Questão 2 e 3

Giovani Valdrighi

23/11/2020

Questão 2

Análise da série

Iremos resolver juntamente os itens **a)**, **b)** e **c)**, para isso iremos plotar os dados e suas diferentes diferenciações. Apesar de o enúnciado falar para analisar os lags 4, vamos utilizar os lags 12, pois esse é o período da série de consumo.

```r
library(readxl)
library(tseries)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```r
library(magrittr)
library(zoo)
```

```
##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```r
library(forecast)
library(fpp2)
```

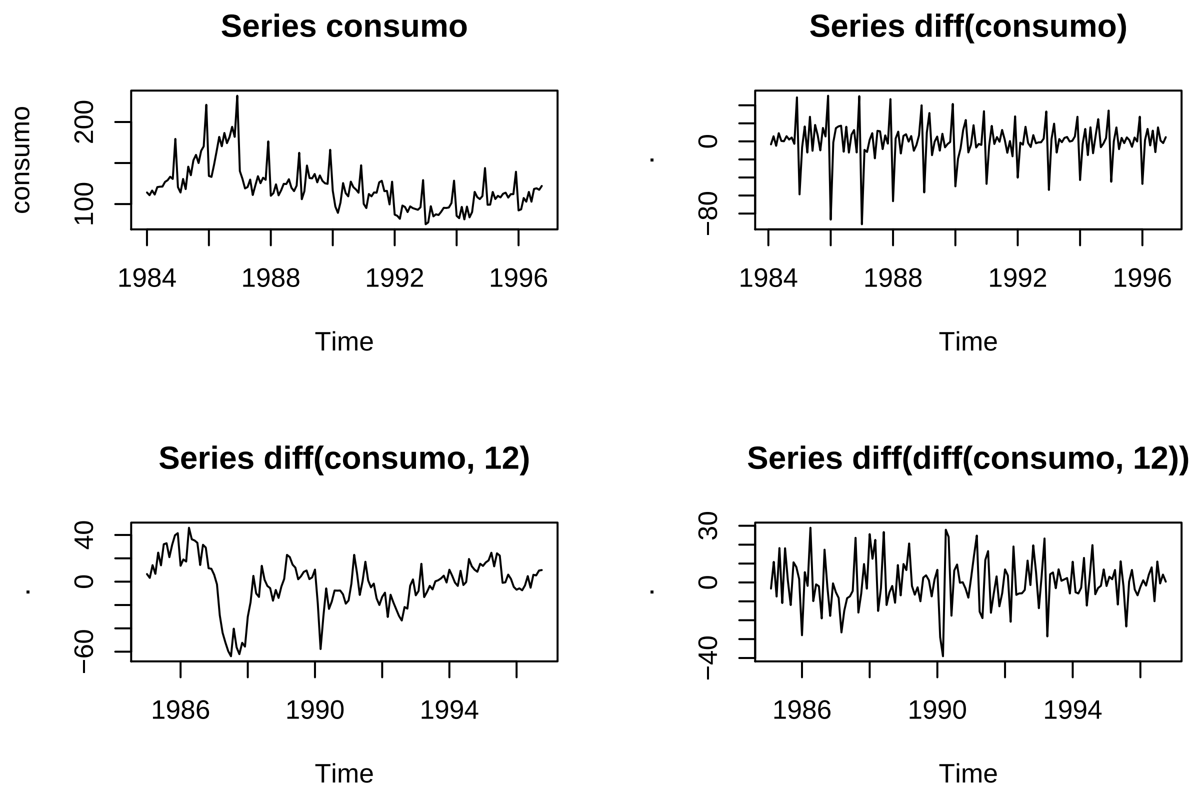
```
## -- Attaching packages ---------------------------------------------------------------------------

## v ggplot2   3.3.2     v expsmooth 2.3
## v fma       2.4

##
```

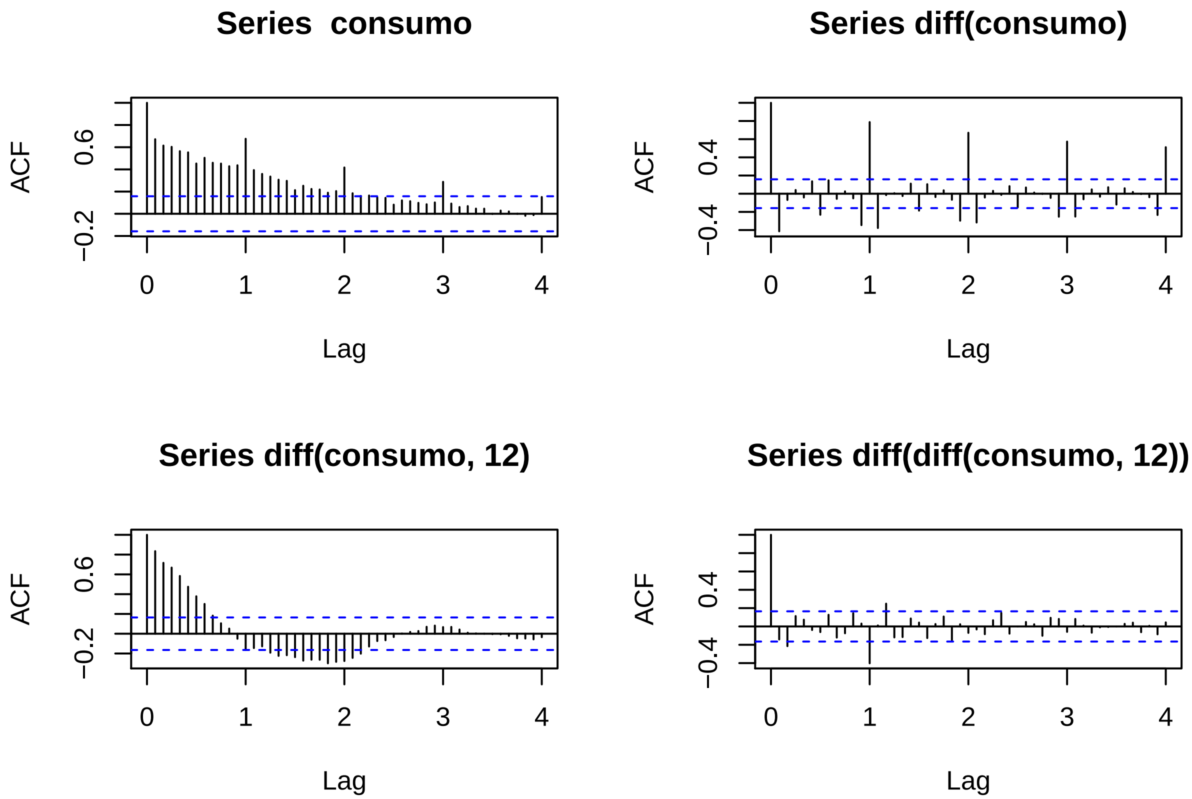
```
CONSUMO <- read_excel('CONSUMO.XLS')
consumo <- ts(CONSUMO$consumo, start = c(1984), frequency = 12)

par(mfrow = c(2, 2))
plot(consumo, main = "Series consumo")
diff(consumo) %>% plot(main = "Series diff(consumo)")
diff(consumo, lag = 12) %>%  plot(main = "Series diff(consumo, 12)")
diff(consumo, lag = 12) %>% diff() %>% plot(main = "Series diff(diff(consumo, 12))")
```



