Time Series Analysis (STAT 560)

Homework 3

Bindu Paudel and Gena Ram Mahato

16 September, 2020

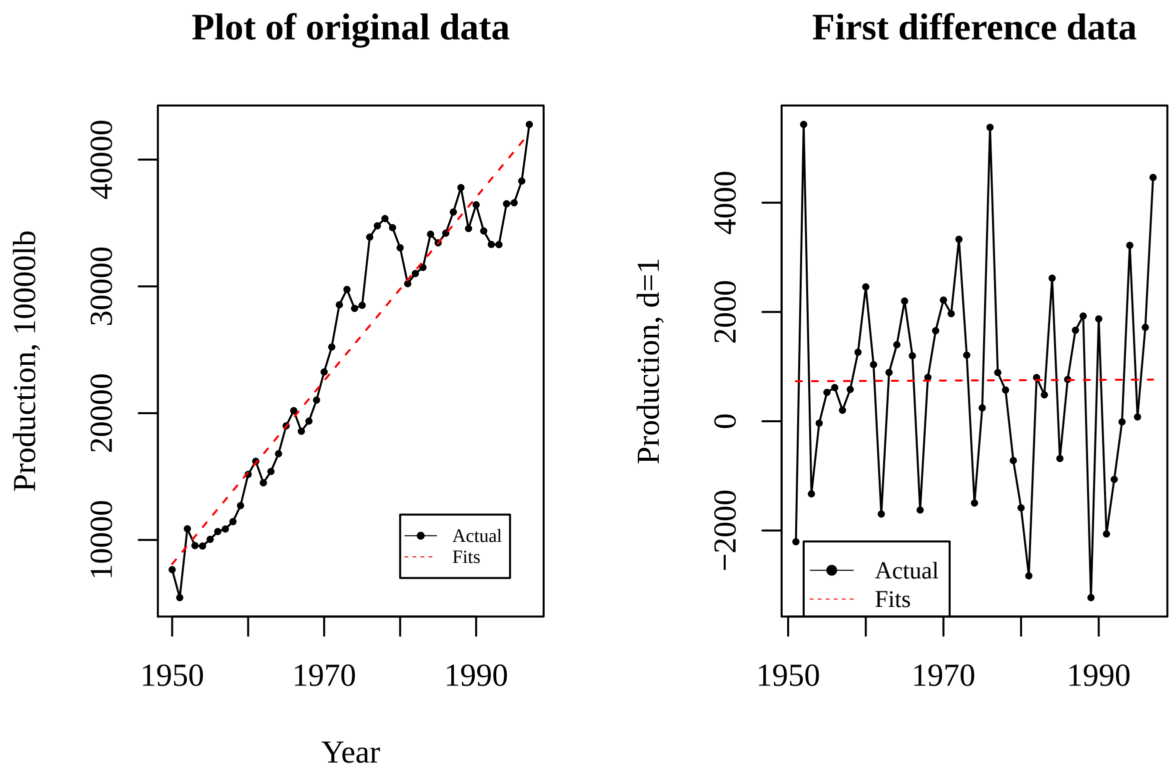
2.2 Consider the data on US production of blue and gongonzola cheeses in Table B.4.

b. Take the first difference of the time series, then find the sample autocorrelation function and the variogram. What conclusion can you draw about the structure and behavior of the time series?

```
setwd("C:/Users/Gena.Mahato/Desktop/Semester/Fall 2020/HW3")
cheese.data <- read.csv("Cheesedata.csv", header=TRUE, sep=",")
head(cheese.data)
```

```
##   ï..Year Production_1000lbs
## 1    1950               7657
## 2    1951               5451
## 3    1952              10883
## 4    1953               9554
## 5    1954               9519
## 6    1955              10047
```

