

Trabalho Séries Temporais

Adriane Akemi nUSP: 11915623 Álvaro Valentim nUSP: 10392150
Emerson Miady nUSP: 11207943

Objetivo

O objetivo do grupo é usar as técnicas de séries temporais para ajustar algum modelo em um conjunto de dados, aplicando as seguintes etapas:

1. Análise exploratória;
2. Ajuste do modelo;
3. Análise de resíduos;
4. Previsão.

Dados

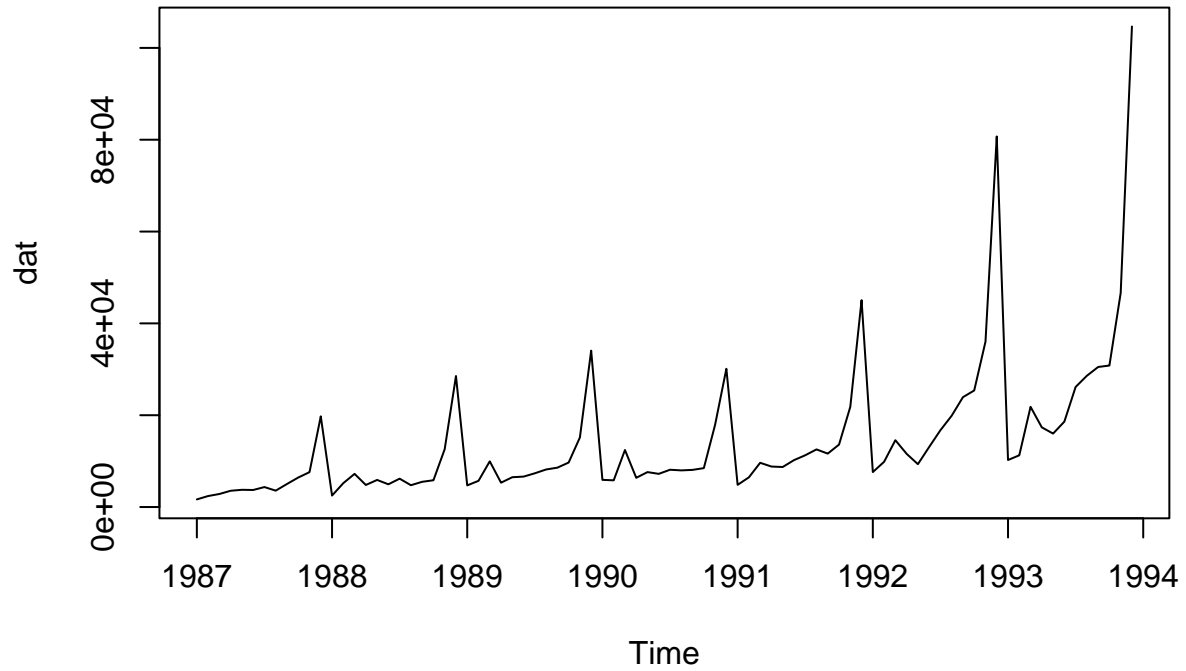
Os dados utilizados são de vendas mensais (em dolar australiano) de uma loja de Souvenir localizada em um beach resort em Queensland, Australia, no período de 1987 a 1993, contendo assim, 84 observações temporais.

```
dat <- scan("http://robjhyndman.com/tsdldata/data/fancy.dat")
dat <- ts(dat, frequency=12, start=c(1987,1))
dat
```

##	Jan	Feb	Mar	Apr	May	Jun	Jul
## 1987	1664.81	2397.53	2840.71	3547.29	3752.96	3714.74	4349.61
## 1988	2499.81	5198.24	7225.14	4806.03	5900.88	4951.34	6179.12
## 1989	4717.02	5702.63	9957.58	5304.78	6492.43	6630.80	7349.62
## 1990	5921.10	5814.58	12421.25	6369.77	7609.12	7224.75	8121.22
## 1991	4826.64	6470.23	9638.77	8821.17	8722.37	10209.48	11276.55
## 1992	7615.03	9849.69	14558.40	11587.33	9332.56	13082.09	16732.78
## 1993	10243.24	11266.88	21826.84	17357.33	15997.79	18601.53	26155.15
##	Aug	Sep	Oct	Nov	Dec		
## 1987	3566.34	5021.82	6423.48	7600.60	19756.21		
## 1988	4752.15	5496.43	5835.10	12600.08	28541.72		
## 1989	8176.62	8573.17	9690.50	15151.84	34061.01		
## 1990	7979.25	8093.06	8476.70	17914.66	30114.41		
## 1991	12552.22	11637.39	13606.89	21822.11	45060.69		
## 1992	19888.61	23933.38	25391.35	36024.80	80721.71		
## 1993	28586.52	30505.41	30821.33	46634.38	104660.67		

