**Time Series Analysis :**

Time Series is a series of data points indexed (Graphs/listed ) in time order

**Data Set Description :**

The data is used for time series analysis from classic box & Jenkins airline data. Dataset contains information about airline passengers travelled monthly between the years 1949 to 1960.

**Project Description :**

The Goal of this project is to analyze airline passenger data through time series analysis. Understanding travel trend of people travelled across different months of the year. The Time series data modeled through ARIMA model. The ARIMA model is used to validate current time series data as well predicting possibilities of people travelled different months of the year.

**Data Set Description :**

**importing dataset in R**

data("AirPassengers")

**Importing data transformation library**

library(dplyr)

**Describing dataset & Type of dataset**

glimpse(AirPassengers)

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

Alternatively we can use below to determine type of dataset

class(AirPassengers)

**Summarizing Dataset**

**Passenger data distributed across the dataset**

summary(AirPassengers)

Min. 1st Qu. Median Mean 3rd Qu. Max.

104.0 180.0 265.5 280.3 360.5 622.0

**here mean and median both are significantly different then we can use median as central tendency if mean and median are significantly (near) same then we can use mean as central tendency.**

**Start of time series**

start(AirPassengers)

[1] 1949

1

**End of time series**

> end(AirPassengers)

[1] 1960 12

**Determining how many cycle in a year**

> frequency(AirPassengers)

>12

**Determining position of the cycle in each observation**

cycle(AirPassengers)

Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec

1949 1 2 3 4 5 6 7 8 9 10 11 12

1950 1 2 3 4 5 6 7 8 9 10 11 12

1951 1 2 3 4 5 6 7 8 9 10 11 12

1952 1 2 3 4 5 6 7 8 9 10 11 12

1953 1 2 3 4 5 6 7 8 9 10 11 12

1954 1 2 3 4 5 6 7 8 9 10 11 12

1955 1 2 3 4 5 6 7 8 9 10 11 12

1956 1 2 3 4 5 6 7 8 9 10 11 12

1957 1 2 3 4 5 6 7 8 9 10 11 12

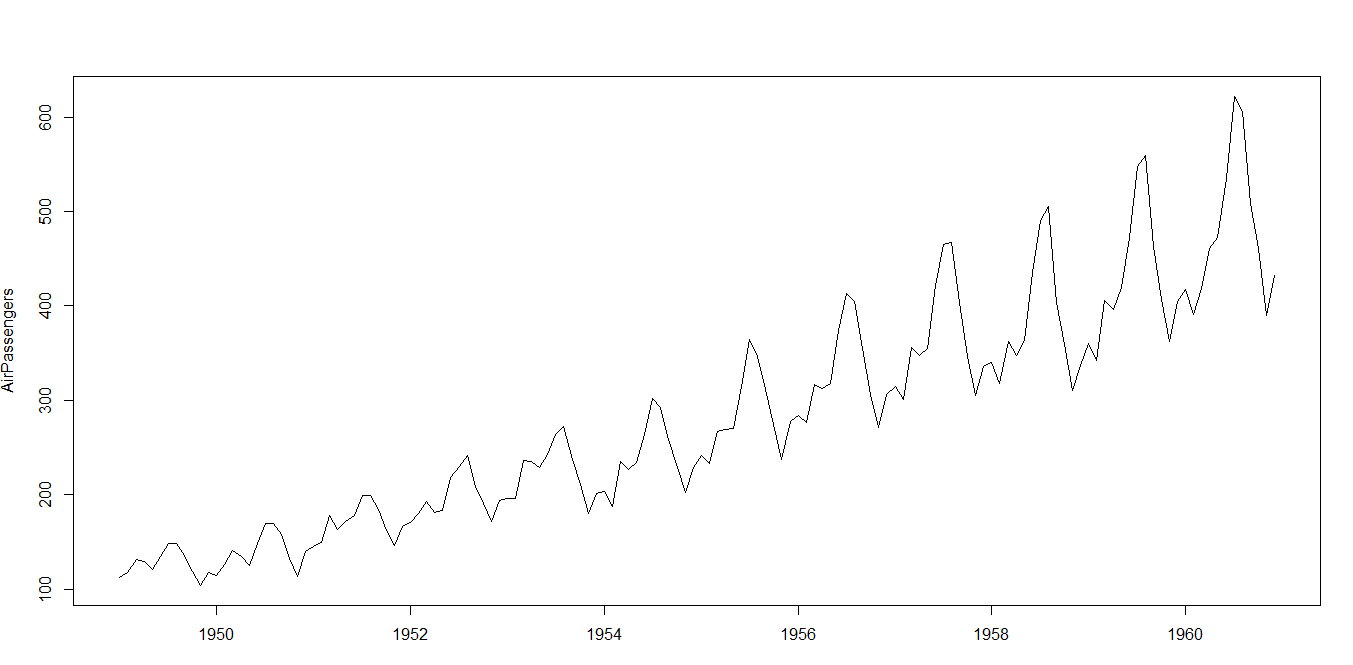
1958 1 2 3 4 5 6 7 8 9 10 11 12

1959 1 2 3 4 5 6 7 8 9 10 11 12

1960 1 2 3 4 5 6 7 8 9 10 11 12

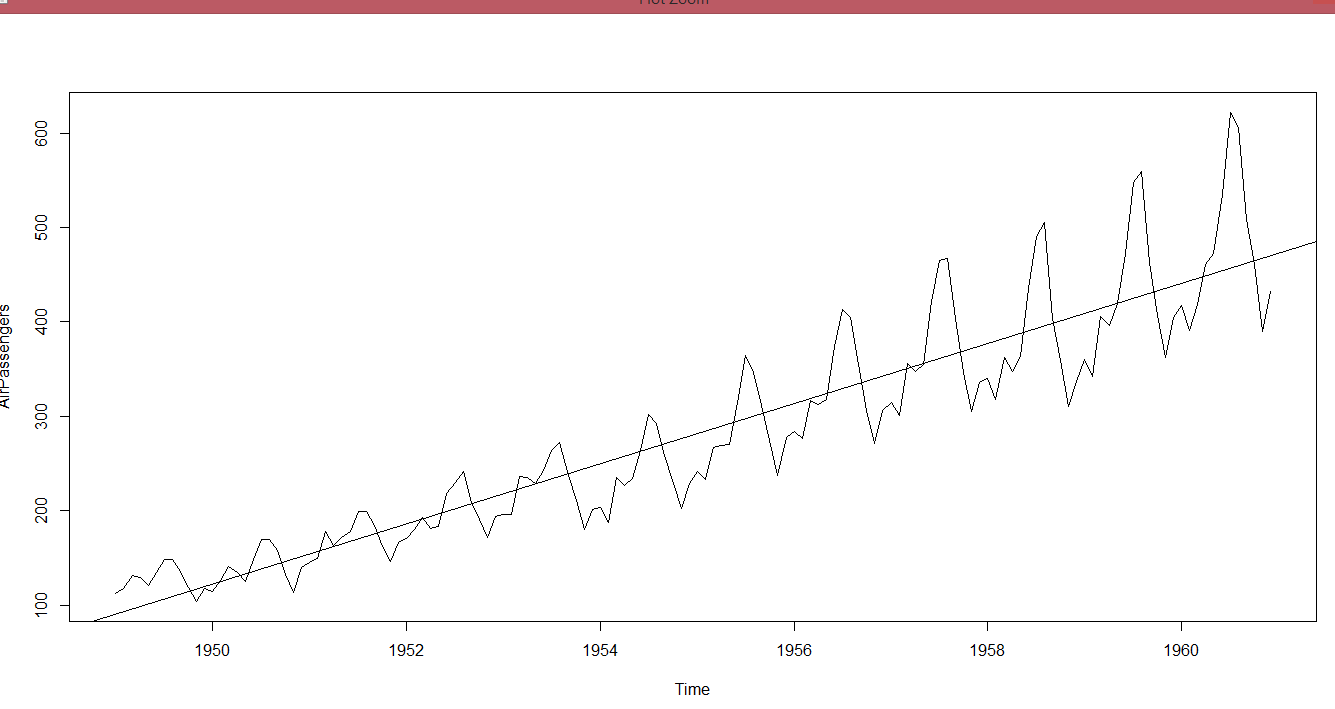
**Representing data in a graphical format**

