ANALYSIS OF THE TIME SERIES BY THE BOX-JENKINS METHOD

Statistical analysis using R language

I. Input file

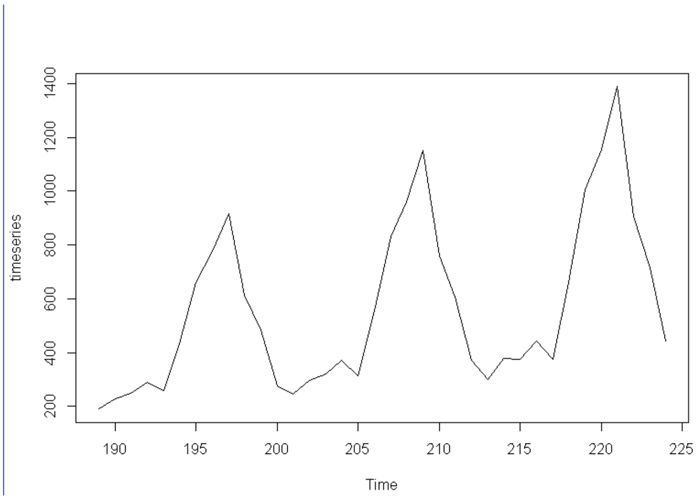
data<- scan(“C:/Documents/cola.dat”) – reading data from file

Read 36 items

> timeseries<-ts(data,frequency=1,start=c(189))
> timeseries

```
> timeseries<-ts(data,frequency=1,start=c(189))
>  timeseries
Time Series:
Start = 189
End = 224
Frequency = 1
 [1]  189  229  249  289  260  431  660  777  915  613  485  277  244  296  319
[16]  370  313  556  831  960 1152  759  607  371  298  378  373  443  374  660
[31] 1004 1153 1388  904  715  441
> 
```

>plot.ts(timeseries)



Picture 1. Time series plot

Series has ascending and seasonal effect. The trend is a 1st order polynomial.

Forming a new time series

Discard the last 12 members of the series and build the model ARIMA.

>dataNew<- scan("C:/Documents/cola_n.dat")

Read 24 items

> timeseriesNew<-ts(dataNew,frequency=1,start=c(189))

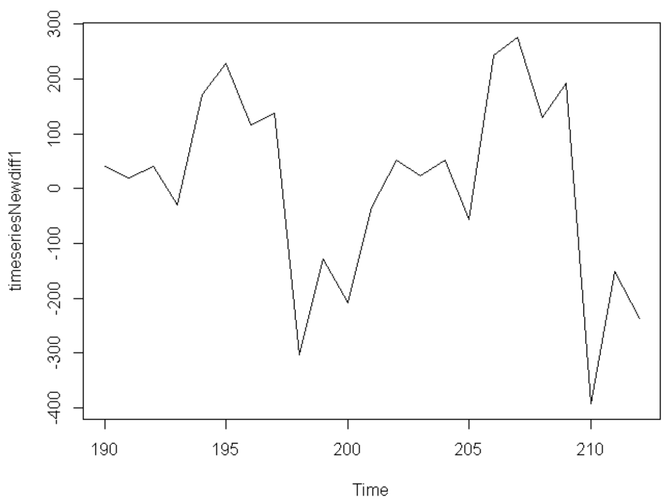
> timeseriesNew

```
> timeseriesNew<-ts(dataNew,frequency=1,start=c(189))
> timeseriesNew
Time Series:
Start = 189
End = 212
Frequency = 1
 [1]  189  229  249  289  260  431  660  777  915  613  485  277  244  296  319
[16]  370  313  556  831  960 1152  759  607  371
> 
```

II. Applying difference of the order of d=1

> timeseriesNewdiff1<-diff(timeseriesNew,differences=1)

>plot.ts(timeseriesNewdiff1)

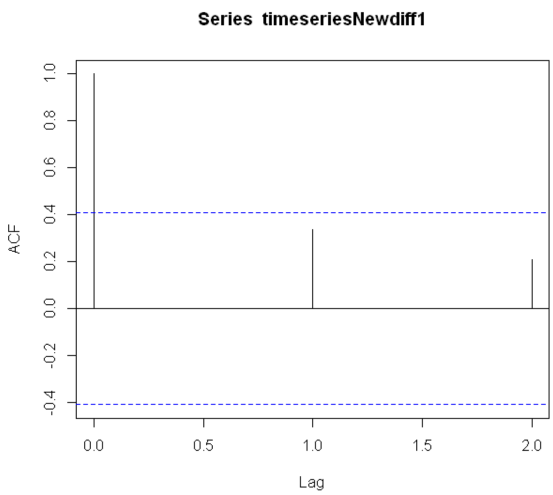


Picture. 2. Plot after applying the difference

Autocorrelation function

>acf(timeseriesNewdiff1, lag.max=2)

