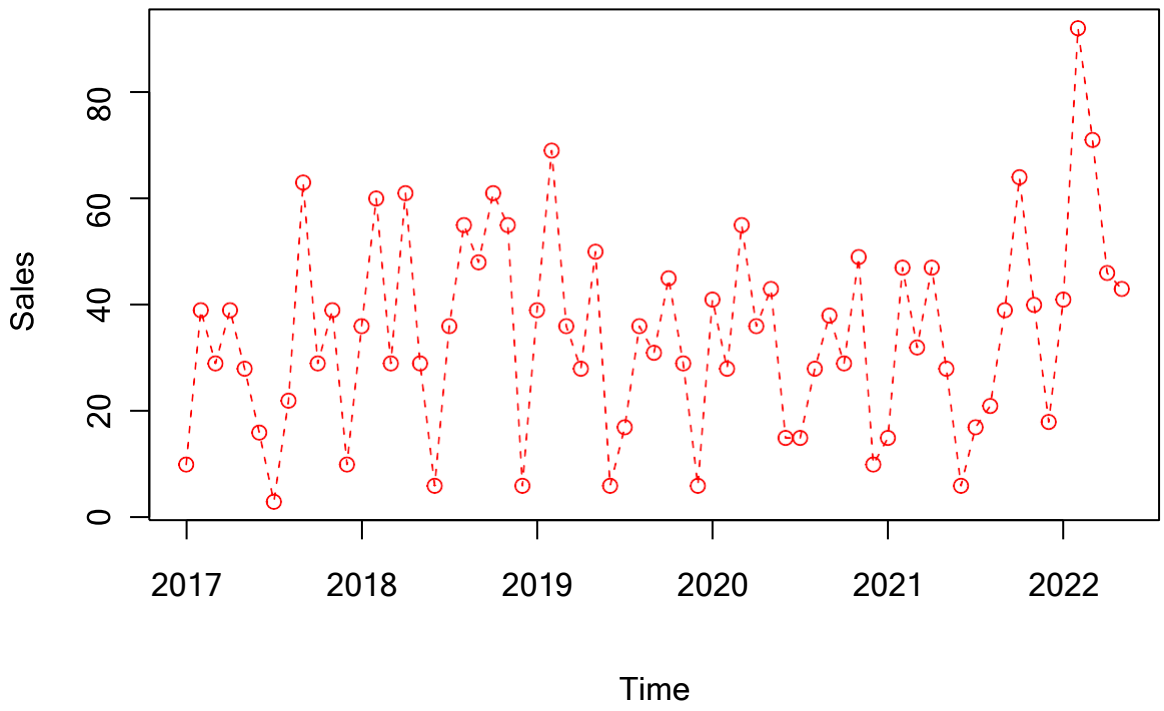
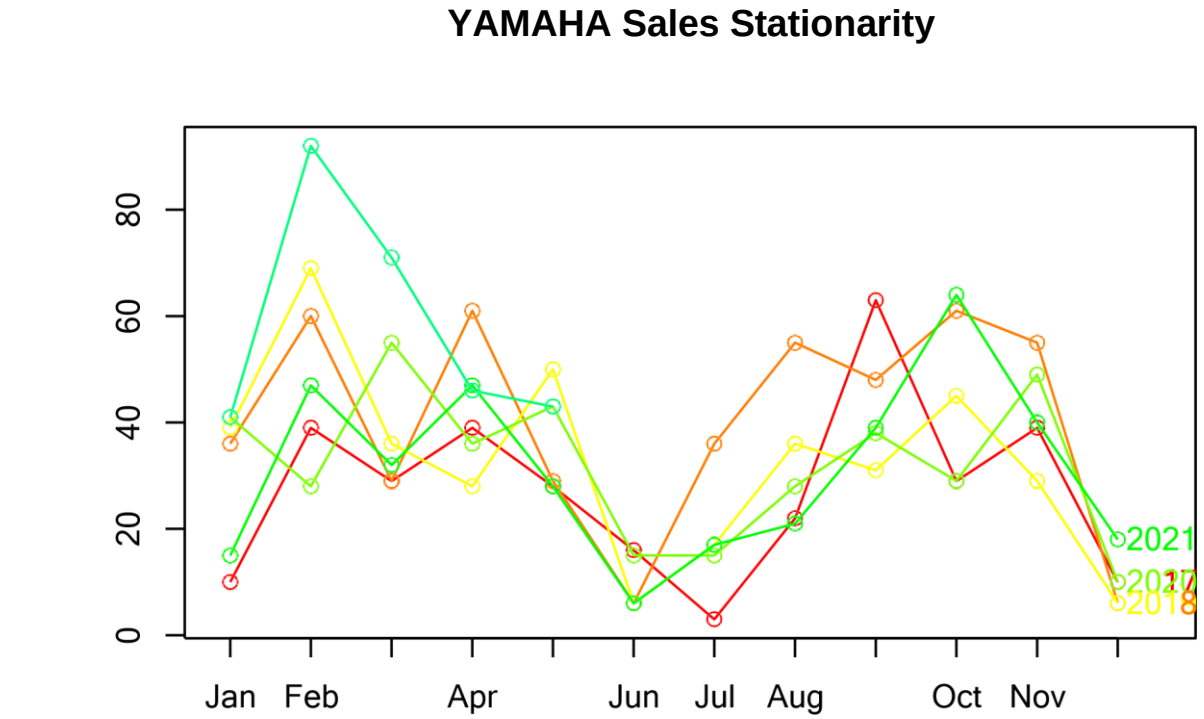
Week 10 Project Submission: Project Report Update

