TESTE\_RECEITA

Juliana Santiago

6 de agosto de 2018

## TESTE RECEITA\_RJ

library(readxl)

## Warning: package 'readxl' was built under R version 3.3.3

library(tseries)

## Warning: package 'tseries' was built under R version 3.3.3

library(ggplot2)  
library(forecast)

## Warning: package 'forecast' was built under R version 3.3.3

##   
## Attaching package: 'forecast'

## The following object is masked from 'package:ggplot2':  
##   
## autolayer

library(foreign)  
library(urca)

## Warning: package 'urca' was built under R version 3.3.3

library(tidyquant)

## Loading required package: lubridate

## Warning: package 'lubridate' was built under R version 3.3.3

##   
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':  
##   
## date

## Loading required package: PerformanceAnalytics

## Warning: package 'PerformanceAnalytics' was built under R version 3.3.3

## Loading required package: xts

## Warning: package 'xts' was built under R version 3.3.3

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 3.3.3

##   
## Attaching package: 'zoo'

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

##   
## Attaching package: 'PerformanceAnalytics'

## The following object is masked from 'package:graphics':  
##   
## legend

## Loading required package: quantmod

## Warning: package 'quantmod' was built under R version 3.3.3

## Loading required package: TTR

## Warning: package 'TTR' was built under R version 3.3.3

## Version 0.4-0 included new data defaults. See ?getSymbols.

## Loading required package: tidyverse

## Warning: package 'tidyverse' was built under R version 3.3.3

## -- Attaching packages -------------------------------------------------------------------------------- tidyverse 1.2.1 --

## v tibble 1.4.2 v purrr 0.2.4  
## v tidyr 0.8.0 v dplyr 0.7.4  
## v readr 1.1.1 v stringr 1.3.1  
## v tibble 1.4.2 v forcats 0.3.0

## Warning: package 'tibble' was built under R version 3.3.3

## Warning: package 'tidyr' was built under R version 3.3.3

## Warning: package 'readr' was built under R version 3.3.3

## Warning: package 'purrr' was built under R version 3.3.3

## Warning: package 'dplyr' was built under R version 3.3.3

## Warning: package 'forcats' was built under R version 3.3.3

## -- Conflicts ----------------------------------------------------------------------------------- tidyverse\_conflicts() --  
## x lubridate::as.difftime() masks base::as.difftime()  
## x lubridate::date() masks base::date()  
## x dplyr::filter() masks stats::filter()  
## x dplyr::first() masks xts::first()  
## x lubridate::intersect() masks base::intersect()  
## x dplyr::lag() masks stats::lag()  
## x dplyr::last() masks xts::last()  
## x lubridate::setdiff() masks base::setdiff()  
## x lubridate::union() masks base::union()

library(TSstudio)  
library(moments)

## Warning: package 'moments' was built under R version 3.3.2

##   
## Attaching package: 'moments'

## The following objects are masked from 'package:PerformanceAnalytics':  
##   
## kurtosis, skewness

getwd()

## [1] "C:/Users/julia/Documents"

## Importando do Excel os dados mensais de receita bruta do Estado do Rio de Janeiro:

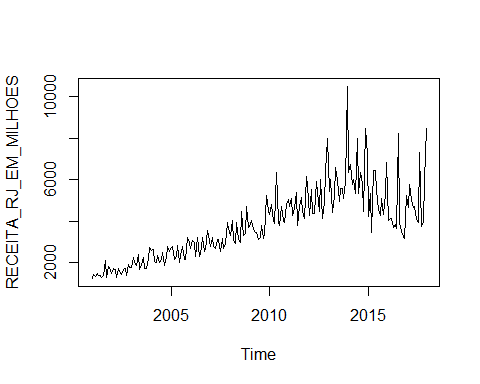
RECEITA\_RJ\_R <- read\_excel("~/UFF MESTRADO/4º SEMESTRE 2018.1/Regressão/RECEITA\_RJ\_R.xlsx")  
  
View(RECEITA\_RJ\_R)

## Avisar ao R que se trata de uma sÃ©rie de tempo com frequÃªncia mensal e inÃ?cio em jan/2001:

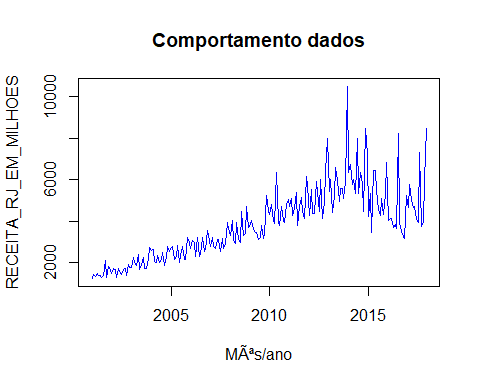
serie\_receita<- ts(RECEITA\_RJ\_R, start = c(2001, 1), frequency = 12)