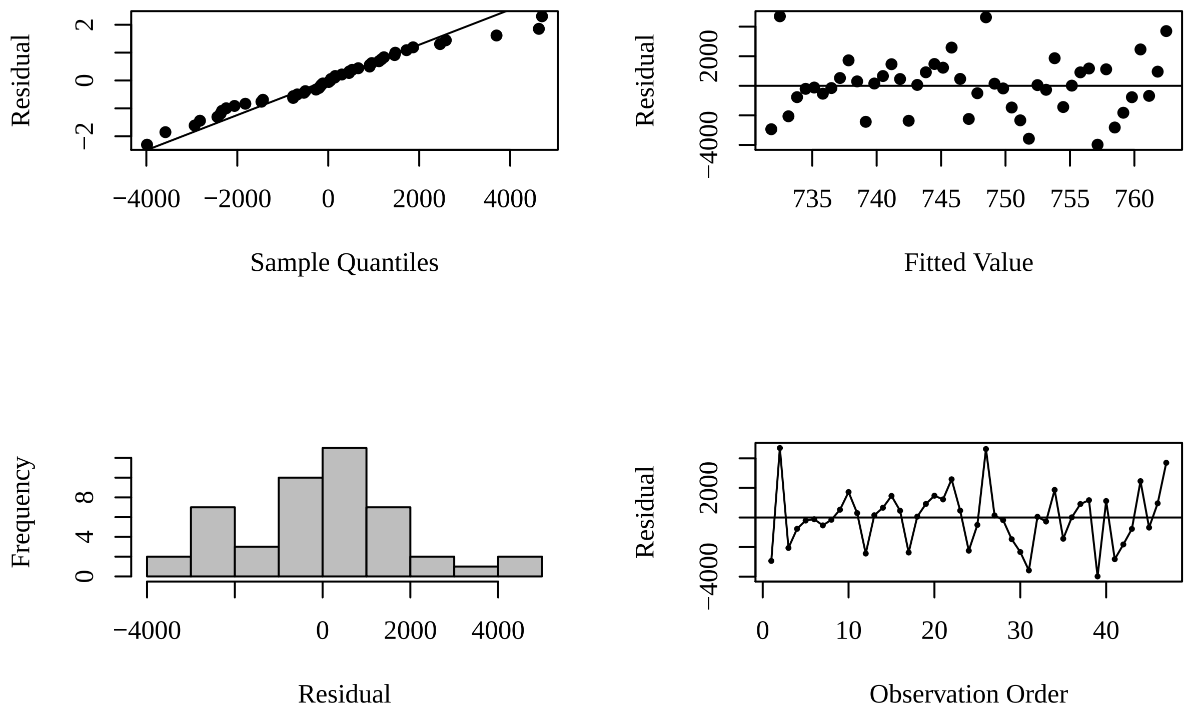
```
#Time Series plot of cheese data
par(mfrow=c(1, 2))
par(family="Times")
fit.cheese<-lm(cheese.data[,2]~cheese.data[,1])
plot(cheese.data,type="l",xlab='Year',ylab='Production, 10000lb', main='Plot of original data')
points(cheese.data,pch=16,cex=.5)
lines(cheese.data[,1], fit.cheese$fit,col="red",lty=2)
legend(1980,12000,c("Actual","Fits"),
pch=c(16,NA),lwd=c(.5,.5),lty=c(1,2),cex=.55,col=c("black","red"))

#Plot first difference of cheese data
nrc<-dim(cheese.data)[1]
diff.cheese.data<-cbind(cheese.data[2:nrc,1],diff(cheese.data[,2]))
fit.diff.cheese<-lm(diff.cheese.data[,2]~diff.cheese.data[,1])
plot(diff.cheese.data,type="l",xlab='',ylab='Production, d=1', main='First difference data')
points(diff.cheese.data,pch=16,cex=.5)
lines(diff.cheese.data[,1], fit.diff.cheese$fit,col="red",lty=2)
legend(1952,-2200,c("Actual","Fits"),
pch=c(16,NA),lwd=c(.5,.5),lty=c(1,2),
cex=.75,col=c("black","red"))
```



Interpretation:

In the time series plot of original cheese data, we can see increasing trend of cheese production. However, when we transform the original data by taking first difference, such trend is removed.

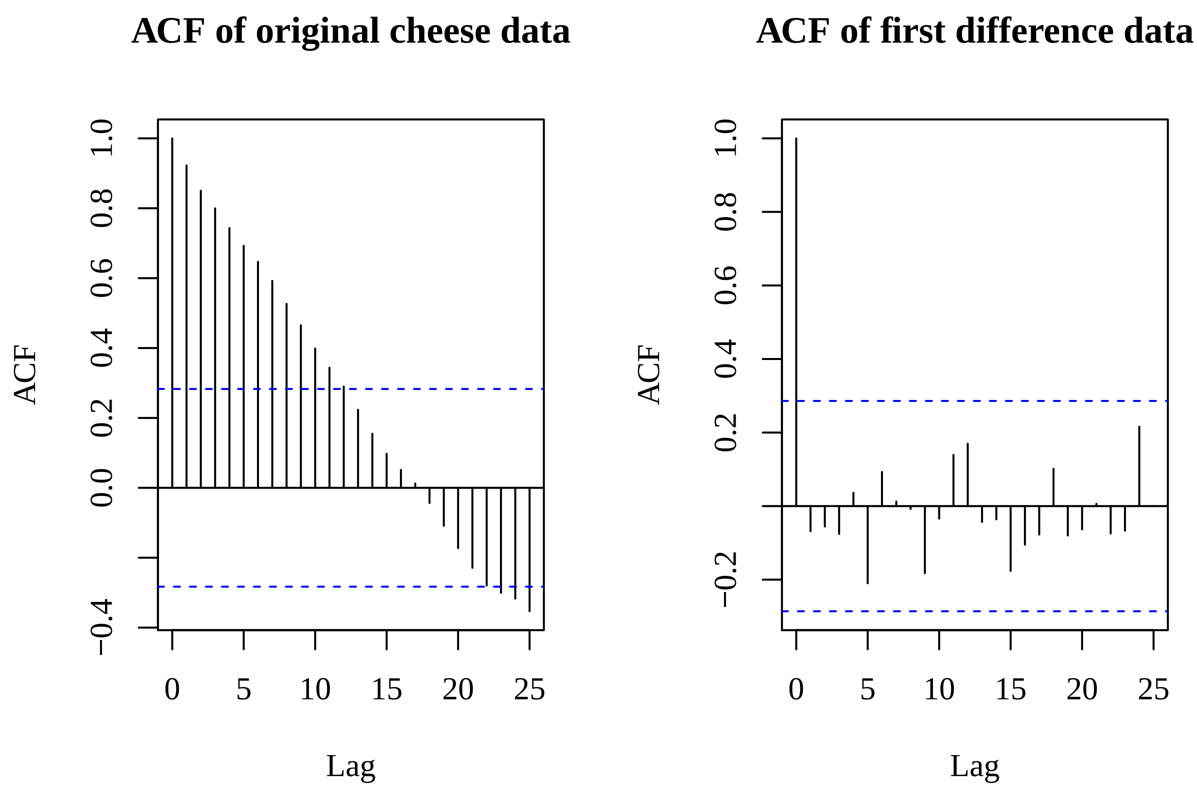
```
par(mfrow=c(2,2),oma=c(0,0,0,0))
par(family="Times")
qqnorm(fit.diff.cheese$res,datax=TRUE,pch=16,xlab='Residual',main='')
qqline(fit.diff.cheese$res,datax=TRUE)
plot(fit.diff.cheese$fit,fit.diff.cheese$res,pch=16, xlab='Fitted Value',
ylab='Residual')
abline(h=0)
hist(fit.diff.cheese$res,col="gray",xlab='Residual',main='')
plot(fit.diff.cheese$res,type="l",xlab='Observation Order',
ylab='Residual')
points(fit.diff.cheese$res,pch=16,cex=.5)
abline(h=0)
```



