#For each of the following series (from the fma package), make a graph of the data.

#If transforming seems appropriate, do so and describe the effect.

#Question 1 ) Monthly total of people on unemployed benefits in Australia (January 1956–July 1992).

#Loading the library

library(fma)

#Loading the Package

dole

#Now taking the Summary of the Dole as to check the mean median and other IQR

summary(dole)

#plotting Dole

plot(dole , sub="Pre-transformation")

#Checking the Lambda for the DOle with the help of Box Cox

Boxcox\_lambda\_dole <- BoxCox.lambda(dole)

#based on the Lambda we will try to plot the Boxcox transformed data

plot(BoxCox(dole,Boxcox\_lambda\_dole),sub="Post-transformation")

#Question 2 ) Monthly total of accidental deaths in the United States (January 1973–December 1978).

#Loading the Package

usdeaths

#plotting the US deaths

plot(usdeaths , sub="Pre-transformation")

#Taking the Summary of the US Deaths

summary(usdeaths)

#Checking the Lambda for the us deaths with the help of Box Cox

Boxcox\_lambda\_usdeaths <- BoxCox.lambda(usdeaths)

#based on the Lambda we will try to plot the Boxcox transformed data

plot(BoxCox(dole,Boxcox\_lambda\_usdeaths),sub="Post-transformation")

#Question 3 ) Quarterly production of bricks (in millions of units) at Portland, Australia (March 1956-September 1994).

#Loading the Package

bricksq

#plotting the Bricks

plot(bricksq ,sub="Pre-transformation")

#Taking the Summary of the Bricks

summary(bricksq)

#Checking the Lambda for the Bricks with the help of Box Cox

Boxcox\_lambda\_Bricks <- BoxCox.lambda(usdeaths)

#based on the Lambda we will try to plot the Boxcox transformed data

plot(BoxCox(dole,Boxcox\_lambda\_Bricks) , sub="Post-transformation")

################################################################################################

#Use the Dow Jones index (data set dowjones) to do the following:

# Question 1 ) Produce a time plot of the series.

#plotting the Dow jones

plot(dowjones)

#ploting the log of the DOw jones

plot(log(dowjones) , sub="log-transformation")

#ploting the log of the Square root

plot(sqrt(dowjones) , sub="Sqrt-transformation")

#Question 2 ) Produce forecasts using the drift method and plot them

#Generating Random Walk Forecasts using drift menthod

drift\_dowjones <- rwf(dowjones , h=15, drift=TRUE)

#Generating Random Walk Forecasts log using drift menthod

drift\_dowjones\_log <- rwf(log(dowjones), h=15, drift=TRUE)

#Generating Random Walk Forecasts sqrt using drift menthod

drift\_dowjones\_sqrt <- rwf(sqrt(dowjones), h=15, drift=TRUE)

#Plotting the Data using Drift menthod

plot(drift\_dowjones , sub="Drift Method-transformation" ,ylab="Index",xlab="Year")

#Plotting the Data using Log menthod

plot(drift\_dowjones\_log , sub="Log Method-transformation" ,ylab="Index",xlab="Year")

#Plotting the Data using SQRT menthod

plot(drift\_dowjones\_sqrt , sub="SQRT Method-transformation" ,ylab="Index",xlab="Year")

#Question 2 ) Show that the graphed forecasts are identical to extending the line drawn between the first and last observations.

#for the First and Last observation we need to do the Mean , Naive and Drift method in the Plot

dowjones\_win <- window(dowjones)

dowjones\_mean <- meanf(dowjones\_win,h=15)

#Generating Random Walk Forecasts with Drift = TRUE

dowjones\_drift <- rwf(dowjones\_win,h=15, drift = TRUE)

#Generating Random Walk Forecasts without Drift

dowjones\_no\_drift <- rwf(dowjones\_win,h=15)

# Generating the Plot for the same

plot(dowjones\_mean, main="Dow Jones")

lines(dowjones\_drift$mean,col=6)

lines(dowjones\_no\_drift$mean,col=3)

lines(dowjones)

legend("topleft", lty=1, col=c(4,6,3), legend=c("Mean ","Naive","Drifit"))

#Question 3 ) Try some of the other benchmark functions to forecast the same data set.

#Which do you think is best? Why?

dowjones\_nw\_win <- window(dowjones)

dowjones\_nw\_mean <- meanf(dowjones\_nw\_win,h=40)

#Generating Random Walk Forecasts with Drift = TRUE

dowjones\_nw\_drift <- rwf(dowjones\_nw\_win,h=30, drift = TRUE)

#Generating Random Walk Forecasts without Drift

dowjones\_no\_nw\_drift <- rwf(dowjones\_nw\_win,h=20)

# Generating the Plot for the same with modified h

plot(dowjones\_nw\_mean, main="Dow Jones")

lines(dowjones\_nw\_drift$mean,col=6)

lines(dowjones\_no\_nw\_drift$mean,col=3)

lines(dowjones)

legend("topleft", lty=1, col=c(4,6,3), legend=c("Mean ","Naive","Drifit"))

######################################################################################################

#Consider the daily closing IBM stock prices (data set ibmclose).

