Time Series Forecasting

Objective: To perform time series forecasting on monthly sunspot dataset as part of Lab Migration Project.

Methods:

(i) Import and load the dataset and check for any missing values.
(ii) Display the summary statistics and plot the data year-wise and month-wise.
(iii) Decompose the data as multiplicative and plot it.
(iv) Visualize the decompose data for trend, seasonal and random patterns.
(v) Stationarize the dataset using ADF test.
(vi) Plot ACF and PACF.
(vii) Build ARIMA model.
(viii) Conclusion

```
library(forecast)
```

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

```
library(tseries)
library("anytime")
data <- read.csv('monthly-sunspots.csv')
```

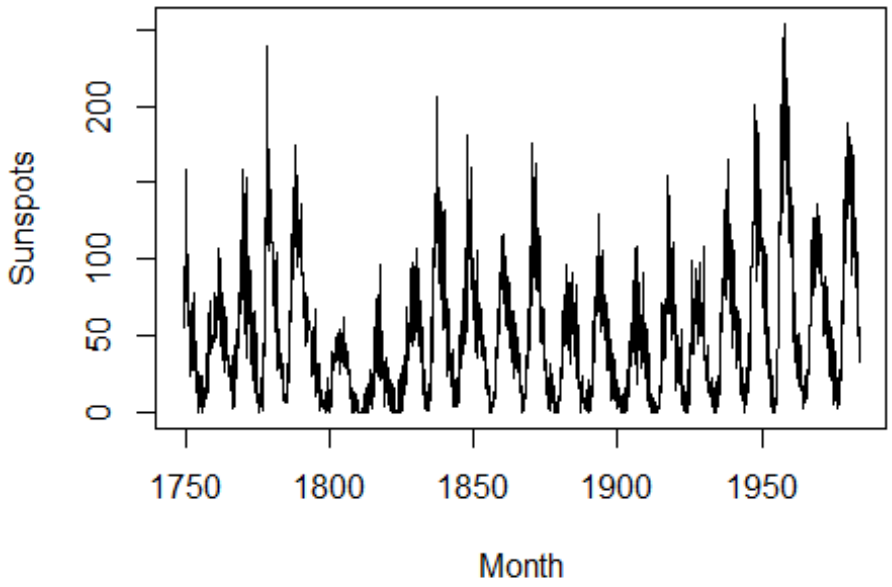
```
#Display head of dataset
head(data)
```

```
##     Month Sunspots
## 1 1749-01     58.0
## 2 1749-02     62.6
## 3 1749-03     70.0
## 4 1749-04     55.7
## 5 1749-05     85.0
## 6 1749-06     83.5
```

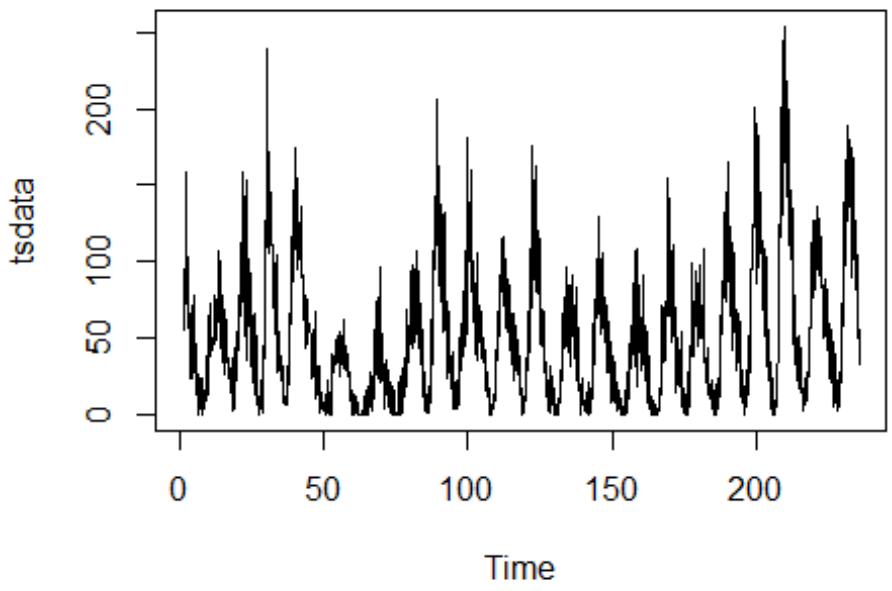
```
#Convert data type into Date type
data$Month <- anydate(data$Month)
data$Month <- as.Date(data$Month)

#Check for unfilled data
sum(is.na(data))
```

