ARIMA

pranit shinde

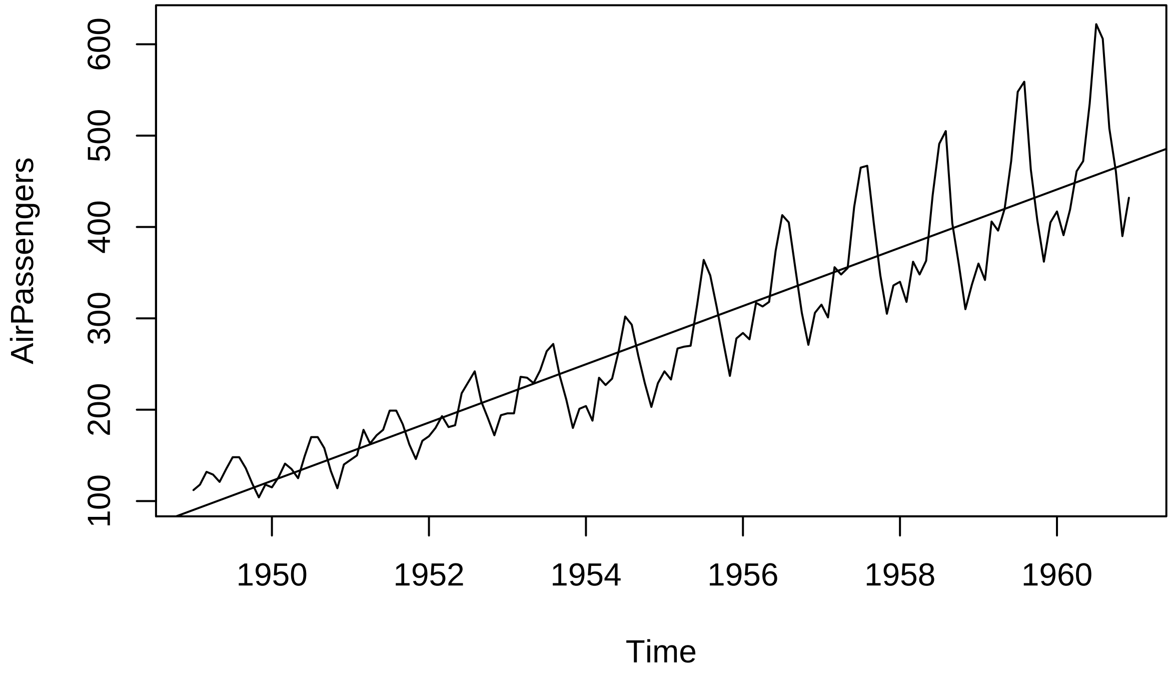
10 December 2018

Time series ARIMA(Auto Regressive Integrated Moving Average)

```
data("AirPassengers")
plot(AirPassengers)
summary(AirPassengers)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   104.0   180.0   265.5   280.3   360.5   622.0
```

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



## the mean for the above plot is not constant, the variance mapped using creast and troughs is also not same above and below the regression line

```
class(AirPassengers)
```

```
## [1] "ts"
```

```
start(AirPassengers)  # start of the time series
```

```
## [1] 1949    1
```

```
end(AirPassengers)  # end of the time series
```

```
## [1] 1960   12
```

```
frequency(AirPassengers) # cycle of the  time series is 12 month a year
```

```
## [1] 12
```

