STAT401_HW1

김정현

```
library(MASS)
library(rgl)
library(lattice)
```

Q1.

Suppose that three random variables, X1, X2, and X3, jointly follow a multivariate normal distribution with the mean vector and the variance-covariance matrix.

```
mu <- c(-1, 2, -5)
sigma <- matrix(c(9, 4, 1,
                  4, 16, -1,
                  1, -1, 14), nrow = 3, byrow = TRUE)
```

(a)

Calculate population correlation matrix.

Note that population correlation matrix equals $\rho = D^{-1/2}\sum D^{1/2}$. Also, since $D^{-1/2} = diag\left(\frac{1}{\sqrt{\sigma_{ii}}}\right)$ and $\sigma_{11} = 9, \sigma_{22} = 16, \sigma_{33} = 4$, We can calculate $D^{-1/2}$.

$$D^{-1/2} = \begin{pmatrix} 1/3 & 0 & 0 \\ 0 & 1/4 & 0 \\ 0 & 0 & 1/2 \end{pmatrix}$$

Therefore, the population correlation matrix equals

$$\rho = D^{-1/2}\sum D^{1/2} = \begin{pmatrix} 1/3 & 0 & 0 \\ 0 & 1/4 & 0 \\ 0 & 0 & 1/2 \end{pmatrix}\begin{pmatrix} 9 & 4 & 1 \\ 4 & 16 & -1 \\ 1 & -1 & 14 \end{pmatrix}\begin{pmatrix} 1/3 & 0 & 0 \\ 0 & 1/4 & 0 \\ 0 & 0 & 1/2 \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 1/3 & 1/6 \\ 1/3 & 1 & -1/8 \\ 1/6 & -1/8 & 1 \end{pmatrix}$$

```
d <- matrix(c(1/3, 0, 0,
              0, 1/4, 0,
              0, 0, 1/2), nrow = 3, byrow = TRUE)
sigma <- matrix(c(9, 4, 1,
                  4, 16, -1,
                  1, -1, 4), nrow = 3, byrow = TRUE)

ans <- d %*% sigma %*% d
print(ans)
```

```
##              [,1]        [,2]        [,3]
## [1,] 1.0000000  0.3333333  0.1666667
## [2,] 0.3333333  1.0000000 -0.1250000
## [3,] 0.1666667 -0.1250000  1.0000000
```

(b)

Let $a' = [1, -1, 1]$ and $x = [x_1, x_2, x_3]$ where $Z_1 = a'X$.

$$E(Z_1) = E(a'X) = a'E(X) = a'\mu = \begin{pmatrix} 1 & -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 2 \\ -5 \end{pmatrix} = -8$$

$$Var(Z_1) = a'\Sigma a = \begin{pmatrix} 1 & -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 9 & 4 & 1 \\ 4 & 16 & -1 \\ 1 & -1 & 4 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} = 25$$

```
a <- matrix(c(1, -1, 1), nrow = 3)
mu <- c(-1, 2, -5)
sigma <- matrix(c(9, 4, 1,
                  4, 16, -1,
                  1, -1, 4), nrow = 3, byrow = TRUE)
e1 = t(a) %*% mu
var1 = t(a) %*% sigma %*% a

print("Mean of Z1:")

## [1] "Mean of Z1:"

print(e1)

##      [,1]
## [1,]   -8

print("Variance of Z1:")

## [1] "Variance of Z1:"

print(var1)

##      [,1]
## [1,]   25
```

(c)

Let $b' = [-1, 2, 3]$ and $x = [x_1, x_2, x_3]$ where $Z_2 = b'X$.

$$E(Z_2) = E(b'X) = b'E(X) = b'\mu = \begin{pmatrix} -1 & 2 & 3 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 2 \\ -5 \end{pmatrix} = -10$$

$$Var(Z_2) = b'\Sigma b = \begin{pmatrix} -1 & 2 & 3 \end{pmatrix} \cdot \begin{pmatrix} 9 & 4 & 1 \\ 4 & 16 & -1 \\ 1 & -1 & 4 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix} = 75$$

```r
b <- matrix(c(-1, 2, 3), nrow = 3)
mu <- c(-1, 2, -5)
sigma <- matrix(c(9, 4, 1,
                  4, 16, -1,
                  1, -1, 4), nrow = 3, byrow = TRUE)
e2 = t(b) %*% mu
var2 = t(b) %*% sigma %*% b

print("Mean of Z2:")

## [1] "Mean of Z2:"

print(e2)

##      [,1]
## [1,]  -10

print("Variance of Z2:")

## [1] "Variance of Z2:"

print(var2)

##      [,1]
## [1,]   75
```