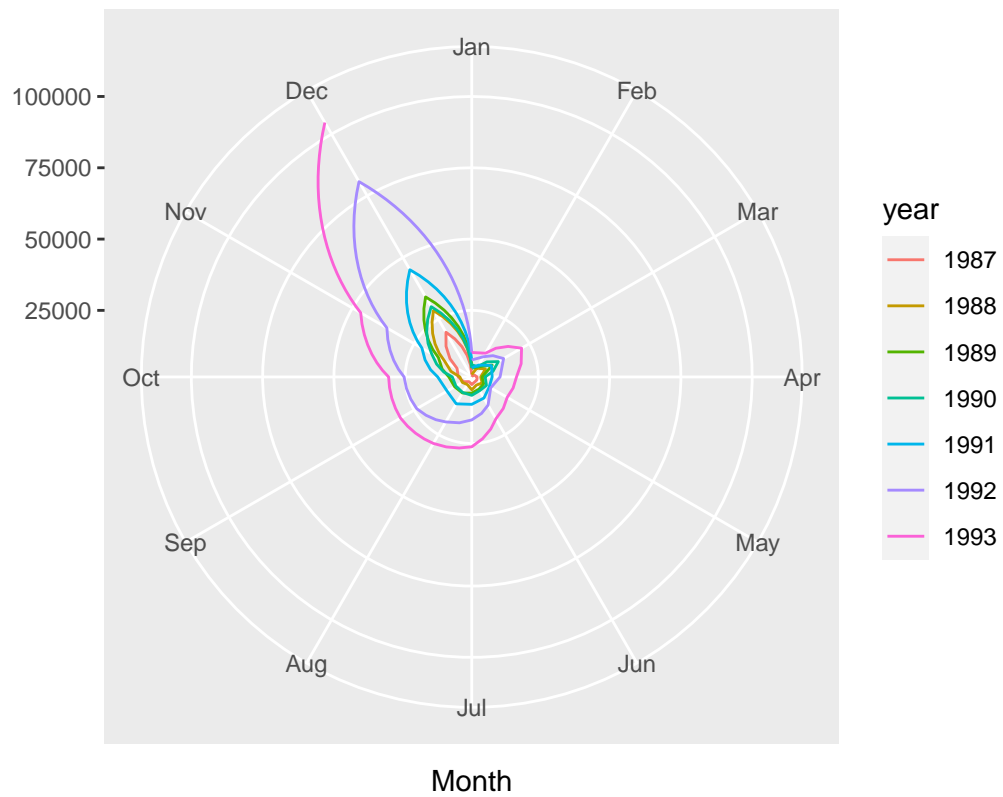
Análise exploratória

```
plot.ts(dat)
```

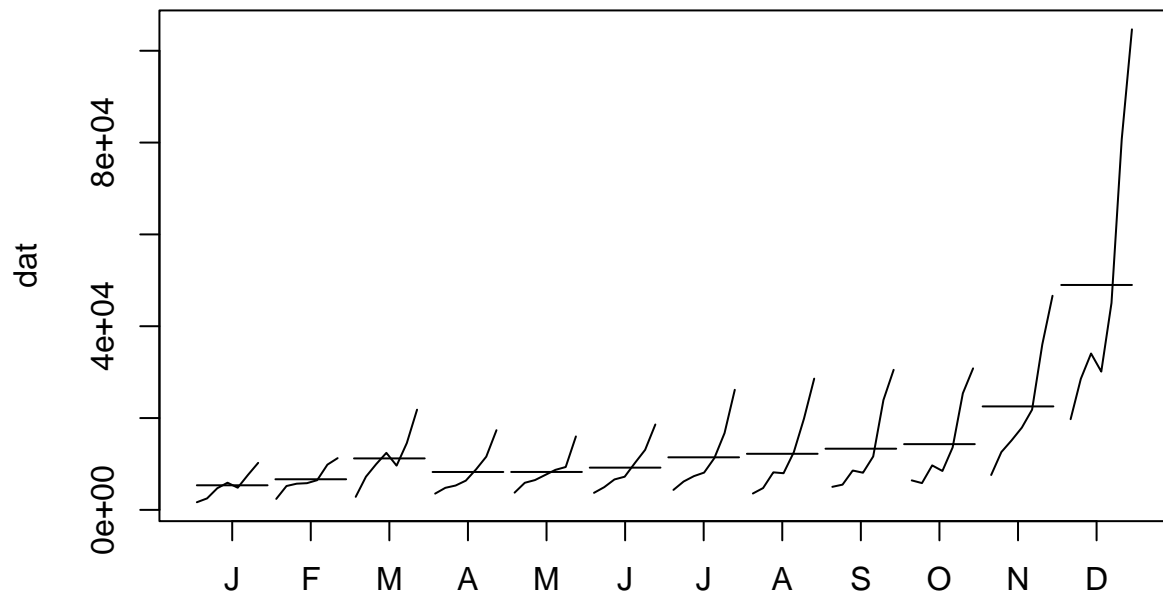


```
ggseasonplot(dat, polar = TRUE)
```