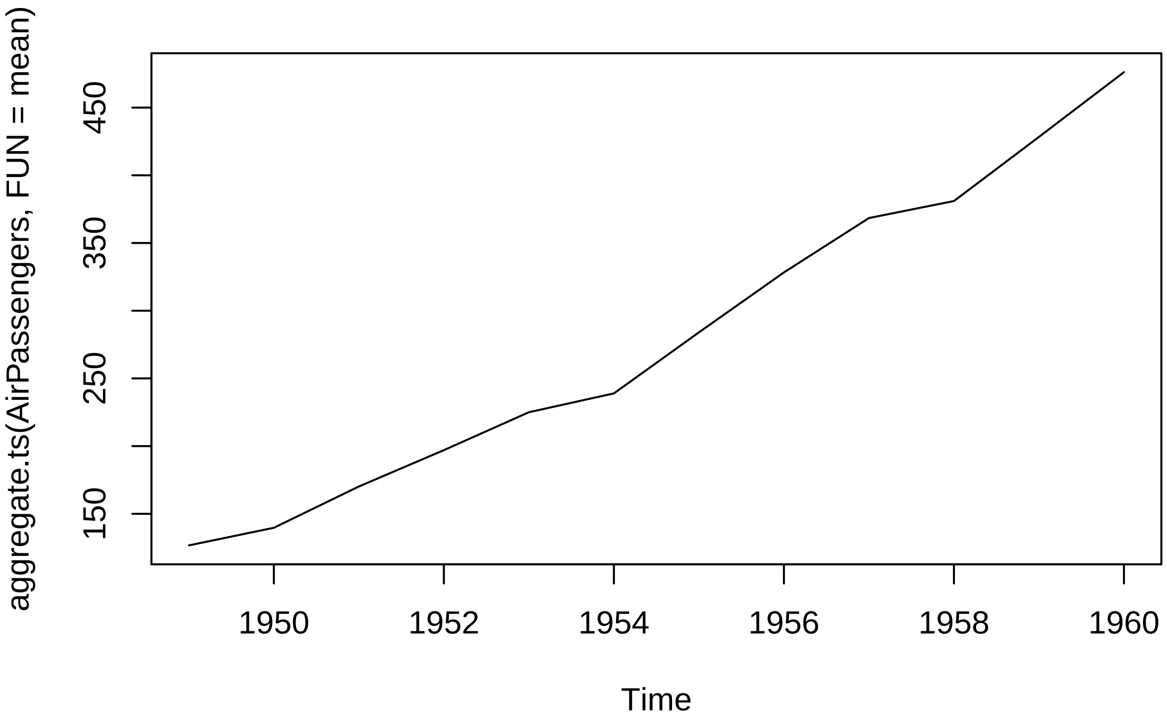
```
summary(AirPassengers) # the no of passengers are distributed across the spectrum
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   104.0   180.0   265.5   280.3   360.5   622.0
```

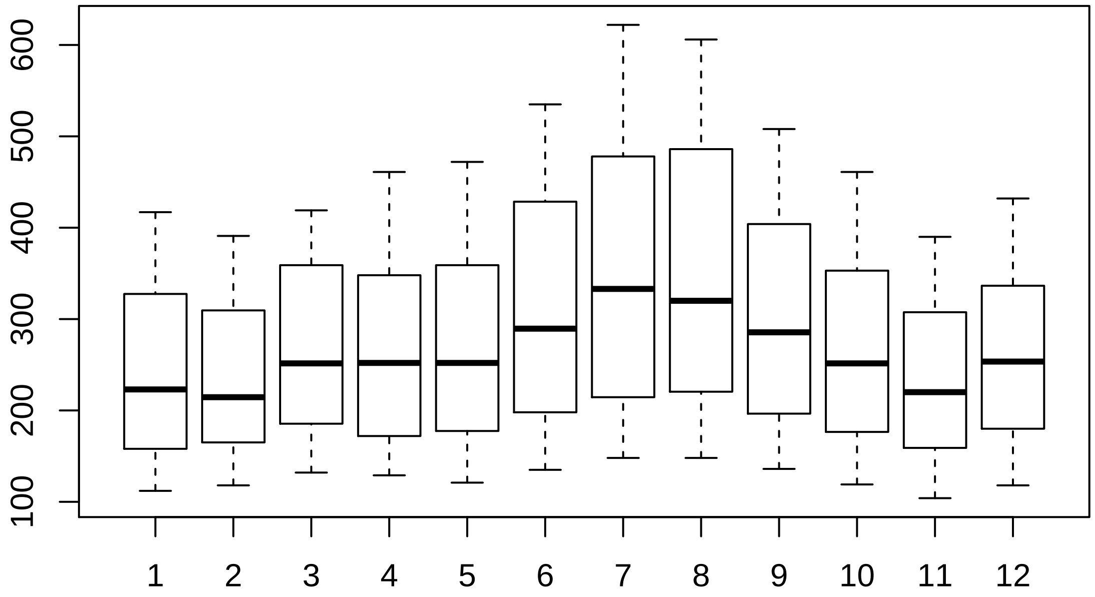
```
cycle(AirPassengers) ## this print the cycle across year
```

```
##      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
```

```r
plot(aggregate.ts(AirPassengers, FUN = mean))
```