Além disso, também vamos analisar a ACF e PACF da série.

```
par(mfrow = c(2, 2))
acf(consumo, lag.max = 48)
diff(consumo) %>% acf(lag.max = 48, main = "Series diff(consumo)")
diff(consumo, lag = 12) %>%  acf(lag.max = 48, main = "Series diff(consumo, 12)")
diff(consumo, lag = 12) %>% diff() %>% acf(lag.max = 48, main = "Series diff(diff(consumo, 12))")
```

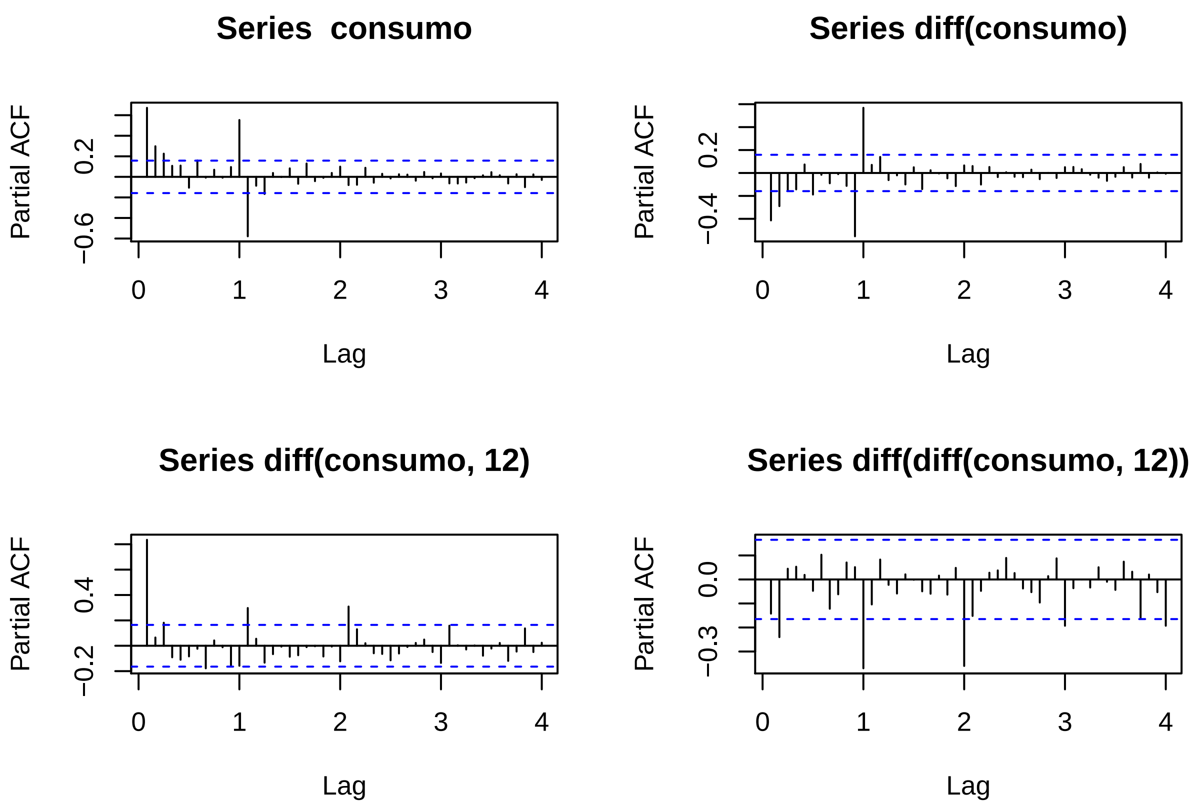



```r
#PACFs
par(mfrow = c(2, 2))
## Z_t
pacf(consumo, lag.max = 48)

## \Delta Z_t
diff(consumo) %>% pacf(lag.max = 48, main = "Series diff(consumo)")

## \Delta_4 Z_t
diff(consumo, lag = 12) %>%  pacf(lag.max = 48, main = "Series diff(consumo, 12)")

## \Delta \Delta_z
diff(consumo, lag = 12) %>% diff() %>% pacf(lag.max = 48, main = "Series diff(diff(consumo, 12))")
```



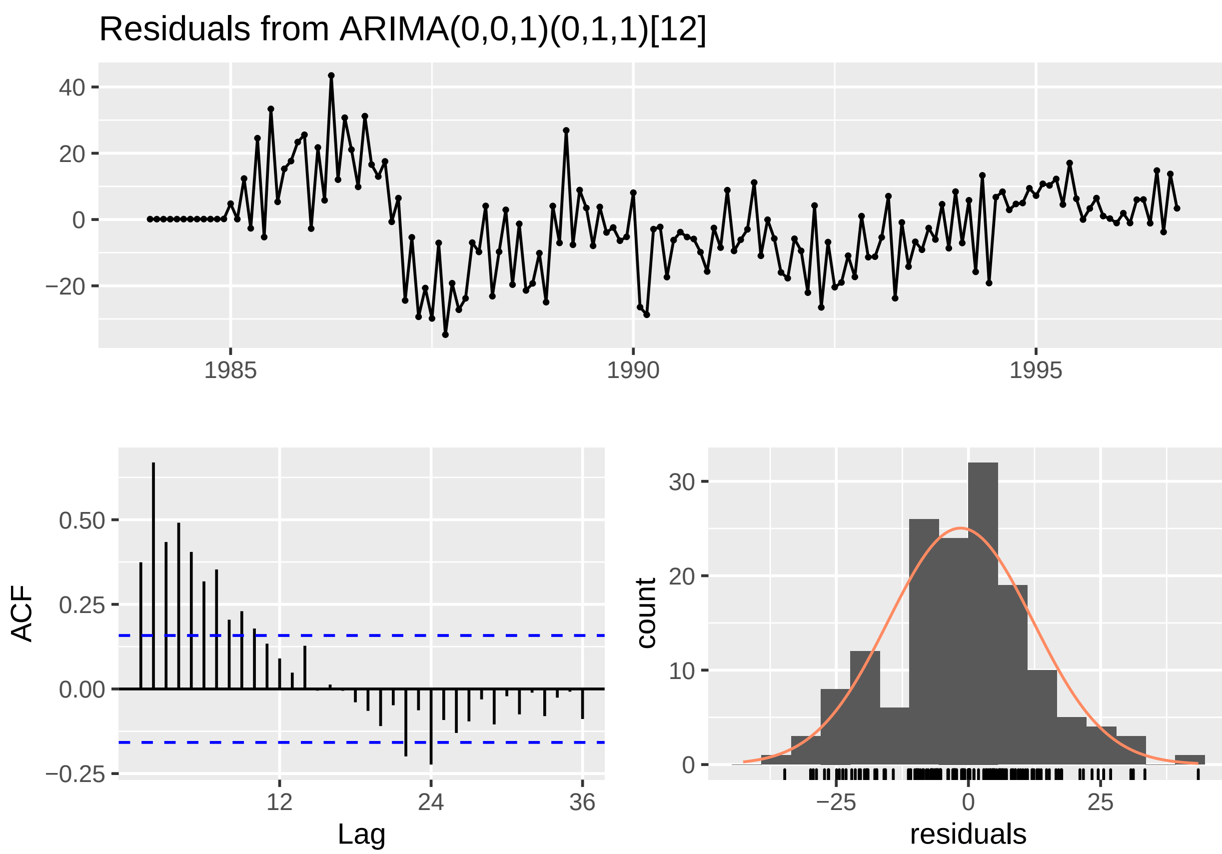
Pela análise dos gráficos das séries diferenciadas podemos ver que apesar da série não apresentar uma clara tendência, os períodos de 12 anos apresentam alguma tendência, o que dá a entender que devemos utilizar $d = 0$ e $D = 1$. Olhando para a ACF, vemos que os lags sazonais caem lentamente, o que indica um modelo AR sazonal, o lag sinificativo da PACF é 1,2 e talvez o 3. Os lags não sazonais caem lentamente na ACF e rápido na PACF, o que indica um modelo AR com lag 1 signficante.

