Q1
10 Points

The ZMTSLA.csv on canvas contains prices of ZM and TSLA over a certain period. Compute daily log returns for both stocks. Denote ZM returns by $\{X_1, \cdots, X_n\}$ with mean μ_{zm}, TSLA returns by $\{Y_1, \cdots, Y_n\}$ with mean μ_{tsla}. Let $D_i = X_i - Y_i$, $1 \le i \le n$, be the differences of the log returns of the two stocks.

Q1.1
1 Point

Conduct a runs test. At significance level $\alpha = 5\%$, is there strong enough evidence that the returns of ZM are not independent? What about D_i's?

Runs Test for log return of ZM
data: log_rtn_ZM
statistic = -0.74305, runs = 172, n1 = 178, n2 = 178, n = 356, p-value = 0.4575
alternative hypothesis: nonrandomness

since p = 0.4575 > alpha = 0.05, we do not reject H0 = random. (so we remain hypothesized that log return of ZM is random and independent.)
So, alpha = 0.05 is not strong enough to proof that ZM are not independent. Rather, p-value of log return of ZM shows a great independency.

Runs Test for D's
data: D
statistic = 0.4246, runs = 183, n1 = 178, n2 = 178, n = 356, p-value = 0.6711
alternative hypothesis: nonrandomness

since p = 0.6711 > alpha = 0.05, we do not reject H0 = random. (so we remain hypothesized that D is random and independent.)
So, alpha = 0.05 is not strong enough to proof that D are not independent. Rather, p-value of D shows a great independency.

Q1-1.R Download

```r
#Q1-1

#Data Preparation
ZMTSLA = read.csv("/Users/yu-chingliao/Desktop/ZMTSLA.csv")
log_rtn_ZM = log(ZMTSLA$ZM[2:nrow(ZMTSLA)] / ZMTSLA$ZM[1:nrow(ZMTSLA)-1])
log_rtn_TSLA = log(ZMTSLA$TSLA[2:nrow(ZMTSLA)] / ZMTSLA$TSLA[1:nrow(ZMTSLA)-1])
mu_logrt_ZM = mean(log_rtn_ZM)
mu_logrt_TSLA = mean(log_rtn_TSLA)
D = log_rtn_ZM - log_rtn_TSLA

#Run Test for testing randomness of log_rtn_ZM
alpha = 0.05
install.packages("randtests")
library(randtests)
runs.test(log_rtn_ZM, alternative = "two.sided", threshold = median(log_rtn_ZM), pvalue = "normal", plot = FALSE)

#Run Test for testing randomness of D
runs.test(D, alternative = "two.sided", threshold = median(D), pvalue = "normal", plot = FALSE)
```

Q1.2
1 Point

Conduct a normality test using the Shapiro-Wilk test and Jarque-Bera test. At significance level $\alpha = 5\%$, is there strong enough evidence that the returns of ZM are not normal?

Shapiro-Wilk normality test of log return of ZM
data: log_rtn_ZM
W = 0.91477, p-value = 2.579e-13

since p = 2.579e-13 < alpha = 0.05, we reject the hypothesis that log return of ZM is normal distributed. As a result, alpha = 0.05 is strong enough to proof that log return of ZM is not normal.

Jarque-Bera Normality Test of log return of ZM
data: log_rtn_ZM
JB = 1209.3, p-value < 2.2e-16
alternative hypothesis: greater

since p < 2.2e-16 < alpha = 0.05, we reject the hypothesis that log return of ZM is normal distributed. As a result, alpha = 0.05 is strong enough to proof that log return of ZM is not normal.

Q1-2.R Download

```r
#Q1-2
#Data Preparation
ZMTSLA = read.csv("/Users/yu-chingliao/Desktop/ZMTSLA.csv")
log_rtn_ZM = log(ZMTSLA$ZM[2:nrow(ZMTSLA)] / ZMTSLA$ZM[1:nrow(ZMTSLA)-1])
log_rtn_TSLA = log(ZMTSLA$TSLA[2:nrow(ZMTSLA)] / ZMTSLA$TSLA[1:nrow(ZMTSLA)-1])
mu_logrt_ZM = mean(log_rtn_ZM)
mu_logrt_TSLA = mean(log_rtn_TSLA)
D = log_rtn_ZM - log_rtn_TSLA

#Testing the Normality of log return of ZM by Shapiro-Wilk normality test
library(moments)
shapiro.test(log_rtn_ZM)

#Testing the Normality of log return of ZM by Jarque-Bera normality test
jarque.test(log_rtn_ZM)
```

Q1.3
1 Point

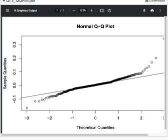
Construct a normal plot for ZM's returns. Based on question 1.2 and the normal plot, are ZM's returns non-normal statistically? or practically? or both?

