Forecasting model

##importing the data set

sales <- read.csv(file.choose())

str(sales)

## in th edata set we have two variable one is month and one is sales

## We will change the number of months into dummy variable

X<- data.frame(outer(rep(month.abb,length = 60), month.abb,"==") + 0 )# Creating dummies for 12 months

View(X)

colnames(X)<-month.abb # Assigning month names

View(X)

##combining the x to the data

newsales <- cbind(sales,X)

newsales <- newsales[,-1]

newsales['t'] <- 1:60

View(newsales)

newsales['log\_sales'] <- log(newsales$Sales)

View(newsales)

newsales['t\_squared'] <- newsales$t\*newsales$t

View(newsales)

attach(newsales)

##spliting the data##

train <- newsales[1:45,]

test <- newsales[45:60,]

########################### LINEAR MODEL #############################

linear\_model<-lm(Sales~t,data=train)

summary(linear\_model)

linear\_pred<-data.frame(predict(linear\_model,interval='predict',newdata =test))

View(linear\_pred)

rmse\_linear<-sqrt(mean((test$Sales-linear\_pred$fit)^2,na.rm = T))

rmse\_linear # 243.6692

######################### Exponential #################################

expo\_model<-lm(log\_sales~t,data=train)

summary(expo\_model)

expo\_pred<-data.frame(predict(expo\_model,interval='predict',newdata=test))

rmse\_expo<-sqrt(mean((test$log\_sales-exp(expo\_pred$fit))^2,na.rm = T))

rmse\_expo # 1409.767

######################### Quadratic ####################################

Quad\_model<-lm(Sales~t+t\_squared,data=train)

summary(Quad\_model)

Quad\_pred<-data.frame(predict(Quad\_model,interval='predict',newdata=test))

rmse\_Quad<-sqrt(mean((test$Sales-Quad\_pred$fit)^2,na.rm=T))

rmse\_Quad # 274.362

######################### Additive Seasonality #########################

sea\_add\_model<-lm(Sales~Jan+Feb+Mar+Apr+May+Jun+Jul+Aug+Sep+Oct+Nov,data=train)

summary(sea\_add\_model)

sea\_add\_pred<-data.frame(predict(sea\_add\_model,newdata=test,interval='predict'))

rmse\_sea\_add<-sqrt(mean((test$Sales-sea\_add\_pred$fit)^2,na.rm = T))

rmse\_sea\_add # 253.816

######################## Additive Seasonality with Linear #################

Add\_sea\_Linear\_model<-lm(Sales~t+Jan+Feb+Mar+Apr+May+Jun+Jul+Aug+Sep+Oct+Nov,data=train)

summary(Add\_sea\_Linear\_model)

Add\_sea\_Linear\_pred<-data.frame(predict(Add\_sea\_Linear\_model,interval='predict',newdata=test))

rmse\_Add\_sea\_Linear<-sqrt(mean((test$Sales-Add\_sea\_Linear\_pred$fit)^2,na.rm=T))

rmse\_Add\_sea\_Linear #114.235

######################## Additive Seasonality with Quadratic #################

Add\_sea\_Quad\_model<-lm(Sales~t+t\_squared+Jan+Feb+Mar+Apr+May+Jun+Jul+Aug+Sep+Oct+Nov,data=train)

summary(Add\_sea\_Quad\_model)