Agora para os items **d)** e **e)**, vamos calcular os parâmetros de alguns modelos selecionados e avaliar os resíduos. O modelo que iremos considerar inicialmente é $SARIMA(0,0,1)(0,1,1)$. Começando com o primeiro modelo:

```
mod1 <- Arima(consumo, order = c(0, 0, 1), seasonal = c(0, 1, 1))
summary(mod1)
```

```
## Series: consumo
## ARIMA(0,0,1)(0,1,1)[12]
##
## Coefficients:
##          ma1     sma1
##       0.7519  -0.4216
## s.e.  0.0505   0.0956
##
## sigma^2 estimated as 207.8:  log likelihood=-580.98
## AIC=1167.96   AICc=1168.13   BIC=1176.83
##
## Training set error measures:
##                     ME     RMSE      MAE       MPE     MAPE      MASE      ACF1
## Training set -1.485802 13.74404 10.35285 -2.225502 8.643511 0.6461826 0.3743327
```

```r
mod1 %>% checkresiduals()
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,0,1)(0,1,1)[12]
## Q* = 273.96, df = 22, p-value < 2.2e-16
##
## Model df: 2.   Total lags used: 24
```

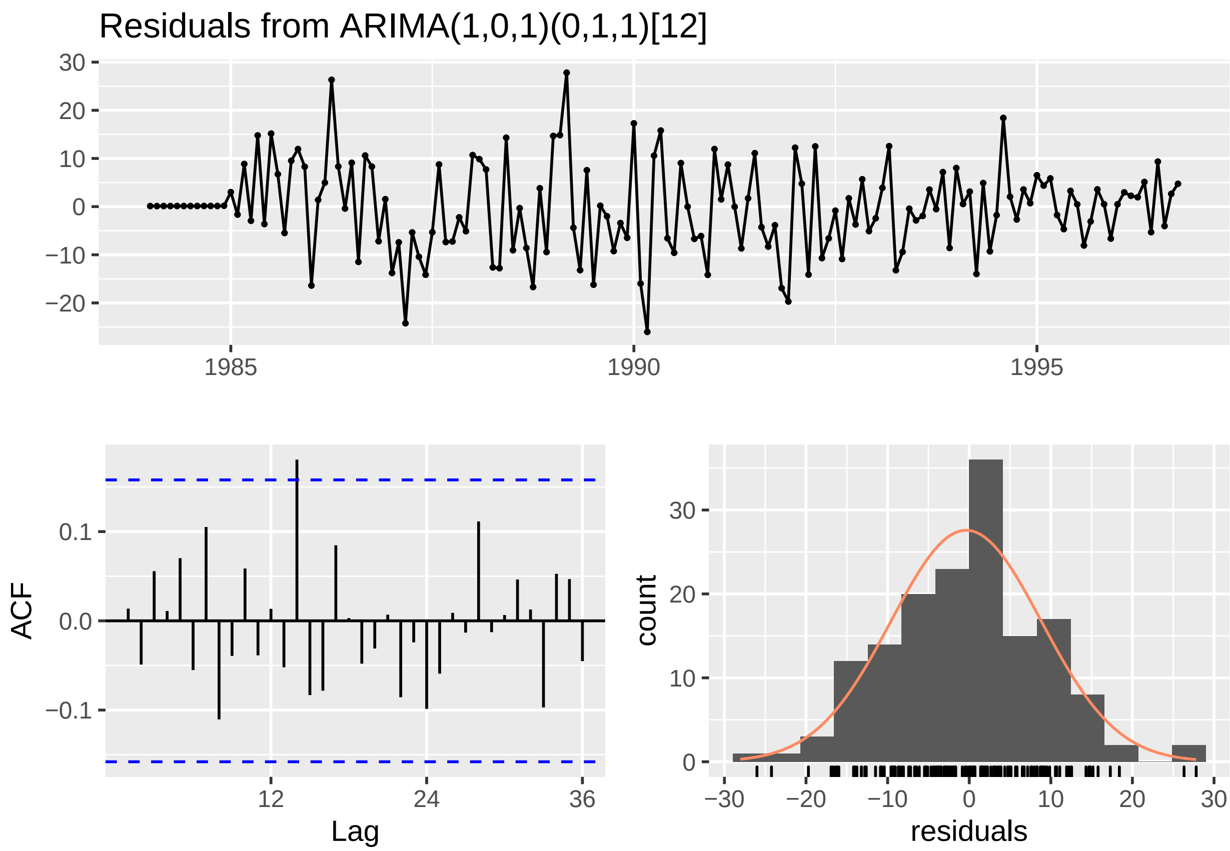
Vemos que a ACF dos resíduos é altamente correlacionada, estamos ignorando algum fator regressivo.

```r
mod2 <- Arima(consumo, order = c(1, 0, 1), seasonal = c(0, 1, 1))
summary(mod2)
```

```
## Series: consumo
## ARIMA(1,0,1)(0,1,1)[12]
##
## Coefficients:
##          ar1      ma1     sma1
##       0.9407  -0.2294  -0.7340
## s.e.  0.0324   0.0928   0.0739
##
## sigma^2 estimated as 93.81:  log likelihood=-527.47
```

```
## AIC=1062.94   AICc=1063.24   BIC=1074.77
##
## Training set error measures:
##                      ME     RMSE      MAE        MPE     MAPE      MASE
## Training set -0.3770845 9.201688 7.100073 -0.5534727 6.053567 0.4431577
##                    ACF1
## Training set 0.01366012
```

```r
mod2 %>% checkresiduals()
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,0,1)(0,1,1)[12]
## Q* = 20.549, df = 21, p-value = 0.4867
##
## Model df: 3.   Total lags used: 24
```