#Question 1 ) Produce some plots of the data in order to become familiar with it.

#Loading the Package

ibmclose

#Now taking the Summary of the ibm as to check the mean median and other IQR

summary(ibmclose)

#Producing some plots

# Generating the Plot Log

plot(log(ibmclose))

# Generating the Plot SQRT

plot(sqrt(ibmclose))

# Generating the Plot Simple

plot(ibmclose)

## Generating the Plot QQ NORM PLOT

qqnorm(ibmclose)

#Question 2 ) Split the data into a training set of 300 observations and a test set of 69 observations.

ibmclose\_train\_data <- window(ibmclose ,Start= 1 ,end=300)

ibmclose\_test\_data <- window(ibmclose ,start=301, end=369)

#Question 3 ) Try various benchmark methods to forecast the training set and compare the

#results on the test set. Which method did best?

ibmclose\_avg <- meanf(ibmclose\_train\_data,h=64)$mean

ibmclose\_naive <- naive(ibmclose\_train\_data ,h=64)$mean

ibmclose\_drift <- rwf(ibmclose\_train\_data ,drift=TRUE,h=64)$mean

#plotting the Data

plot(ibmclose\_train\_data)

lines(ibmclose\_naive,col=2)

lines(ibmclose\_avg,col=4)

lines(ibmclose\_drift,col=3)

lines(ibmclose\_test\_data,col=8)

legend("topleft",lty=1,col=c(4,2,3,8), legend=c("Mean Method","Naive Method","Drift Method","test Data"))

##################################################################################################

#Consider the sales of new one-family houses in the USA, Jan 1973 – Nov 1995 (data set hsales).

#Question 1 ) Produce some plots of the data in order to become familiar with it.

#Loading the Package

hsales

#Now taking the Summary of the ibm as to check the mean median and other IQR

summary(hsales)

#Producing some plots

# Generating the Plot Log

plot(log(hsales))

# Generating the Plot Log

plot(log(hsales))

# Generating the Plot SQRT

plot(sqrt(hsales))

# Generating the Plot Simple

plot(hsales)

## Generating the Plot QQ NORM PLOT

qqnorm(hsales)

#Question 2 ) Split the hsales data set into a training set and a test set, where the test set is the last two years of data.

hsales\_ts <- ts(hsales,start=1,end=275)

hsales\_train\_data <- window(hsales\_ts,start=1,end=251)

hsales\_test\_data <- window(hsales\_ts,start=251)

#Question 3 ) Try various benchmark methods to forecast the training set and compare the results on the test set. Which method did best?

hsales\_avg <- meanf(hsales\_train\_data,h=34)$mean

hsales\_naive <- naive(hsales\_train\_data ,h=34)$mean

hsales\_drift <- rwf(hsales\_train\_data ,drift=TRUE,h=34)$mean

#plotting the Data

plot(hsales\_train\_data)

lines(hsales\_naive,col=2)

lines(hsales\_avg,col=4)

lines(hsales\_drift,col=3)

lines(hsales\_test\_data,col=8)

legend("topleft",lty=1,col=c(4,2,3,8), legend=c("Mean Method","Naive Method","Drift Method","test Data"))

######################################################################################################

# ##

# QUESTION 4 ##

# ##

######################################################################################################

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

#Mwh 16.3 16.8 15.5 18.2 15.2 17.5 19.8 19.0 17.5 16.0 19.6 18.0

#temp 29.3 21.7 23.7 10.4 29.7 11.9 9.0 23.4 17.8 30.0 8.6 11.8

#Question 1 ) Plot the data and find the regression model for Mwh with temperature as an explanatory variable. Why is there a negative relationship?

#Loading the Package

econsumption

#Now taking the Summary of the ibm as to check the mean median and other IQR

summary(econsumption)

#Producing some plots

# Generating the Plot Log

plot(log(econsumption))

#question 2) a. Plot the data and find the regression model for Mwh with temperature as an explanatory variable. Why is there a negative relationship?

plot(Mwh ~ temp, data = econsumption)

fit\_eco = lm(formula = Mwh ~ temp, data = econsumption)

abline(fit\_eco, col=5)

#Question 3 ) Produce a residual plot. Is the model adequate? Are there any outliers or influential observations?

plot(fit\_eco)

#Question 4) Use the model to predict the electricity consumption that you would expect for a day with maximum temperature 10 and a day with maximum temperature 35. Do you believe these predictions?

