TIME SERIES ANALYSIS USING R

We will be installing the library packages forecast, sweep, timetk and ggplot. The forecast function provides automatic selection for ARIMA models.

library(forecast)

## Registered S3 method overwritten by 'xts':  
## method from  
## as.zoo.xts zoo

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

## Registered S3 methods overwritten by 'forecast':  
## method from   
## fitted.fracdiff fracdiff  
## residuals.fracdiff fracdiff

library(sweep)  
library(timetk)  
library(ggplot2)

here we are taking the air passengers data from the year 1949 to 1961

data("AirPassengers")  
MF <- AirPassengers  
str(MF)

## Time-Series [1:144] from 1949 to 1961: 112 118 132 129 121 135 148 148 136 119 ...

head(MF)

## Jan Feb Mar Apr May Jun  
## 1949 112 118 132 129 121 135

ts(MF, frequency = 12, start = c(1949,1))

## Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec  
## 1949 112 118 132 129 121 135 148 148 136 119 104 118  
## 1950 115 126 141 135 125 149 170 170 158 133 114 140  
## 1951 145 150 178 163 172 178 199 199 184 162 146 166  
## 1952 171 180 193 181 183 218 230 242 209 191 172 194  
## 1953 196 196 236 235 229 243 264 272 237 211 180 201  
## 1954 204 188 235 227 234 264 302 293 259 229 203 229  
## 1955 242 233 267 269 270 315 364 347 312 274 237 278  
## 1956 284 277 317 313 318 374 413 405 355 306 271 306  
## 1957 315 301 356 348 355 422 465 467 404 347 305 336  
## 1958 340 318 362 348 363 435 491 505 404 359 310 337  
## 1959 360 342 406 396 420 472 548 559 463 407 362 405  
## 1960 417 391 419 461 472 535 622 606 508 461 390 432

attributes(MF)

## $tsp  
## [1] 1949.000 1960.917 12.000  
##   
## $class  
## [1] "ts"

AP <- log(MF)  
plot(MF)

A close up of a logo

Description automatically generated

From the graph we can see that it’s a multiplicate seasonal changes, here the magnitude of seasonal change increases over time as data value increases

#Decomposition of additive time series   
decomp <- decompose(MF)  
decomp$figure

## [1] -24.748737 -36.188131 -2.241162 -8.036616 -4.506313 35.402778  
## [7] 63.830808 62.823232 16.520202 -20.642677 -53.593434 -28.619949

plot(decomp$figure,  
 type = 'b',  
 xlab = 'Month',  
 ylab = 'Seasonality Index',  
 col = 'blue',  
 las = 2)

we use decomposition

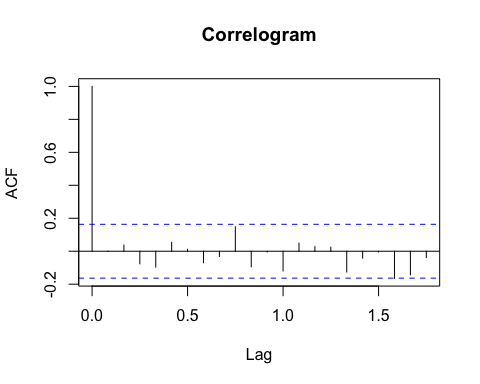
A close up of a map

Description automatically generated

#ARIMA - Autoregressive Integrated Moving Avergae  
  
x <- forecast::auto.arima(MF)  
attributes(forecast::auto.arima(MF))

## $names  
## [1] "coef" "sigma2" "var.coef" "mask" "loglik"   
## [6] "aic" "arma" "residuals" "call" "series"   
## [11] "code" "n.cond" "nobs" "model" "bic"   
## [16] "aicc" "x" "fitted"   
##   
## $class  
## [1] "forecast\_ARIMA" "ARIMA" "Arima"

#ACF and PACF Plots  
# We use residuals values as it is the difference between actual and fitted values  
  
acf(x$residuals, main = 'Correlogram')



pacf(x$residuals, main = 'Partial Correlogram')

A screenshot of a cell phone

Description automatically generated

#Ljung-Box Test  
  
Box.test(x$residuals, lag = 20, type = 'Ljung-Box')

##   
## Box-Ljung test  
##   
## data: x$residuals  
## X-squared = 22.524, df = 20, p-value = 0.3128

#Residual Plot  
  
hist(x$residuals,  
 col = 'red',  
 xlab = 'Error',  
 main = 'Histogram of residuals',  
 freq = FALSE)  
lines(density(x$residuals))

we can see that the acf is centered around 3

A close up of text on a black background

Description automatically generated

#Forecast  
  
ff <- forecast::forecast(x,48)  
ggplot2::autoplot(ff)

The residual plot is centred around 0, which indicates that the ARIMA model is a good fit

A screenshot of a cell phone

Description automatically generated

forecast::accuracy(ff)

## ME RMSE MAE MPE MAPE MASE  
## Training set 1.342306 10.84619 7.867539 0.4206996 2.800458 0.245628  
## ACF1  
## Training set

CONCLUSION

In this we use forecast and ARIMA MODEL, and understand the time-series analysis and plot it for air-passenger.