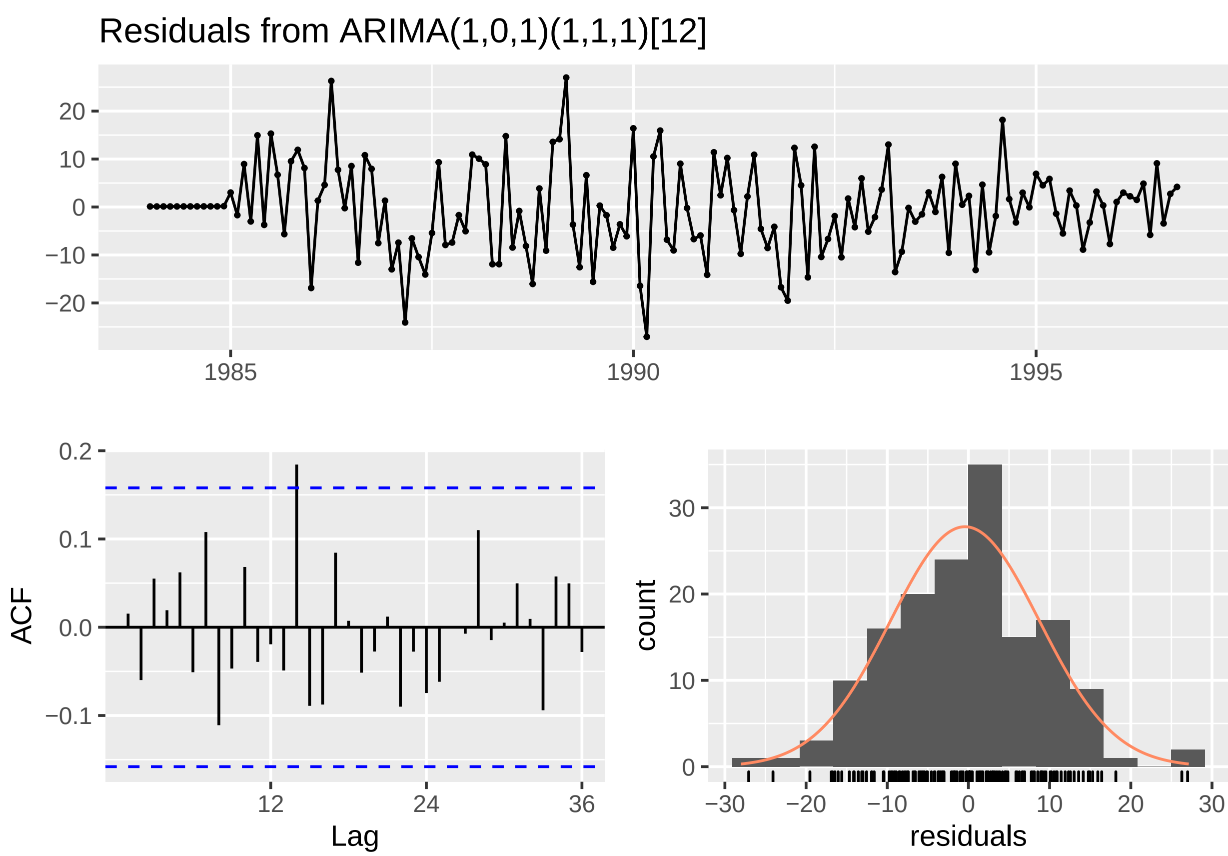
Vemos que os resíduos se assemelham a um ruído normal, apenas com o lag 12 significativo. Podemos testar aumentar o termo regressivo sazonal.

```r
mod3 <- Arima(consumo, order = c(1, 0, 1), seasonal = c(1, 1, 1))
summary(mod3)
```

```
## Series: consumo
```

```
## ARIMA(1,0,1)(1,1,1)[12]
##
## Coefficients:
##          ar1      ma1    sar1     sma1
##       0.9398  -0.2280  0.0756  -0.7805
## s.e.  0.0330   0.0942  0.1284   0.1101
##
## sigma^2 estimated as 93.86:  log likelihood=-527.29
## AIC=1064.58   AICc=1065.02   BIC=1079.36
##
## Training set error measures:
##                      ME     RMSE      MAE        MPE     MAPE      MASE
## Training set -0.4318177 9.171038 7.091462 -0.5982687 6.055918 0.4426202
##                    ACF1
## Training set 0.01540069
```

```
mod3 %>% checkresiduals()
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,0,1)(1,1,1)[12]
## Q* = 21.056, df = 20, p-value = 0.3938
##
## Model df: 4.   Total lags used: 24
```

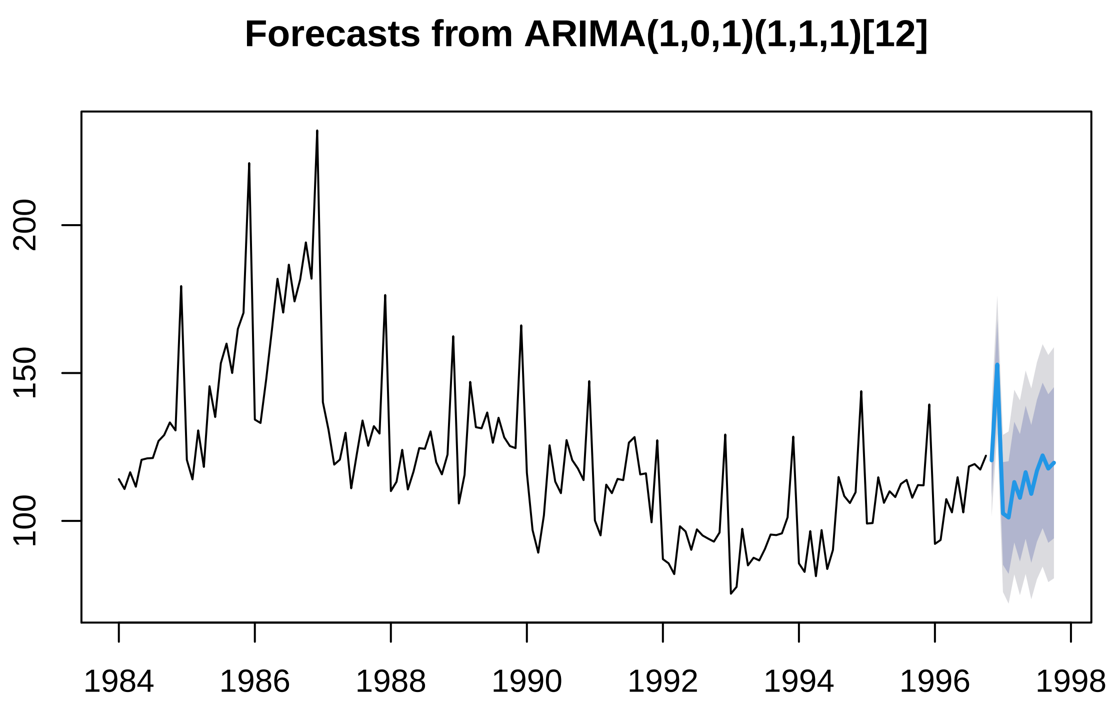
Vamos comparar com o modelo da função auto.Arima.

```
mod4 <- auto.arima(consumo)
summary(mod4)
```

```
## Series: consumo
## ARIMA(1,0,1)(0,1,1)[12]
##
## Coefficients:
##          ar1      ma1     sma1
##       0.9407  -0.2294  -0.7340
## s.e.  0.0324   0.0928   0.0739
##
## sigma^2 estimated as 93.81:  log likelihood=-527.47
## AIC=1062.94   AICc=1063.24   BIC=1074.77
##
## Training set error measures:
##                      ME     RMSE      MAE        MPE     MAPE      MASE
## Training set -0.3770845 9.201688 7.100073 -0.5534727 6.053567 0.4431577
##                    ACF1
## Training set 0.01366012
```

Vemos que o modelo considerado pelo auto.Arima é o $SARIMA(1,0,1)(0,1,1)$, que também é o nosso modelo que obtive o menor AIC, vamos utilizar dele. Para o item **f)** vamos fazer a previsão do ano seguinte.