coeffs = coefficients(fit\_eco)

pred\_temp = c(10, 35)

p\_temp = coeffs[1] + coeffs[2]\*pred\_temp

p\_temp

#Question 5 Give prediction intervals for your forecasts. The following R code will get you started:

forecast <- forecast(fit\_eco , newdata=data.frame(temp=10))

plot(forecast)

forecast2 <- forecast(fit\_eco, newdata=data.frame(temp=35))

plot(forecast2)

temp10 = data.frame(temp=10)

temp35 = data.frame(temp=35)

predict(fit\_eco, temp10, interval="predict")

predict(fit\_eco, temp35, interval="predict")

##########################################################################################################

#The following table gives the winning times (in seconds) for the men’s 400 meters final in each Olympic Games from 1896 to 2012 (data set `olympic`).

#question 1) Update the data set `olympic` to include the winning times from the last few Olympics.

olympic

#question 2) Plot the winning time against the year. Describe the main features of the scatterplot.

plot(time ~ Year, data = olympic)

#Question 3) Fit a regression line to the data. Obviously the winning times have been decreasing, but at what average rate per year?

oly\_fit = lm(formula = time ~ Year, data = olympic)

plot(time ~ Year, data = olympic)

abline(oly\_fit, col=5)

olympic\_ts <- ts(olympic,start=1,end=28)

#Question 4) Plot the residuals against the year. What does this indicate about the suitability of the fitted line?

plot(oly\_fit)

#Question 5) Predict the winning time for the men’s 400 meters final in the 2000, 2004, 2008 and 2012 Olympics. Give a prediction interval for each of your forecasts. What assumptions have you made in these calculations?

coeff = coefficients(oly\_fit)

pred\_time = c(2000, 2004, 2008, 2012)

p\_time = coeff[1] + coeff[2]\*pred\_time

p\_time

fcast3 <- forecast(oly\_fit, newdata=data.frame(Year=2000))

plot(fcast3, xlab="Year", ylab="time")

fcast4 <- forecast(oly\_fit, newdata=data.frame(Year=2004))

plot(fcast4, xlab="Year", ylab="time")

fcast5 <- forecast(oly\_fit, newdata=data.frame(Year=2008))

plot(fcast5, xlab="Year", ylab="time")

fcast6 <- forecast(oly\_fit, newdata=data.frame(Year=2012))

plot(fcast6, xlab="Year", ylab="time")

time2000 = data.frame(Year=2000)

time2004 = data.frame(Year=2004)

time2008 = data.frame(Year=2008)

time2012 = data.frame(Year=2012)

predict(oly\_fit, time2000, interval="predict")

predict(oly\_fit, time2004, interval="predict")

predict(oly\_fit, time2008, interval="predict")

predict(oly\_fit, time2012, interval="predict")

#Question 6 ) Find out the actual winning times for these Olympics (see www.databaseolympics.com). How good were your forecasts and prediction intervals?

coeffs1 = coefficients(oly\_fit)

pred\_time = c(2000, 2004, 2008, 2012)

p\_time = coeffs1[1] + coeffs1[2]\*pred\_time

p\_time

#######################################################################################################

#An elasticity coefficient is the ratio of the percentage change in the forecast variable (yy) to the percentage change in the predictor variable (xx). Mathematically, the elasticity is defined as (dy/dx)×(x/y)(dy/dx)×(x/y). Consider the log-log model,

## logy=β0+β1logx+ε.

#Express yy as a function of xx and show that the coefficient β1β1 is the elasticity coefficient.

# logy=B0+B1logx+e

#Taking Diffrential

#dy/y = dx/x\*B1 (dy/y / dx/x) = B1

# change in y divided by y over the the change in x divided by x

#100(dy/y) = 100(dx/x)B1 % change y = % change x B1

###############################################################################################

###The data below (data set fancy) concern the monthly sales figures of a shop which opened in January 1987 and sells gifts, souvenirs, and novelties. The shop is situated on the wharf at a beach resort town in Queensland, Australia. The sales volume varies with the seasonal population of tourists. There is a large influx of visitors to the town at Christmas and for the local surfing festival, held every March since 1988. Over time, the shop has expanded its premises, range of products, and staff.

#Question 1 ) ) Produce a time plot of the data and describe the patterns in the graph.

#Identify any unusual or unexpected fluctuations in the time series.

library(fma)

library(fpp)

plot(fancy)

#Question 2 ) Explain why it is necessary to take logarithms of these data before fitting a model.