Interpretation:

From the above plot of sample quantiles, we can see that the residuals follow normal distribution. Similarly, scatterplot of residuals are described by fitted value and observation order plot. Furthermore, the residual plot above shows the residuals' distribution, noramlity and skewneses of model.

```
#Sample autocorrelation function of original cheese data
par(mfrow=c(1, 2))
par(family="Times")
ACF_cheese.data <-acf(cheese.data[,2], lag.max=25,type="correlation",
                      main="ACF of original cheese data")
#Sample autocorrelation function of first difference cheese data
ACF_cheese.data <-acf(diff.cheese.data[,2], lag.max=25,type="correlation",
                      main="ACF of first difference data")
```

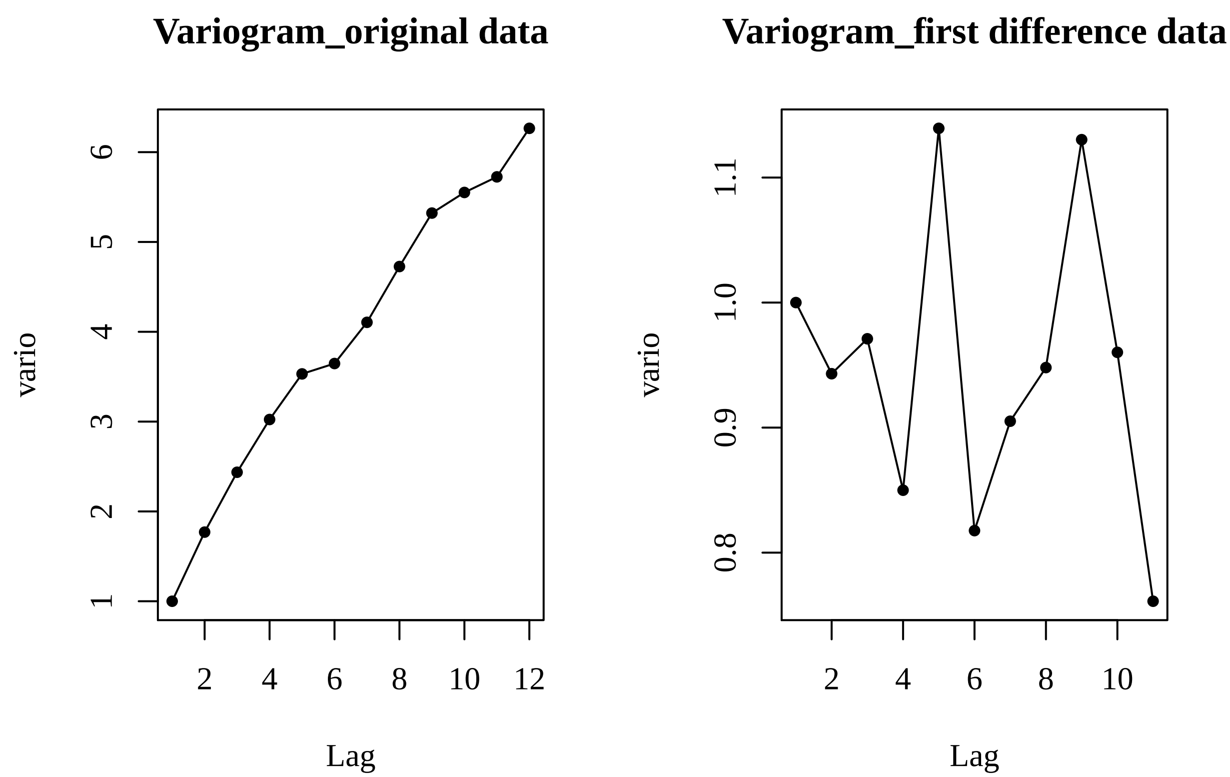


Interpretation:

From the plot above, we can see that the ACF plot of transformed data quickly decreases to zero and is within the significant range. So, we can say that the data transformed taking first difference is stationary.

```
#Defining variogram function
variogram <- function(x, lag){
  Lag <- NULL
vark <- NULL
vario <- NULL
for (k in 1:lag){
Lag[k] <- k
vark[k] = sd(diff(x,k))^2
vario[k] = vark[k]/vark[1]
}
return(as.data.frame(cbind(Lag,vario)))
}

#Variogram of  original cheese time series data
par(mfrow=c(1, 2))
par(family="Times")
y_cheese <- as.matrix(cheese.data[, 2])
vario_ <- variogram(y_cheese, length(y_cheese)/4)
plot(vario_, type="l", main="Variogram_original data", pch=19,cex=.8, col="black" )
points(vario_ ,pch=16,cex=.8)

#Variogram of first difference cheese data
```

```r
diff_y_cheese <- as.matrix(diff.cheese.data[, 2])
diff_vario_ <- variogram(diff_y_cheese, length(diff_y_cheese)/4)
plot(diff_vario_, type="l", main="Variogram_first difference data", pch=19,cex=.8,
     col="black" )
points(diff_vario_ ,pch=16,cex=.8)
```



Interpretation:

From the plot above, we can see that the variogram of original cheese data shows a increasing trend. However, that trend is removed when we take first difference of time series data. Similarly, on comparing the variograms of original and transformed data, we see less variance in transformed data. Variogram of transformed data seems stable which depicts that the transformed data is stationary.