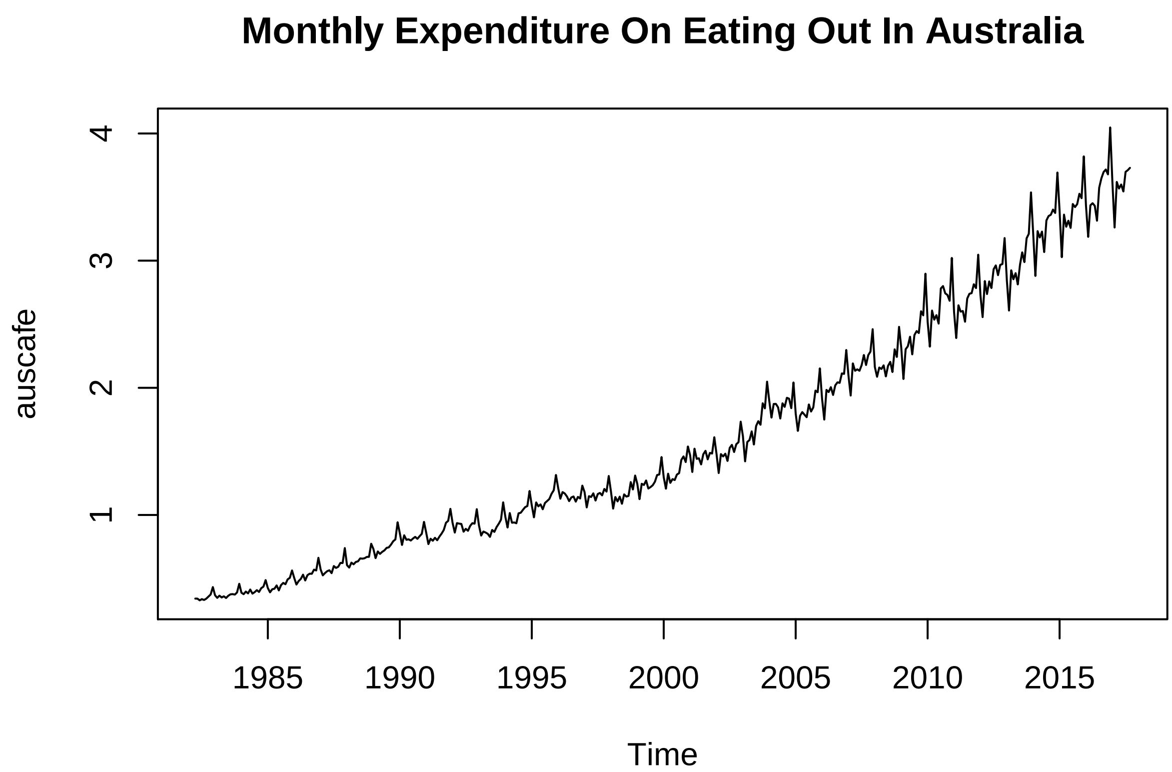
```
plot(forecast(mod3, h = 12))
```



Questão 3

Iremos agora modelar a série Auscafe que contém os valores mensais gastos em cafés e restaurantes na Austrália.

```
plot(auscafe, main = "Monthly Expenditure On Eating Out In Australia")
```



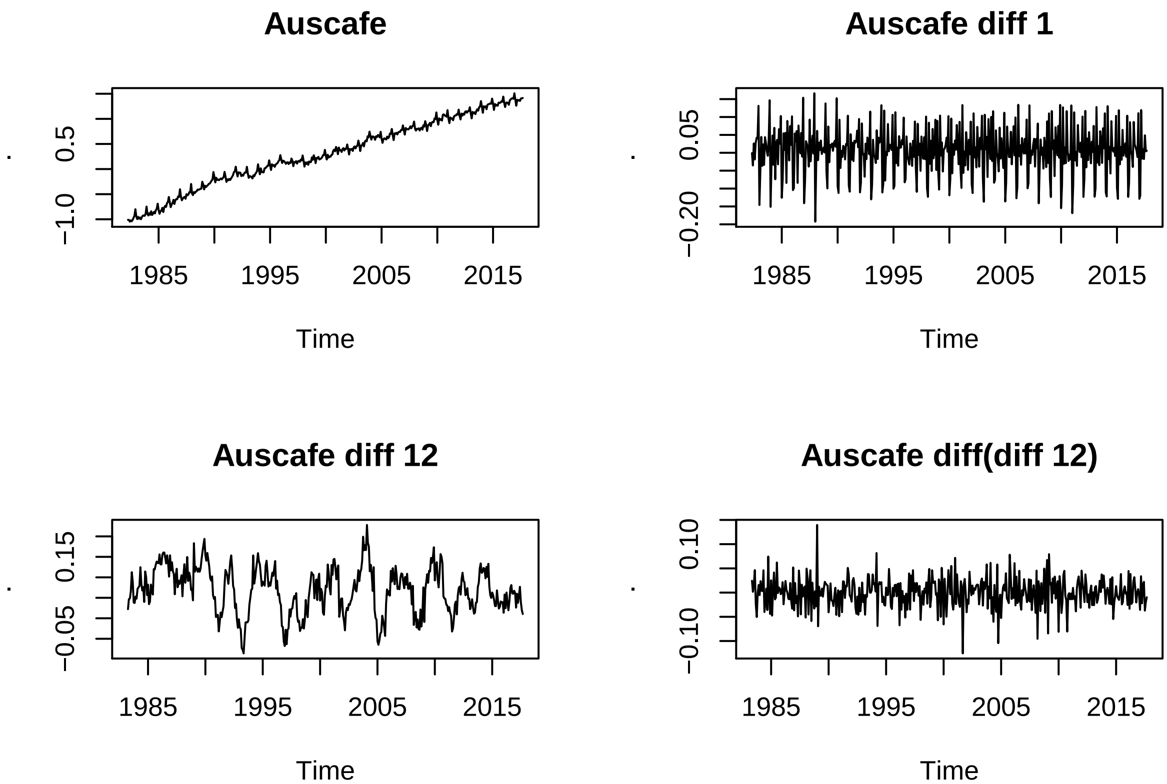
Além de ser notável que a série possui tendência, vemos que a série precisa ter a variância estabilizada, para isso vamos computar o lambda da transformação de BoxCox.

```
lambda <- BoxCox.lambda(auscafe)
lambda
```

```
## [1] 0.109056
```