Inference: Here 0 denotes that there are no missing values in the given dataset. So, no data cleaning is required.

```
## [1] 0

#Statistical info
summary(data)

##      Month                Sunspots
##  Min.   :1749-01-01   Min.   :  0.00
##  1st Qu.:1807-09-23   1st Qu.: 15.70
##  Median :1866-06-16   Median : 42.00
##  Mean   :1866-06-16   Mean   : 51.27
##  3rd Qu.:1925-03-08   3rd Qu.: 74.92
##  Max.   :1983-12-01   Max.   :253.80

plot(data,type="l")
```




```
#Plot as timeseries data
tsdata <- ts(data$Sunspots,frequency=12)
plot(tsdata)
```



Inference: The graphs denote that the monthly count of the number of observed sunspots increases and decreases over time with each year.

```
#Decompose the data as multiplicative
ddata <- decompose(tsdata,"multiplicative")
```

We will decompose the time series for estimates of trend, seasonal, and random components using moving average method.

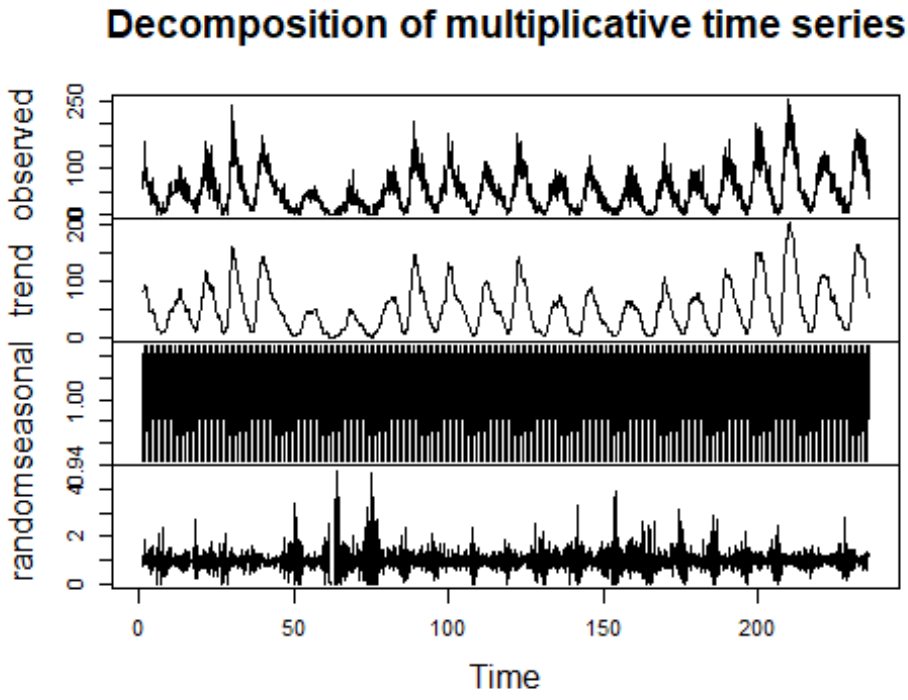
The multiplicative model is:

$Y[t]=T[t]*S[t]*e[t]$

where

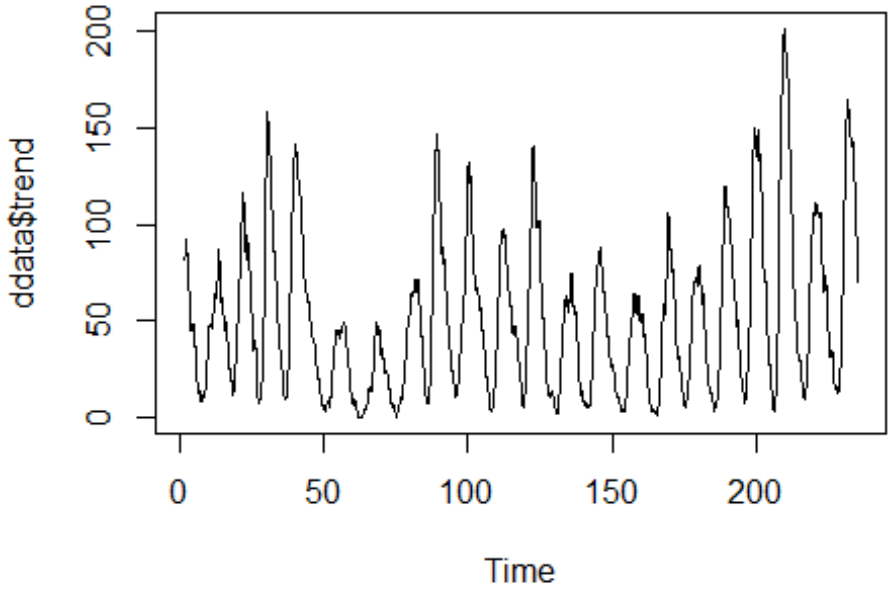
- Y(t) is the number of passengers at time t,
- T(t) is the trend component at time t,
- S(t) is the seasonal component at time t,
- e(t) is the random error component at time t.

```
plot(ddata)
```

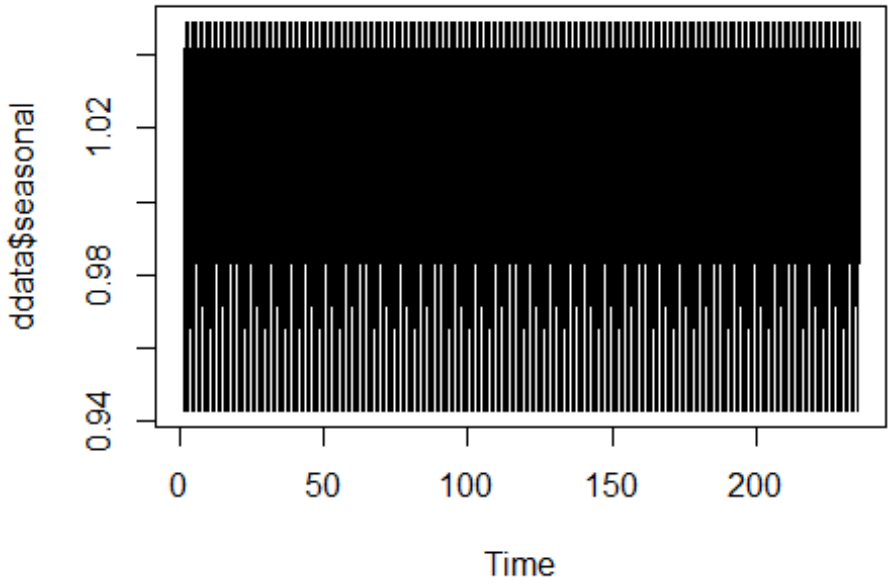


Inference: The decomposed plots show the trend and seasonality.