# Answer 2 ) It is because of the increasing seasonal fluctuations

log\_fancy <- log(fancy)

#question 3 ) c) Use R to fit a regression model to the logarithms of these sales data with a linear trend, seasonal dummies and a "surfing festival" dummy variable.

log\_fancy <- log(fancy)

dummy\_fest = rep(0, length(fancy))

dummy\_fest[seq\_along(dummy\_fest)%%12 == 3] <- 1

dummy\_fest[3] <- 0

dummy\_fest <- ts(dummy\_fest, freq = 12, start=c(1987,1))

my\_data <- data.frame(

log\_fancy,

dummy\_fest

)

fit <- tslm(log\_fancy ~ trend + season + dummy\_fest, data=my\_data)

future\_data <- data.frame(

dummy\_fest = rep(0, 12)

)

future\_data[3,] <- 1

forecast(fit, newdata=future\_data)

#question 4 d) Plot the residuals against time and against the fitted values. Do these plots reveal any problems with the model?

plot(residuals(fit), type='p')

plot(as.numeric(fitted(fit)), residuals(fit), type='p')

#Question 5 ) e) Do boxplots of the residuals for each month. Does this reveal any problems with the model?

boxplot(resid(fit) ~ cycle(resid(fit)))

#Question 5 ) ; f) What do the values of the coefficients tell you about each variable?

# The value of the coefficients show how much the model thinks each month contributes

# to the conditional mean of the model.

#Question 5 g) What does the Durbin-Watson statistic tell you about your model?

dwtest(fit)

# Question 6 h ) Regardless of your answers to the above questions, use your regression model to predict the monthly sales for 1994, 1995, and 1996. Produce prediction intervals for each of your forecasts.

future\_data <- data.frame(

dummy\_fest = rep(0, 36)

)

preds <- forecast(fit, newdata=future\_data)

# Question 5 i ) Transform your predictions and intervals to obtain predictions and intervals for the raw data.

df\_pred <- as.data.frame(preds)

df\_pred <- exp(df\_pred)

# Question 5 j ) How could you improve these predictions by modifying the model?

#ANswers 5 J) We Could use consider using a dynamic-regression model

#######################################################################################################

# Question 5b ) (1) The data below (data set texasgas) shows the demand for natural gas and the price of natural gas for 20 towns in Texas in 1969.

# a) Do a scatterplot of consumption against price. The data are clearly not linear. Three possible nonlinear models for the data are given below

library(fma)

library(fpp)

library(segmented)

df\_gas <- (texasgas)

plot(df\_gas$price, df\_gas$consumption)

# b) Can you explain why the slope of the fitted line should change with price?

# Ans The data is not linear so the slope needs to change in order to get the Data from our model

# c) Fit the three models and find the coefficients, and residual variance in each case.

# First Model

fit\_gas <- lm(consumption ~ exp(price), df\_gas)

fit\_gas

# Residual Variance

(summary(fit\_gas)$sigma)\*\*2

# Second model - piecewise linear regression

lin.mod <- lm(consumption ~ price, df\_gas)

lin.mod

segmented.mod <- segmented(lin.mod, seg.Z = ~price, psi=60)

slope(segmented.mod)

# Residual variance

(summary(segmented.mod)$sigma)\*\*2

# Third model - polynomial regression

poly\_fit <- lm(consumption ~ poly(price, 2), df\_gas)

# Residual variance

(summary(poly\_fit)$sigma)\*\*2

#d) For each model, find the value of R2 and AIC, and produce a residual plot. Comment on the adequacy of the three models.

# First model - basic linear regression

# Adjusted R-squared: -0.004

# AIC: 200.736

resid <- residuals(fit\_gas)

plot(fit\_gas$fitted.values, resid, ylab='residuals', xlab='fitted values', main='linear regression')

abline(0,0)

# Second model - piecewise linear regression

# Adjusted R-squared: 0.847

# AIC: 164.756

resid <- residuals(segmented.mod)

plot(segmented.mod$fitted.values, resid, ylab='residuals', xlab='fitted values',

main='piecewise linear regression')

abline(0,0)

# Third model - polynomial regression.

# Adjusted R-squared: 0.812

# AIC: 168.116

resid <- residuals(poly\_fit)

plot(poly\_fit$fitted.values, resid, ylab='residuals', xlab='fitted values',

main='polynomial linear regression')

abline(0,0)

# e) For prices 40, 60, 80, 100, and 120 cents per 1,000 cubic feet, compute the forecasted per capita demand using the best model of the three above.

new.data <- data.frame(price=c(40, 60, 80, 100, 120))

predict(segmented.mod, new.data)

# F ) Compute 95% prediction intervals. Make a graph of these prediction intervals and discuss their interpretation.

newx <- seq(min(new.data), max(new.data), length.out=5)

intervals <- predict(segmented.mod, new.data, interval="predict")

plot(consumption ~ price, data = df\_gas, type = 'n')

polygon(c(rev(newx), newx), c(rev(intervals[ ,3]), intervals[ ,2]), col = 'grey80', border = NA)