2.11 Reconsider the data on the number of airline miles flown in the United Kingdom from Exercise 2.10. Take the natural logarithm of the data and plot this new time series.

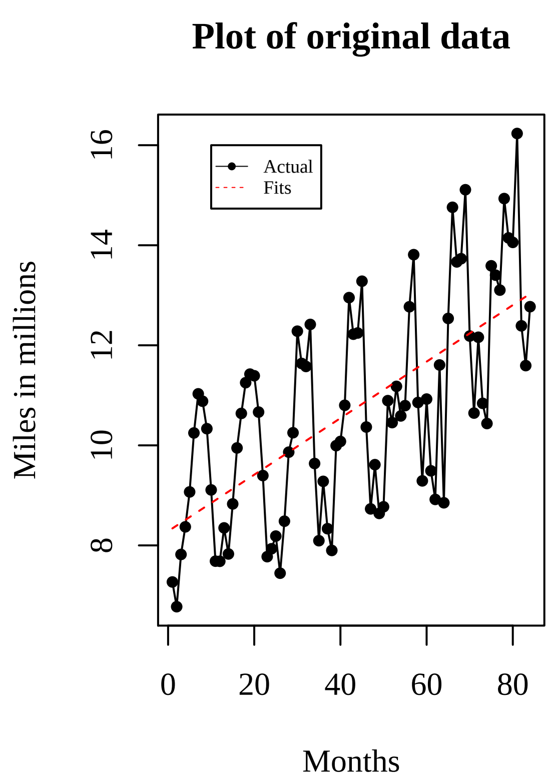
a. What impact has the log transformation had on the time series?

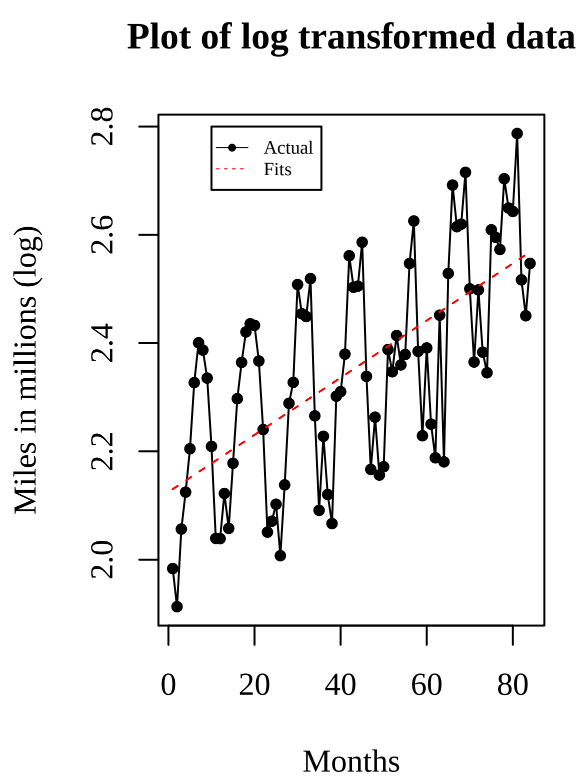
```r
Airlines.data <- read.csv("Airlinesdata.csv", header=TRUE, sep=",")
head(Airlines.data)
```

```
##   ï..Month Miles..in.Millions
## 1   Jan-64              7.269
## 2   Feb-64              6.775
## 3   Mar-64              7.819
## 4   Apr-64              8.371
## 5   May-64              9.069
```

```
## 6   Jun-64            10.248
```

```r
#Detrend the data
nrb<-dim(Airlines.data)[1]
tt<-1:nrb

#Time series plot of original airlines data
par(mfrow=c(1, 2))
par(family="Times")
fit.airlines <- lm(Airlines.data[,2]~tt)
plot(Airlines.data[,2], type="l", main="Plot of original data", ylab="Miles in millions",
     xlab = "Months")
points(Airlines.data[,2] ,pch=16,cex=.8)
lines(tt, fit.airlines$fit,col="red",lty=2)
legend(10, 16 ,c("Actual","Fits"),
pch=c(16,NA),lwd=c(.5,.5),lty=c(1,2),cex=.55,col=c("black","red"))

#Time series plot of log transformed airlines data
fit.log.airlines <- lm(log(Airlines.data[,2])~tt)
plot(log(Airlines.data[,2]), type="l", main="Plot of log transformed data",
     ylab="Miles in millions (log)", xlab = "Months")
points(log(Airlines.data[,2]) ,pch=16,cex=.8)
lines(tt, fit.log.airlines$fit,col="red",lty=2)
legend(10, 2.8 ,c("Actual","Fits"),
pch=c(16,NA),lwd=c(.5,.5),lty=c(1,2),cex=.55,col=c("black","red"))
```