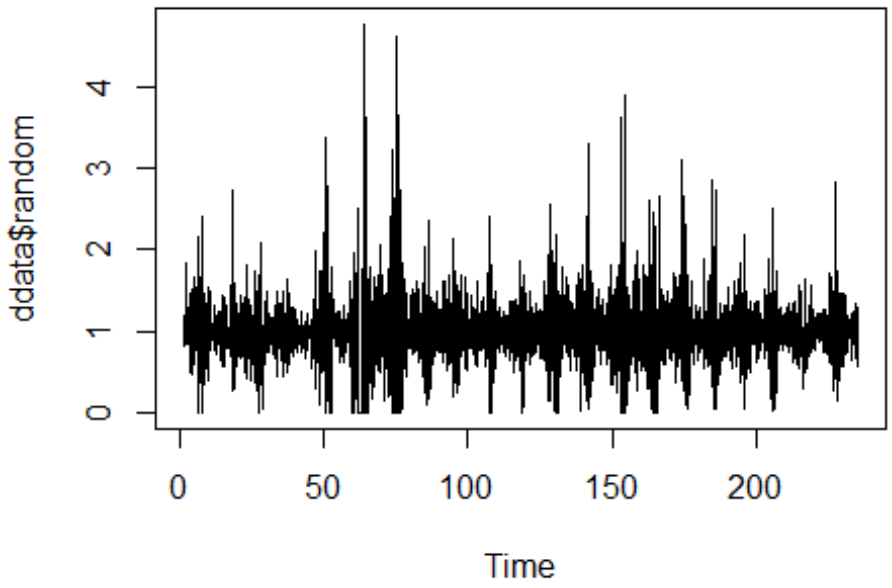
```
plot(ddata$trend)
```



```
plot(ddata$seasonal)
```




```
plot(ddata$random)
```




```
#ADF test for stationarity
adf.test(data$Sunspots)

## Warning in adf.test(data$Sunspots): p-value smaller than printed p-value

##
##  Augmented Dickey-Fuller Test
##
## data:  data$Sunspots
```

```
## Dickey-Fuller = -6.494, Lag order = 14, p-value = 0.01
## alternative hypothesis: stationary
```

To test the stationarity of the time series, we run the Augmented Dickey-Fuller Test using the adf.test() function from the tseries package.

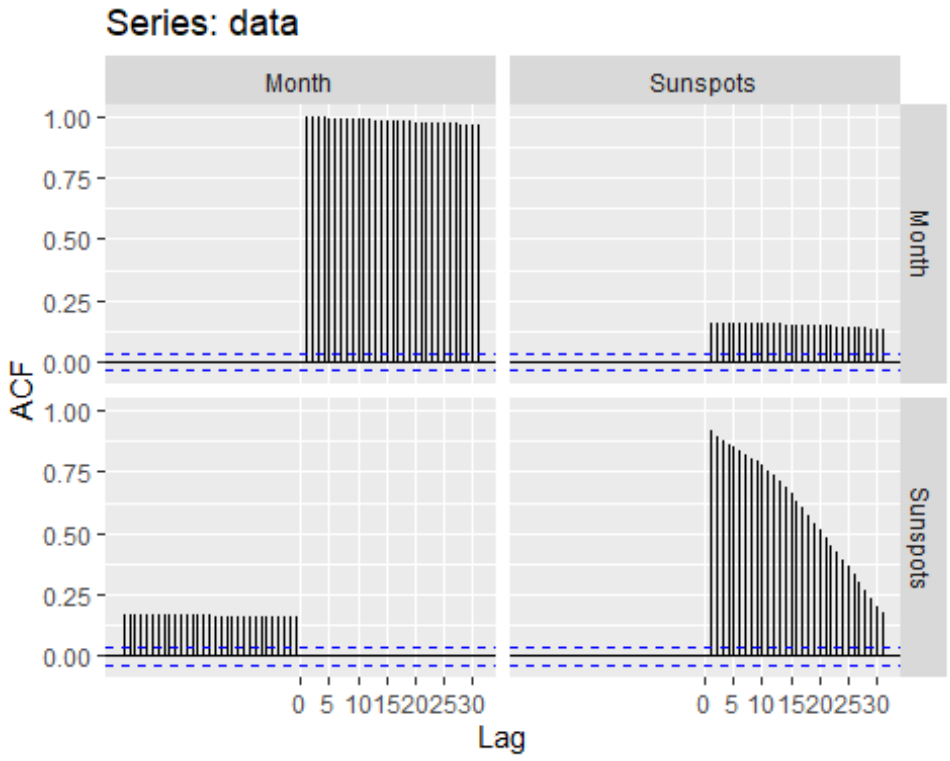
First set the hypothesis test:

The null hypothesis H_0: that the time series is non stationary

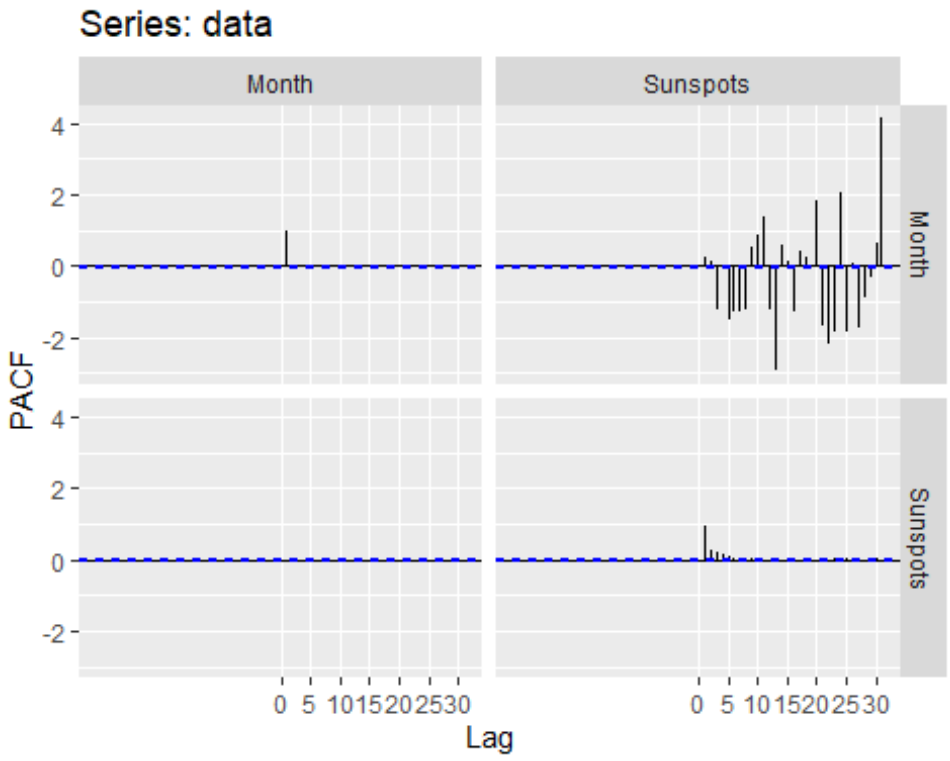
The alternative hypothesis H_A: that the time series is stationary

Inference: Here, the p-value is 0.01 which is less than 0.05, so we reject the null hypothesis, and go with the alternative hypothesis that the time series is stationary.

```
autoplot(acf(data,plot=FALSE))
```




```
autoplot(pacf(data,plot=FALSE))
```




```
#Model using ARIMA
auto.arima(data$Sunspots)

## Series: data$Sunspots
## ARIMA(2,1,2)
##
## Coefficients:
##          ar1      ar2      ma1     ma2
##       1.3467  -0.3963  -1.7710  0.8103
## s.e.  0.0303   0.0287   0.0205  0.0194
##
## sigma^2 = 243.8:  log likelihood = -11745.5
## AIC=23500.99   AICc=23501.01   BIC=23530.71
```

Inference: The ARIMA(2,1,2) model parameters are lag 1 differencing (d), an autoregressive term of second lag (p) and a moving average model of order 2 (q).

Conclusion:

The sunspot numbers increase and decrease over time with each year. The monthly sunspot data shows the pattern of seasonality with large differences between seasons.