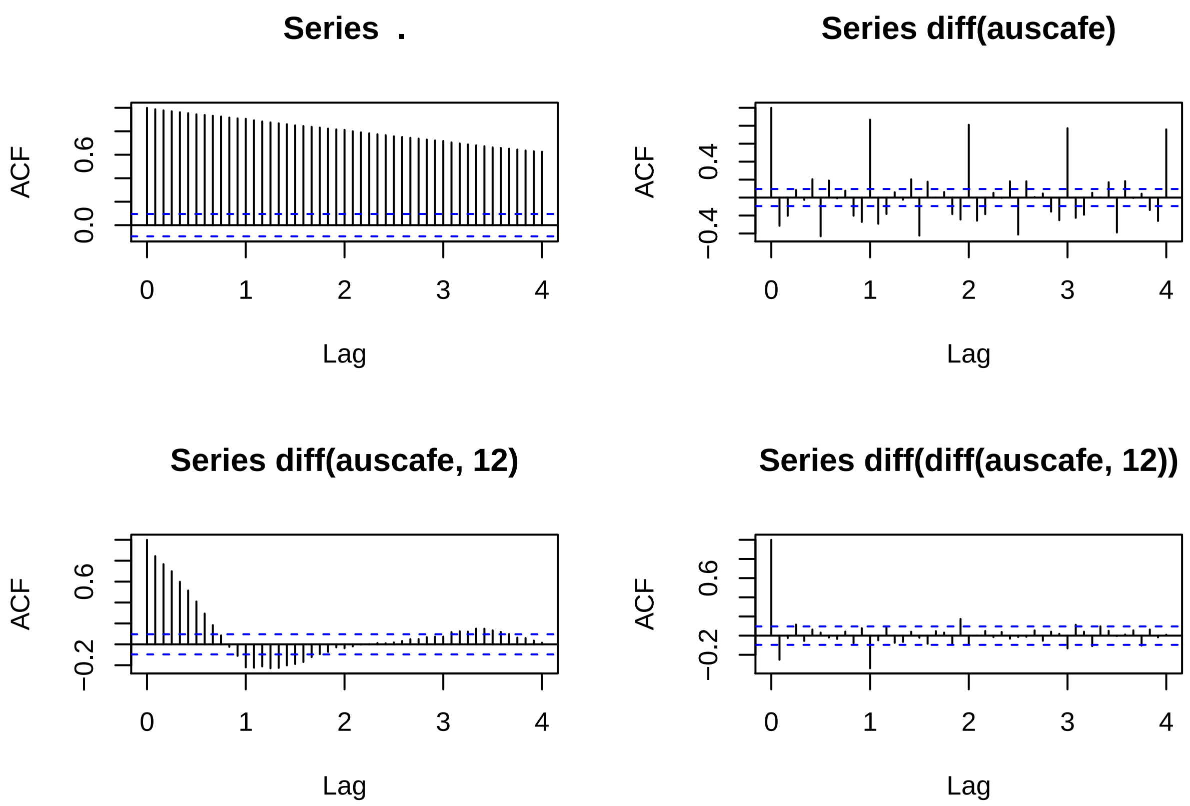
Agindo de forma semelhante aos dados anteriores, vamos analisar a série diferenciada com lag 1 e com lag 12.

```
par(mfrow = c(2, 2))
auscafe %>% BoxCox(lambda) %>%  plot(main = "Auscafe")
auscafe %>% BoxCox(lambda) %>% diff() %>% plot(main = "Auscafe diff 1")
auscafe %>% BoxCox(lambda) %>% diff(12) %>% plot(main = "Auscafe diff 12")
auscafe %>% BoxCox(lambda) %>% diff(12) %>% diff() %>% plot(main = "Auscafe diff(diff 12)")
```

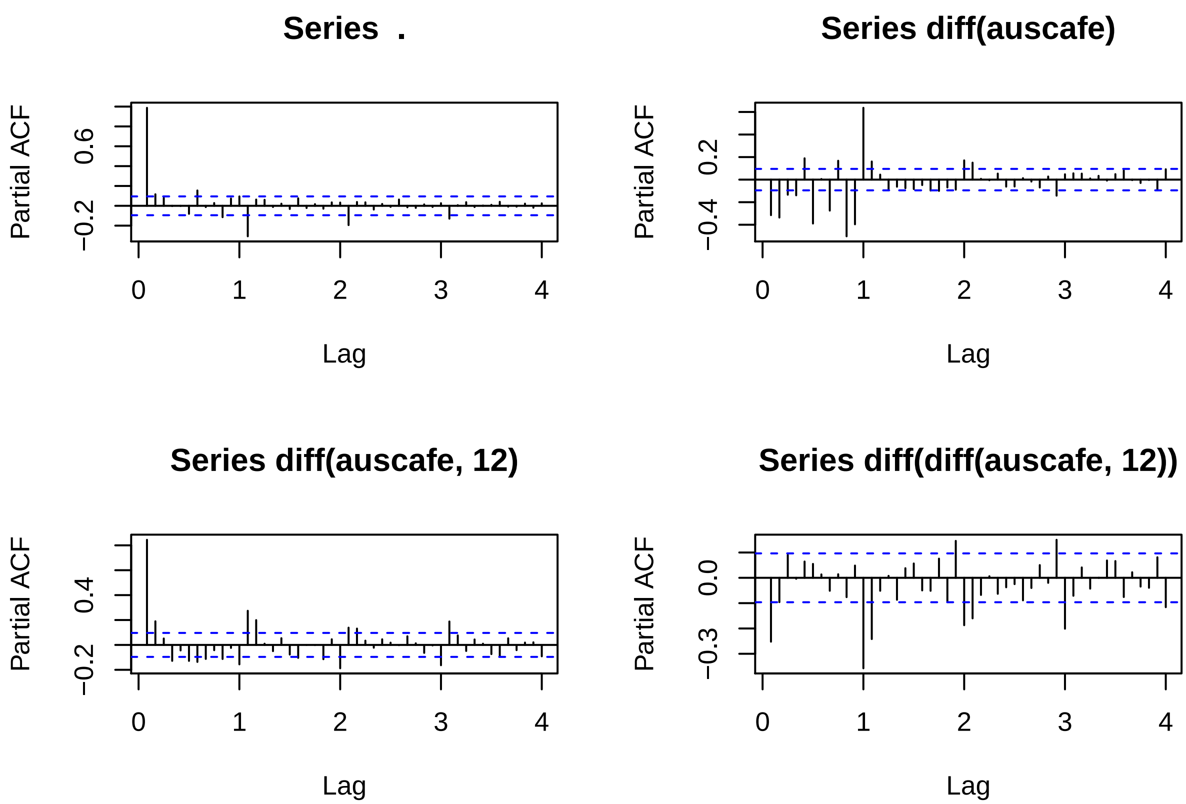


Também iremos analisar as ACF e PACF.

```r
par(mfrow = c(2, 2))
auscafe %>% BoxCox(lambda) %>% acf(lag.max = 48)
auscafe %>% BoxCox(lambda) %>% diff() %>% acf(lag.max = 48, main = "Series diff(auscafe)")
auscafe %>% BoxCox(lambda) %>% diff(lag = 12) %>%  acf(lag.max = 48, main = "Series diff(auscafe, 12)")
auscafe %>% BoxCox(lambda) %>% diff(lag = 12) %>% diff() %>% acf(lag.max = 48, main = "Series diff(diff
```




```
par(mfrow = c(2, 2))
auscafe %>% BoxCox(lambda) %>% pacf(lag.max = 48)
auscafe %>% BoxCox(lambda) %>% diff() %>% pacf(lag.max = 48, main = "Series diff(auscafe)")
auscafe %>% BoxCox(lambda) %>% diff(lag = 12) %>%  pacf(lag.max = 48, main = "Series diff(auscafe, 12)"
auscafe %>% BoxCox(lambda) %>% diff(lag = 12) %>% diff() %>% pacf(lag.max = 48, main = "Series diff(dif
```