The ability of decision makers to predict the number of sales in the coming period is essential to be able to determine the procurement of goods more precisely. There is a method called Auto-Regressive Integrated Moving Average (ARIMA). This method is one model that can be used to forecast sales based on sales time series data in previous periods. Therefore, in this week the goal is perform an Autoregressive Integrated Moving Average (ARIMA) (p, d, q) forecasting model in time series analysis for predicting the sales volume of motorcycle in Colombia. This model uses monthly sales data from January 2017 to May 2022.

Motorcycles sales Jan 2017 – May 2022



First, we check whether the time series is stationary or not stationary.

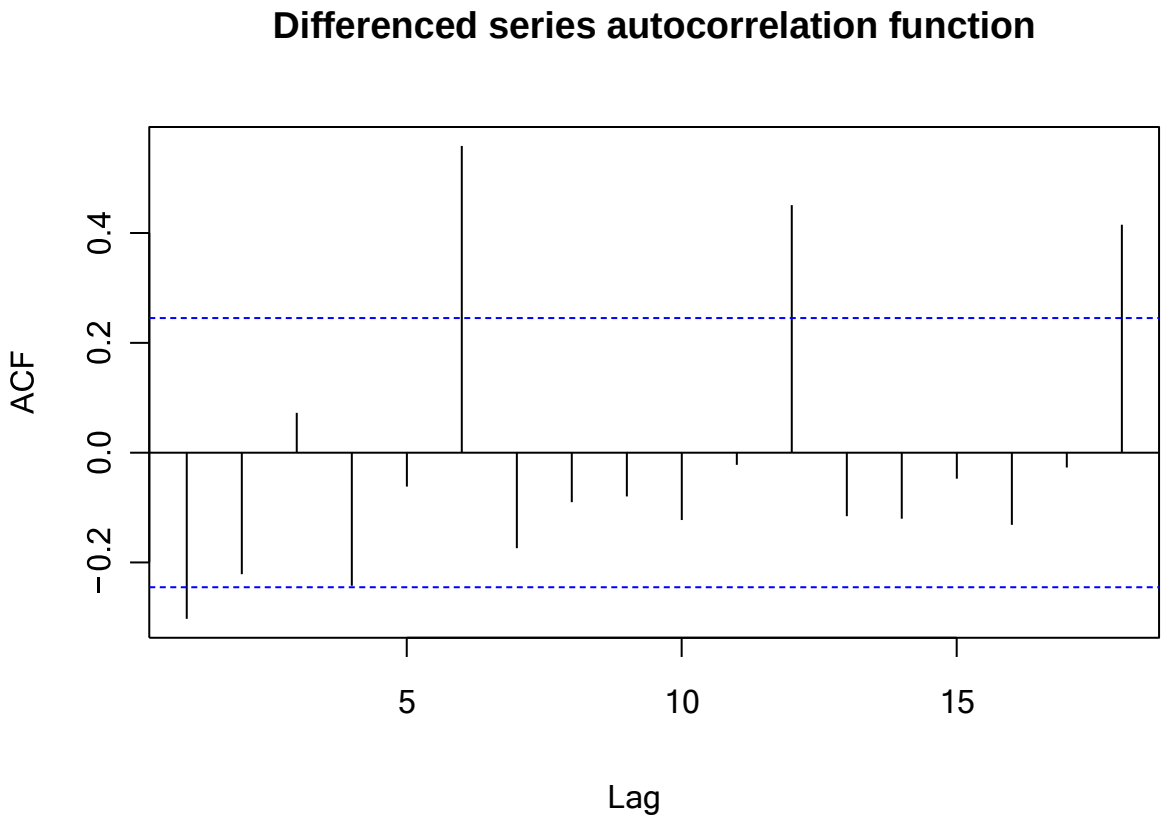