plot(AirPassengers)

****

**below command will return mean of time series data**

abline(reg=lm(AirPassengers~time(AirPassengers)))

****

From the observation it is clear that mean is not constant (keeps on increasing : represented in a straight line) and variance and co variance also not equally distance from the mean at different time intervals. Time series is not stationary. For a Time series to be stationary it should have below properties

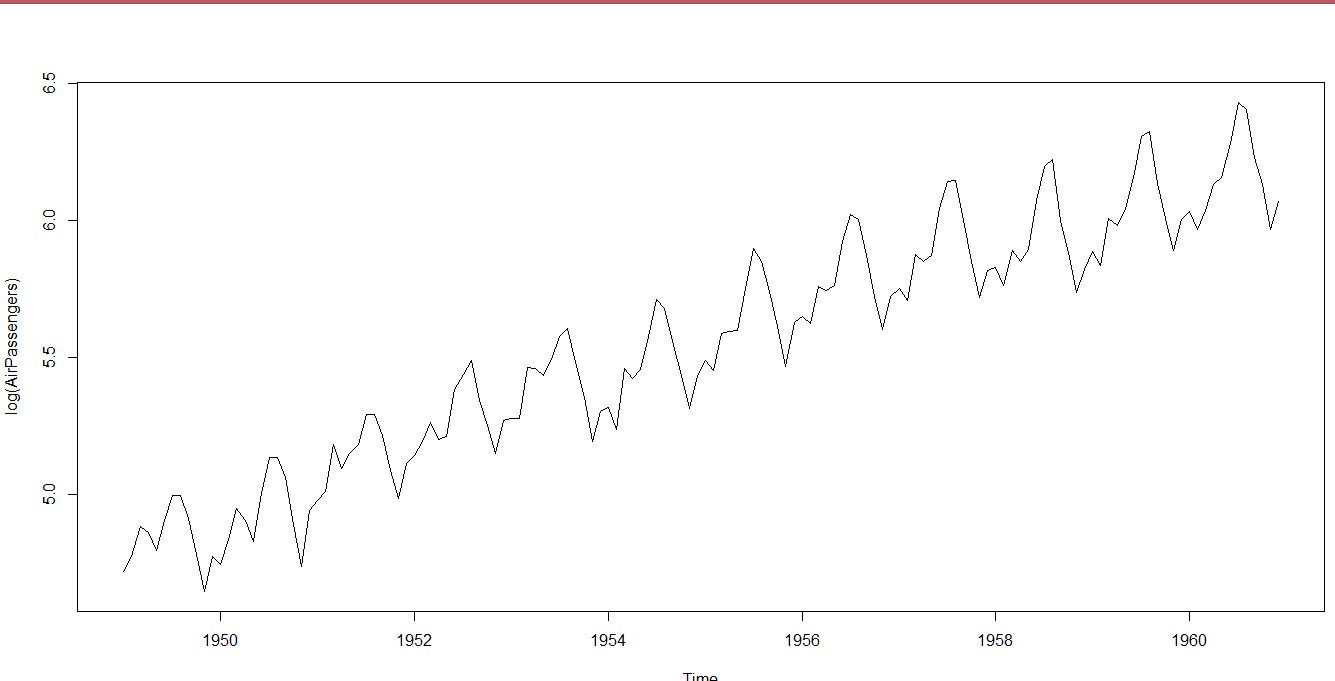
**1** Mean should be constant in the entire series.