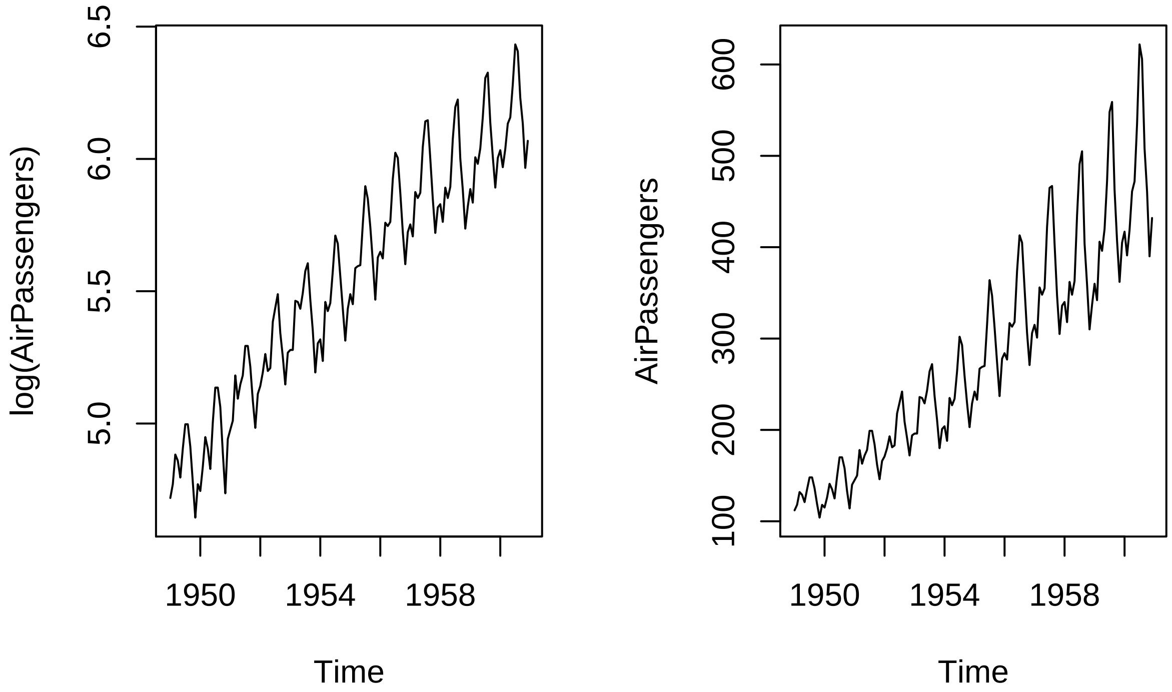


```r
boxplot(AirPassengers ~ cycle(AirPassengers))
```



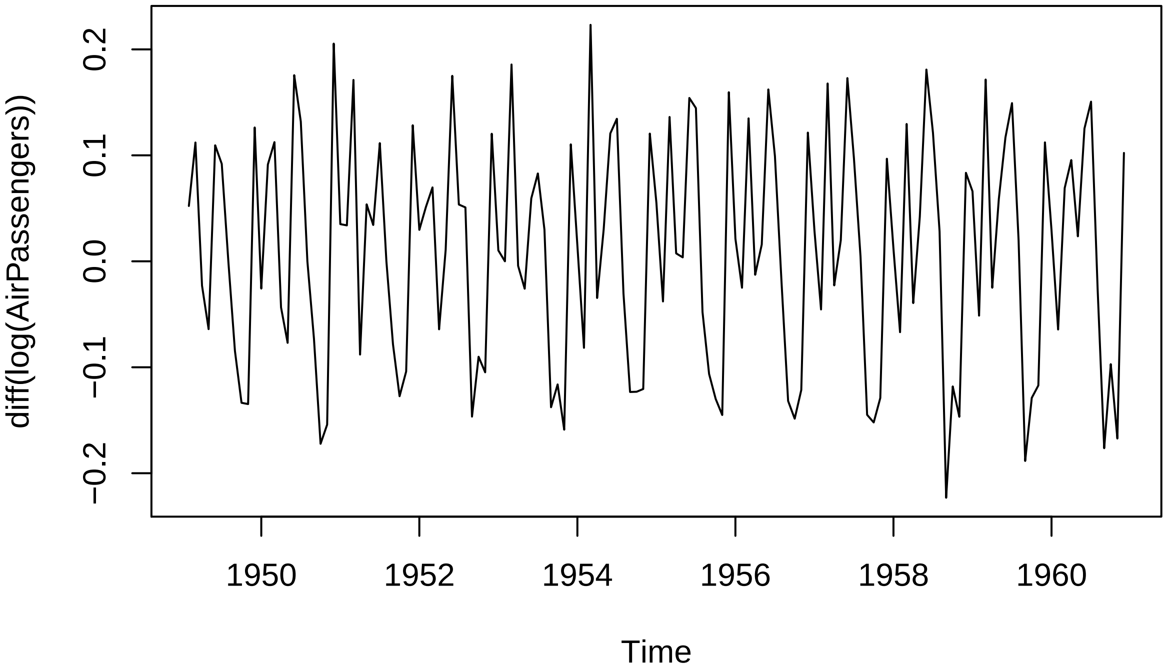
To make sure that the variance is constant through time series, we take log of the data

```
par(mfrow = c(1,2))
plot(log(AirPassengers))
plot(AirPassengers)
```