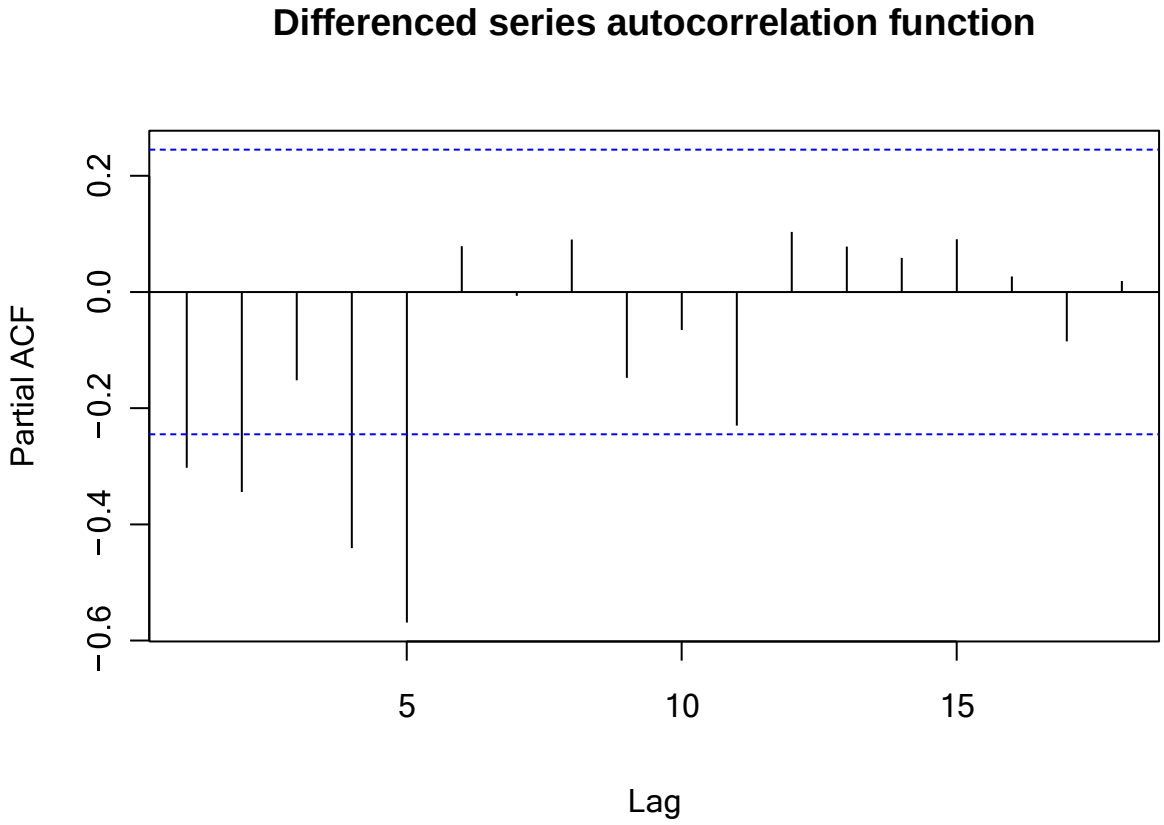
Graphically it seems to be stationary due to the ups and downs at the same months of the year. However, an augmented Dickey-Fuller test is performed to verify the stationarity.

```
Warning in adf.test(Ventasts): p-value      smaller than printed p-value

Augmented Dickey-Fuller Test

data: Ventasts
Dickey-Fuller = -4.2685, Lag order = 3,
alternative hypothesis: stationary
                                          p-value = 0.01
```

```
adf.test(varporcentual)
## Warning in adf.test(varporcentual): p-value smaller than printed p-value ##
## Augmented Dickey-Fuller Test
##
## data: varporcentual
## Dickey-Fuller = -8.1822, Lag order = 3, p-value = 0.01
## alternative hypothesis: stationary
```