**2** Variance and co variance should be equally distance from the mean during different time intervals

**Transforming series data into stationary**

**variance become equal for different time intervals in the time series. but mean still keeps on increasing in time seris**

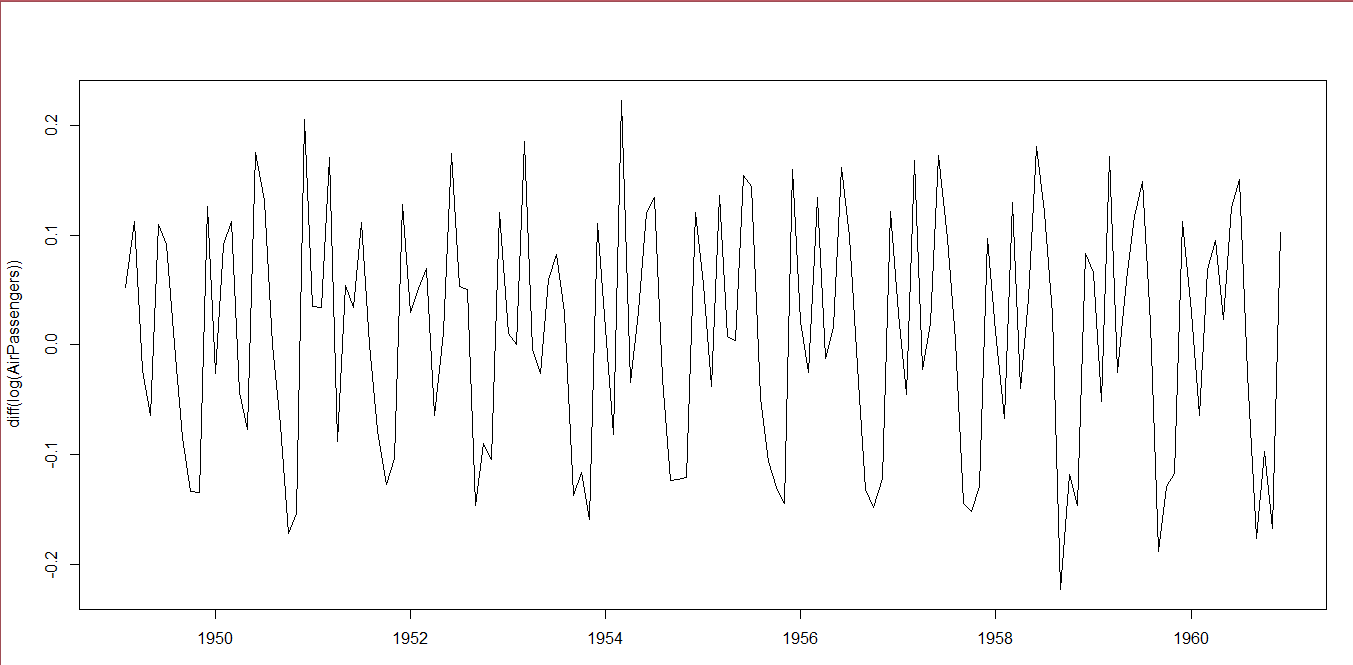
plot(log(AirPassengers))



**Converting mean constant in the time series**

Through below transformation data become stationary in the time series (mean become constant)

