**1-step Prediction and Correlation between Pepsi and Coca-Cola**

* Hridesh Kapur - 111105706
* Junjie He - 110171483

## **Introduction:-**

* The goals of this study were to determine
* Whether the expected log returns of each stock index were equal to 0.
* Whether the expected log returns differed from each other significantly.
* The correlation between the two stocks, and find the significance.
* The linear model between the two stocks, along with the R squared value.
* Predict the future data by a month, then compare it with the actual results to plot a graph.

## **Appendix:-**

For the log returns, we use excel. In excel we use the function ln().=ln(cell2/cell1) for the exact log returns.

#a)

KO<-file.choose()

PEP <- file.choose()

KO<-read.csv(KO)

PEP <- read.csv(PEP)

x=PEP$logreturn

y=KO$logreturn

mean(x)

mean(y)

t.test(x, alternative = "two.sided", mu = 0)

t.test(y, alternative = "two.sided", mu = 0)

#b)

t.test(x,y,alternative = "two.sided",var.equal = TRUE)

#c)

cor(x, y)

cor.test(x,y)

#d)

results <- lm(y~x)

summary(results)

confint(results,"x", level=0.95)

#e)

xx = data.frame(x=c(6,7))

predict(results, xx, interval="predict")

fitted <- results$fitted.values

fitted

plot(x, fitted)

newdata = data.frame(x)

pred.int = predict(results, newdata, interval="predict")

#graph

fitted.values = pred.int[,1]

pred.lower = pred.int[,2]

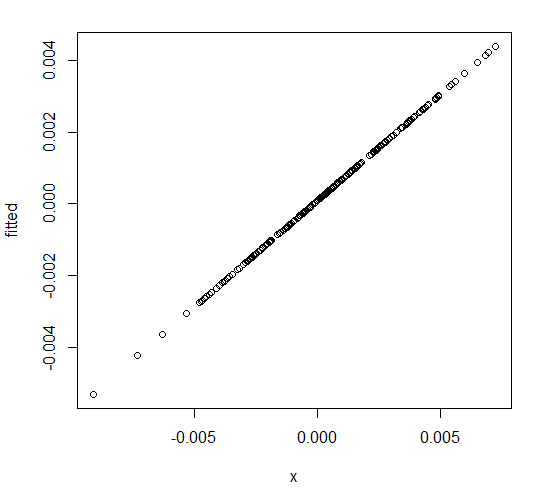
pred.upper = pred.int[,3]

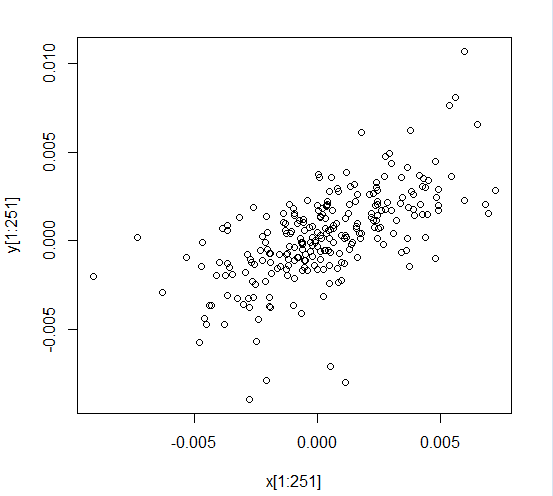
plot(x[1:251],y[1:251])

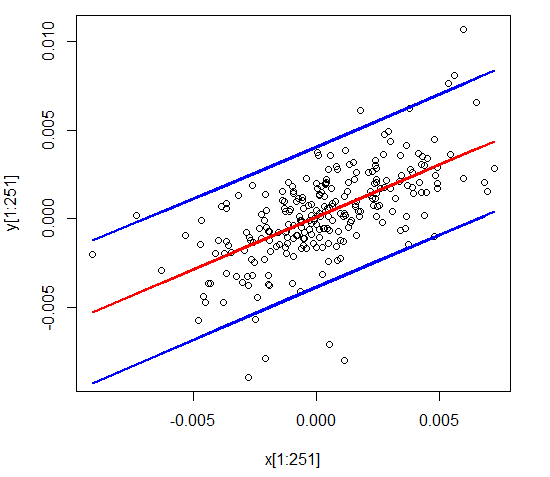
lines(x[1:251],fitted.values[1:251],col="red",lwd=2)

lines(x[1:251],pred.lower[1:251],lwd=2,col="blue")

lines(x[1:251],pred.upper[1:251],lwd=2,col="blue")

**1.02i:**

**1.02ii:-**

**1.02iii:-**

## **Methods:-**

Taking 95% confidence interval for all cases.

### To determine the expected log returns of each stock, we find the mean of each using mean(stock1) and mean(stock2) in R. Then we use the two-sided t-test to determine the p-value. We take H0 as true mean for each stock is equal to 0.

### To determine if the difference in both the stocks is equal to 0, we use the two-sided t-test with H0 as true difference in mean equal to 0.

### To determine the correlation between the two stocks, we use cor(stock1,stock2). We also use cor.test(stock1,stock2) to find the significance.

### To determine the linear model and other statistics, we use the summary() function. We also predict the confidence interval using confint() function.

### To predict the future values, the fitted equation and the confidence interval, we use the predict() function and the fitted() function respectively.

## **Results:-**

### The t-statistic for Pepsi from the t-test was 1.9248 = 1.93 which is less than t alpha/2 that is 1.969. Therefore, we conclude that H0 is true and we accept it. The true mean is close to 0.

### The t-statistic for Coca-Cola from the t-test was 1.7518 = 1.752 which is less than t alpha/2 that is 1.969. Therefore, we conclude that H0 is true and we accept it. The true mean is close to 0.

### The t-statistic for both the stocks was 0.1959 = 0.196 which is less than t alpha/2 that is 1.964. Therefore, we conclude that H0 is true and we accept it. The true difference in mean may be equal to 0.

### The correlation between Pepsi and Coca-Cola is 0.626. That means that the stocks are highly dependent on each other. Using the p-value which is 2.2e-16 which is less than 0.05 we can conclude that we reject H0 and the true correlation is not equal to 0.

### The R squared value is 0.392. The Confidence interval is (0.499, 0.684).

### Graph 1.02i shows the linear model fitted equation. Graph 1.02ii shows the predicted values along with the actual values of the stocks. Graph 1.02iii shows the actual and predicted values of the stocks with a 95% confidence interval.

## **Conclusion:-**

We conclude from the graphs that the stocks Pepsi and Coca-Cola are highly correlated with each other. That is, any effect on either of the stocks has consequences on the other. The predicted and actual value using 1-step prediction with a 95% confidence interval is quite narrow therefore the fitted equation is correct and seems to fit the stocks adequately.