Add\_sea\_Quad\_pred<-data.frame(predict(Add\_sea\_Quad\_model,interval='predict',newdata=test))

rmse\_Add\_sea\_Quad<-sqrt(mean((test$Sales-Add\_sea\_Quad\_pred$fit)^2,na.rm=T))

rmse\_Add\_sea\_Quad # 146.5779

######################## Multiplicative Seasonality #########################

multi\_sea\_model<-lm(log\_sales~Jan+Feb+Mar+Apr+May+Jun+Jul+Aug+Sep+Oct+Nov,data = train)

summary(multi\_sea\_model)

multi\_sea\_pred<-data.frame(predict(multi\_sea\_model,newdata=test,interval='predict'))

rmse\_multi\_sea<-sqrt(mean((test$Sales-exp(multi\_sea\_pred$fit))^2,na.rm = T))

rmse\_multi\_sea # 258.3827

######################## Multiplicative Seasonality Linear trend ##########################

multi\_add\_sea\_model<-lm(log\_sales~t+Jan+Feb+Mar+Apr+May+Jun+Jul+Aug+Sep+Oct+Nov,data = train)

summary(multi\_add\_sea\_model)

multi\_add\_sea\_pred<-data.frame(predict(multi\_add\_sea\_model,newdata=test,interval='predict'))

rmse\_multi\_add\_sea<-sqrt(mean((test$Sales-exp(multi\_add\_sea\_pred$fit))^2,na.rm = T))

rmse\_multi\_add\_sea # 129.1332

# Preparing table on model and it's RMSE values

table\_rmse<-data.frame(c("rmse\_linear","rmse\_expo","rmse\_Quad","rmse\_sea\_add","rmse\_Add\_sea\_Linear","rmse\_Add\_sea\_Quad","rmse\_multi\_sea","rmse\_multi\_add\_sea"),c(rmse\_linear,rmse\_expo,rmse\_Quad,rmse\_sea\_add,rmse\_Add\_sea\_Linear,rmse\_Add\_sea\_Quad,rmse\_multi\_sea,rmse\_multi\_add\_sea))

View(table\_rmse)

colnames(table\_rmse)<-c("model","RMSE")

View(table\_rmse)

#Addictive seasonality linear trend has least RMSE value

new\_model <- lm(Sales~t+Jan+Feb+Mar+Apr+May+Jun+Jul+Aug+Sep+Oct+Nov,data = newsales)

# Building Autoregressive model on residuals consider lag-1

resid <- residuals(new\_model)

resid[1:10]

windows()

acf(resid,lag.max = 10)

# By principal of parcimony we will consider lag - 1 as we have so

# many significant lags

k <- arima(resid, order=c(1,0,0))

str(k)

View(data.frame(res=resid,newresid=k$residuals))

windows()

acf(k$residuals,lag.max = 15)

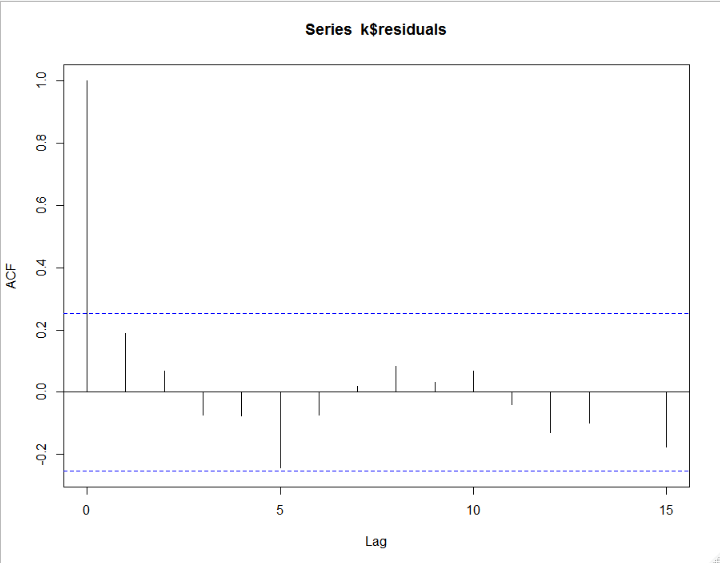
pred\_res<- predict(arima(k$residuals,order=c(1,0,0)),n.ahead = 12)

str(pred\_res)

pred\_res$pred

acf(k$residuals)

|  |  |  |
| --- | --- | --- |
| **1** | rmse\_linear | 243.6692 |
| **2** | rmse\_expo | 1409.7670 |
| **3** | rmse\_Quad | 274.3620 |
| **4** | rmse\_sea\_add | 253.8168 |
| **5** | rmse\_Add\_sea\_Linear | 114.2350 |
| **6** | rmse\_Add\_sea\_Quad | 146.5779 |
| **7** | rmse\_multi\_sea | 258.3827 |
| **8** | rmse\_multi\_add\_sea | 129.1332 |