## Visualizando os dados ao longo do tempo:

plot(serie\_receita)



plot(serie\_receita, col='blue', main = 'Comportamento dados', xlab = 'MÃªs/ano')



## Observando a plotagem dos dados, a série parece ser não estacionária, pois não apresenta uma regularidade ao longo do tempo.

## EstatÃ?stica bÃ¡sica:

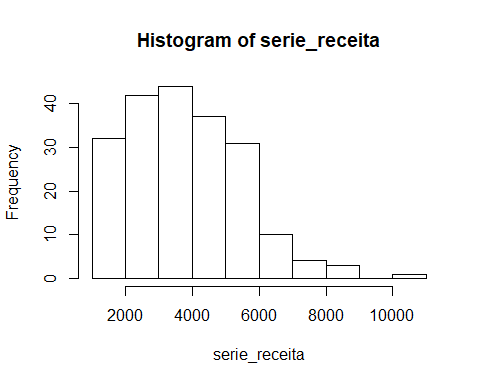
summary(serie\_receita)

## RECEITA\_RJ\_EM\_MILHOES  
## Min. : 1239   
## 1st Qu.: 2457   
## Median : 3678   
## Mean : 3799   
## 3rd Qu.: 4836   
## Max. :10493

## A proximidade dos valores de mÃ©dia e mediana sugere uma simetria da distribuiÃ§Ã£o dos dados.

## HISTOGRAMA

hist(serie\_receita)



## ASSIMETRIA

## Ãndice de assimetria (skewness). Se |Ia|<= 0,15, a distribuiÃ§Ã£o dos dados Ã©, aproximadamente, simÃ©trica. Se 0,15<|Ia|<=1, a distribuiÃ§Ã£o Ã© moderadamente assimÃ©trica. Se |Ia|>1, a distribuiÃ§Ã£o Ã© altamente assimÃ©trica.

skewness(serie\_receita)

## RECEITA\_RJ\_EM\_MILHOES   
## 0.7381396

## |Ia|=0,7381 (0,15<|Ia|<=1), o que sugere que a distribuiÃ§Ã£o de Receita\_RJ Ã© moderadamente assimÃ©trica.

## CURTOSE

## Curtose (Ic) tambÃ©m Ã© uma medida de assimetria. Se Ic = 0,263, a distribuiÃ§Ã£o Ã© mesocÃºrtica. Se Ic < 0,263, a distribuiÃ§Ã£o Ã© leptocÃºrtica. Se Ic>0,263, a distribuiÃ§Ã£o Ã© platicÃºrtica.

kurtosis(serie\_receita)

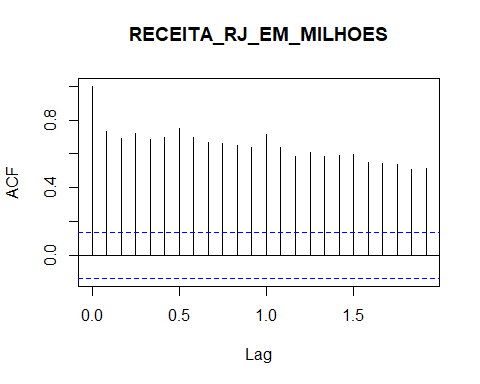
## RECEITA\_RJ\_EM\_MILHOES   
## 3.582738

## Ic = 3,58 (>0,263), o que sugere que a distribuiÃ§Ã£o Ã© platicÃºrtica (com curva mais achatada).

## CORRELOGRAMA

## O correlograma Ã© o grÃ¡fico utilizado em sÃ©ries temporais para traÃ§ar as autocorrelaÃ§Ãµes (tambÃ©m chamadas em inglÃªs de ACF = autocorrelation function) em diversas defasagens. A anÃ¡lise desse grÃ¡fico permite entender se a sÃ©rie Ã© aleatÃ³ria ou possui alguma tendÃªncia ou sazonalidade.

acf(serie\_receita)



## No grÃ¡fico, o eixo vertical indica a autocorrelaÃ§Ã£o e o horizontal a defasagem. A linha tracejada azul indica onde Ã© significativamente diferente de zero. Como Ã© possÃ?vel ver na imagem,a maior parte dos valores ACF estÃ£o fora do limite da linha tracejada azul, Ou seja, autocorrelaÃ§Ã£o diferente de zero, indicando que a sÃ©rie nÃ£o Ã© aleatÃ³ria e que pode haver sazonalidade.