Seasonal plot: dat



```
monthplot(dat)
```



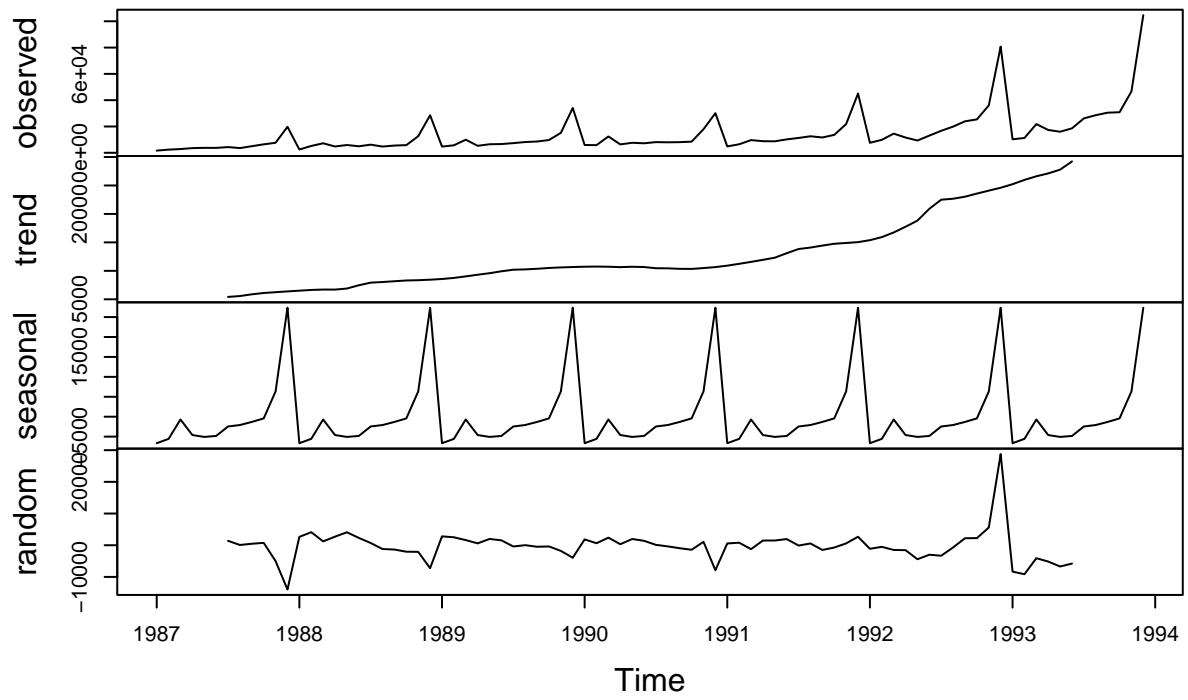
O primeiro plot mostra uma tendência de crescimento da série conforme se passa os anos, indicando que a loja está em constante crescimento. Além disso, existe uma certa periodicidade nas vendas mensais.

O segundo plot é uma confirmação do primeiro, já que em todos os anos o formato da curva se mantém aproximadamente, porém vemos um crescimento de sua área.