(d)

$$E(Z_3) = E\begin{pmatrix} Z_1 \\ -Z_2 \end{pmatrix} = \begin{pmatrix} -8 \\ -10 \end{pmatrix}$$

$Var(Z_1) = 25$ and $Var(-Z_2) = Var(Z_2) = 75$

$$Cov(Z_1, -Z_2) = Cov(a'x, -b'x) = -a'\Sigma b = -1 \cdot \begin{pmatrix} 1 & -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 9 & 4 & 1 \\ 4 & 16 & -1 \\ 1 & -1 & 4 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix} = 14$$

```r
cov_z1z2 = -t(a) %*% sigma %*% b
print("Covariance for Z1, Z2")

## [1] "Covariance for Z1, Z2"

print(cov_z1z2)

##      [,1]
## [1,]   14
```

Therefore

$$Cov(Z_3) = \begin{pmatrix} 25 & -14 \\ -14 & 75 \end{pmatrix}$$

(e)

```r
n <- 50
sample_data <- mvrnorm(n, mu, sigma)
print("Sample mean")
```

```
## [1] "Sample mean"
```

```r
print(colMeans(sample_data))
```

```
## [1] -1.1800521  0.8789824 -4.8219785
```

```r
print("Sample Variance")
```
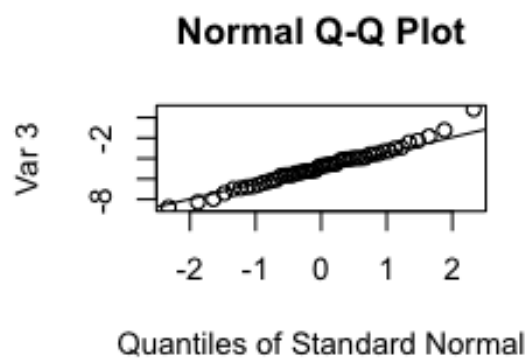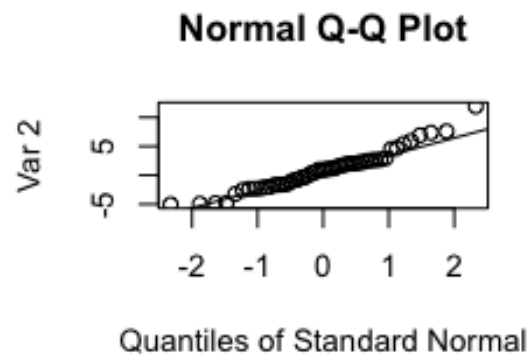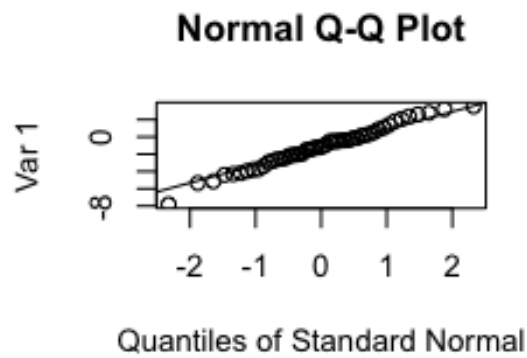
```
## [1] "Sample Variance"
```

```r
print(cov(sample_data))
```

```
##            [,1]      [,2]     [,3]
## [1,] 5.760218  2.241794 1.293143
## [2,] 2.241794 12.098062 1.040807
## [3,] 1.293143  1.040807 3.330100
```