>acf(timeseriesNewdiff1, lag.max=2,plot=FALSE)



Picture 3. Autocorrelation function

```
> acf(timeseriesNewdiff1, lag.max=2)
> acf(timeseriesNewdiff1, lag.max=2,plot=FALSE)

Autocorrelations of series 'timeseriesNewdiff1', by lag

    0     1     2
1.000 0.337 0.209
>
```

Partial autocorrelation function

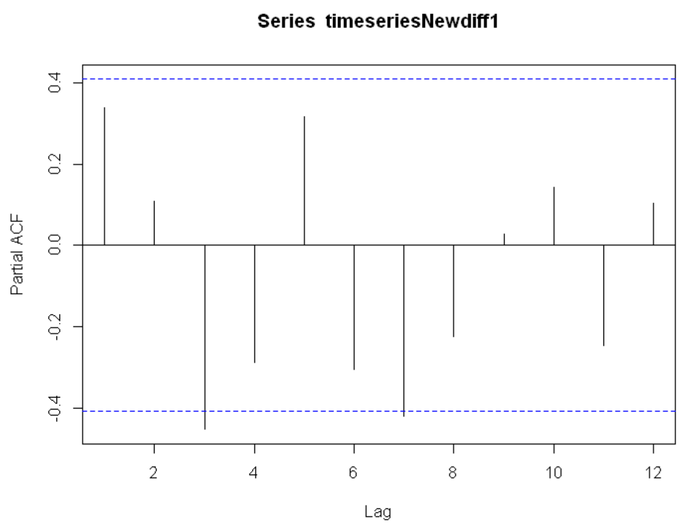
>pacf(timeseriesNewdiff1, lag.max=12)

>pacf(timeseriesNewdiff1, lag.max=12, plot=FALSE)

```
> pacf(timeseriesNewdiff1, lag.max=12)
> pacf(timeseriesNewdiff1, lag.max=12, plot=FALSE)

Partial autocorrelations of series 'timeseriesNewdiff1', by lag

     1      2      3      4      5      6      7      8      9     10     11
 0.337  0.107 -0.453 -0.287  0.316 -0.305 -0.420 -0.223  0.028  0.144 -0.247
    12
 0.104
>
```

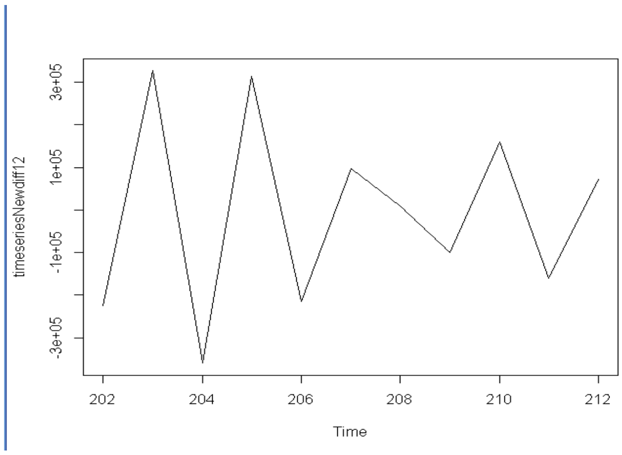


Picture 4. Partial autocorrelation function

III. Applying difference of the order of d=12

> timeseriesNewdiff12<-diff(timeseriesNewdiff1,differences=12)

> plot.ts(timeseriesNewdiff12)