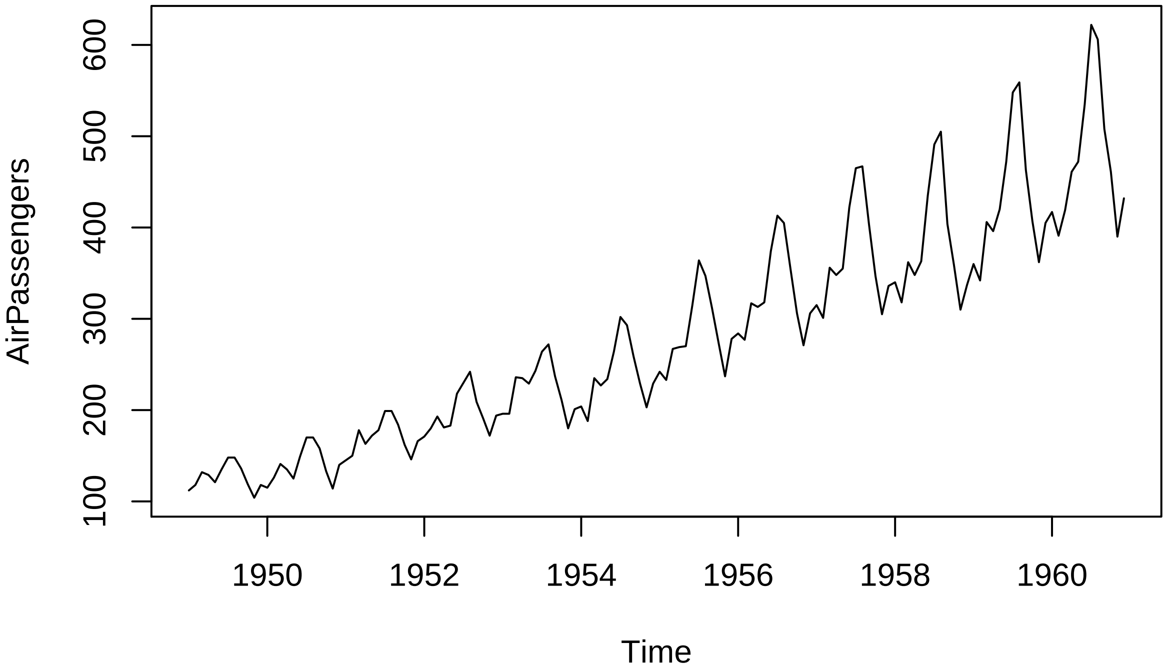


To make sure that the mean is constant through the time series , we take the derivative of the log of data

```r
plot(diff(log(AirPassengers)))  # constant mean around zero
```



```r
plot(AirPassengers)
```