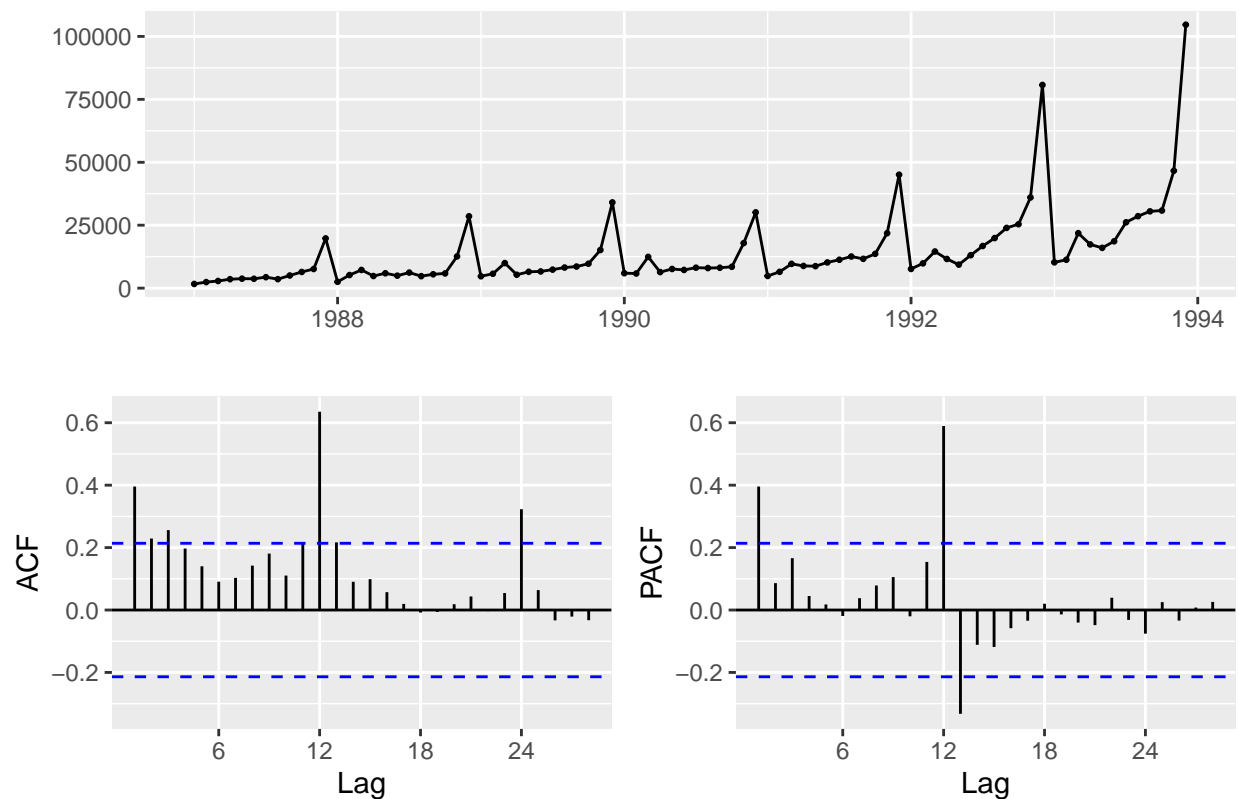
O último plot indica que, as médias de vendas (linhas horizontais) aumentam no final do ano, atingindo um pico em dezembro.

```
plot(decompose(dat))
```

Decomposition of additive time series



```
ggtsdisplay(dat)
```



Da decomposição da série, vemos que existe sazonalidade e uma clara tendência de crescimento nas vendas (série não estacionária), corroborando com as observações feitas anteriormente. Para tratar a tendência, podemos utilizar a diferença de lags entre as observações, ou seja, alterando o argumento “d” do ARIMA.

Vamos fazer o teste ADF (raiz unitária) com $d = 1$ para verificar se a série ainda é não estacionária.

```
adf.test(diff(dat))
```

```
##
## Augmented Dickey-Fuller Test
##
## data: diff(dat)
## Dickey-Fuller = -3.7374, Lag order = 4, p-value = 0.02655
## alternative hypothesis: stationary
```

