Stat 577 Homework 2

Name: Lian Zuo

## Question #1

a)

The mean vector is:

X Y

9.546947 19.442551

The sample covariance matrix for D is:

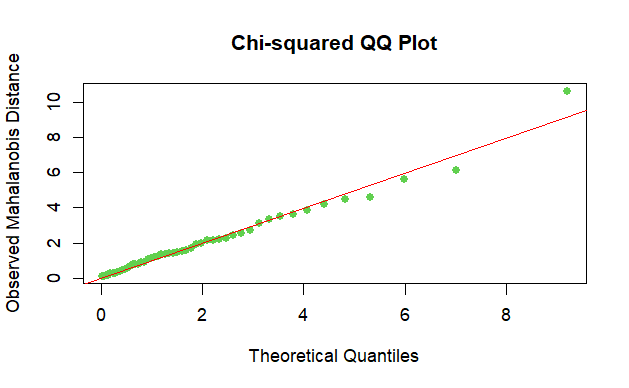
X Y

X 0.783302 1.184362

Y 1.184362 3.333838

b) The Chi-squared QQ plot of Mahalanobis distance versus the corresponding chi-squared quantiles is as follows.

Comment: Scattering of points around a straight line and the QQ plot for data D shows reasonably a linear fit, suggest that (X, Y) jointly follows a bivariate normal distribution. Since linearity is subjective particularly with graphics, let us check the normality of the data using a formal test.



c) The null hypothesis: the data is coming from a bivariate normal distribution.

Then alternative hypothesis: the data is not coming from a bivariate normal distribution.

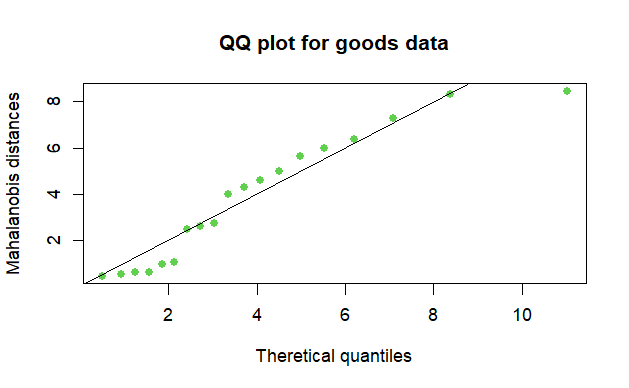
p-value for the test: p-value = 0.06778

conclusion: at , the multivariate Shapiro-Wilk p-value is not significant. Therefore, we accept the null hypothesis and conclude that the data is coming from a bivariate normal distribution.

## Question # 2

1. The QQ plot for the Mahalanobis distance and theoretical quantiles from a chi-squared distribution is as follows.

Comment: The QQ plot for the goods data shows reasonably a linear fit (except one point), indicating that it jointly follows a multivariate normal distribution. Since linearity is subjective particularly with graphics, let us check the normality of the data using a formal test.



b)

The null hypothesis: the data is coming from a multivariate normal distribution.

Then alternative hypothesis: the data is not coming from a multivariate normal distribution.

test statistic value: 0.89405

p-value for the test: p-value = 0.03803

conclusion: at , the multivariate Shapiro-Wilk p-value is significant. Therefore, we reject the null hypothesis and conclude that the data is not coming from a multivariate normal distribution.

c)

the p-value of Shapiro-Wilk test for x1(length of cycle) is 0.3017, at , the p-value is not significant which indicate that the data of variable x1 comes from a normal distribution.

the p-value of Shapiro-Wilk test for x2(percentage of rising prices) is 0.285, at , the p-value is not significant which indicate that the data of variable x2 comes from a normal distribution.

the p-value of Shapiro-Wilk test for x3(cyclical amplitude) is 0.06417, at , the p-value is significant which indicate that the data of variable x2 is not coming from a normal distribution.

the p-value of Shapiro-Wilk test for x4(rate of change) is 0.3151, at , the p-value is not significant which indicate that the data of variable x4 comes from a normal distribution.

Appendix-- R Code

# Stat 577 Homework 2

# Question 1

set.seed(1234)

z1 <- rnorm(50, 0,1)

z2 <- rnorm(50, 0,1)

mu\_x <- 10

mu\_y <- 20

sigma\_x <- 1

sigma\_y <- 2

rho <- 0.8

X <- sigma\_x \* z1 + mu\_x

Y <- sigma\_y \* (rho \* z1 + sqrt(1 - rho^2) \* z2) + mu\_y

D <- cbind(X, Y)

mean\_vec<- colMeans(D)

mean\_vec

cov\_matrix <- cov(D)

cov\_matrix

# Calculate Mahalanobis distance

mahal\_D <- mahalanobis(D, center = mean\_vec, cov = cov\_matrix)

mahal\_D

# Degrees of freedom for Chi-squared distribution

df <- 2 # Since we have 2 variables

# Calculate the expected quantiles

theo.quan<- qchisq(ppoints(50), df)

# Create the QQ plot

qqplot(theo.quan, mahal\_D,

xlab = "Theoretical Quantiles",

ylab = "Observed Mahalanobis Distance",

main = "Chi-squared QQ Plot",

pch=16,col=3)

# Draw a reference line

abline(0, 1, col = "red")

# Show the plot

#Install. packages("mvnormtest)

library(mvnormtest)

mshapiro.test(t(D))

#The multivariate Shapiro-Wilk p-value is not significant.Therefore, we accept the null hypothesis and conclude that the data is coming from a bivariate normal distribution.

# Question 2

goods <- read.table("C:/Users/lian/Desktop/goods.txt",header=F)

goods <- goods[,-1]

colnames(goods) <- c("x1","x2","x3","x4")

smean <- colMeans(goods)

scov <- cov(goods)

print (list(smean,scov))

n <- nrow(goods)

mhd.goods <- mahalanobis(goods,smean,scov)

mhd.goods

theo.quan2 <- qchisq(ppoints(n),4)# df=number of variables = 4

qqplot(theo.quan2,mhd.goods,xlab = "Theretical quantiles",

ylab="Mahalanobis distances",pch=16,col=3,main="QQ plot for goods data")

abline(0,1)# drawing a reference line

#install.packages("mvnormtest")

library(mvnormtest)

mshapiro.test(t(goods))

# Formal test for checking normality:Shapiro-Wilk test

attach(goods)

shapiro.test(x1)

shapiro.test(x2)

shapiro.test(x3)

shapiro.test(x4)