#Amulya Geereddy

**Preprocessing**

library(rio)

library(moments)

library(car)

mydata=import("6304 Time Series Assignment Data.xlsx")

colnames(mydata)=tolower(make.names(colnames(mydata)))

mydata$item=seq(1:nrow(mydata))

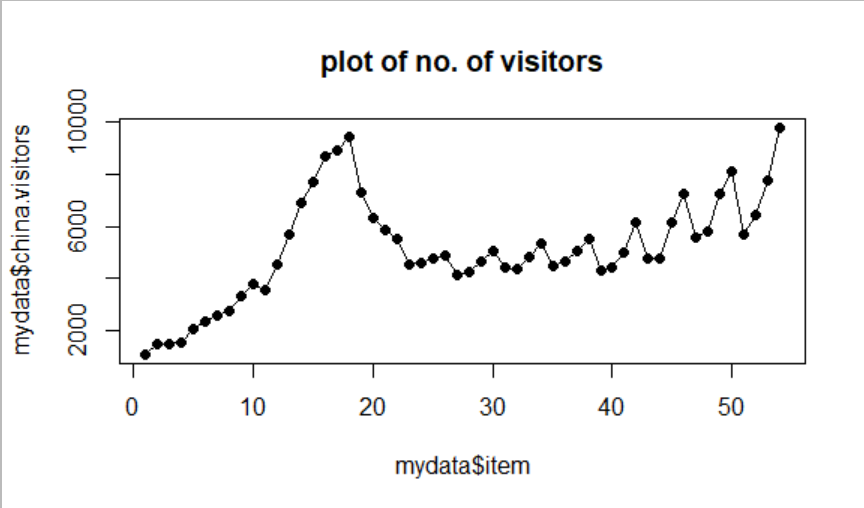
attach(mydata)

**Analysis**

1 .

plot(mydata$item,mydata$china.visitors,type='o', main="plot of no. of visitors",pch=19)

**Output:**

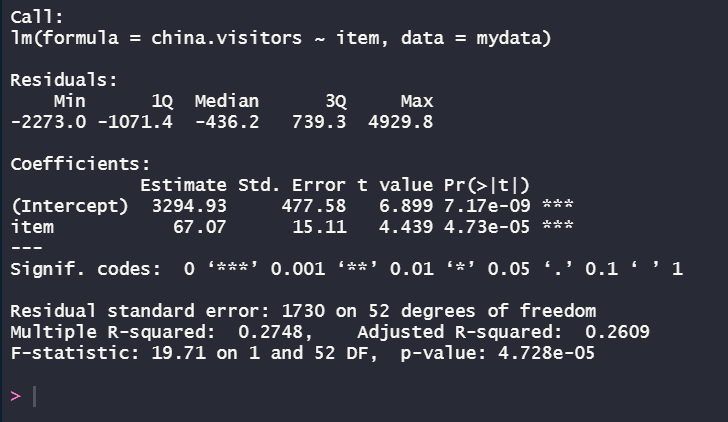


2 .

base\_out=lm(china.visitors~item,data=mydata)

summary(base\_out)

**Output:**



From the above data we can say that the beta coefficients are 3294.93 and 67.07 and both are statistically significant because the p-values are alot less than 5 percent.

The regression equation represents that china visitors increases by 67.07 every quarter.

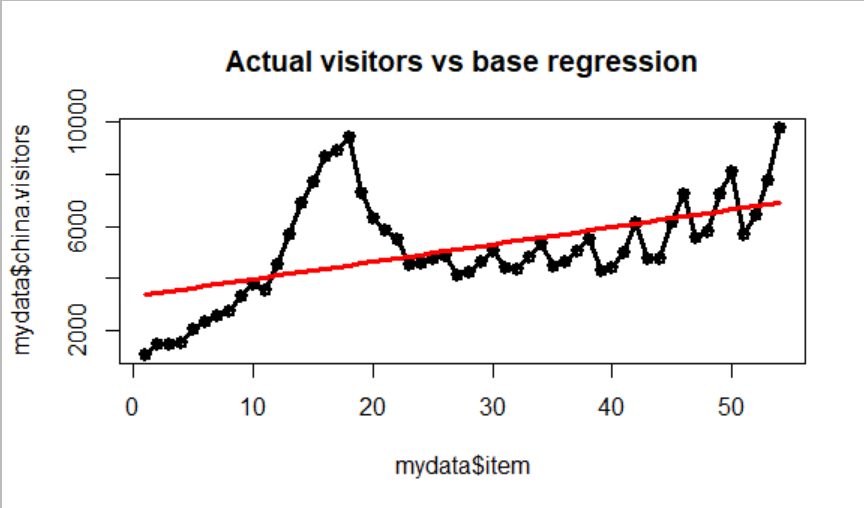
The R-squared value is **27%** for the data.