As p value is lower than 0.05 null hypothesis is rejected so both the series and its difference are stationary.

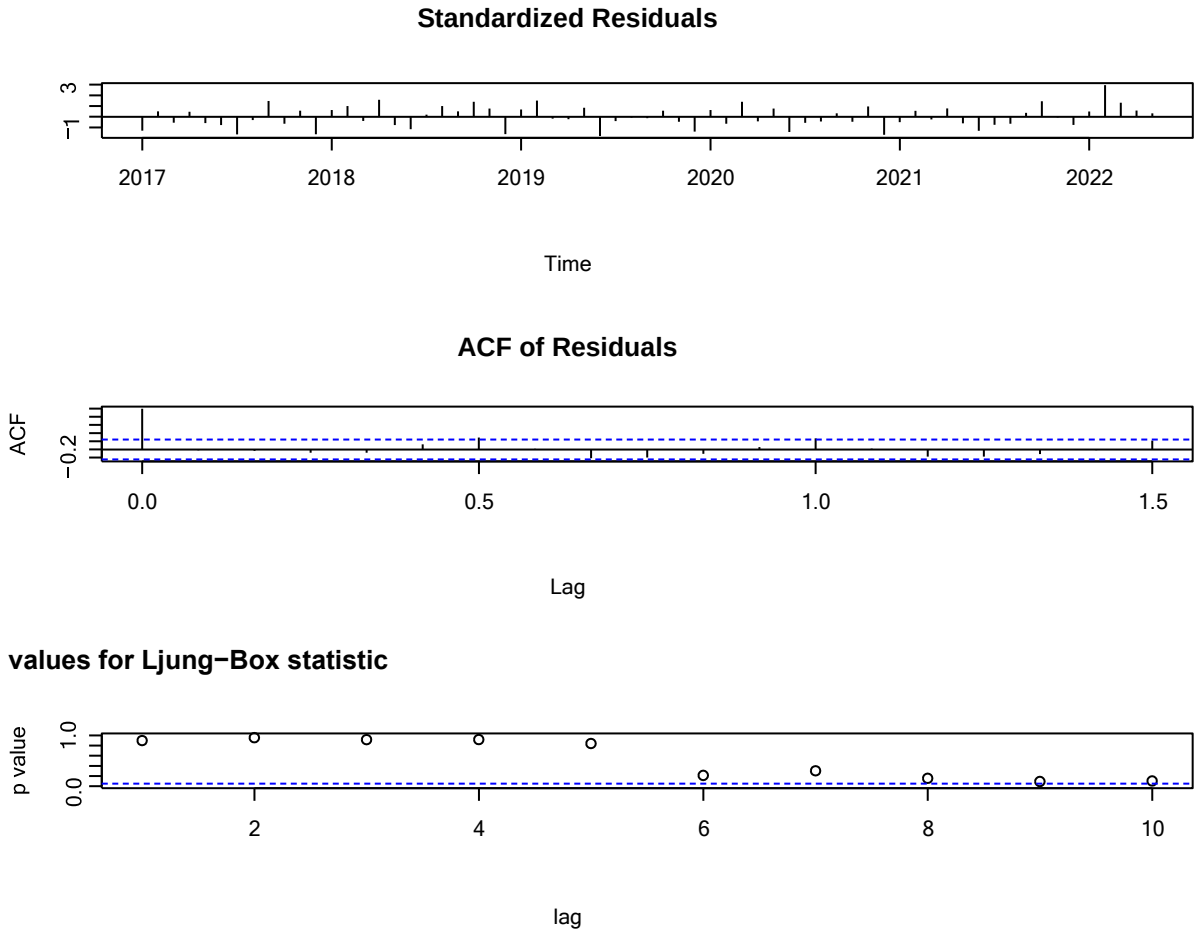

Differenced series autocorrelation function
ACF
0.4
0.2
0.0
−0.2
5
10
15
Lag


Differenced series autocorrelation function
Partial ACF
0.2
0.0
−0.2
−0.4
−0.6
5
10
15
Lag

Sales time series is transformed (its percentage change or difference), autocorrelation