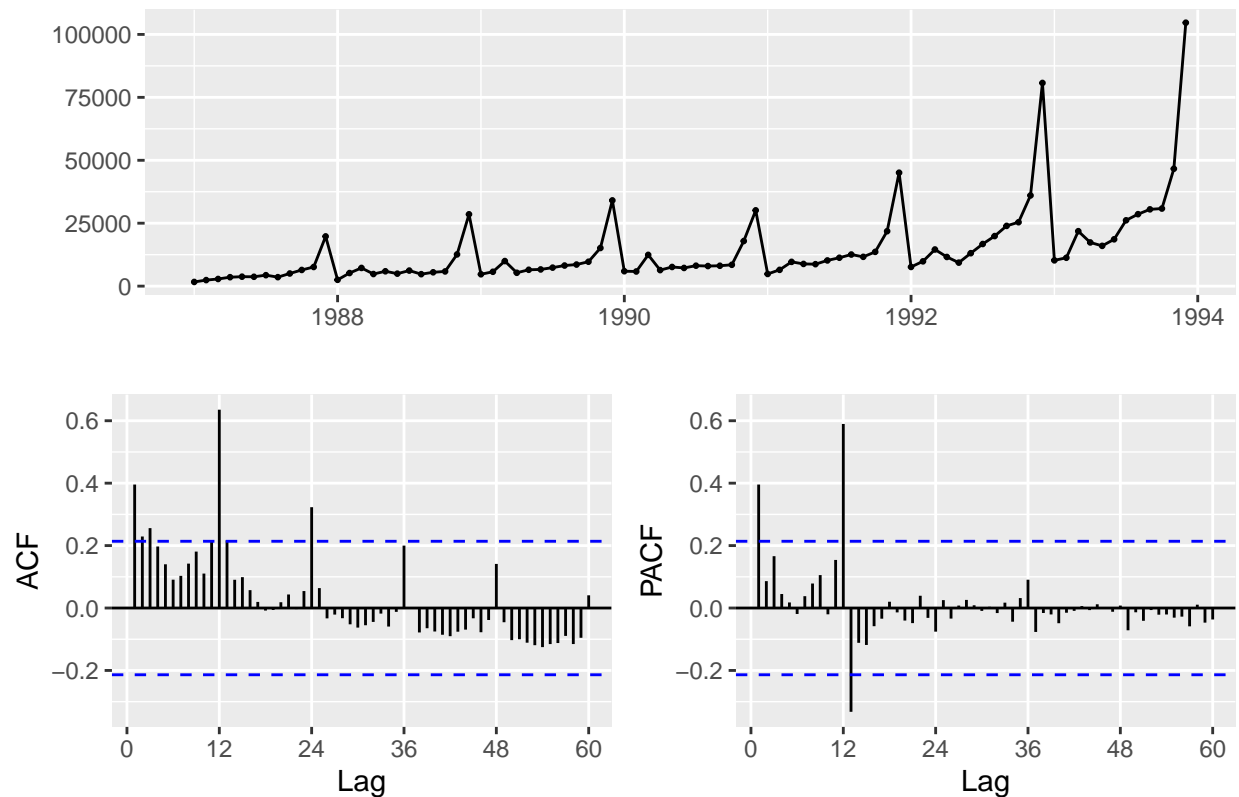
O $d = 1$ já indica que a série é estacionária, sendo assim, usaremos este parâmetro para o ARIMA.

Para achar o “p” do AR, olhamos para a PACF, e um valor que queremos testar é 1. Isto porque existe um valor significativo na primeira autocorrelação parcial. Já para descobrir o “q” do MA, observamos a FAC, e analogamente, escolhemos o “q” igual a 1.

Então nosso modelo até o momento é $ARIMA(1, 1, 1)$.

Dado que existe sazonalidade, vamos incluir a componente sazonal neste modelo.

```
ggtsdisplay(dat, lag.max=60)
```



Os gráficos acima são os mesmos que os anteriores, só alterando o número de lags. No PACF, vemos um truncamento periódico (período = 12) em $t = 12$, e além disso, temos os lags periódicos do ACF decaindo lentamente, em $t = 12, 24, 36 \dots$

Sendo assim, supomos que o parâmetro “P” do SARIMA deve ser 1 e o “Q” é 0.