3 .

plot(mydata$item,mydata$china.visitors,type="o",lwd=3 ,main="Actual visitors vs base regression")

points(mydata$item,base\_out$fitted.values,type="l",col="red",lwd=3)

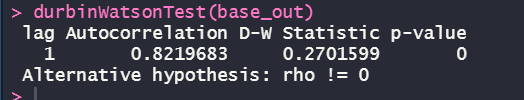
**Output:**



**4 .**

durbinWatsonTest(base\_out)

**Output:**



From the above Durbin Watson test values, The value shows that there exists positive auto correlation for the given data.

5 .

for(i in 1:4) {

count=0

for(j in 1:nrow(mydata)) {

if(i==mydata$quarter[j]) {

indices$average[i]=indices$average[i]+mydata$china.visitors[j]

count=count+1

}

}

indices$average[i]=indices$average[i]/count

indices$index[i]=indices$average[i]/mean(mydata$china.visitors)}

#Deseasonalizing the original data

for(i in 1:4){

for(j in 1:nrow(mydata)){

if(i==mydata$quarter[j]){

mydata$deseason.visitors[j]=mydata$china.visitors[j]/indices$index[i]

}

}

}

6 .

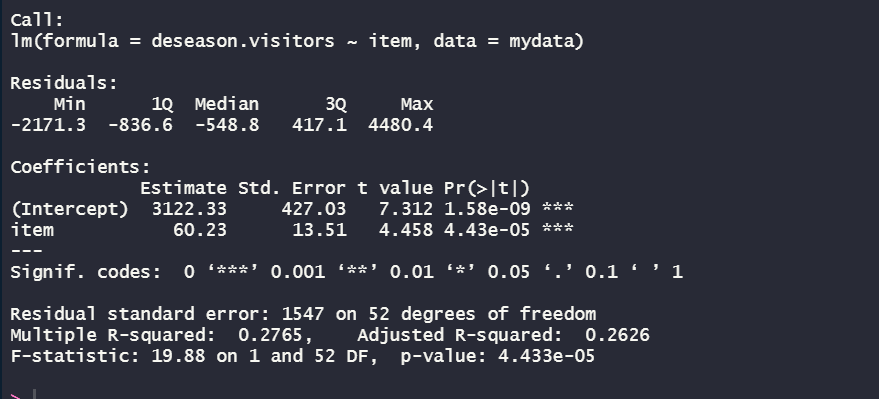
des\_out=lm(deseason.visitors~item,data=mydata)

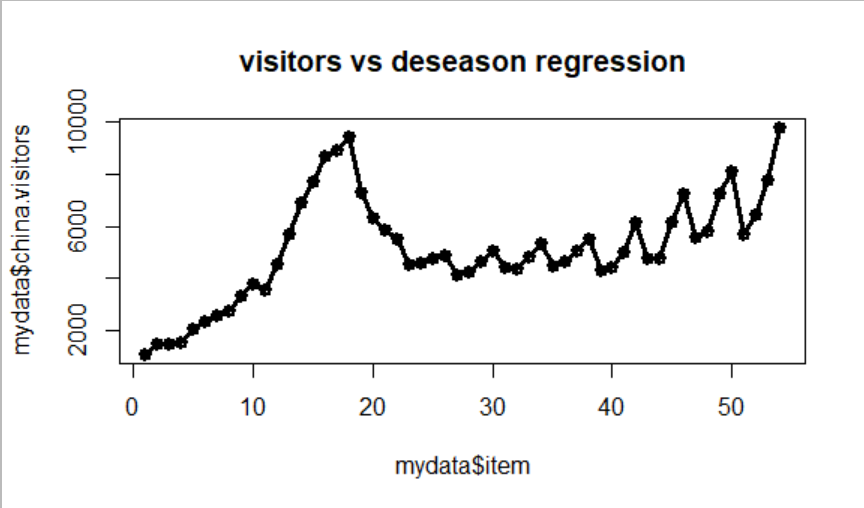
summary(des\_out)

plot(mydata$item,mydata$china.visitors,type="o",lwd=3 ,main="visitors vs deseason regression")

points(mydata$item,desreg\_out$fitted.values,type="l",col="red",lwd=3)

