#READING THE CSV DATA

ftse<-read.csv("C:/Users/pc/Desktop/akano-documents/ehi r assignment/ftse.csv")

US<-read.csv("C:/Users/pc/Desktop/akano-documents/ehi r assignment/spx.csv")

UKnew <- UK

USnew <- US

#CONVERT THE DATE IN THE DATA TO THE DATE R CAN UNDERSTAND

library(lubridate)

> US$new\_date<-dmy(US$Date)

> UK$new\_date<-dmy(UK$Date)

> head(UK)

> UK1<-UK %>% arrange(Date)

> UK1<-UK %>% arrange(new\_date)

> head(UK1)

UKnew <- data.frame(Date=as.Date(UK$Date , "%d/%m/%Y"), ftse=UK$ftse)

USnew <- data.frame(Date=as.Date(US$Date , "%d/%m/%Y"), spx=US$spx)

#generate the year column

UKnew$year <- strftime(UKnew$Date, "%Y")

#generate the month column

UKnew$month <- strftime(UKnew$Date, "%m")

#generate the year column

USnew$year <- strftime(USnew$Date, "%Y")

#generate the year column

USnew$month <- strftime(USnew$Date, "%m")

#DESCRIPTIVE STATISTICS

library(ggplot2)

#Generate time series plot

tsplot\_uk<- ggplot(UKnew, aes(x=Date, y=ftse)) + geom\_line()

tsplot\_uk

tsplot\_us <- ggplot(USnew, aes(x=Date, y=spx)) + geom\_line()

tsplot\_us

#MAKE THE NUMBER OF ROWS IN BOTH DATA EQUAL

nrow(UKnew)

nrow(USnew)

UKnew1<- head(UKnew , 1260)

#GENERATE SCATTER PLOT

scatter.smooth(USnew$spx ,UKnew1$ftse, main="UK VS US")

#GENERATE BOXPLOTS

boxplot(UKnew1$ftse)

boxplot(UKnew1$ftse , main="BOXPLOT FOR UK FTSE")

boxplot(USnew$spx , main="BOXPLOT FOR US SPX")

#GENERATE HISTOGRAM

hist(UKnew1$ftse , main="HISTOGRAM FOR UK FTSE")

hist(USnew$spx , main="HISTOGRAM FOR US SPX")

#VIEW SUMMARY OF DATA

summary(UKnew$ftse)

summary(USnew$spx)

# RESGRESSION MODEL

#FIT LINEAR REGRESSION MODEL

model <- lm(UKnew1$ftse~USnew$spx)

#VIEW MODEL SUMMARY

summary(model)

# The regression equation is R(UK)=6728 + 0.094US(spx)

#ARITHMETRIC MEAN AND GEOMETRIC MEAN

#CALCULATE THE ARITHMETIC MEAN PER YEAR

USnew\_mean <- aggregate(spx ~ year, USnew1, FUN=mean)

#install.package('psych')

library('psych')

USnew\_mean <- aggregate(spx ~ year, USnew, FUN=mean)

USnew\_geomean <- aggregate(spx ~ year, USnew, FUN=geometric.mean)

UKnew\_mean <- aggregate(ftse ~ year, UKnew , FUN=mean)

UKnew\_geomean <- aggregate(ftse ~ year, UKnew, FUN=geometric.mean)

#WRITE DATAFRAME TO EXCEL FILE

#install package writexl

library("writexl")

write\_xlsx(UKnew1 , "C:/Users/pc/Desktop/akano-documents/ehi r assignment/uknew.xlsx")

write\_xlsx(USnew , "C:/Users/pc/Desktop/akano-documents/ehi r assignment/usnew.xlsx")

#NOTE: OPEN THE NEWLY GENERATED EXCEL FILE uknew.xlsx

#ALSO OPEN THE NEWLY GENERATED EXCEL FILE usnew.xlsx

#USE EXCEL FORMULA TO GENERATE RUWAR #AND COVID COLUMNS

#FOR RUWAR THE EXCEL FORMULA IS =IF(A2>DATE(2022,2,24),1,0)

#FOR COVID THE EXCEL FORMULAR IS =IF(A2>DATE(2020,1,29),1,0)

#READ THE DATA INTO A DATA FRAME

#Installing and loading readxl package

#install.packages("readxl")

#Load

library("readxl")

uk<- read\_excel("C:/Users/pc/Desktop/akano-documents/ehi r ssignment/uknew.xlsx")

us<- read\_excel("C:/Users/pc/Desktop/akano-documents/ehi r ssignment/usnew.xlsx")

data<-

data<-read\_excel("C:/Users/pc/Desktop/akano-documents/ehi r assignment/uknew.xlsx")

data<-subset(data , select=-c(Date))

head(data)

#VIEW PAIRPLOT OF THE DATA

pairs(UKnew\_data, pch = 18, col = "steelblue")

#FIT THE REGRESSION MODEL

model2 <- lm(UKnew\_data$ftse ~ UKnew\_data$spx+ UKnew\_data$RUWAR + UKnew\_data$COVID)

#CHECK ASSUMPTIONS OF THE MODEL

#The distribution of model residuals should be approximately normal.

#We can check if this assumption is met by creating a simple histogram of residuals.

#PLOT HISTOGRAM

hist(residuals(model2) , col="steelblue")

#The variance of the residuals should be consistent for all observations.

#This preferred condition is known as homoskedasticity. Violation of this assumption is known as #heteroskedasticity.

#To check if this assumption is met we can create a fitted value vs. residual plot:

plot(fitted(model2), residuals(model2))

#add horizontal line at 0

abline(h = 0, lty = 500)

VIEW MODEL SUMMARY

summary(model2)