## De acordo com a função de autocovariância, a série é integrada, ou seja, não estacionária.

## ESTACIONARIDADE

## 1. Augmented Dickey-Fuller (ADF)

## H0 : tem raiz unitÃ¡ria (nÃ£o Ã© estacionÃ¡ria)

## H1 : nÃ£o tem raiz unitÃ¡ria (Ã© estacionÃ¡ria)

## Assim, se o p-valor for menor do que 5%, p<0,05, entÃ£o rejeita-se a nÃ£o-estacionariedade. JÃ¡ se p>0,05, aceita-se a nÃ£o-estacionariedade.

## Kwiatkowski, Phillips, Schmidt and Shin (1992) propuseram um teste para identificação da tendência e/ou nível estacionário.

## Temos então:

## H0 : estacionária

## HA : não estacionária

adf.test(serie\_receita)

##   
## Augmented Dickey-Fuller Test  
##   
## data: serie\_receita  
## Dickey-Fuller = -2.3543, Lag order = 5, p-value = 0.4276  
## alternative hypothesis: stationary

kpss.test(serie\_receita)

## Warning in kpss.test(serie\_receita): p-value smaller than printed p-value

##   
## KPSS Test for Level Stationarity  
##   
## data: serie\_receita  
## KPSS Level = 4.4723, Truncation lag parameter = 3, p-value = 0.01

## De acordo com o teste ADF, a série é não estacionária, pois o p-valor é alto (>0,05), logo, aceita-se a hipótese nula de que a série é não estacionária. E o teste KPSS corrobora isso, uma vez que o p-valor é baixo (<0,05), logo aceita-se a hipótese alternativa de não estacionaridade da série.

library(fpp)

## Warning: package 'fpp' was built under R version 3.3.3

## Loading required package: fma

## Warning: package 'fma' was built under R version 3.3.3

## Loading required package: expsmooth

## Warning: package 'expsmooth' was built under R version 3.3.3

## Loading required package: lmtest

## Warning: package 'lmtest' was built under R version 3.3.3

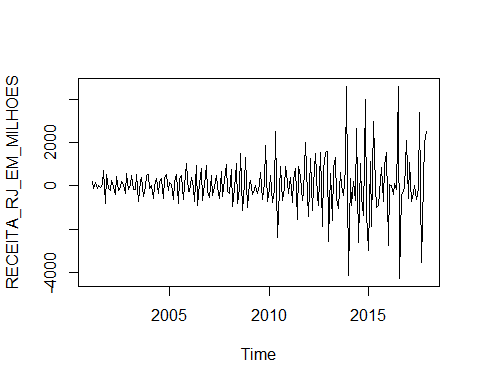
modelo <- auto.arima(serie\_receita)  
modelo

## Series: serie\_receita   
## ARIMA(0,1,1)(0,0,2)[12] with drift   
##   
## Coefficients:  
## ma1 sma1 sma2 drift  
## -0.8455 0.2683 0.2210 22.5468  
## s.e. 0.0380 0.0796 0.0838 13.7891  
##   
## sigma^2 estimated as 737143: log likelihood=-1658.83  
## AIC=3327.66 AICc=3327.97 BIC=3344.23

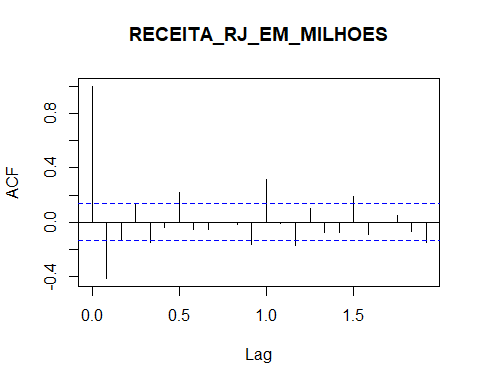
## Aplicando-se a função autoarima, obtém-se um arima (0,1,1).