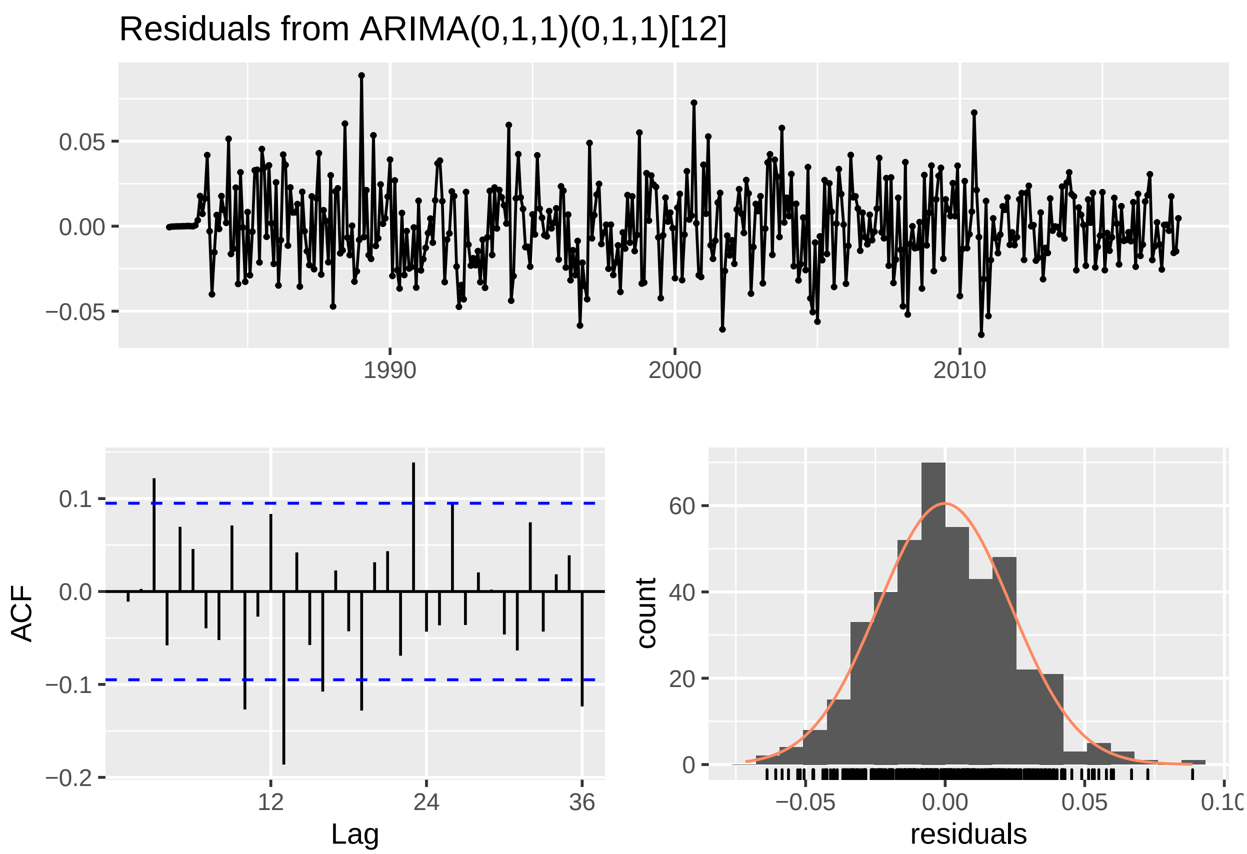


Com a análise dos gráficos, vemos inicialmente que devemos utilizar $d = 1$ e $D = 1$ pois a série apresenta tendência mensal e também anual. Vemos que a ACF tanto sazonal quando não sazonal decresce gradulmente, enquanto as PACF decrescem mais rapidamente. Vamos considerar inicialmente o modelo $SARIMA(0,1,1)(0,1,1)$.

```
mod1 <- Arima(auscafe, order = c(0, 1, 1), seasonal = c(0, 1, 1), lambda = lambda)
summary(mod1)
```

```
## Series: auscafe
## ARIMA(0,1,1)(0,1,1)[12]
## Box Cox transformation: lambda= 0.109056
##
## Coefficients:
##           ma1     sma1
##       -0.3649  -0.8204
## s.e.   0.0431   0.0325
##
## sigma^2 estimated as 0.0005856:  log likelihood=945.15
## AIC=-1884.3   AICc=-1884.24   BIC=-1872.23
##
## Training set error measures:
##                         ME       RMSE        MAE         MPE     MAPE      MASE
## Training set -0.0007623818 0.03690703 0.02723568 -0.03770534 1.833843 0.2606623
##                      ACF1
## Training set -0.006901538
```

```r
checkresiduals(mod1)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,1)(0,1,1)[12]
## Q* = 69.58, df = 22, p-value = 7.7e-07
##
## Model df: 2.   Total lags used: 24
```

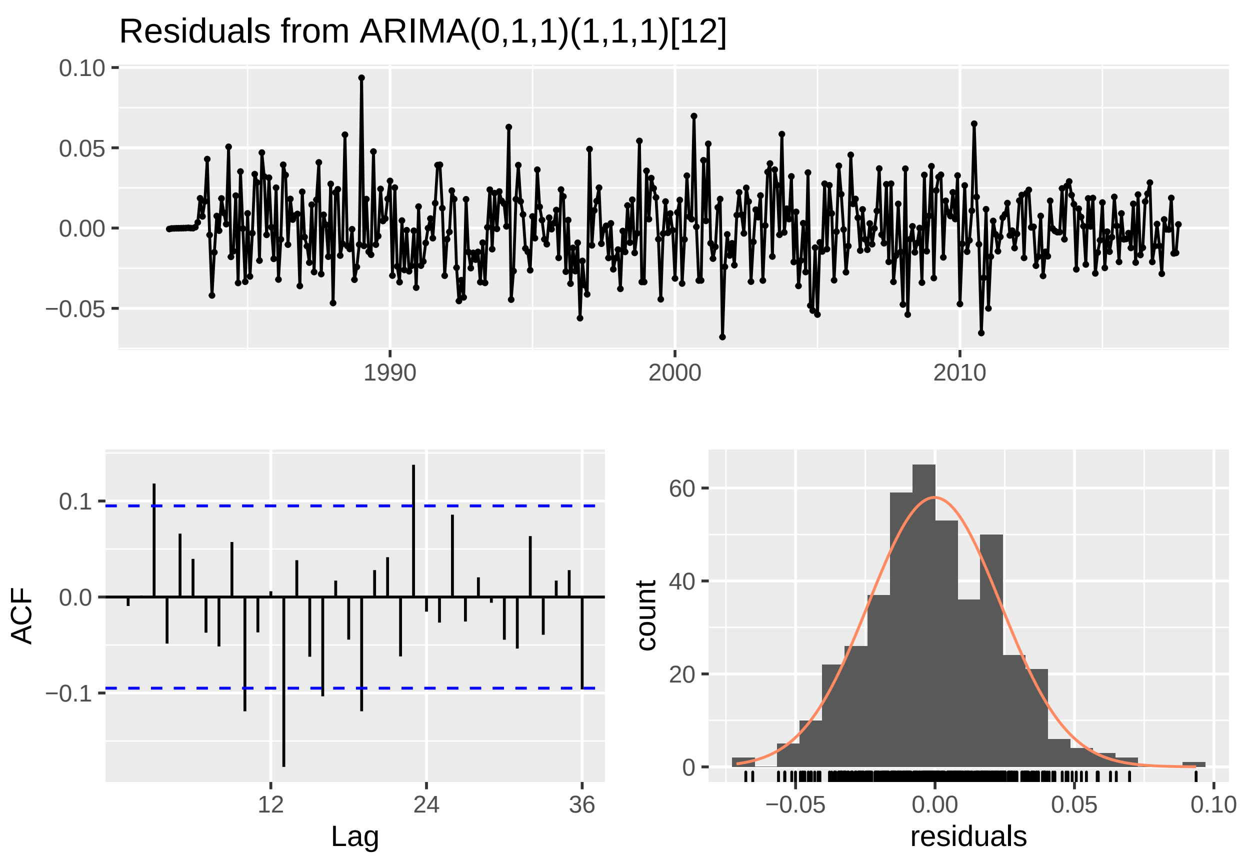
Note que a ACF dos resíduos apresenta alguns lags significativos, vamos tentar inserir um termo regressivo sazonal.

```r
mod2 <- Arima(auscafe, order = c(0, 1, 1), seasonal = c(1, 1, 1), lambda = lambda)
summary(mod2)
```

```
## Series: auscafe
## ARIMA(0,1,1)(1,1,1)[12]
## Box Cox transformation: lambda= 0.109056
##
## Coefficients:
##           ma1    sar1     sma1
##       -0.3414  0.1251  -0.8606
## s.e.   0.0451  0.0611   0.0343
```

```
##
## sigma^2 estimated as 0.0005808:  log likelihood=947.26
## AIC=-1886.51   AICc=-1886.42   BIC=-1870.42
##
## Training set error measures:
##                          ME       RMSE       MAE         MPE     MAPE      MASE
## Training set -0.0008127811 0.03702901 0.0270468 -0.03920067 1.809829 0.2588545
##                     ACF1
## Training set 0.001043405
```

```
checkresiduals(mod2)
```



