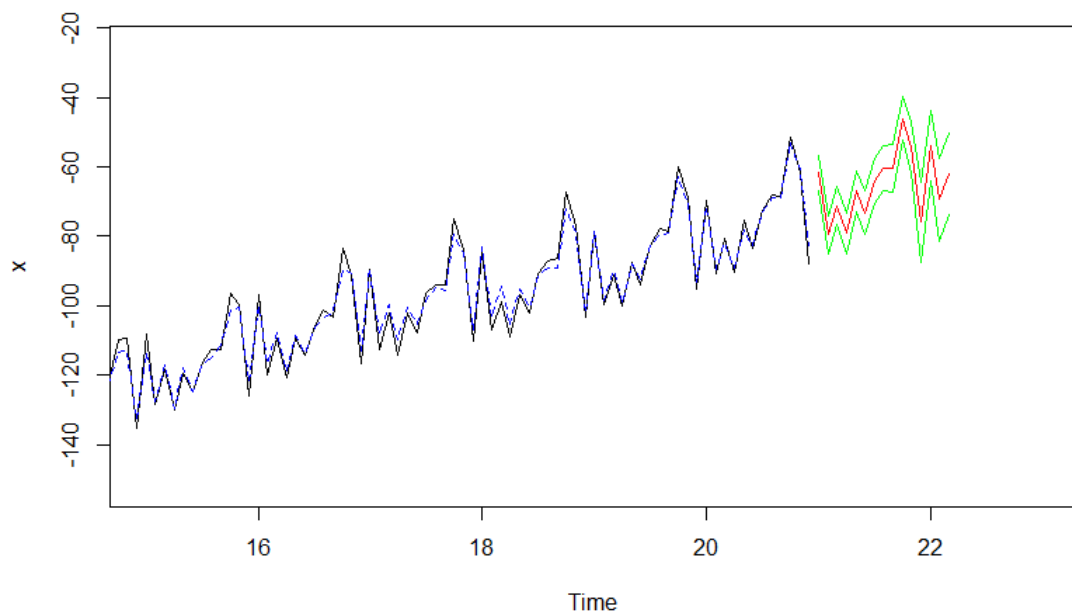


Figura 2.13: Predição de 15 meses pelo método Holt-Winter

Em que as linhas verdes representam o intervalos de confiança de 95% e a linha vermelha os valores preditos, e como a linha vermelha está entre as verdes conclui-se que a previsão é adequada.

2.5.3 Comparação

Para poder comparar os valores preditos pelos dois métodos, plotei o grafico representado na Figura 2.14.

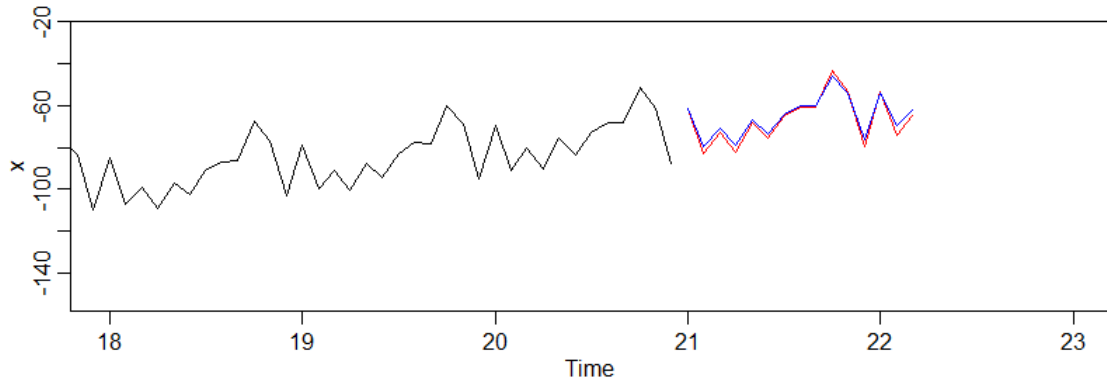


Figura 2.14: Predição de 15 meses pelo método SARIMA e Holt-Winter

Em que a linha vermelha representa os valores preditos pelo método SARIMA, e a linha azul os valores preditos pelo método Holt-Winters.

