STAT 413 Data Science HW 5

Hyo Young Ana Lim

3/6/2021

#Statistics Review   
#An investor sued his broker for lack of diversification. Below are the rates of return (in percent) for the investor’s portfolio over 39 months (data from Moore, McCabe, and Craig (2017)). The average of the S&P 500 stock index for the same period was 0.95%. Does the broker perform worse than average?   
  
#Copy and paste the data below into R  
  
Stockreturns <- c(-8.36, 1.63, -2.27, -2.93, -2.70,   
 -2.93, -9.14, -2.64, 6.82, -2.35,   
 -3.58, 6.13, 7.00, -15.25, -8.66,  
 -1.03, -9.16, -1.25, -1.22, -10.27,  
 -5.11, -0.80, -1.44, 1.28, -0.65,  
 4.34, 12.22, -7.21, -0.09, 7.34,   
 5.04, -7.24, -2.14, -1.01, -1.41,   
 12.03, -2.53, 4.33, 1.35)  
Stockreturns

## [1] -8.36 1.63 -2.27 -2.93 -2.70 -2.93 -9.14 -2.64 6.82 -2.35  
## [11] -3.58 6.13 7.00 -15.25 -8.66 -1.03 -9.16 -1.25 -1.22 -10.27  
## [21] -5.11 -0.80 -1.44 1.28 -0.65 4.34 12.22 -7.21 -0.09 7.34  
## [31] 5.04 -7.24 -2.14 -1.01 -1.41 12.03 -2.53 4.33 1.35

#1) Use and show R code to find the mean of the sample data.  
mean(Stockreturns)

## [1] -1.124615

#2) Use and show R code to find the standard deviation of the sample data  
sd(Stockreturns)

## [1] 5.977673

#3) Assuming a normal distribution, use and show R code to find the proportion of returns that are less than -1.5.  
pnorm(q = -1.5, mean = -1.124615, sd = 5.977673)

## [1] 0.4749637

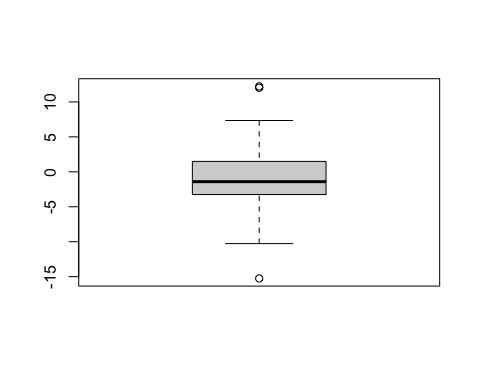
#4) Assuming a normal distribution, use and show R code to find return value that is above 70% of the returns  
pnorm(q = 0.70, mean = -1.124615, sd = 5.977673)

## [1] 0.6199077

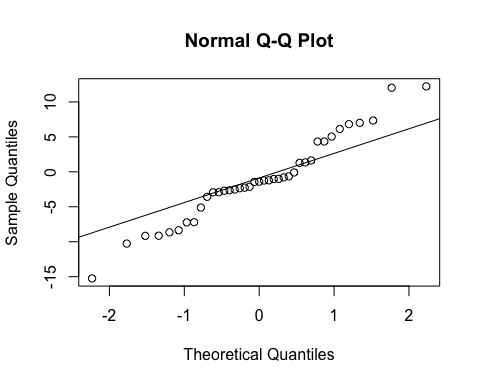
#5) Use and show R code to find Q1 for the data.  
quantile(Stockreturns, 0.25)

## 25%   
## -3.255

#6) Explore the data by producing and examining a boxplot and checking for normality.  
  
boxplot(Stockreturns)



qqnorm(Stockreturns)  
qqline(Stockreturns)



#7) State the appropriate null and alternative hypothesis required for the appropriate t test.  
  
# H(0) : population mean = 0.95  
# H(A) : population mean does not equal 0.95  
  
#8) Use and show R code that will output the needed p value and confidence interval to determine if the null hypothesis should be rejected.  
t.test(Stockreturns, mu = 0, var.equal = FALSE)

##   
## One Sample t-test  
##   
## data: Stockreturns  
## t = -1.1749, df = 38, p-value = 0.2473  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## -3.0623529 0.8131221  
## sample estimates:  
## mean of x   
## -1.124615

#The Null hypothesis cannot be rejected because the p-value is > than .05.Furthermore, 0 is not in my confidence interval, which means I can reject the Null hypothesis.   
  
#9) Now answer the question originally stated. Does the broker perform worse than average? (Explain or justify in two or three sentences)  
  
#The mean monthly return of investment was x=-1.124615. This means the broker is performing worse than average.