function is used to try to infer the order of the moving average component. On the other hand, the partial autocorrelation function is used to try to infer the order of the autoregressive component. A Box-Jung test is performed to validate the assumptions

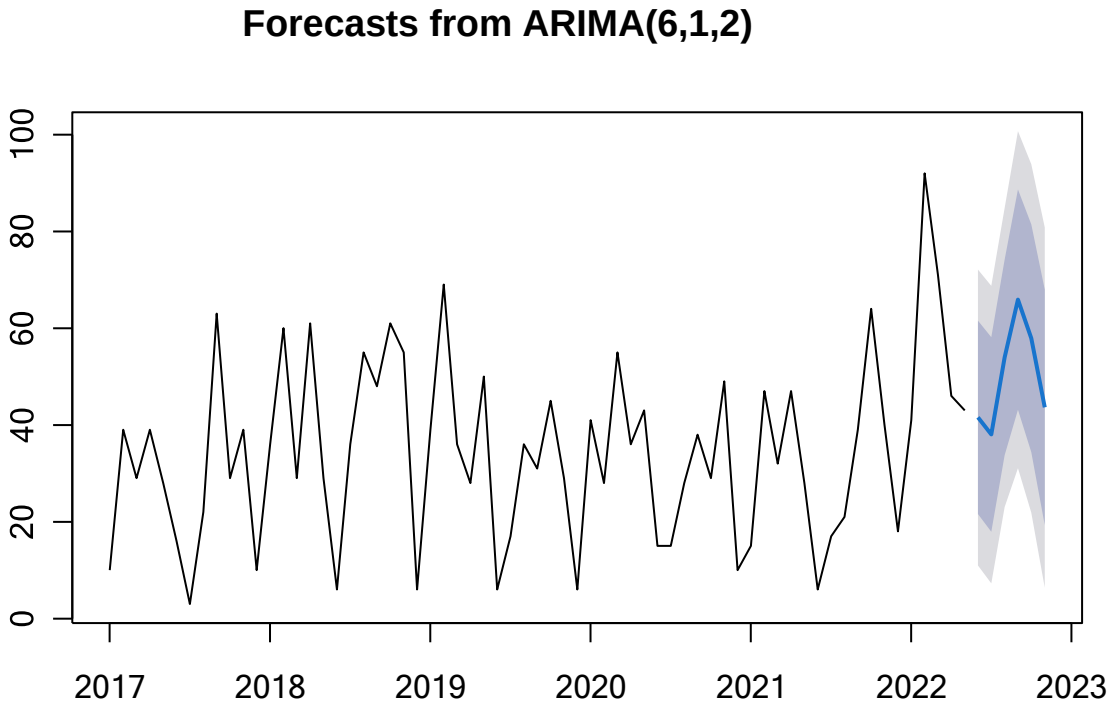
```
    Box-Ljung test
## data: residuals(modelo1)
## X-squared = 0.015958, df = 1, p-value = 0.8995
```