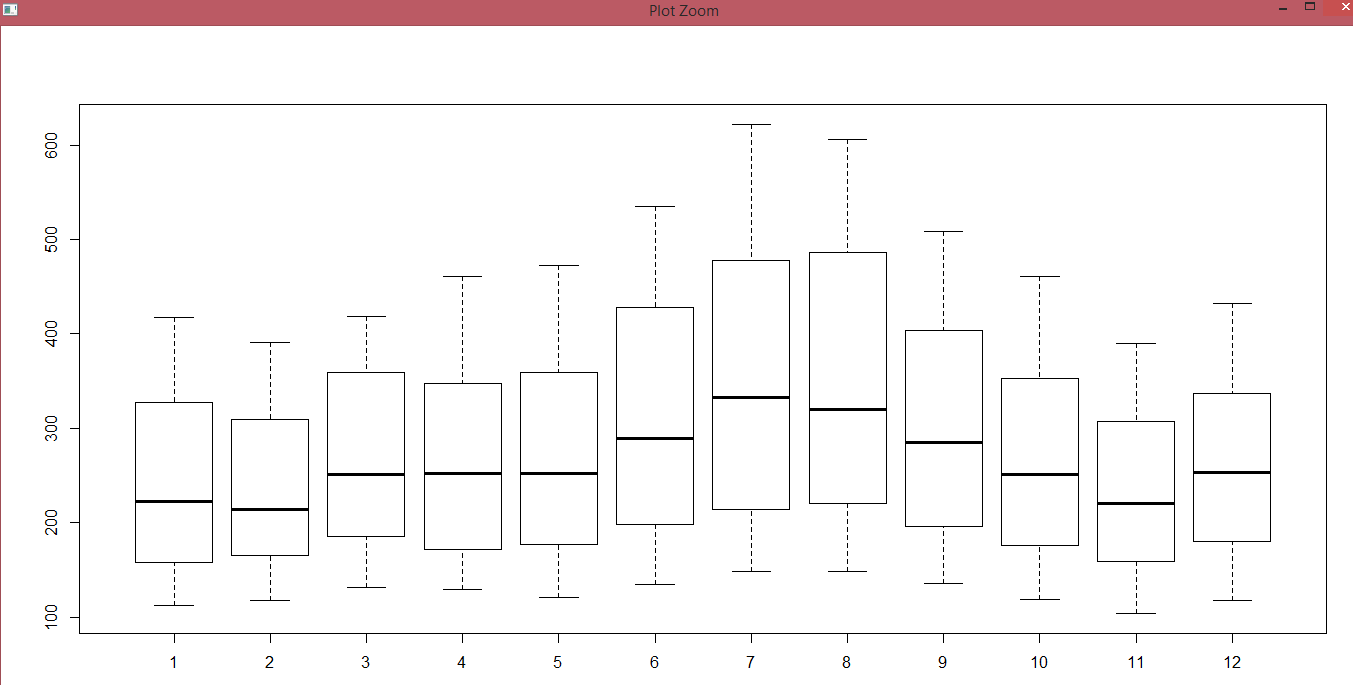
plot(diff(log(AirPassengers)))