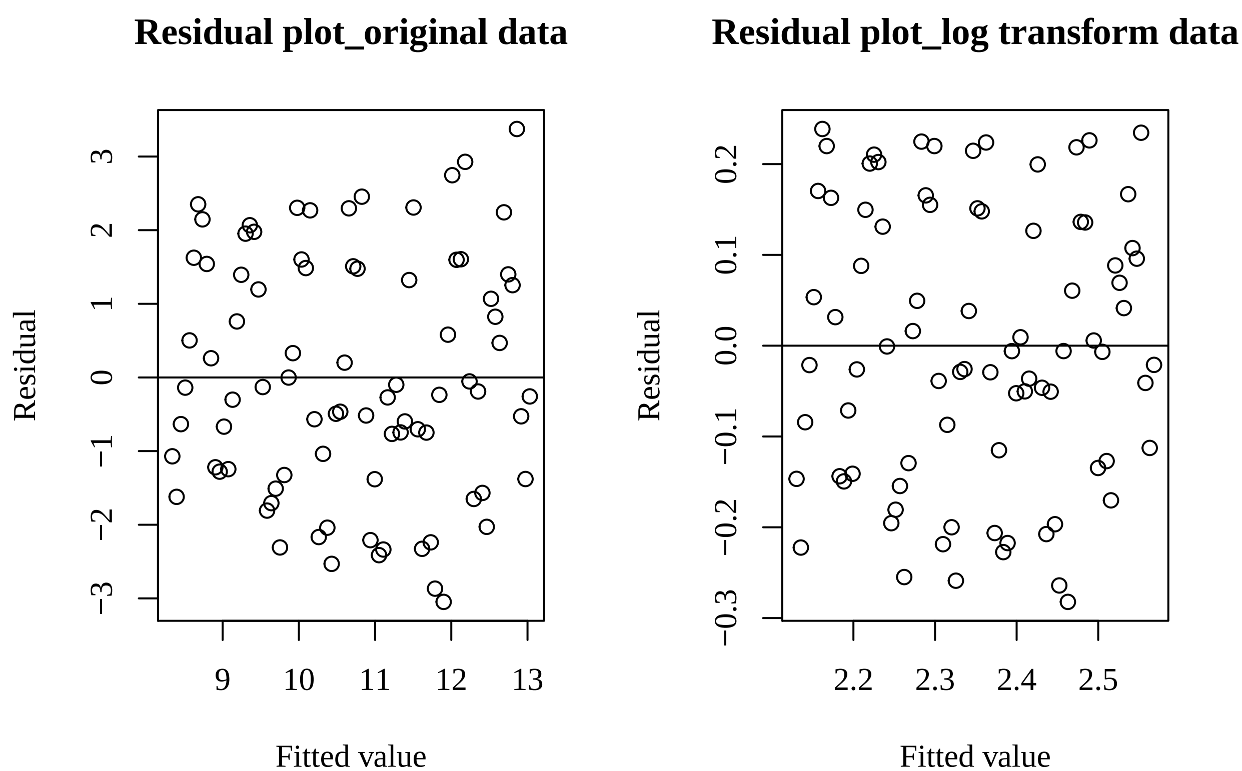




Interpretation:

From the above plot, we can see that there is a presence of trend and variance in original data showing its spread from the range of 6 to 16. While log transforming the data, the variation of time series data is reduced. However, there is still a trend in the data with log transformation.

```
#Residual plots of model from original data and model from log transformed data
par(mfrow=c(1, 2))
par(family="Times")
plot(fit.airlines$fitted.values, fit.airlines$residuals, pch=1, xlab = "Fitted value",
     ylab = "Residual", main="Residual plot_original data")
abline(h=0)
plot(fit.log.airlines$fitted.values, fit.log.airlines$residuals, pch=1, xlab = "Fitted value",
     ylab = "Residual", main="Residual plot_log transform data")
abline(h=0)
```



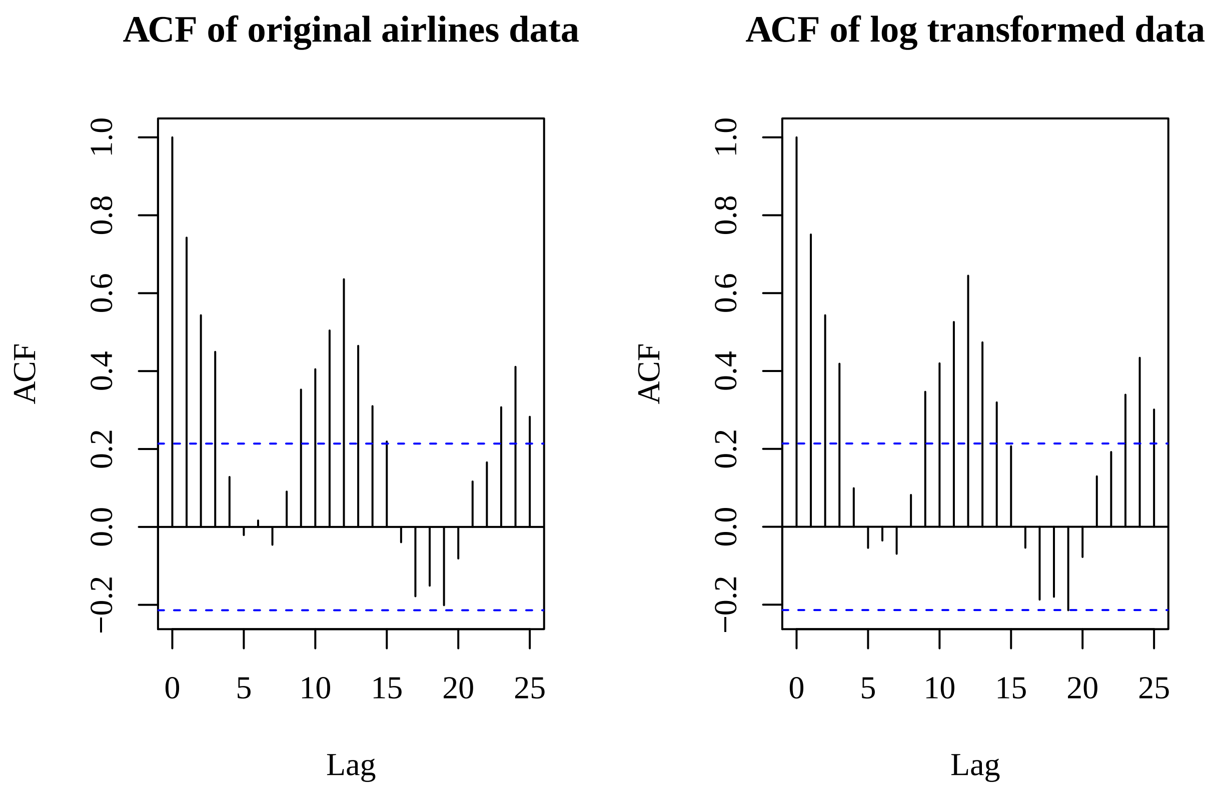
Interpretation:

Also, variance is compared by observing the residuals plot. Here, we can see that the variance in transformed data is very low compared to original data.

b. Find the autocorrelation function for this time series.

```
#Sample autocorrelation function of original airlines data
par(mfrow=c(1, 2))
par(family="Times")
ACF_airlines.data <-acf(Airlines.data[,2], lag.max=25,type="correlation",
                        main="ACF of original airlines data")
```

```r
#Sample autocorrelation function of log transformed airlines data
ACF_log.airlines.data <-acf(log(Airlines.data[,2]), lag.max=25,
                            type="correlation",main="ACF of log transformed data")
```




```r
#Computing ACF values for log transformed data
ACF_log.airlines.data$acf[1:25]
```

```
##  [1]  1.00000000  0.75054359  0.54324989  0.41873794  0.09902608 -0.05424545
##  [7] -0.03532738 -0.06937698  0.08184879  0.34663346  0.41969845  0.52589269
## [13]  0.64478676  0.47356223  0.31944526  0.20677895 -0.05379615 -0.18705174
## [19] -0.17979997 -0.21428802 -0.07756771  0.12957065  0.19205635  0.33913957
## [25]  0.43392200
```

c. Interpret the sample autocorrelation function.

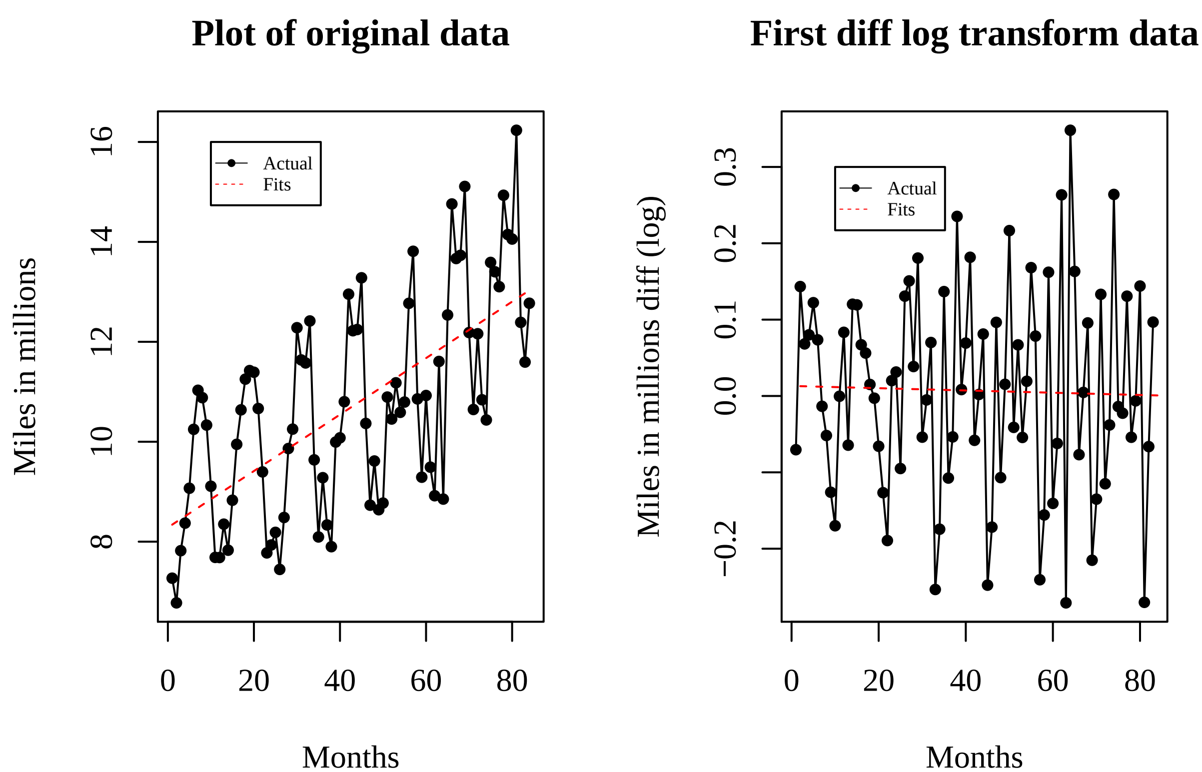
Interpretation:

From the plot of ACF, we can see that the ACF value of original data quickly reduced to zero and fluctuated around it. However, the ACF values exceed beyond the range of line. Looking this, we cannot say that given original airlines data is stationary. Futhermore, conducting log transformation in airlines data decreased variance of data. But, this could not change trend in data. Due to this, we also see high fluctuation in ACF of log transformed data as like in original data. Considering all these issues in log transformation, we need some other transformations like seasonal or differencial transformaions to make this time series data stationary.

