READING THE CSV DATA

UK<-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

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

USnew1$year <- strftime(USnew$Date, "%Y")

#generate the year column

USnew1$month <- strftime(USnew$Date, "%m")

#descriptive statistics

library(ggplot2)

#Generate time series plot

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

tsplot

tsplot1 <- ggplot(USnew, aes(x=Date, y=spx)) + geom\_line()

tsplot1

UKnew1<- head(UKnew , 1260)

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

boxplot(UKnew1$ftse)

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

boxplot(USnew1$spx , main="BOXPLOT FOR US SPX")

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

hist(USnew1$spx , main="HISTOGRAM FOR US SPX")

summary(UKnew1$ftse)

Min. 1st Qu. Median Mean 3rd Qu. Max.

4994 6885 7191 7054 7437 7877

summary(USnew1$spx)

Min. 1st Qu. Median Mean 3rd Qu. Max.

2237 2843 3275 3449 4028 4797

#fit simple linear reression model

model <- lm(UKnew2$ftse~USnew$spx)

#VIEW MODEK SUMMARY

summary(model)

Call:

m(formula = UKnew2$ftse ~ USnew$spx)

---------------------not a code------

esiduals:

Min 1Q Median 3Q Max

1964.2 -174.0 138.6 370.3 893.5

Coefficients:

Estimate Std. Error t value Pr(>|t|)

Intercept) 6.728e+03 7.884e+01 85.34 < 2e-16 \*\*\*

Snew$spx 9.402e-02 2.244e-02 4.19 2.98e-05 \*\*\*

--

ignif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 532.7 on 1258 degrees of freedom

ultiple R-squared: 0.01376, Adjusted R-squared: 0.01298

-statistic: 17.56 on 1 and 1258 DF, p-value: 2.984e-05

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

# Multiple R-squared: This number tells us the percentage of the variation in the stock index return can be explained by the US equity returns. In general, the larger the R-squared value of a regression model the better the explanatory variables are able to predict the value of the response variable. In this case,only 1% of the variation in the UK stock index return can be explained by US equity return.

---------------------------------------------------------------------------------------------------------------------

#calculate the arithmetic mean per year

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

# install.package('psych')

#click packages , select the package name and install

library('psych')

USnew\_geomean <- aggregate(spx ~ year, USnew1, FUN=geometric.mean)

USnew\_mean <- aggregate(spx ~ year,USnew\_geomean

UKnew\_mean <- aggregate(ftse ~ year, UKnew1 , FUN=mean)

UKnew\_geomean <- aggregate(ftse ~ year, UKnew1 , FUN=geometric.mean)

#install package writexl

library("writexl")

write\_xlsx(UKnew2 , "C:/Users/pc/Desktop/akano-documents/ehi r assignment/uknew2.xlsx")

write\_xlsx(USnew1 , "C:/Users/pc/Desktop/akano-documents/ehi r assignment/usnew1.xlsx")

---- To read your , open the excel file UKnew2.xlsx , highlight and copy all the data there excluding the dates , then run the code below in r

UKnew\_data<-read.table(file="clipboard" ,sep="\t" , header=TRUE)

head(UKnew\_data)

head(UKnew\_data)

ftse spx year month RUWAR COVID

1 7451.74 3839.50 2022 12 1 1

2 7512.72 3849.28 2022 12 1 1

3 7497.19 3783.22 2022 12 1 1

4 7473.01 3829.25 2022 12 1 1

5 7469.28 3844.82 2022 12 1 1

6 7497.32 3822.39 2022 12 1 1

tail(UKnew\_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)

#Checking 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.

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

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)

Call:

lm(formula = UKnew\_data$ftse ~ UKnew\_data$spx + UKnew\_data$RUWAR +

UKnew\_data$COVID)

Residuals:

Min 1Q Median 3Q Max

-688.62 -145.80 -27.12 98.83 1209.90

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 5.107e+03 5.048e+01 101.17 <2e-16 \*\*\*

UKnew\_data$spx 7.802e-01 1.727e-02 45.17 <2e-16 \*\*\*

UKnew\_data$RUWAR 5.013e+02 2.169e+01 23.11 <2e-16 \*\*\*

UKnew\_data$COVID -1.419e+03 2.336e+01 -60.75 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 263 on 1256 degrees of freedom

Multiple R-squared: 0.76, Adjusted R-squared: 0.7594

F-statistic: 1325 on 3 and 1256 DF, p-value: < 2.2e-16