****

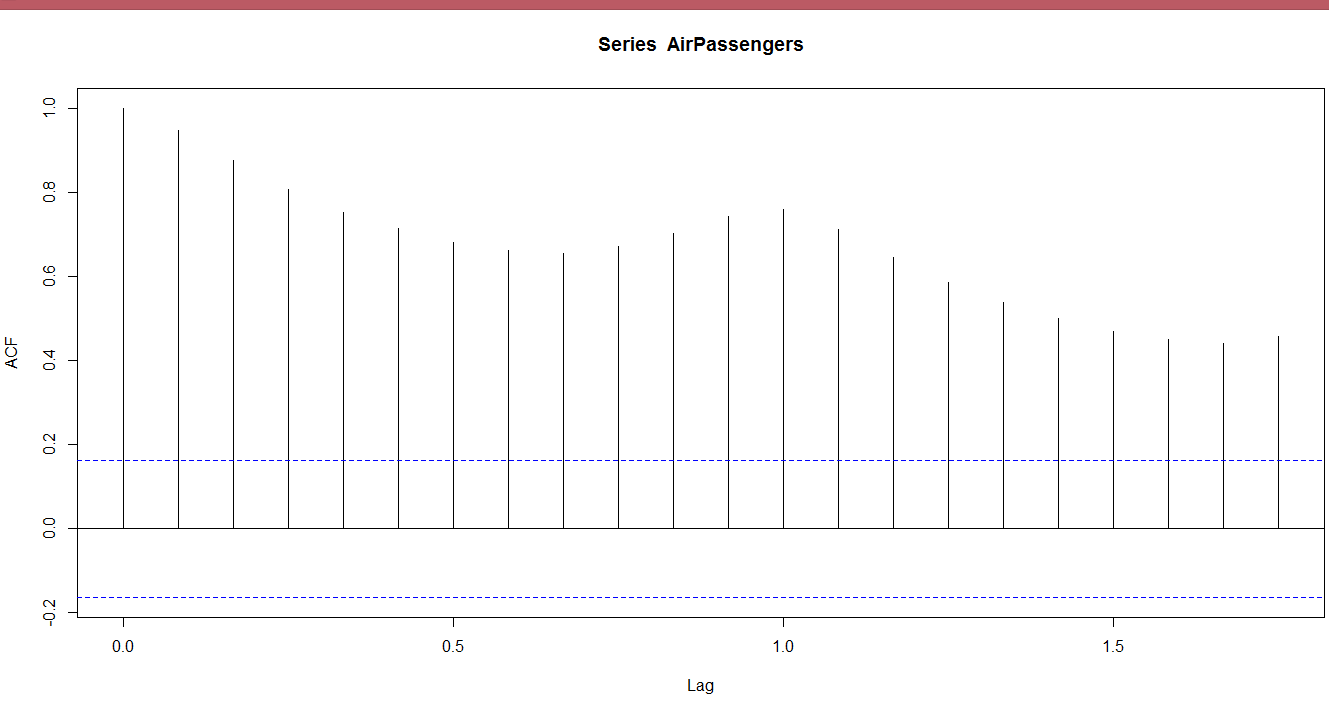
**Understanding the seasonality of data in time series**

boxplot(AirPassengers~cycle(AirPassengers))

From the graph it is clear the people travelled most in the 7th & 8th month of the year

****

**Analysis of Dataset when it is not converted into stationary dataset :-**

****

from above dataset it is clear that values are exceeding from the blue line , for this we need to normalize dataset before using it in a time series analysis.