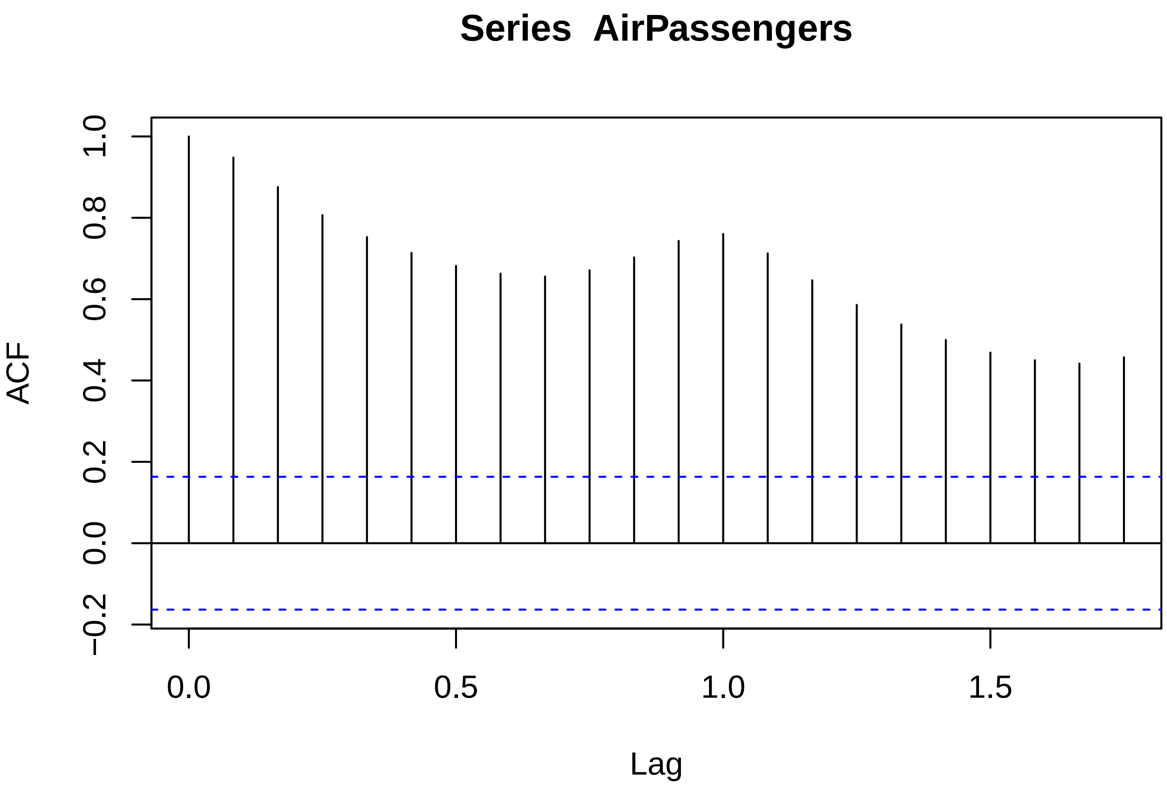


AR I MA

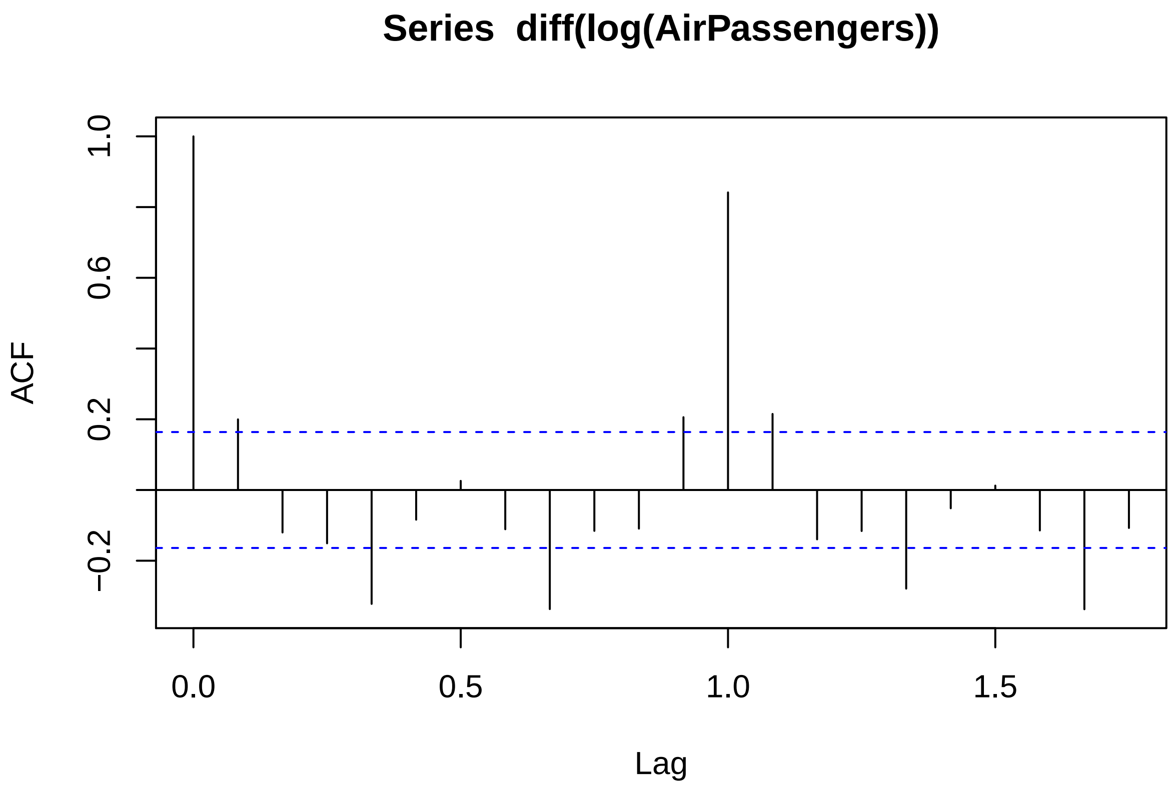
p d q

p, d and q are three parametersr that are used to build a time series model using ARIMA

```
acf(AirPassengers)  ## acf function to be directly applied if the time series is stationary