Statistically, based on the output given on question 1.2, the log return of ZM is not normally distributed at all. However, based on the Normal Q-Q Plot, we can noticed that, though it is not completely normal distributed, log return of ZM is not significantly deviated from normal distribution.
Thus, statistically, log return of ZM is not normal. However, practically, I would say it is partially normal.

Q1-3.R Download

```r
#Q1-3
#Data Preparation
ZMTSLA = read.csv("/Users/yu-chingliao/Desktop/ZMTSLA.csv")
log_rtn_ZM = log(ZMTSLA$ZM[2:nrow(ZMTSLA)] / ZMTSLA$ZM[1:nrow(ZMTSLA)-1])
log_rtn_TSLA = log(ZMTSLA$TSLA[2:nrow(ZMTSLA)] / ZMTSLA$TSLA[1:nrow(ZMTSLA)-1])
mu_logrt_ZM = mean(log_rtn_ZM)
mu_logrt_TSLA = mean(log_rtn_TSLA)
D = log_rtn_ZM - log_rtn_TSLA

#Plotting for normality
qqnorm(log_rtn_ZM)
qqline(log_rtn_ZM)
```

Q1_3_QQPlot.pdf Download



Q1.4
1 Point

Suppose you want to test whether there is strong enough evidence that $\mu_{zm} < \mu_{tsla}$. Write down the null hypothesis H_0 and the alternative hypothesis H_1.

Null hypothesis H0: muZM - muTSLA = muD >= mu0 = 0
alternative hypothesis H1: muZM - muTSLA = muD < mu0 = 0

No files uploaded

Q1.5
1 Point

Assume that D_i's are i.i.d. and the sample size is large enough for the central limit theorem to apply. Give the test statistic and the corresponding sampling distribution when H_0 is true.

Since the sample size is large enough to apply CLT, the sampling distribution will be N(0,1) if H0 is true. (Usually, we should use t-dist, but contributed by CLT, we can test as Z.)

The test statistics Z0 is equal to (muD -mu0)/(Sn/sqrt(n)).
for muD is mean of D's,
mu0 = 0,
and Sn/sqrt(n) is standard error of muD.

No files uploaded

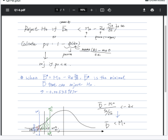
Q1.6
1 Point

For significance level α, determine the criteria for rejecting H_0 in support of H_1 in terms of $\bar{D}_n$: i.e., for what values of $\bar{D}_n$, will H_0 be rejected? For the data given, will H_0 be rejected at significance level $\alpha = 5\%$?

the minimal Dbar that can reject H0 is -0.005359872.

Under significance level of 5%, not reject H0.
So, we can hypothesize that, since we not reject that muD muZM - muTSLA >= 0, there is no significant evidence to show that muZM < muTSLA under 95% confidence.
(Criteria and Code shown below.)

Q1-6.pdf Download



Q1-6.R Download

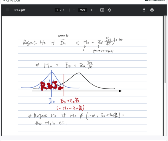
```r
#Q1-6
#Data Preparation
ZMTSLA = read.csv("/Users/yu-chingliao/Desktop/ZMTSLA.csv")
log_rtn_ZM = log(ZMTSLA$ZM[2:nrow(ZMTSLA)] / ZMTSLA$ZM[1:nrow(ZMTSLA)-1])
log_rtn_TSLA = log(ZMTSLA$TSLA[2:nrow(ZMTSLA)] / ZMTSLA$TSLA[1:nrow(ZMTSLA)-1])
mu_logrt_ZM = mean(log_rtn_ZM)
mu_logrt_TSLA = mean(log_rtn_TSLA)
D = log_rtn_ZM - log_rtn_TSLA

#Calculate the minimal D_bar that can reject H0.
muD = mean(D)
mu0 = 0
n = length(D)
alpha = 0.05
Zalpha = qnorm(1-alpha)
Z0 = (muD - mu0)/se
se = sd(D)/sqrt(n)
D_star = mu0 - Zalpha*se
D_star

#Testing by Testing Statistics
#Reject if Z0 < -Za
if (Z0 < -Zalpha){
  print("Reject H0")
}else {
  print("Not Reject H0")
}
muD
Z0
-Zalpha
```