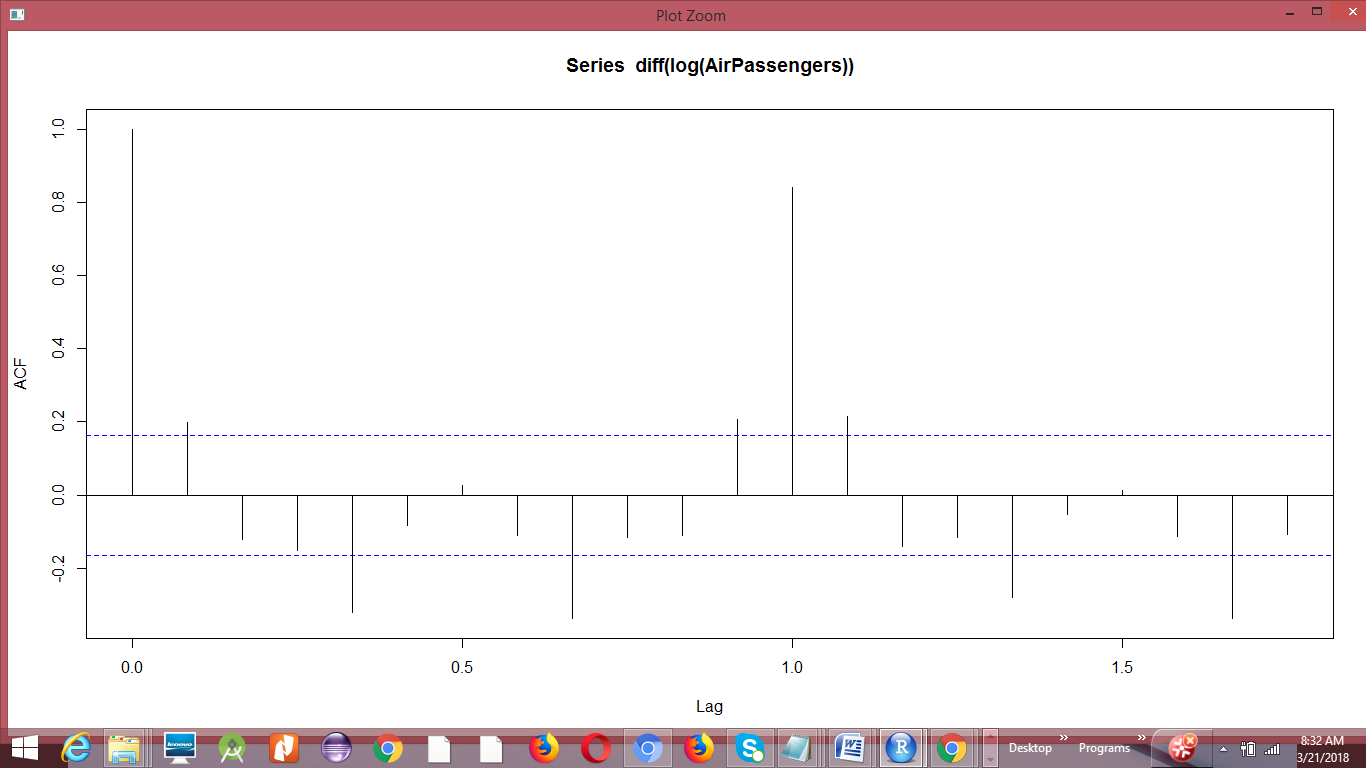
**calculating the value of q**

acf(diff(log(AirPassengers))) = q

from the graph it is clear that first two lines are well above the blue line

so we take the 3rd one from the left (counting zero th from the first row)