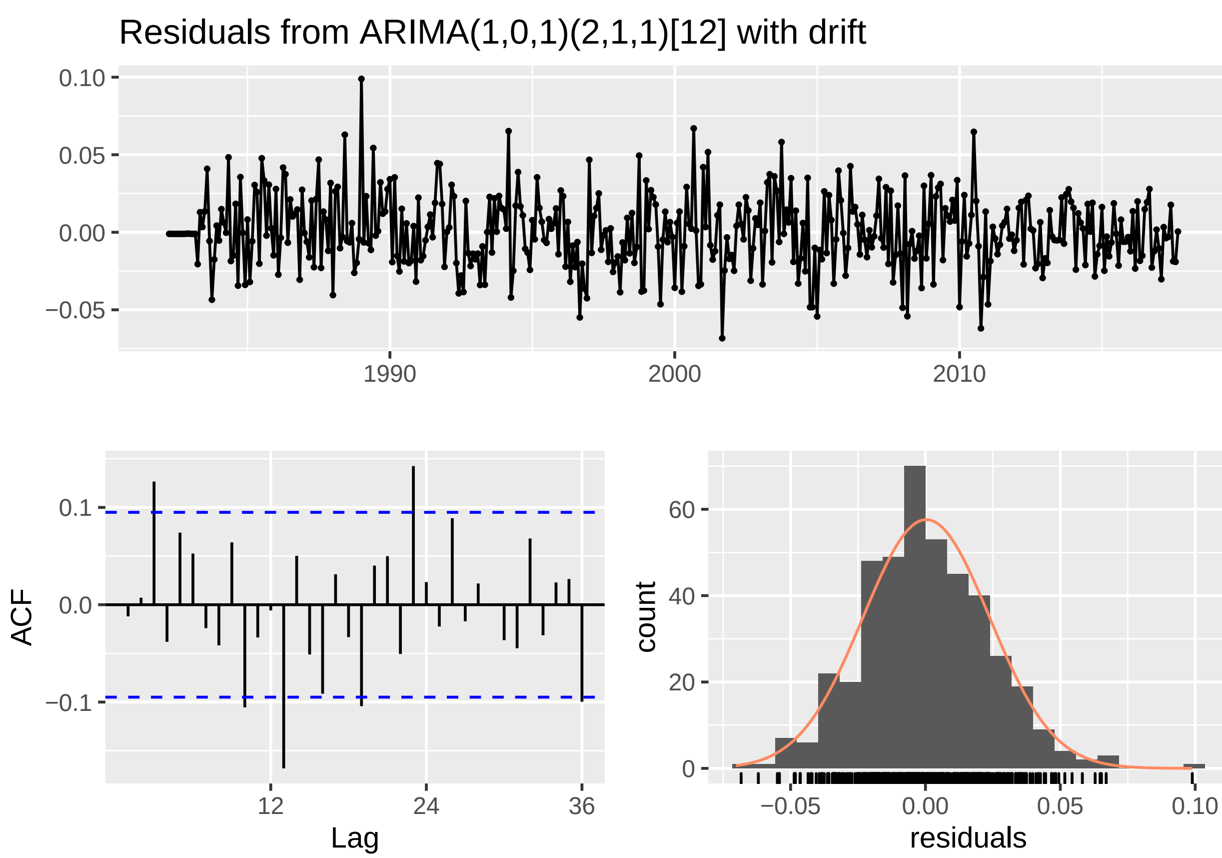

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,1)(1,1,1)[12]
## Q* = 59.51, df = 21, p-value = 1.515e-05
##
## Model df: 3.   Total lags used: 24
```

Comparando agora com o resultado do método auto.Arima.

```
mod3 <- auto.arima(auscafe, lambda = lambda)
summary(mod3)
```

```
## Series: auscafe
## ARIMA(1,0,1)(2,1,1)[12] with drift
## Box Cox transformation: lambda= 0.109056
##
## Coefficients:
##          ar1      ma1     sar1      sar2     sma1   drift
##       0.9718  -0.3190  0.1270  -0.0527  -0.8423  0.0056
## s.e.  0.0131   0.0478  0.0649   0.0585   0.0431  0.0004
##
## sigma^2 estimated as 0.0005754:  log likelihood=952.96
## AIC=-1891.92   AICc=-1891.65   BIC=-1863.74
##
## Training set error measures:
##                         ME       RMSE        MAE        MPE    MAPE      MASE
## Training set -0.0008622186 0.03664217 0.02686113 0.01635299 1.79394 0.2570775
##                     ACF1
## Training set -0.00872097
```

```
checkresiduals(mod3)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,0,1)(2,1,1)[12] with drift
## Q* = 56.069, df = 18, p-value = 8.692e-06
```

```
##
## Model df: 6.   Total lags used: 24
```

Vemos que diferentemente do que modelamos, o método auto.Arima obtive $p = 1$ e $P = 2$ sendo o maior AIC obtido. Vamos utilizar então do modelo $SARIMA(1,0,1)(2,1,1)$ para prever o próximo ano.

```
plot(forecast(mod3, h = 12), include =  48)
```