## Como a série é não estacionária, efetua-se a diferenciação da mesma.

dados\_ts.diff <- diff(serie\_receita)  
plot.ts(dados\_ts.diff)



acf(dados\_ts.diff)



adf.test(dados\_ts.diff)

## Warning in adf.test(dados\_ts.diff): p-value smaller than printed p-value

##   
## Augmented Dickey-Fuller Test  
##   
## data: dados\_ts.diff  
## Dickey-Fuller = -10.593, Lag order = 5, p-value = 0.01  
## alternative hypothesis: stationary

kpss.test(dados\_ts.diff)

## Warning in kpss.test(dados\_ts.diff): p-value greater than printed p-value

##   
## KPSS Test for Level Stationarity  
##   
## data: dados\_ts.diff  
## KPSS Level = 0.029726, Truncation lag parameter = 3, p-value = 0.1

## Observando a plotagem da série em primeira diferença, a mesma parece ser estacionária, pois aparenta um ruído.

## De acordo com o teste ADF, a série é estacionária, pois o p-valor é baixo (<0,05), logo, rejeita-se a hipótese nula de que a série é não estacionária. E o teste KPSS corrobora isso, uma vez que o p-valor é 0,1 (>0,05), logo aceita-se a hipótese nula de estacionaridade da série. Conforme indicado pelo resultado do autoarima.

## Testando um modelo adjacente

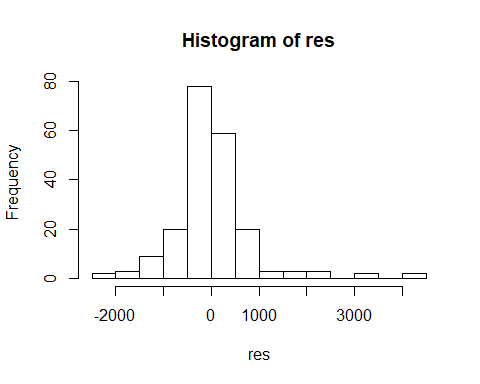
adj <- arima(serie\_receita, order=c(0,0,2))  
adj

##   
## Call:  
## arima(x = serie\_receita, order = c(0, 0, 2))  
##   
## Coefficients:  
## ma1 ma2 intercept  
## 0.6126 0.2666 3806.9441  
## s.e. 0.0760 0.0546 167.4304  
##   
## sigma^2 estimated as 1628989: log likelihood = -1748.62, aic = 3505.25

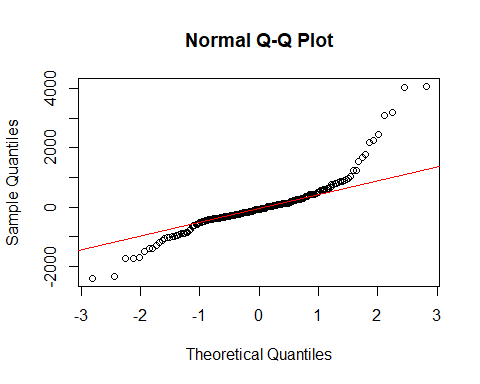
## Ao testar um modelo adjacente de arrima, percebe-se que este obteve um pior resultado. Portanto, usarei o modelo sugerido pelo autorima (0,1,1).

## RESÍDUOS

res <- modelo$residuals  
hist(res,20)



qqnorm(res)  
qqline(res, col="red")



## Observando o histograma dos resíduos, os mesmos parecem ter alguma discrepância nas caldas da distribuição, sugerindo a não normalidade da série. E olhando para o resultando do qqplot, percebe-se que os pontos seguem a linha, mas há maior discrepância nas caldas da distribuição, o que corrobora pra que a série não tenha distribuição normal.

jarque.bera.test(res)

##   
## Jarque Bera Test  
##   
## data: res  
## X-squared = 405.86, df = 2, p-value < 2.2e-16

## O teste de Jarque-Bera tem como hipÃ³tese nula a normalidade. Assim, se o p-valor for menor do que 5%, p<0,05, entÃ£o rejeita-se a normalidade. JÃ¡ se p>0,05, aceita-se a normalidade.

## Nesse caso, o pvalor é muito baixo (>0,05), ou seja, para um nÃ?vel de confianÃ§a de 95%, a sÃ©rie nÃ£o possui distribuiÃ§Ã£o normal.

Box.test(res, type="Ljung")

##   
## Box-Ljung test  
##   
## data: res  
## X-squared = 0.078632, df = 1, p-value = 0.7792

## Como o p-valor do Box test é alto (>0,05), aceita-se a hipótese nula de que os valores de autocorrelação dos resíduos são nulos, ou seja, não são significantes.

## PREVISÃO

modelo.forescast<-forecast(modelo, h=24)  
plot(modelo.forescast)