Data set Coca-Cola

library(readxl)

#loading the data in excel into the system##

data <- read\_excel(file.choose())

View(data)## quarter seasonality in the dataset##

head(data)

windows()

attach(data)

plot(data$Sales,type='o')

##creating four dummy variable because four quarters##

Q1 <- ifelse(grepl("Q1",data$Quarter),'1','0')

Q2 <- ifelse(grepl("Q2",data$Quarter),'1','0')

Q3 <- ifelse(grepl("Q3",data$Quarter),'1','0')

Q4 <- ifelse(grepl("Q4",data$Quarter),'1','0')

##combining the dummy variable with the data##

data <- cbind(data,Q1,Q2,Q3,Q4)

colnames(data)

data['t'] <- 1:42

View(data)

data["log\_Sales"]<-log(data["Sales"])

data["t\_square"]<-data["t"]\*data["t"]

attach(data)

View(data)

train<-data[1:36,]

test<-data[36:42,]

########################### LINEAR MODEL #############################

linear\_model<-lm(Sales~t,data=train)

summary(linear\_model)

linear\_pred<-data.frame(predict(linear\_model,interval='predict',newdata =test))

View(linear\_pred)

rmse\_linear<-sqrt(mean((test$Sales-linear\_pred$fit)^2,na.rm = T))

rmse\_linear # 620.7862

######################### Exponential #################################

expo\_model<-lm(log\_Sales~t,data=train)

summary(expo\_model)

expo\_pred<-data.frame(predict(expo\_model,interval='predict',newdata=test))

rmse\_expo<-sqrt(mean((test$log\_Sales-exp(expo\_pred$fit))^2,na.rm = T))

rmse\_expo # 4238.53

######################### Quadratic ####################################

Quad\_model<-lm(Sales~t+t\_square,data=train)

summary(Quad\_model)

Quad\_pred<-data.frame(predict(Quad\_model,interval='predict',newdata=test))

rmse\_Quad<-sqrt(mean((test$Sales-Quad\_pred$fit)^2,na.rm=T))

rmse\_Quad # 458.8969

######################### Additive Seasonality #########################

sea\_add\_model<-lm(Sales~Q1+Q2+Q3+Q4,data=train)

summary(sea\_add\_model)

sea\_add\_pred<-data.frame(predict(sea\_add\_model,newdata=test,interval='predict'))

rmse\_sea\_add<-sqrt(mean((test$Sales-sea\_add\_pred$fit)^2,na.rm = T))

rmse\_sea\_add # 1822.849

######################## Additive Seasonality with Linear #################

Add\_sea\_Linear\_model<-lm(Sales~t+Q1+Q2+Q3+Q4,data=train)

summary(Add\_sea\_Linear\_model)

Add\_sea\_Linear\_pred<-data.frame(predict(Add\_sea\_Linear\_model,interval='predict',newdata=test))

rmse\_Add\_sea\_Linear<-sqrt(mean((test$Sales-Add\_sea\_Linear\_pred$fit)^2,na.rm=T))

rmse\_Add\_sea\_Linear #525.0682

######################## Additive Seasonality with Quadratic #################

Add\_sea\_Quad\_model<-lm(Sales~t+t\_square+Q1+Q2+Q3+Q4,data=train)

summary(Add\_sea\_Quad\_model)