```



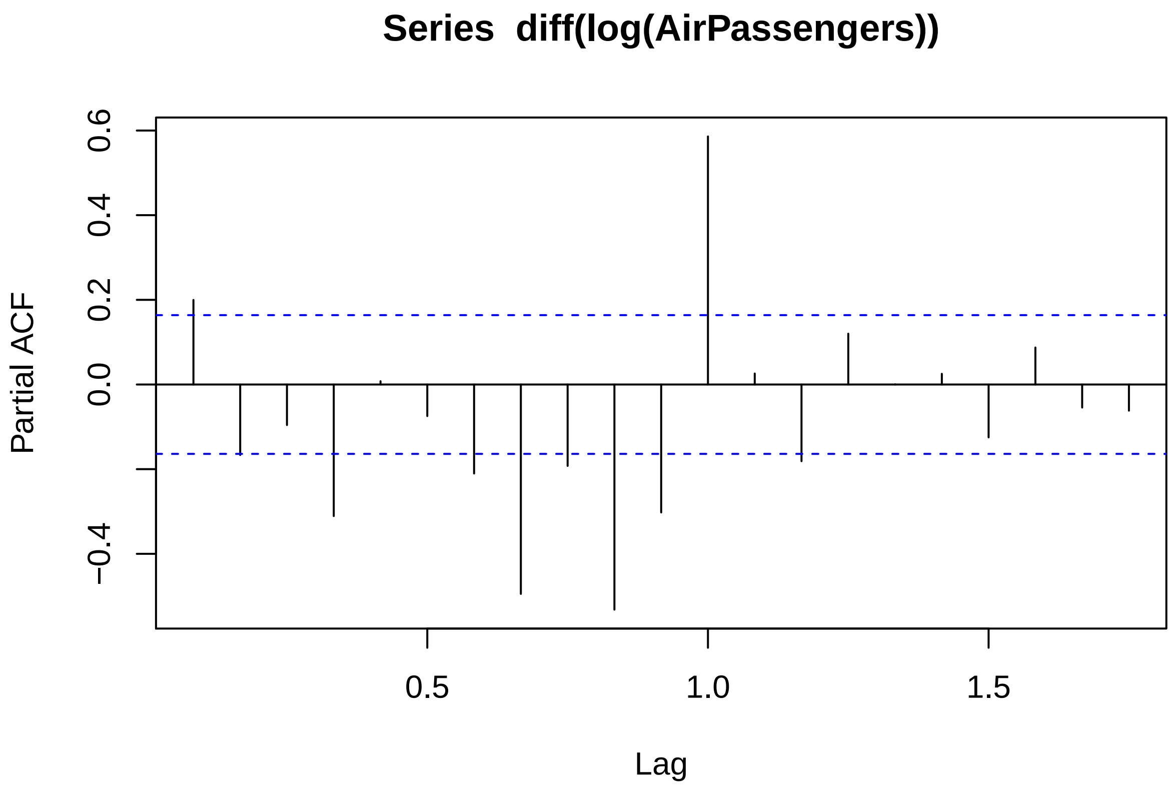
```
acf(diff(log(AirPassengers)))  ## acf function to be applied after transformation, if the series is
```



```
## not statinary , this determine the value of q.
```

