Multivariate Analysis

Homework 2

M052040003 鍾冠毅

4.26. (a)

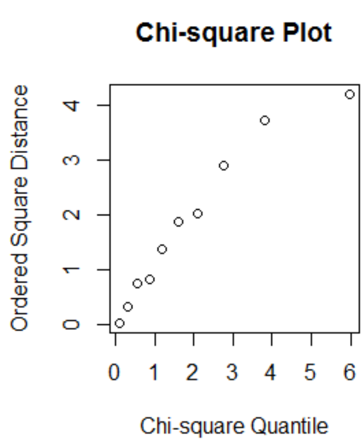
```
> dii
   diag.D.sq.
1   1.8753045
2   2.0203262
3   2.9009088
4   0.7352659
5   0.3105192
6   0.0176162
7   3.7329012
8   0.8165401
9   1.3753379
10  4.2152799
```

(b)

```
> chisq.2.0.5 <- dii[dii <= qchisq(0.5, 2)]
> length(chisq.2.0.5)/length(dii)
[1] 0.5
```

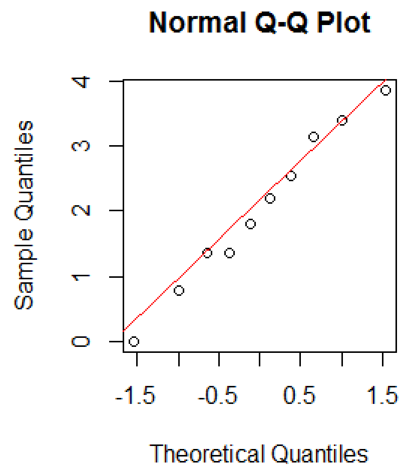
(c)

```
> order.d
 [1] 0.0176162 0.3105192 0.7352659 0.8165401 1.3753379 1.8753045 2.0203262
 [8] 2.9009088 3.7329012 4.2152799
```

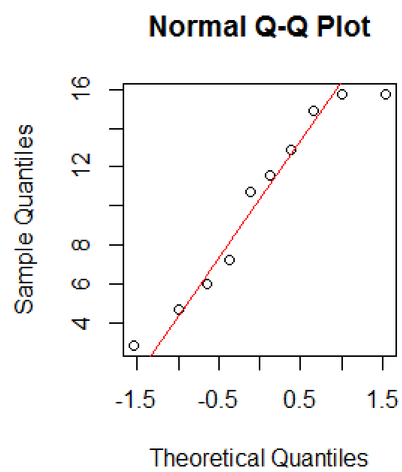


(d) 由(b)(c)結果可知，兩數據服從雙變數常態分配。

4.30. (a) $\hat{\lambda}_1 = 0.3708906$



(b) $\hat{\lambda}_2 = 0.9361967$



(c) $(\hat{\lambda}_1, \hat{\lambda}_2) = (1.2732157, 0.0310405)$

若個別計算最佳的$\hat{\lambda}$，則會使兩變數個別較趨近常態分佈。若同時計$\hat{\lambda}_1$與$\hat{\lambda}_2$，則會找到最佳的$\hat{\lambda}$，使得兩變數較趨近於雙變量常態分佈。

4.39.

```
> df.norm
                   Indep         Supp       Benev      Conform       Leader
df.ks.p    0.009751414 0.0002120256 0.03796667 0.005667049 5.349873e-05
df.lambda  0.523772578 1.3962618614 1.26187573 1.039266313 3.815471e-01
df.ks.p.n  0.125018888 0.0014273401 0.11409417 0.008039632 7.012327e-02
```