Add\_sea\_Quad\_pred<-data.frame(predict(Add\_sea\_Quad\_model,interval='predict',newdata=test))

rmse\_Add\_sea\_Quad<-sqrt(mean((test$Sales-Add\_sea\_Quad\_pred$fit)^2,na.rm=T))

rmse\_Add\_sea\_Quad # 266.4439

######################## Multiplicative Seasonality #########################

multi\_sea\_model<-lm(log\_Sales~Q1+Q2+Q3+Q4,data = train)

summary(multi\_sea\_model)

multi\_sea\_pred<-data.frame(predict(multi\_sea\_model,newdata=test,interval='predict'))

rmse\_multi\_sea<-sqrt(mean((test$Sales-exp(multi\_sea\_pred$fit))^2,na.rm = T))

rmse\_multi\_sea # 1906.603

######################## Multiplicative Seasonality Linear trend ##########################

multi\_add\_sea\_model<-lm(log\_Sales~t+Q1+Q2+Q3+Q4,data = train)

summary(multi\_add\_sea\_model)

multi\_add\_sea\_pred<-data.frame(predict(multi\_add\_sea\_model,newdata=test,interval='predict'))

rmse\_multi\_add\_sea<-sqrt(mean((test$Sales-exp(multi\_add\_sea\_pred$fit))^2,na.rm = T))

rmse\_multi\_add\_sea # 309.995

# Preparing table on model and it's RMSE values

table\_rmse<-data.frame(c("rmse\_linear","rmse\_expo","rmse\_Quad","rmse\_sea\_add","rmse\_Add\_sea\_Quad","rmse\_multi\_sea","rmse\_multi\_add\_sea"),c(rmse\_linear,rmse\_expo,rmse\_Quad,rmse\_sea\_add,rmse\_Add\_sea\_Quad,rmse\_multi\_sea,rmse\_multi\_add\_sea))

View(table\_rmse)

colnames(table\_rmse)<-c("model","RMSE")

View(table\_rmse)

# Additive seasonality with Quadratic has least RMSE value

new\_model <- lm(Sales~t+t\_square+Q1+Q2+Q3+Q4,data=data)

resid <- residuals(new\_model)

resid[1:10]

windows()

acf(resid,lag.max = 10)

# By principal of parcimony we will consider lag - 1 as we have so

# many significant lags

# Building Autoregressive model on residuals consider lag-1

k <- arima(resid, order=c(1,0,0))

str(k)

View(data.frame(res=resid,newresid=k$residuals))

windows()

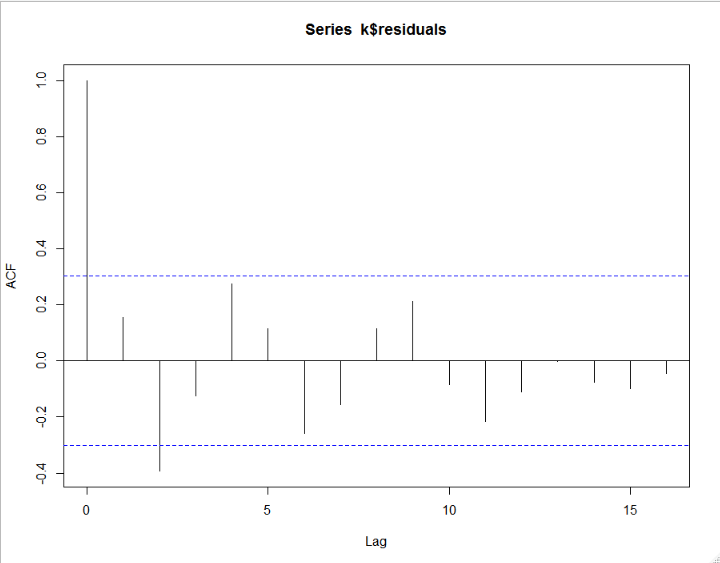
acf(k$residuals,lag.max = 15)