Q1.7
1 Point

Construct the one-sided confidence interval for $\mu_{zm} - \mu_{tsla}$ corresponding to question 1.6. Is the confidence interval consistent with your conclusion in question 1.6?

the ond-side CI of muD = muZM - muTSLA is (-infinity, 0.004783379].
mu0 is in this CI, thus we not reject H0.
So, we can hypothesize that, since we not reject that muD muZM - muTSLA = 0, there is no significant evidence to show that muZM < muTSLA under 95% confidence.

Q1-7.pdf Download



Q1-7.R Download

```r
#Q1-7
#Data Preparation
ZMTSLA = read.csv("/Users/yu-chingliao/Desktop/ZMTSLA.csv")
log_rtn_ZM = log(ZMTSLA$ZM[2:nrow(ZMTSLA)] / ZMTSLA$ZM[1:nrow(ZMTSLA)-1])
log_rtn_TSLA = log(ZMTSLA$TSLA[2:nrow(ZMTSLA)] / ZMTSLA$TSLA[1:nrow(ZMTSLA)-1])
mu_logrt_ZM = mean(log_rtn_ZM)
mu_logrt_TSLA = mean(log_rtn_TSLA)
D = log_rtn_ZM - log_rtn_TSLA

#Testing by Confidence Interval
muD = mean(D)
mu0 = 0
n = length(D)
alpha = 0.05
Zalpha = qnorm(1-alpha)
se = sd(D)/sqrt(n)
ub = muD + Zalpha * se

#Reject if mu0 "not in" CI.
if (mu0 > ub){
  print("Reject H0")
}else {
  print("Not Reject H0")
}
```

Q1.8
1 Point

Give the formula for the p-value of the test in question 1.4. Compute the p-value. Does the p-value confirm your conclusion in question 1.6?

P- value = 0.5702128
Since, P > a, not reject H0.
So, we can hypothesize that, since we not reject that muD muZM - muTSLA = 0, there is no significant evidence to show that muZM < muTSLA under 95% confidence.

Q1-8.R Download

```r
#Q1-8
#Data Preparation
ZMTSLA = read.csv("/Users/yu-chingliao/Desktop/ZMTSLA.csv")
log_rtn_ZM = log(ZMTSLA$ZM[2:nrow(ZMTSLA)] / ZMTSLA$ZM[1:nrow(ZMTSLA)-1])
log_rtn_TSLA = log(ZMTSLA$TSLA[2:nrow(ZMTSLA)] / ZMTSLA$TSLA[1:nrow(ZMTSLA)-1])
mu_logrt_ZM = mean(log_rtn_ZM)
mu_logrt_TSLA = mean(log_rtn_TSLA)
D = log_rtn_ZM - log_rtn_TSLA

#Testing by P-Value
Z0 = (muD - mu0)/se
P = 1 - pnorm(Z0)

#Reject if P < a
if (P < alpha){
  print("Reject H0")
}else {
  print("Not Reject H0")
}
P
```

Q1.9
1 Point

When does the type I error occur in the test in question 1.4? What's the probability of type I error in question 1.6?

When the H0 actually correct, however we mistakenly reject it. To be more specific, if muZM actually equals to muTSLA, however, under (1-alpha)*100% confidence, we mistakenly reject this fact, then it is type 1 error.
the probability of type 1 error is alpha = 0.05.

No files uploaded

Q1.10
1 Point

If H_0 is rejected, does that mean H_1 is true? If H_0 is not rejected, does that mean H_0 is true?

If we conclude 'do not reject H0', this does not necessarily mean that the null hypothesis is true, it only suggests that there is not sufficient evidence to reject H0.
If we reject the null hypothesis, then it suggests that the alternative hypothesis may be true

No files uploaded