(f)

```r
p <- ncol(sample_data)
par(mfrow=c(2,2))
for(i in 1:p){qqnorm(sample_data[,i],xlab="Quantiles of Standard
Normal",ylab=paste("Var",i))
    qqline(sample_data[,i])}
```
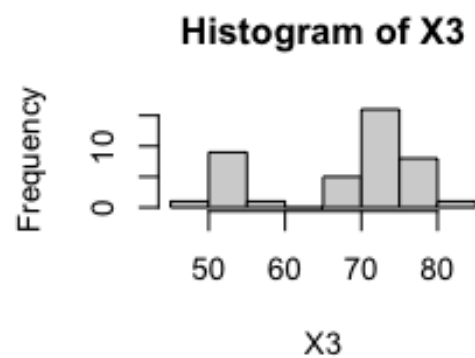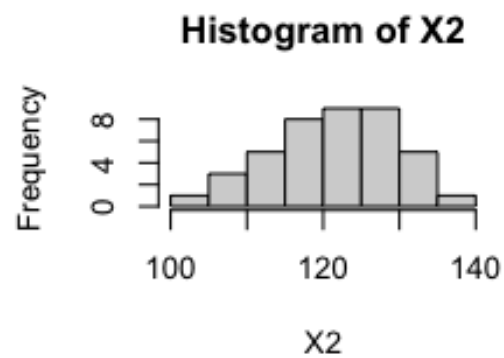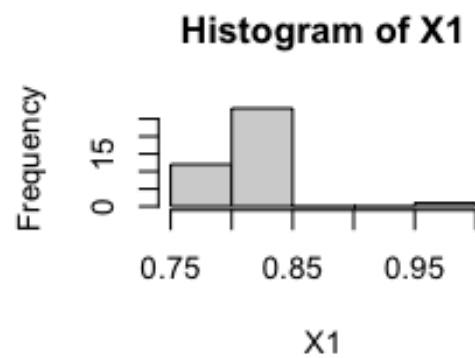
Normal Q-Q Plot

Normal Q-Q Plot

Normal Q-Q Plot

The plot shows that the distribution follows the multivariate normal distribution, which means that the plot follows the straight line.
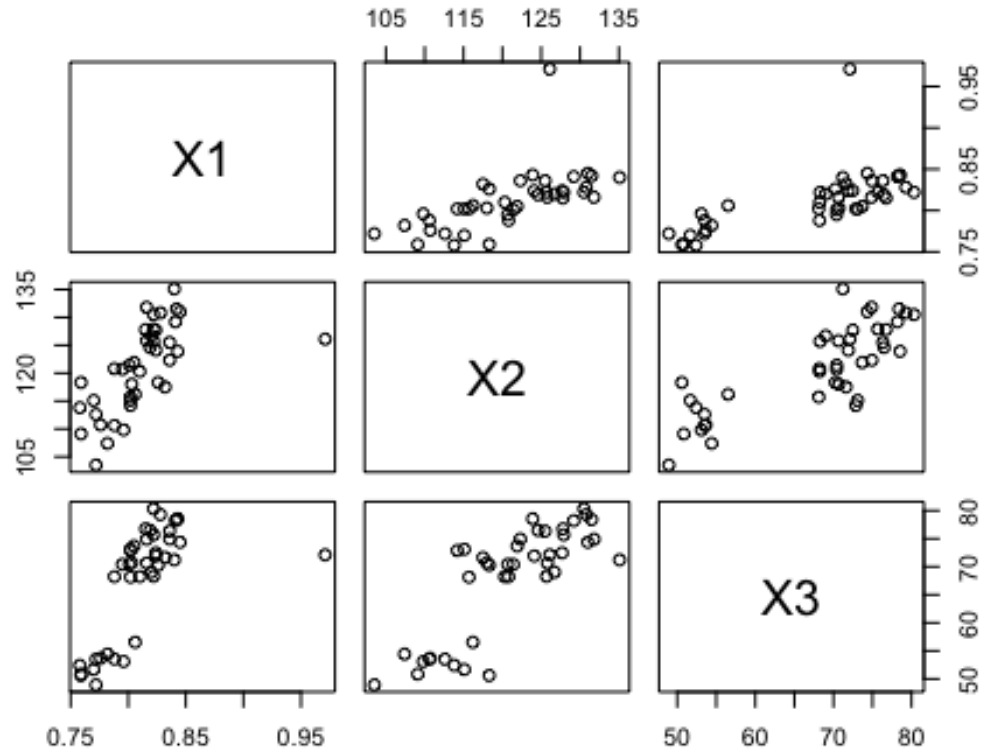
(2)

(a)

```r
par(mfrow=c(2,2))
data <- read.table('Paper.dat', header = T)
attach(data)
hist(X1)
hist(X2)
hist(X3)
```

Histogram of X1

Histogram of X2

Histogram of X3

One can observe that X1 (Density) has an outlier by boxplot. Also by histogram, We can find that X2 follows the normal distribution similarly and X3 does not.
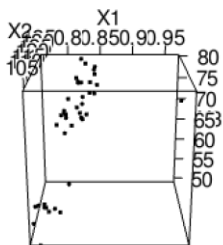
(b)
```
pairs(data)
```

X1 and X2 are strongly related (except 1 outlier). X1 and X3 are strongly related (except 1 outlier). X2 and X3 are strongly related.