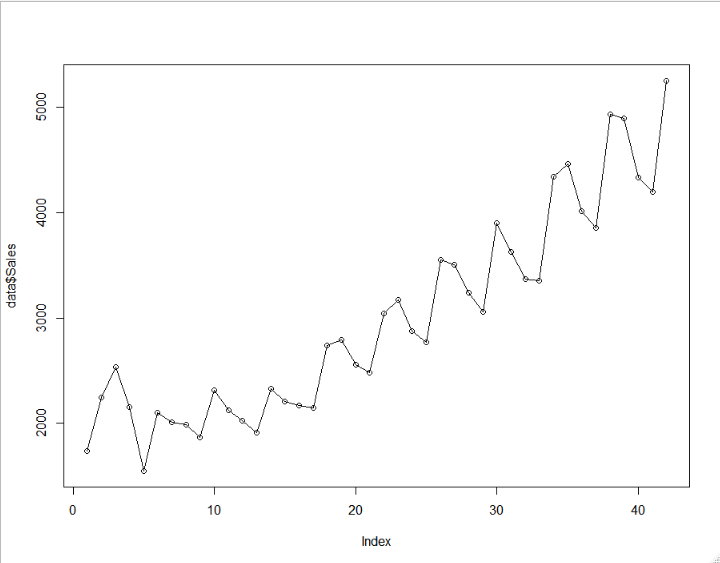
pred\_res<- predict(arima(k$residuals,order=c(1,0,0)),n.ahead = 12)

str(pred\_res)

pred\_res$pred

acf(k$residuals)





|  | **model** | **RMSE** |
| --- | --- | --- |
|  |  |  |
| **1** | rmse\_linear | 620.7862 |
| **2** | rmse\_expo | 4238.5334 |
| **3** | rmse\_Quad | 458.8969 |
| **4** | rmse\_sea\_add | 1822.8497 |
| **5** | rmse\_Add\_sea\_Quad | 266.4439 |
| **6** | rmse\_multi\_sea | 1906.6029 |
| **7** | rmse\_multi\_add\_sea | 309.9950 |

Showing 1 to 7 of 7 entries, 2 total columns

Airlines data set

library(readxl)

#loading the data in excel into the system##

data <- read\_excel(file.choose())

View(data)## Month seasonality in the data set##

head(data)

windows()

plot(data$Passengers,type='o')

# So creating 12 dummy variables

X<- data.frame(outer(rep(month.abb,length = 96), month.abb,"==") + 0 )# Creating dummies for 12 months

View(X)

colnames(X)<-month.abb # Assigning month names

View(X)

##combining data to x##

data <- cbind(data,X)

View(data)

colnames(data)

data['t'] <- 1:96

View(data)

data['log\_passengers'] <- log(data$Passengers)

View(data)

data['t\_squared'] <- data$t\*data$t

View(data)

attach(data)

##spliting the data##

train <- data[1:70,]

test <- data[70:96,]

########################### LINEAR MODEL #############################

linear\_model<-lm(Passengers~t,data=train)

summary(linear\_model)

linear\_pred<-data.frame(predict(linear\_model,interval='predict',newdata =test))

View(linear\_pred)

rmse\_linear<-sqrt(mean((test$Passengers-linear\_pred$fit)^2,na.rm = T))

rmse\_linear # 47.6409

######################### Exponential #################################

expo\_model<-lm(log\_passengers~t,data=train)

summary(expo\_model)

expo\_pred<-data.frame(predict(expo\_model,interval='predict',newdata=test))

rmse\_expo<-sqrt(mean((test$log\_passengers-exp(expo\_pred$fit))^2,na.rm = T))

rmse\_expo # 301.5812

######################### Quadratic ####################################

Quad\_model<-lm(Passengers~t+t\_squared,data=train)

summary(Quad\_model)