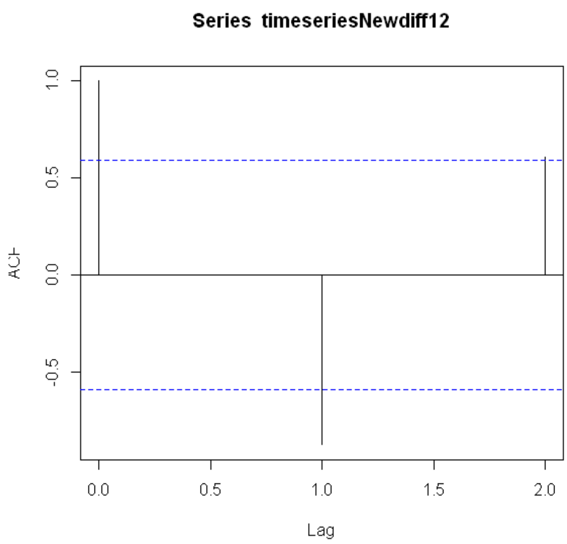


Picture 5. Plot after applying the difference

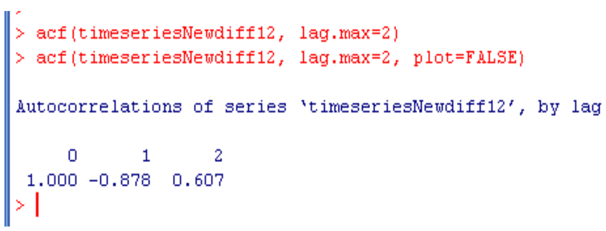
Autocorrelation function

>acf(timeseriesNewdiff12, lag.max=2)

>acf(timeseriesNewdiff12, lag.max=2, plot=FALSE)



Picture 6. Autocorrelation function



```
> acf(timeseriesNewdiff12, lag.max=2)
> acf(timeseriesNewdiff12, lag.max=2, plot=FALSE)

Autocorrelations of series 'timeseriesNewdiff12', by lag

     0      1      2
 1.000 -0.878  0.607
>
```

Partial autocorrelation function

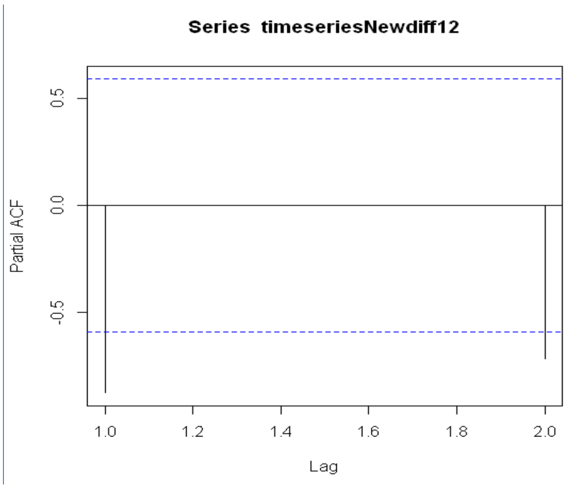
>pacf(timeseriesNewdiff12, lag.max=2)

>pacf(timeseriesNewdiff12, lag.max=2, plot=FALSE)

```
> pacf(timeseriesNewdiff12, lag.max=2)
> pacf(timeseriesNewdiff12, lag.max=2, plot=FALSE)

Partial autocorrelations of series ‘timeseriesNewdiff12’, by lag

     1      2
-0.878 -0.717
>
```



Picture 7. Partial autocorrelation function

Let’s predict previously discarded 12 series members using ARIMA.

>library("forecast")

>timeseriesNewArima<-arima(timeseriesNew,order=c(0,1,1))

```
> timeseriesNewArima<-arima(timeseriesNew,order=c(0,1,1))
> timeseriesNewArima

Call:
arima(x = timeseriesNew, order = c(0, 1, 1))

Coefficients:
         ma1
      0.2393
s.e.  0.1624

sigma^2 estimated as 27897:  log likelihood = -150.38,  aic = 304.76
>
```

>timeseriesNewforecasts<-forecast.Arima(timeseriesNewArima,h=12)

>timeseriesNewforecasts

```
> timeseriesNewforecasts<-forecast.Arima(timeseriesNewArima,h=12)
> timeseriesNewforecasts
    Point Forecast      Lo 80     Hi 80      Lo 95     Hi 95
213       317.2846  103.23383  531.3354   -10.0778  644.6470
214       317.2846  -23.57213  658.1413  -204.0108  838.5800
215       317.2846 -114.62845  749.1976  -343.2694  977.8386
216       317.2846 -189.58270  824.1519  -457.9020 1092.4712
217       317.2846 -254.79934  889.3685  -557.6423 1192.2115
218       317.2846 -313.30688  947.8761  -647.1218 1281.6910
219       317.2846 -366.82883 1001.3980  -728.9766 1363.5458
220       317.2846 -416.45702 1051.0262  -804.8764 1439.4455
221       317.2846 -462.93482 1097.5040  -875.9580 1510.5272
222       317.2846 -506.79544 1141.3646  -943.0371 1577.6062
223       317.2846 -548.43677 1183.0059 -1006.7220 1641.2912
224       317.2846 -588.16505 1222.7342 -1067.4811 1702.0503
>
```

Conclusion. The time series is predictable if you look at the graphs, since there are significant values that exceed critical ones - control boundaries.