2.12 Reconsider the data on the number of airline miles flown in the United Kingdom from Exercises 2.10 and 2.11. Take the first difference of the natural logarithm of the data and plot this new time series.

a. What impact has the log transformation had on the time series?

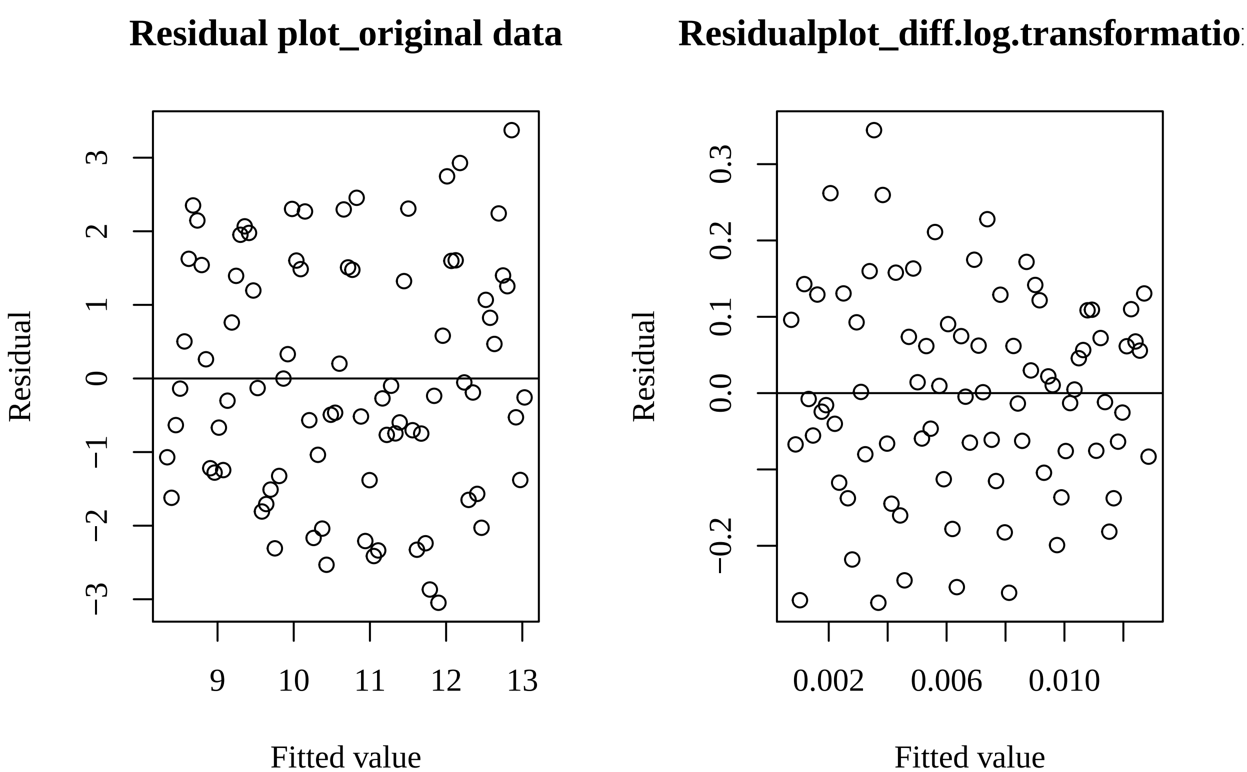
```
#Time series plot of original airlines data
par(mfrow=c(1, 2))
par(family="Times")
fit.airlines <- lm(Airlines.data[,2]~tt)
plot(Airlines.data[,2], type="l", main="Plot of original data", ylab="Miles in millions",
     xlab = "Months")
points(Airlines.data[,2] ,pch=16,cex=.8)
lines(tt, fit.airlines$fit,col="red",lty=2)
legend(10, 16 ,c("Actual","Fits"),
pch=c(16,NA),lwd=c(.5,.5),lty=c(1,2),cex=.55,col=c("black","red"))

#Time series plot of first difference log transformed airlines data
log <- log(Airlines.data[,2])
diff.log.airlines.data <- cbind(2:nrb,diff(log))
fit.diff.log.airlines <- lm(diff.log.airlines.data[,2]~diff.log.airlines.data[,1])
plot(diff.log.airlines.data[,2], type="l", main="First diff log transform data",
     ylab="Miles in millions diff (log)", xlab = "Months")
points((diff.log.airlines.data[,2]), pch=16,cex=.8)
lines(diff.log.airlines.data[,1], fit.diff.log.airlines$fit,col="red",lty=2)
legend(10, 0.3 ,c("Actual","Fits"),
pch=c(16,NA),lwd=c(.5,.5),lty=c(1,2),cex=.55,col=c("black","red"))
```



Interpretation:

From the above plot, we can see that there is a presence of trend and variance in original data showing its spread from the range of 6 to 16. While transforming the data by log transformation and taking first difference, the variation in time series data is reduced and there is no trend in data with fist difference log transformed data.

```
#Residual plots of model from original data and model from first difference log transformed data
par(mfrow=c(1, 2))
par(family="Times")
par(font.axis=6)
plot(fit.airlines$fitted.values, fit.airlines$residuals, pch=1, xlab = "Fitted value",
     ylab = "Residual", main="Residual plot_original data")
abline(h=0)
plot(fit.diff.log.airlines$fitted.values, fit.diff.log.airlines$residuals, pch=1,
     xlab = "Fitted value", ylab = "Residual", main="Residualplot_diff.log.transformation")
abline(h=0)
```



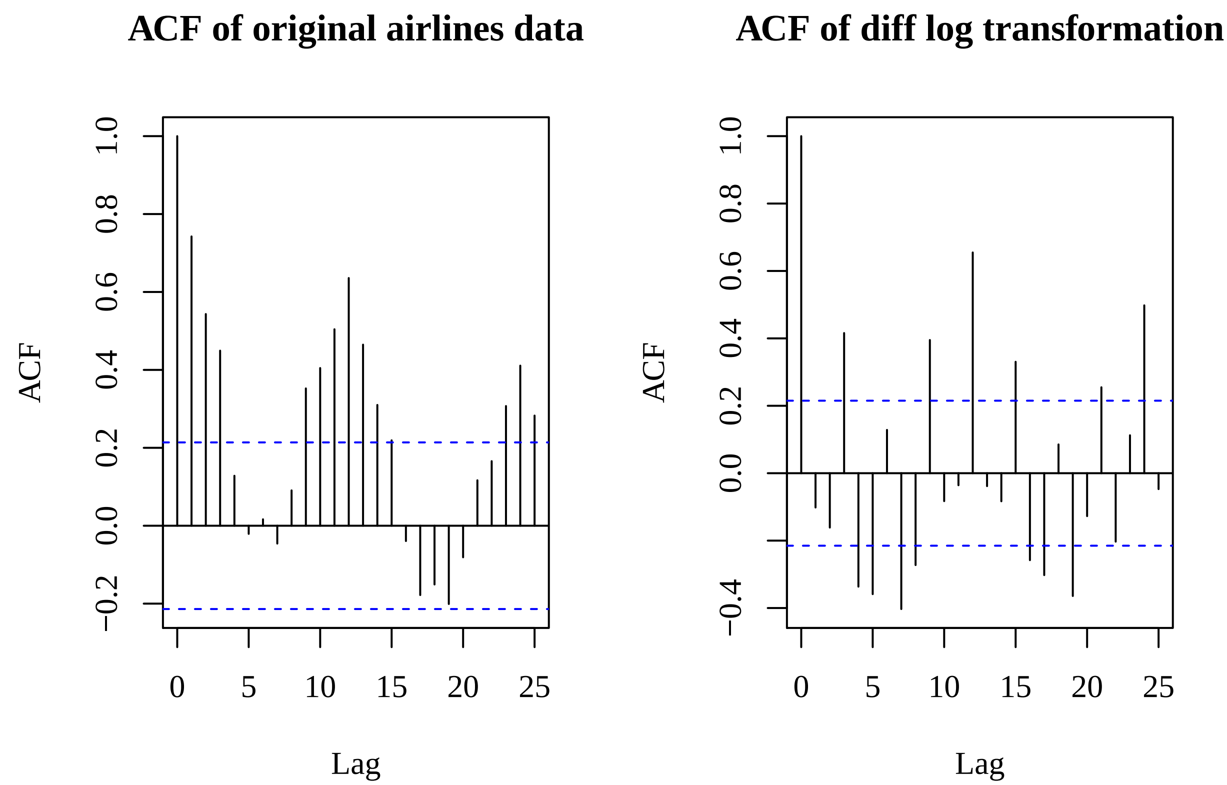
Interpretation:

Also, variance is compared by having a look at residuals plot. Here, we can see that the variance in first difference log transformed data is reduced and is more constant compared to original data.

b. Find the autocorrelation function for this time series.

```
#Sample autocorrelation function of original airlines data
par(mfrow=c(1, 2))
par(family="Times")
```

```
ACF_airlines.data <-acf(Airlines.data[,2], lag.max=25,type="correlation",
                        main="ACF of original airlines data")
#Sample autocorrelation function of first difference log transformed airlines data
ACF_diff.log.airlines.data <-acf(diff.log.airlines.data[,2], lag.max=25,type="correlation",
                                 main="ACF of diff log transformation")
```




```
#Computing ACF values for log transformed data
ACF_diff.log.airlines.data$acf[1:25]
```

```
##  [1]  1.00000000 -0.10174145 -0.16131694  0.41582795 -0.33656967 -0.35878708
##  [7]  0.12845827 -0.40298769 -0.27255118  0.39516512 -0.08293518 -0.03570862
## [13]  0.65489754 -0.03835996 -0.08328607  0.33057908 -0.25800213 -0.30240211
## [19]  0.08538301 -0.36430811 -0.12734191  0.25500487 -0.20307139  0.11255443
## [25]  0.49804880
```

c. Interpret the sample autocorrelation function.

Interpretation:

From the plot of ACF, we can see that the ACF value of original data quickly reduced to zero and fluctuated around it. However, the ACF values exceed beyond the range of line. Looking this, we cannot say that given original airlines data is stationary. Futhermore, conducting log transformation and taking first difference of airlines data decreased variance of data and change trend in data. But, still ACF values are not stable and are fluctuating outside the range of line after being reduced to zero. Considering all these issues in first difference log transformation data, still it is very hard to say that the data is stationary.