Quad\_pred<-data.frame(predict(Quad\_model,interval='predict',newdata=test))

rmse\_Quad<-sqrt(mean((test$Passengers-Quad\_pred$fit)^2,na.rm=T))

rmse\_Quad # 43.45248

######################### Additive Seasonality #########################

sea\_add\_model<-lm(Passengers~Jan+Feb+Mar+Apr+May+Jun+Jul+Aug+Sep+Oct+Nov,data=train)

summary(sea\_add\_model)

sea\_add\_pred<-data.frame(predict(sea\_add\_model,newdata=test,interval='predict'))

rmse\_sea\_add<-sqrt(mean((test$Passengers-sea\_add\_pred$fit)^2,na.rm = T))

rmse\_sea\_add # 123.0927

######################## Additive Seasonality with Linear #################

Add\_sea\_Linear\_model<-lm(Passengers~t+Jan+Feb+Mar+Apr+May+Jun+Jul+Aug+Sep+Oct+Nov,data=train)

summary(Add\_sea\_Linear\_model)

Add\_sea\_Linear\_pred<-data.frame(predict(Add\_sea\_Linear\_model,interval='predict',newdata=test))

rmse\_Add\_sea\_Linear<-sqrt(mean((test$Passengers-Add\_sea\_Linear\_pred$fit)^2,na.rm=T))

rmse\_Add\_sea\_Linear #33.87741

######################## Additive Seasonality with Quadratic #################

Add\_sea\_Quad\_model<-lm(Passengers~t+t\_squared+Jan+Feb+Mar+Apr+May+Jun+Jul+Aug+Sep+Oct+Nov,data=train)

summary(Add\_sea\_Quad\_model)

Add\_sea\_Quad\_pred<-data.frame(predict(Add\_sea\_Quad\_model,interval='predict',newdata=test))

rmse\_Add\_sea\_Quad<-sqrt(mean((test$Passengers-Add\_sea\_Quad\_pred$fit)^2,na.rm=T))

rmse\_Add\_sea\_Quad # 29.86647

######################## Multiplicative Seasonality #########################

multi\_sea\_model<-lm(log\_passengers~Jan+Feb+Mar+Apr+May+Jun+Jul+Aug+Sep+Oct+Nov,data = train)

summary(multi\_sea\_model)

multi\_sea\_pred<-data.frame(predict(multi\_sea\_model,newdata=test,interval='predict'))

rmse\_multi\_sea<-sqrt(mean((test$Passengers-exp(multi\_sea\_pred$fit))^2,na.rm = T))

rmse\_multi\_sea # 127.7214

######################## Multiplicative Seasonality Linear trend ##########################

multi\_add\_sea\_model<-lm(log\_passengers~t+Jan+Feb+Mar+Apr+May+Jun+Jul+Aug+Sep+Oct+Nov,data = train)

summary(multi\_add\_sea\_model)

multi\_add\_sea\_pred<-data.frame(predict(multi\_add\_sea\_model,newdata=test,interval='predict'))

rmse\_multi\_add\_sea<-sqrt(mean((test$Passengers-exp(multi\_add\_sea\_pred$fit))^2,na.rm = T))

rmse\_multi\_add\_sea # 11.64118

# Preparing table on model and it's RMSE values

table\_rmse<-data.frame(c("rmse\_linear","rmse\_expo","rmse\_Quad","rmse\_sea\_add","rmse\_Add\_sea\_Quad","rmse\_multi\_sea","rmse\_multi\_add\_sea"),c(rmse\_linear,rmse\_expo,rmse\_Quad,rmse\_sea\_add,rmse\_Add\_sea\_Quad,rmse\_multi\_sea,rmse\_multi\_add\_sea))

View(table\_rmse)

colnames(table\_rmse)<-c("model","RMSE")

View(table\_rmse)

#Multiplicative Seasonality Linear trend has least RMSE value

new\_model <- lm(log\_passengers~t+Jan+Feb+Mar+Apr+May+Jun+Jul+Aug+Sep+Oct+Nov,data = data)

