Atiya Mahboob- Assessed Coursework 2022

1) mean=-10.243

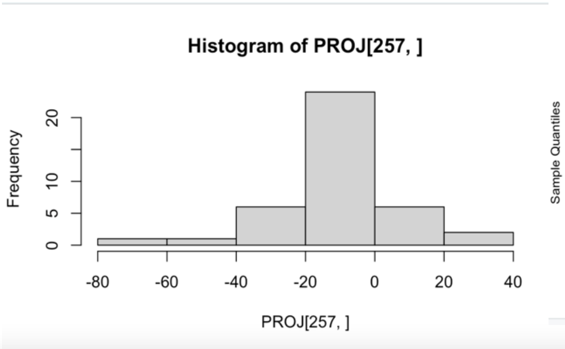
```
> mx=mean(PROJ[257,])
> mx
[1] -10.243
```

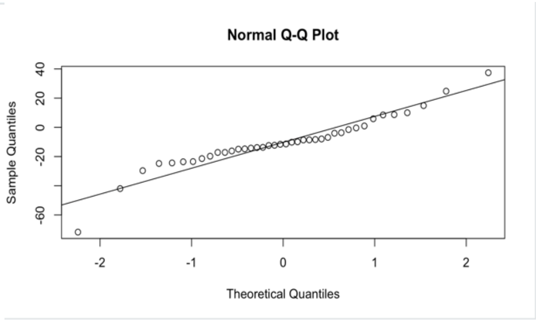
2)Sample standard deviation = 17.747

```
> sampleSD=sd(PROJ[257,])
> sampleSD
[1] 17.74729
```

3)

```
> histogram=hist(PROJ[257,])
> histogram
$breaks
[1] -80 -60 -40 -20   0  20  40

$counts
[1]  1  1  6 24  6  2

$density
[1] 0.00125 0.00125 0.00750 0.03000 0.00750 0.00250

$mids
[1] -70 -50 -30 -10  10  30

$xname
[1] "PROJ[257, ]"

$equidist
[1] TRUE

attr(,"class")
[1] "histogram"
>
```





The histogram appears to be normally distributed as it is unimodal and symmetric. The data points line up with the line , but as we approach the tails of the distribution there appears to be 1-2 outliers where the sample quantiles are larger than the theoretical quantiles or the theoretical quantiles are larger than the sample quantiles. However, as the majority of the data matches up with line these outliers do not appear to be significant. So we can argue that the normal distribution is not the appropriate model for this set of data.

4)

First condition: a sample size of 40 is used, which is greater than 30, so the sample size is sufficiently large.
Second condition: the sample size was taken randomly from 40 companies from the S&P 500 index, so we can assume independence.
Third condition: the data is not heavily skewed, the distribution is relatively normal, although there are 1-2 outliers in the data, they're not common and the sufficiently large sample size counteracts the significance of these outliers.

5)standard deviation=2.806

```
> se=(sampleSD)/sqrt(n)
> se
[1] 2.806093
```

6)95% confidence interval
Zstar1=1.96
Confidence Interval=(-15.743, -4.743)

```
> zstar1=qnorm(0.975)
> zstar1
[1] 1.959964
> CI1=c(mx-zstar1*se, mx+zstar1*se)
> CI1
[1] -15.742841  -4.743159
>
```

7)99% confidence interval
Zstar2=2.58
Confidence Interval=(-17.471, -3.0150)

```
> zstar2=qnorm(0.995)
> zstar2
[1] 2.575829
> CI2=c(mx-zstar2*se,mx+zstar2*se)
> CI2
[1] -17.471016  -3.014984
>
```

8)

H_0: μ=0
H_a: μ≠ 0 (two- tailed test)
Significance level =0.05
X~N(0, 2.806^(2))
P(x<-10.243)
z=(sampleMean- μ)/se
z=(-10.243-0)/2.806
z=-3.650271
P(Z<z)=0.0002619633
So, p value = 0.000262

```
> z=(mx-mu)/se
> z
[1] -3.650271
```

```
> pValue=2*pnorm(z)
> pValue
[1] 0.0002619633
```

Assuming H_0 is true. As 0.000262<0.05, and as -3.65 is outside of the 95% confidence interval, we reject H_0 in favour H_a. There is sufficient statistical evidence to indicate that the average year to date stock return of S&P 500 companies is not equal to 0.

9)) Assuming H_0 is true , as 0.000262<0.01, we can reject the null hypothesis in favour of the alternative hypothesis. There is sufficient statistical evidence to indicate that the stock return of S&P companies is not equal to 0.

10) Assuming H_0 is true , as 0.000262<0.05, we can reject the null hypothesis. There is sufficient evidence to indicate that the average year to date stock return of S&P companies is not equal to zero.

```r
load(file = "PROJ.Rdata")
mydata = PROJ[257,]

#number of data points
n=40
n
#calculate the sample mean
mx=mean(PROJ[257,])
mx
#calculate the sample standard deviation
sampleSD=sd(PROJ[257,])
sampleSD
#plot histogram of data against normal distribution
histogram=hist(PROJ[257,], prob= TRUE)
lines(density(PROJ[257,]))
#plot Q-Q plot of data against the normal distribution
QQPlot=qqnorm(PROJ[257,]);abline(mx,sampleSD)
QQPlot
#Calculate the standard error of the mean
se=(sampleSD)/sqrt(n)
se
#calculate a 95% confidence interval
zstar1=qnorm(0.975)
zstar1
CI1=c(mx-zstar1*se, mx+zstar1*se)
CI1
#calculate z-star value for 99% confidence interval
zstar2=qnorm(0.995)
zstar2
#calculate the 99& confidence interval
CI2=c(mx-zstar2*se,mx+zstar2*se)
CI2
#Setting up hypothesis test
n=40
se
mx
mu=0
mu
#calculate the z score
z=(mx-mu)/se
z
#calculate the p-value
pValue=2*pnorm(z)
pValue
```

Feedback comments

1. 1/1

2. 1/1

3. 4/4

4. 3/3

5 1/1

6. 2/2

7. 1.5/2 3 d.p.

8. 4/4

9. 1/1

10. 1/1

Total 19.5/20