Interpretaion : The line that get inverted , the index of the line just before that is q :in our case line is 1 hence q is 1

```
pacf(diff(log(AirPassengers)))
```



# Intepretation: p is one line before the line that gets imvented, here the # function that is used to generate this graph iscalled auto correlation # function'; in our case the value of p is 0 # value od d : d determines the number of times you do differentiation to # stationarize the time series, in our case we did differentiiation just once # hence d will be 1.

lets fit the ARIMA model and predict the next 10years

```
fit <- arima(log(AirPassengers), c(0,1,1), seasonal = list(order = c(0,1,1),
                                                          period = 12))

pred <- predict(fit, n.ahead = 10*12)  ## for 10 year prediction
pred1 <- 2.718 ^ pred$pred
```

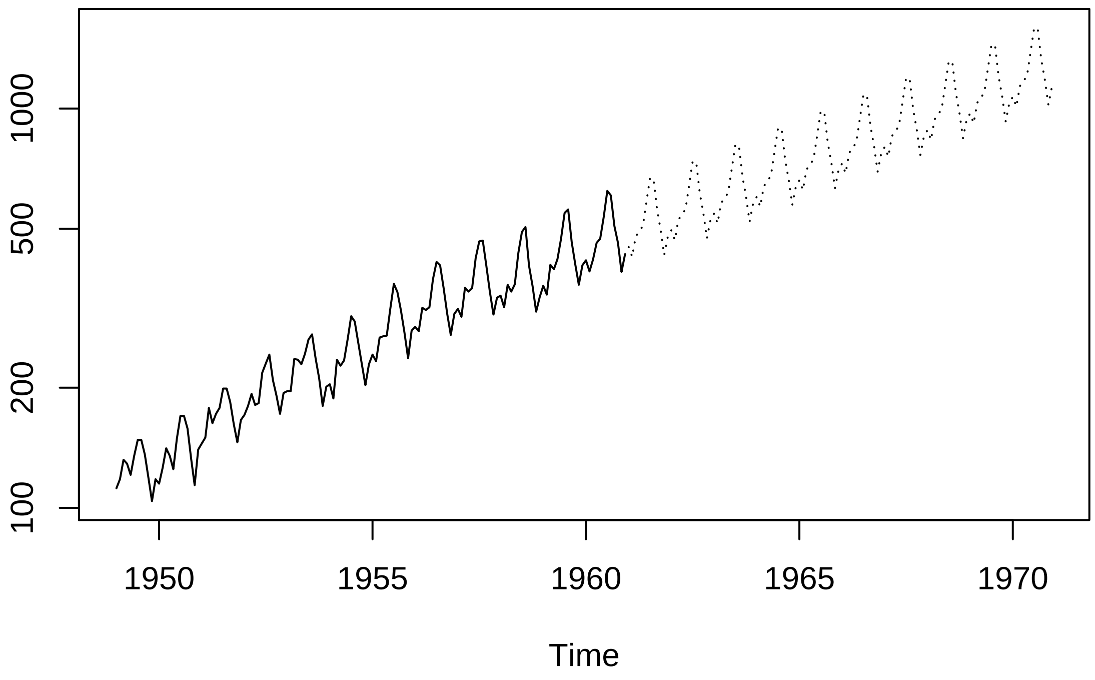
prediction are in the log form , hence to convert to them to decimal interpretable values

we use the shown formula, the value of e is 2.718

```
par(mfrow= c(1,1))
```

plotting the model

```
ts.plot(AirPassengers, 2.718 ^ pred$pred, log = "y", lty= c(1,3))
```



# tesing the model

```
datawide <- ts(AirPassengers, frequency = 12, start = c(1949,1),
                           end = c(1959,12))
fit <- arima(log(datawide), c(0,1,1), seasonal = list(order = c(0,1,1),
                                                      period = 12))
pred <- predict(fit, n.ahead = 10*12)
pred1 <- 2.718^pred$pred
data1 <- head(pred1, 12)
predict_1960 <- round(data1, digits = 0)
original_1960 <- tail(AirPassengers, 12)
predict_1960
```

```
##  [1] 419 399 466 454 473 547 622 630 526 462 406 452
```

```
original_1960
```

```
##  [1] 417 391 419 461 472 535 622 606 508 461 390 432
```