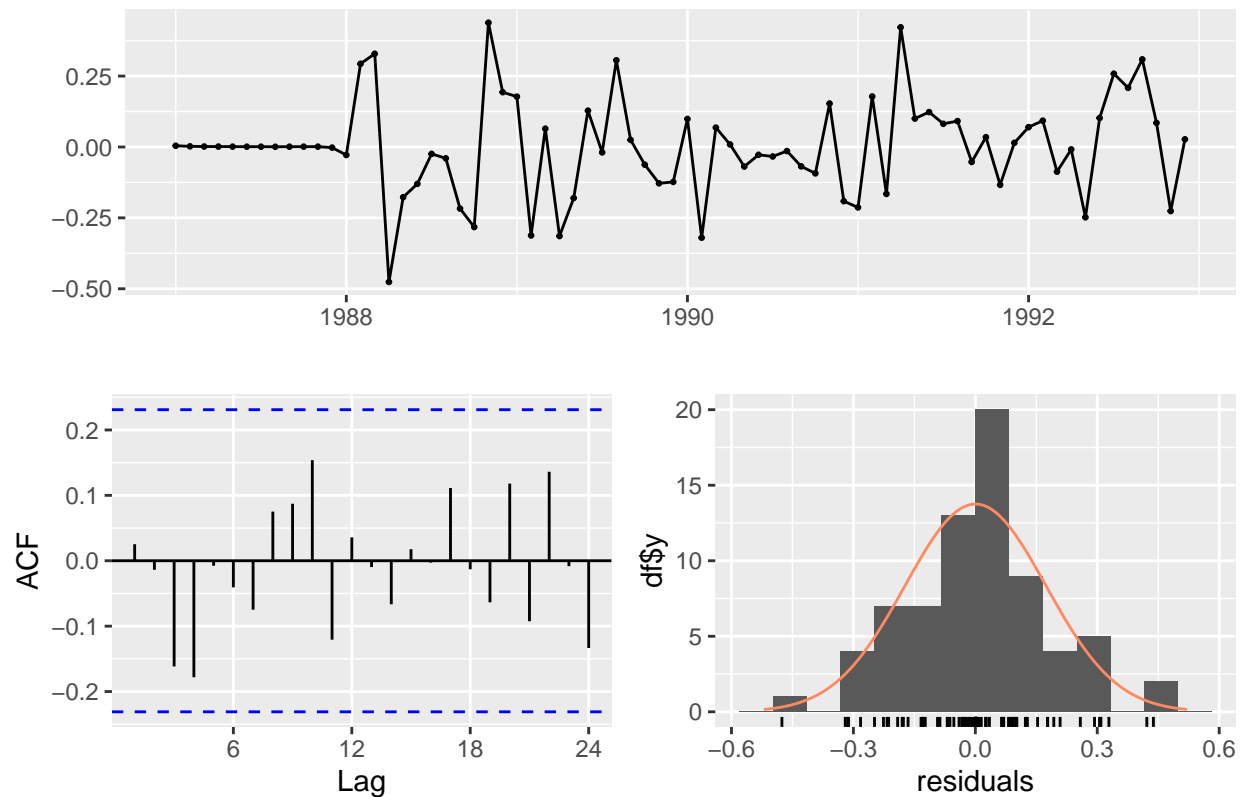
Logo, o modelo que supomos foi um $SARIMA(1, 1, 1)(1, 1, 0)_{12}$

Ajuste

```
training <- window(dat, end = c(1992,12))
test <- window(dat, start = c(1993,1))
fit <- Arima(training, order = c(1,1,1), seasonal = c(1,1,0),
             method = "ML", lambda = 0)

checkresiduals(fit)
```

Residuals from ARIMA(1,1,1)(1,1,0)[12]



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,1,1)(1,1,0)[12]
## Q* = 10.112, df = 11, p-value = 0.5204
##
## Model df: 3.   Total lags used: 14
```

O teste de Ljung-Box verifica se a autocorrelação dos resíduos é igual a 0. Como o p-valor obtido foi maior que o nível de significância de 5%, não rejeitamos a hipótese nula e portanto, os resíduos não são autocorrelacionados.

Vemos também que, as autocorrelações do ACF não passam da banda de confiança e a distribuição dos resíduos aparenta ser Normal.

Para testar a normalidade deles, vamos usar o teste de Shapiro Wilk.