so the value of q = 1



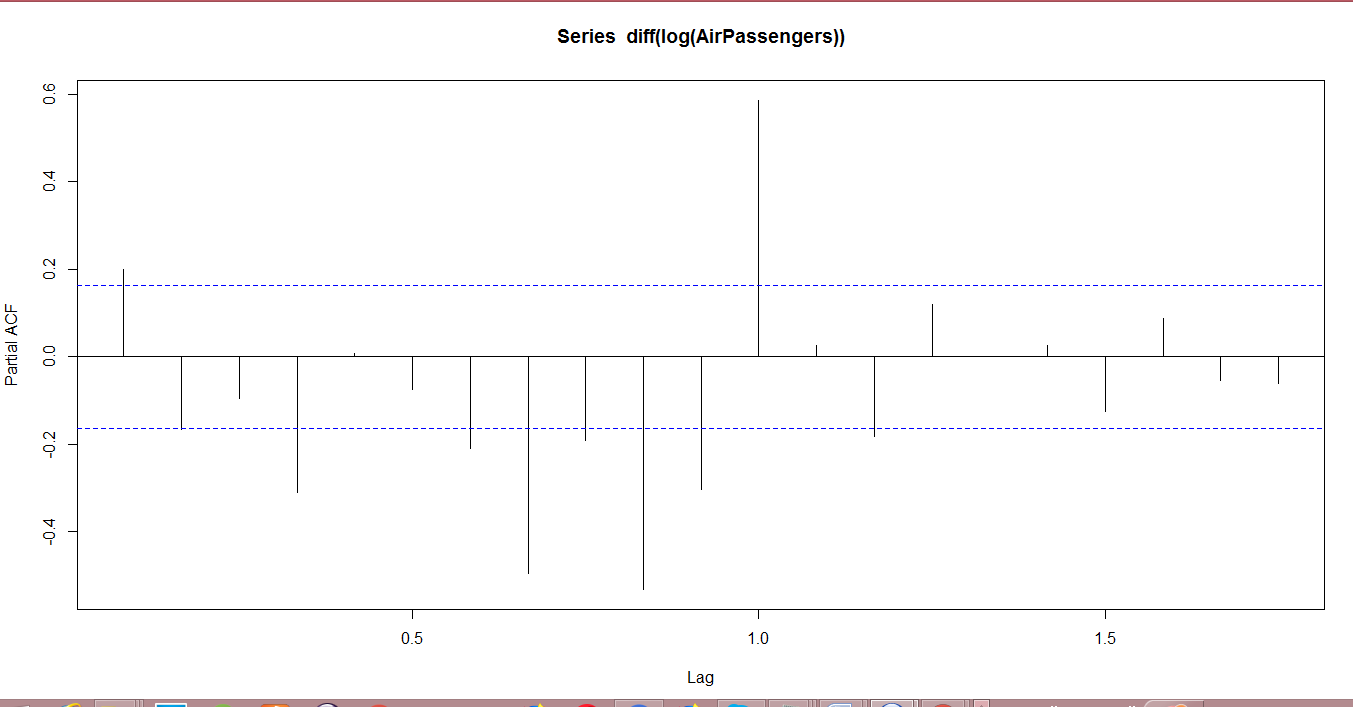
**Calculating the value of p**

pacf(diff(log(AirPassengers))) = p

from the below graph it is clear that second line from the left

well below the below the blue line (count zero th from the first element)

so the value of p = 0 (because before zero th line , the first line has gone below the blue line)



**Calculate the value of d**

here d = 1 because we have use differentiation only to get mean constant

**Applying ARIMA Model**

model = arima(AirPassengers,c(0,1,1),seasonal = list(order=c(0,1,1)),period=12))

Above is a Arima model can be described as below c(0,1,1) = values of p,d,q respectively.

**Predicting from Arima model**

pred= predict(model,n.ahead= 10\*12)

predicting for next 10 years from the model

pred1 = 2.718\*pred$pred

converting from log value

pred1 =

Jan Feb Mar Apr May Jun Jul Aug Sep

1961 1215.091 1146.663 1232.684 1331.550 1364.935 1533.563 1766.144 1730.590 1464.787

1962 1303.503 1235.075 1321.097 1419.963 1453.347 1621.975 1854.556 1819.003 1553.200

1963 1391.916 1323.488 1409.509 1508.376 1541.760 1710.388 1942.969 1907.416 1641.612

1964 1480.328 1411.901 1497.922 1596.788 1630.173 1798.800 2031.381 1995.828 1730.025

1965 1568.741 1500.313 1586.335 1685.201 1718.585 1887.213 2119.794 2084.241 1818.437

1966 1657.154 1588.726 1674.747 1773.613 1806.998 1975.626 2208.207 2172.653 1906.850

1967 1745.566 1677.138 1763.160 1862.026 1895.410 2064.038 2296.619 2261.066 1995.263

1968 1833.979 1765.551 1851.572 1950.439 1983.823 2152.451 2385.032 2349.479 2083.675

1969 1922.391 1853.964 1939.985 2038.851 2072.236 2240.863 2473.444 2437.891 2172.088

1970 2010.804 1942.376 2028.398 2127.264 2160.648 2329.276 2561.857 2526.304 2260.500

Oct Nov Dec

1961 1334.721 1149.236 1263.197

1962 1423.133 1237.649 1351.610

1963 1511.546 1326.062 1440.023

1964 1599.959 1414.474 1528.435

1965 1688.371 1502.887 1616.848

1966 1776.784 1591.299 1705.260

1967 1865.196 1679.712 1793.673

1968 1953.609 1768.125 1882.086

1969 2042.022 1856.537 1970.498

1970 2130.434 1944.950 2058.911

**Rounding off data : This will remove decimal points from data**

data3<-round(pred1,digits = 0)

> data3