**Output:**





7 .

data$deseason.forecast=desreg\_out$fitted.values

for(i in 1:4){

for(j in 1:nrow(data)){

if(i==data$quarter[j]){

data$reseason.forecast[j]=data$deseason.forecast[j]\*indices$index[i]

}

}

}

plot(data$item,data$china.visitors,type="o",pch=19,

main="Original Data and Reseasonalized Forecasts")

par(new = TRUE)

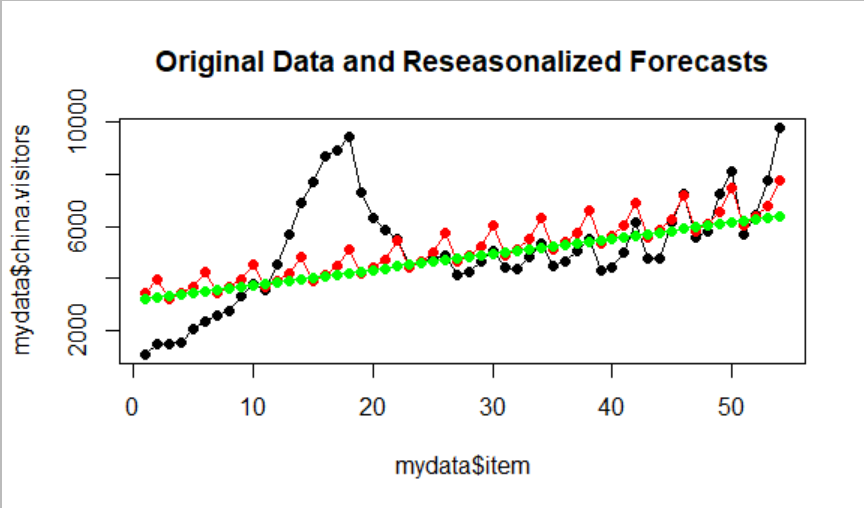
points(data$item,data$reseason.forecast,

type="o",pch=19,col="red")

points(data$item,data$deseason.forecast,

type="o",pch=19,col="green")

**Output**:



From the above plot, we can see that reseasonalized values have a better fit than the original data.

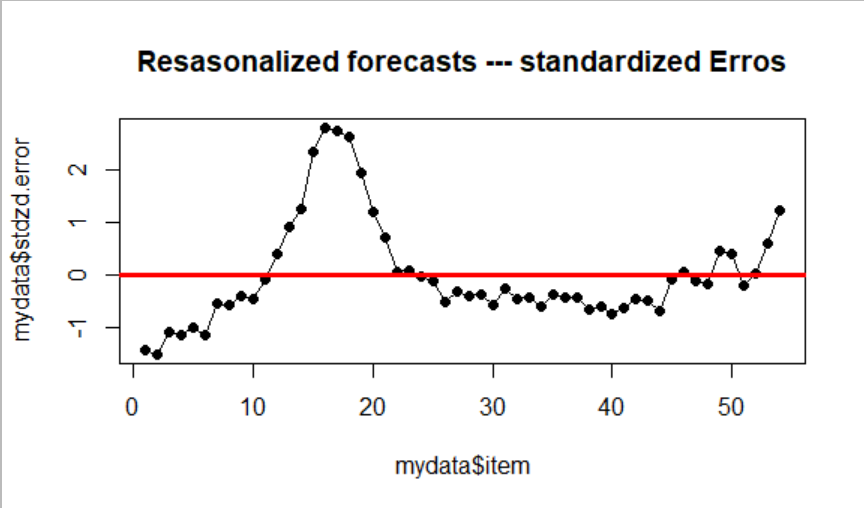
8 .

mydata$error=mydata$china.visitors-mydata$reseason.forecast

mydata$stdzd.error=scale(mydata$error)

plot(datanew$index,datanew$stdzd.error,pch=19,type="o",main="Resasonalized forecasts --- standardized Erros")

abline(0,0,col="red",lwd=3)



From the above plot we can say that the plot has the outliers on the graph. And there is a pattern visible for the errors.