(c)

```
# plot3d(data)
```


Image Title

We can identify one outlier by drawing the graph.

(d)

```
print("Sample mean")
```

```
## [1] "Sample mean"
```

```
xm <- colMeans(data)
print(xm)

##          X1          X2          X3
##   0.8118537 120.9534146  67.7231707

print("Sample Variance")

## [1] "Sample Variance"

Sx <- cov(data)
print(Sx)

##              X1          X2          X3
## X1 0.001264578  0.1684468   0.225248
## X2 0.168446762 59.3211480  60.992531
## X3 0.225247976 60.9925314  95.856667
```
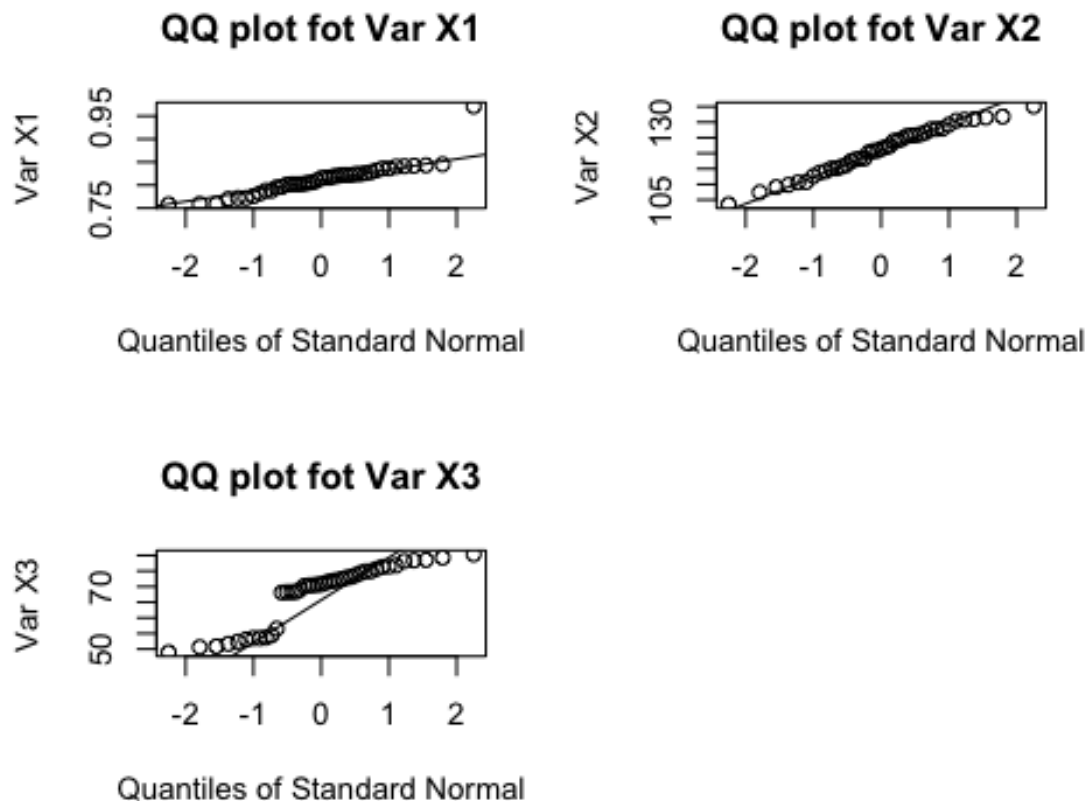
(e)
```
p <- ncol(data)
par(mfrow=c(2,2))
for(i in 1:p){qqnorm(data[,i],xlab="Quantiles of Standard Normal",
                     ylab=paste("Var", colnames(data)[i]),
                     main=paste("QQ plot fot Var", colnames(data)[i]))
    qqline(data[,i])}
```

QQ plot fot Var X1

QQ plot fot Var X2

QQ plot fot Var X3

The plot for X1 shows that the distribution follows the multivariate normal distribution (except 1 outlier), which means that the plot follows the straight line. This has same result with the box plot drew in (a).