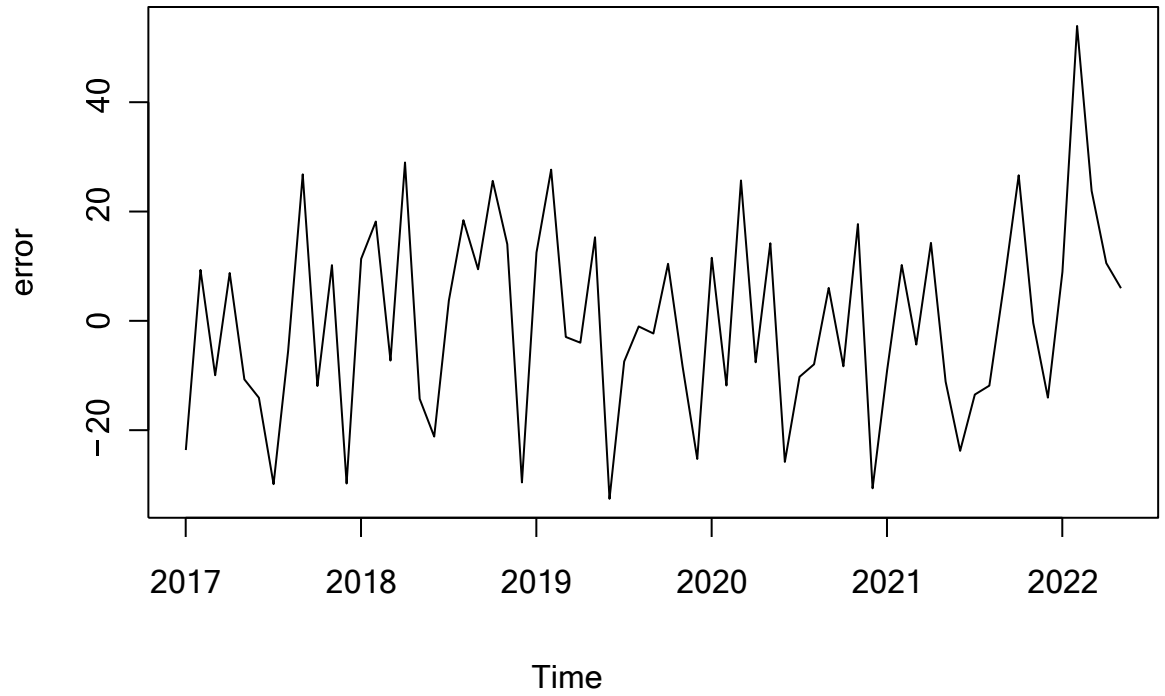
As p value is greater than the 5% alpha there is not enough statistical evidence to reject the null hypothesis, so the series is not autocorrelated.



Having found the likely candidates for p, d, and q in the autocorrelation functions, an ARIMA (6,1,2) model is estimated

```
## Series: Ventasts
## ARIMA(6,1,2)
##
## Coefficients:
##           ar1      ar2      ar3      ar4      ar5      ar6      ma1
##        0.1317  -0.8586  -0.3406  -0.5914  -0.1279  -0.0657  -1.0403
## s.e.  0.1385   0.1386   0.1599   0.1562   0.1412   0.1422   0.0676
##           ma2
## sigma^2 = 235.9:  log likelihood = -265.76
## AIC=549.53   AICc=552.86   BIC=568.96
        0.9998
        0.0838
```





As residuals are white noise (independent and identically distributed random variables) the model is correctly specified.

The Forecasts from ARIMA(6,1,2) for the next six months with highs and lows at 80 and 95 percent of confidence.

##	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jun 2022	41.56982	21.58192	61.55773	11.000964	72.13868
Jul 2022	38.03235	17.92995	58.13475	7.288386	68.77631
Aug 2022	53.93781	33.74299	74.13263	23.052492	84.82312
Sep 2022	65.90709	43.14662	88.66756	31.097953	100.71623
Oct 2022	57.90296	34.36593	81.44000	21.906172	93.89975
Nov 2022	43.62679	19.30035	67.95322	6.422710	80.830