```
shapiro.test(fit$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  fit$residuals
## W = 0.9775, p-value = 0.2239
```



```
mean(fit$residuals)
```

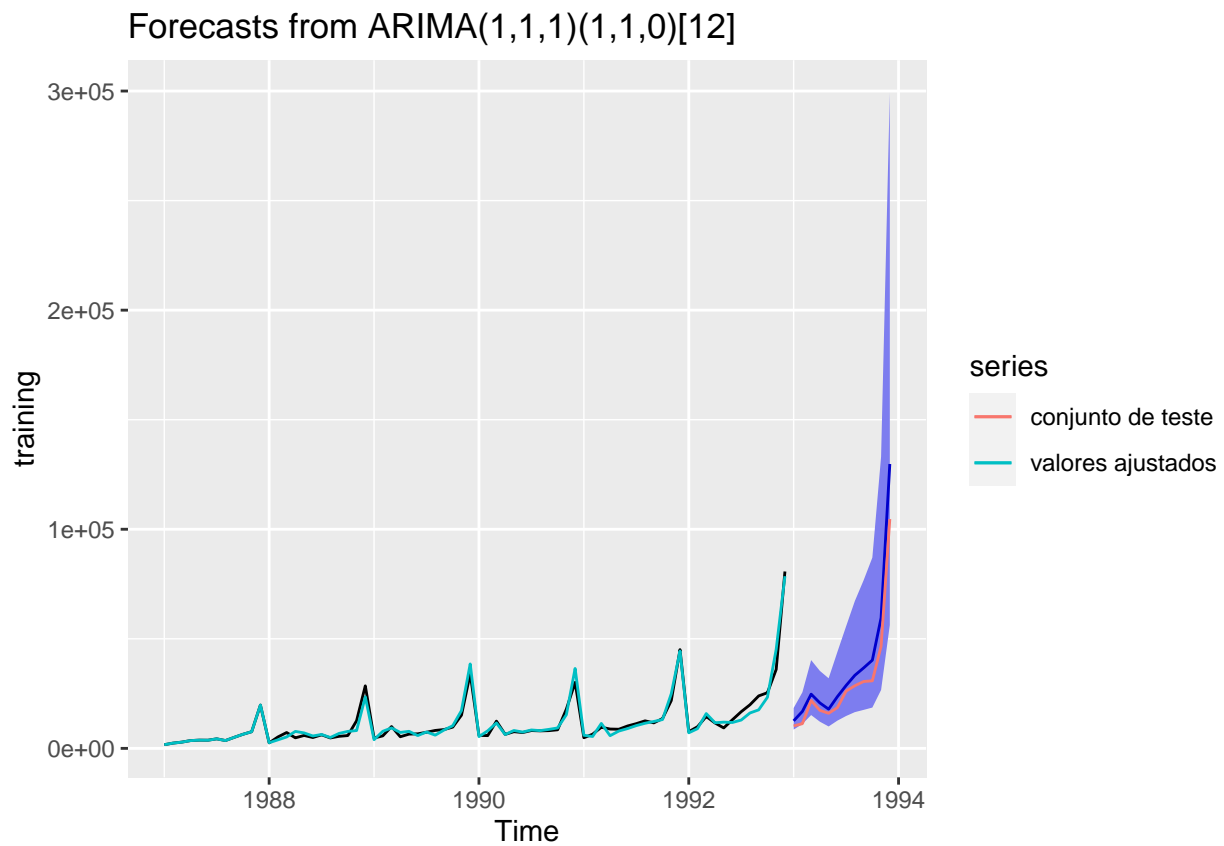
```
## [1] 0.0003959808
```

A média dos resíduos foi praticamente 0, mostrando evidências de que a distribuição deles é $N(0, \sigma^2)$.

Predição

```
forc <- forecast(object = fit, h=12, level = 0.95)
```

```
autoplot(forc) + autolayer(test, series = "conjunto de teste") + autolayer(fitted(forc), series = "valores ajustados")
```



```
accuracy(forc, test)
```

```
##           ME      RMSE      MAE      MPE      MAPE      MASE
## Training set  -8.001654 2247.371 1436.015 -1.450601 12.56400 0.3834691
## Test set     -6839.544870 9309.593 6839.545 -22.730360 22.73036 1.8264110
##           ACF1 Theil's U
## Training set 0.05633702    NA
## Test set     0.40528354 0.6266895
```

Aparentemente o modelo realizou um bom ajuste, com uma boa predição. Além de que as métricas MPE e MAPE dos conjuntos de treino e teste foram bem satisfatórios.

Conclusão

Passamos por todas as etapas e um bom modelo a ser ajustado foi o $SARIMA(1, 1, 1)(1, 1, 0)_{12}$, com boas predições e métricas de avaliação.