As duas previsões tem valores muito próximos, e indicar qual predição é a melhor se torna um pouco irrelevante (tendo em vista que são muito próximas), mas seria possível calcular algumas medidas de acurácia para ambos os métodos, podendo assim definir qual previsão é a melhor.

Apêndice A

Código

Os seguinte código feito em R foi utilizado nessa análise:

```
1
2 # DADOS
3 library(readxl)
4 df = read_excel("E:/Analise2/Analise2_DADOS1.xlsx")
5 df2 = df$`752728`
6 df2
7
8 ts.plot(df2)
9
10 x = ts(df2, start = c(1,1), frequency = 12)
11 x
12 ts.plot(x)
13
14
15 # Analise descritiva
16 # Plotar o Grafico da serie
17 ts.plot(x, xlab = "Tempo", ylab = "Valores", main = "Grafico da S rie ")
18
19 plot(decompose(x, type = c("multiplicative")))
20
21 # Amplitude media
22 med.var<-function(x,k)
23 {N<-length(x)
24 x.m<-rep(0,(N-k))
25 x.r<-rep(0,(N-k))
26 for (i in 1:(N-k)) x.m[i]<-mean(x[i:(i+k)])
27 for (i in 1:(N-k)) x.r[i]<-max(x[i:(i+k)])-min(x[i:(i+k)])
28 plot(x.m,x.r,xlab="medias",ylab="amplitude", main = "Grafico de
    Amplitude")
29 aa1<-lm(x.r~x.m)
30 abline(aa1$coef[1],aa1$coef[2],col=2)
31 summary(aa1)
32 }
```

```

33
34 med.var(x,12)
35
36 library(astsa)
37 acf2(x)
38
39 adf.test(x, k = 12)
40
41 cox.stuart.test(x)
42
43 # S rie diferenciada
44 x2 = diff(x)
45
46 ts.plot(x2)
47 plot(decompose(x2, type = c("multiplicative")))
48
49 acf2(x2)
50
51 adf.test(x2, k = 12)
52 cox.stuart.test(x2)
53
54 #
55 x3 = diff(diff(x))
56
57 ts.plot(x3)
58 acf2(x3)
59
60 adf.test(x3, k = 12)
61
62
63 #####
64 ### MODELO ###
65 #####
66 ## Ajuste do modelo arima
67 # Auto
68 modelo = auto.arima(x)
69 modelo
70 tsdiag(modelo)
71
72 S1 = sarima(x, 0,1,2,0,1,1,12)
73
74 M1 = arima(x, order = c(0,1,2), seasonal= c(0,1,1))
75 detectA0(M1)
76 detectI0(M1)
77 detectA0(modelo_autoarima)
78
79
80 # Previs o
81 # Sarima

```

```

82 P1 = sarima.for(x,15,0,1,2,0,1,1,12)
83 P1
84 P1$pred
85
86
87 # Holt
88 HW = HoltWinters(x, seasonal = "multiplicative")
89 HW.pred <- predict(HW, 15, prediction.interval = TRUE, level=0.95)
90 HW.pred
91
92 ts.plot(x)
93 plot(x, ylab="x", xlim = c(15, 23))
94 lines(HW3$fitted[,1], lty=2, col="blue")
95 lines(HW3.pred[,1], col="red")
96 lines(HW3.pred[,2], col="green")
97 lines(HW3.pred[,3], col="green")
98
99
100 # Compara o
101 plot(x, ylab="x", xlim = c(18, 23))
102 lines(P1$pred, col = "red")
103 lines(HW3.pred[,1], col="blue")

```