The plot for X2 shows that the distribution follows the multivariate normal distribution, which means that the plot follows the straight line. This has same result with the histogram drew in (a) which follow the normal distribution similarly.

The plot for X3 shows that the distribution does not follow the multivariate normal distribution, which means that the plot does not follow the straight line. This has same result with the histogram drew in (a) which did not follow the normal distribution.

(f)
```
di2 = mahalanobis(data, xm, Sx)
di2
```

```
##  [1]   0.5007665   0.9300751   1.2763678   3.0528182   7.1035022   1.8763221
##  [7]   1.1995473   4.6709142   1.9771076   1.1547018   1.2031490   5.3194161
## [13]   0.0539644   1.5027361   2.3411077   3.0074340   3.2297978   3.7373589
## [19]   7.1113692   2.2063399   2.0176036   1.3774709   2.2696565   0.9579655
## [25]  30.7154056   0.5823940   0.8512959   1.4444657   2.4181352   1.0008221
## [31]   2.1338584   0.1984087   0.8615855   3.1915160   0.8750852   1.1904194
## [37]   1.2271369   2.3648853   5.3240380   2.1450363   3.3980191
```

```
cnt <- length(xm)
qqmath(di2,distribution = function(p) qchisq(p,df=cnt))
```


function(p) qchisq(p, df = cnt)
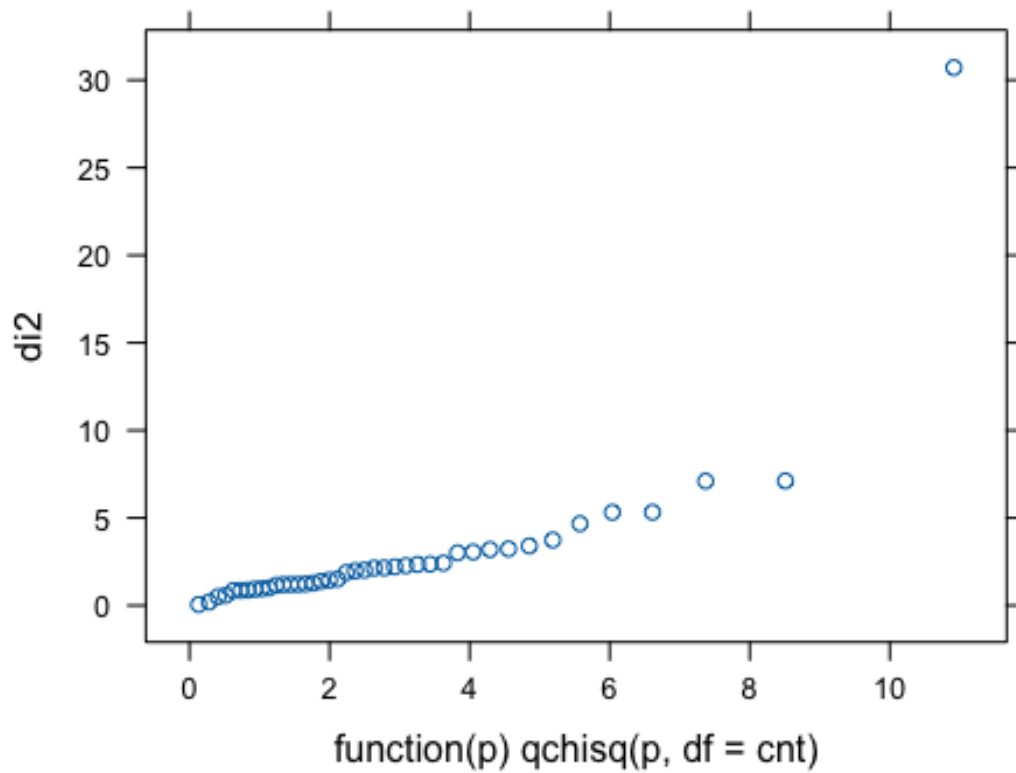
This data set follows the multivariate normal distribution since the plot follows the chi-Squared distribution (except 1 outlier).