# Building Autoregressive model on residuals consider lag-1

resid <- residuals(new\_model)

resid[1:10]

windows()

acf(resid,lag.max = 10)

# By principal of parcimony we will consider lag - 1 as we have so

# many significant lags

k <- arima(resid, order=c(1,0,0))

str(k)

View(data.frame(res=resid,newresid=k$residuals))

windows()

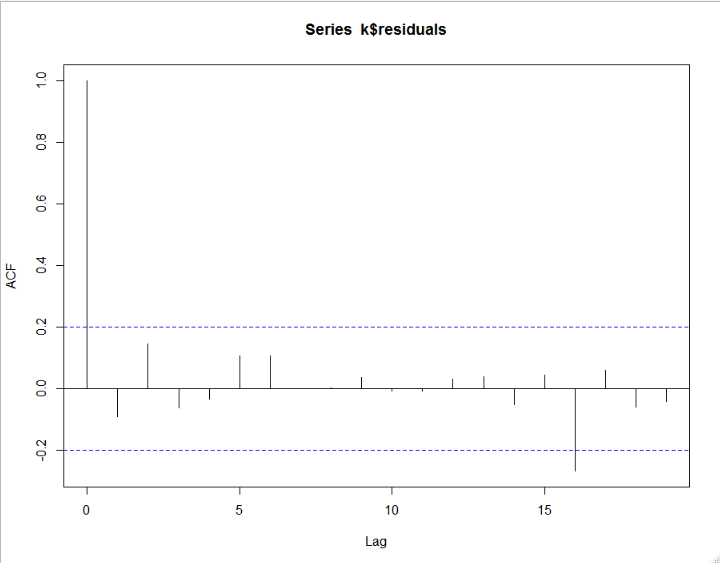
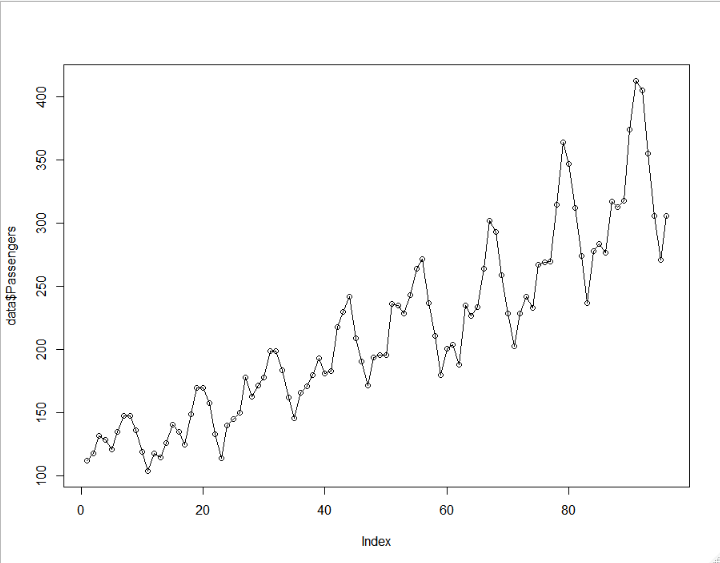
acf(k$residuals,lag.max = 15)

pred\_res<- predict(arima(k$residuals,order=c(1,0,0)),n.ahead = 12)

str(pred\_res)

pred\_res$pred

acf(k$residuals)



|  | **model** | **RMSE** |
| --- | --- | --- |
|  |  |  |
| **1** | rmse\_linear | 47.64090 |
| **2** | rmse\_expo | 301.58115 |
| **3** | rmse\_Quad | 43.45248 |
| **4** | rmse\_sea\_add | 123.09267 |
| **5** | rmse\_Add\_sea\_Quad | 29.86647 |
| **6** | rmse\_multi\_sea | 127.72137 |
| **7** | rmse\_multi\_add\_sea | 11.64118 |

Showing 1 to 7 of 7 entries, 2 total columns