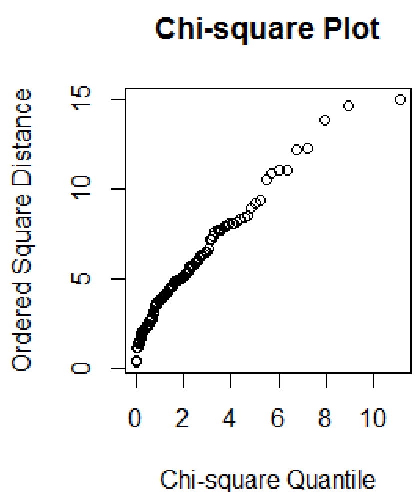
df.ks.p ：轉換前的 Kolmogorov-Smirnov 檢定之 p-value

df.lamda ：轉換所使用的最佳λ

df.ks.p ：轉換後的 Kolmogorov-Smirnov 檢定之 p-value

(a) 由上表可知所有的 p-value 皆小於 0.05，故拒絕常態之假設。

(b) 雖然在 Chi-square 圖中各點分布接近一斜直線，但是由(a)可知個變數之邊際分佈皆不符從常態分佈，故此資料不符從多變量常態分佈。



(c) 由上表可知，各變數經轉換後重新檢定之 p-value 皆有成長的趨勢，甚至大於 0.05，即服從常態分佈，如：Indep、Benev、Leader 等。

Appendix

```
### 4.26 ###

# (a) #

x1 <- c(1, 2, 3, 3, 4, 5, 6, 8, 9, 11)

x2 <- c(18.95, 19.00, 17.95, 15.54, 14.00,

       12.95, 8.94, 7.49, 6.00, 3.99)


df <- data.frame(cbind(x1, x2))

S <- cov(df)

x1.bar <- mean(x1)

x2.bar <- mean(x2)

A <- t(cbind((x1 - x1.bar), (x2 - x2.bar)))

D.sq <- t(A)%*%solve(S)%*%A

dii <- diag(D.sq)


# (b) #

chisq.2.0.5 <- dii[dii <= qchisq(0.5, 2)]

length(chisq.2.0.5)/length(dii)


# (c) #

order.d <- dii[order(dii)]

chi.10 <- sapply(1:10,

              function(j)

                qchisq((10-j+0.5)/10 ,2, lower.tail
= F))

plot(order.d ~ chi.10, xlab = "Chi-square
Quantile",

    ylab = "Ordered Square Distance", main =
"Chi-square Plot")


### 4.30. ###

install.packages("MASS")

library(MASS)

library(car)


# (a) #

lambda1 <- powerTransform(x1)

l1 <- lambda1$lambda

x1.n <- (x1^l1-1)/l1

qqnorm(x1.n)

qqline(x1.n, col="red")

# (b) #

lambda2 <- powerTransform(x2)

l2 <- lambda2$lambda

x2.n <- (x2^l2-1)/l2

qqnorm(x2.n)

qqline(x2.n, col="red")


# (c) #

lambda12 <- powerTransform(cbind(x1,x2))

l1.c <- lambda12$lambda[1]

l2.c <- lambda12$lambda[2]

x1.c <- (x1^l1.c-1)/l1.c
```

```
x2.c <- (x1^l2.c-1)/l2.c


### 4.39 ###

# (a) #

#install.packages("nortest")

library(nortest)

df <- read.csv("4.39.csv", header = F, sep =
",")

names(df) <- c("Indep", "Supp", "Benev",
"Conform",

            "Leader", "Gender", "Socio")

df.ks.p <- sapply(1:5, FUN = function(x)
lillie.test(df[,x])$p.value)

df.sw.p <- sapply(1:5, FUN = function(x)
shapiro.test(df[,x])$p.value)


# (b) #

df <- df[, 1:5]

S <- cov(df)

A <- t(apply(df, MARGIN = 2, FUN =
function(x) x-mean(x)))

D.sq <- t(A)%*%solve(S)%*%A

dii <- diag(D.sq)

order.d <- dii[order(dii)]

chi.130 <- sapply(1:130,

              FUN = function(j)

               qchisq((length(dii)-
j+0.5)/length(dii) ,2, lower.tail = F))

plot(order.d~chi.130, xlab = "Chi-square
Quantile",

    ylab = "Ordered Square Distance", main =
"Chi-square Plot")


# (c) #

library(car)

sapply(1:5, function(x)
powerTransform(df[,x])$lambda)


df.lambda <- sapply(1:5, function(x)
powerTransform(df[, x])$lambda)

df.n <-

apply(df, MARGIN = 2,

     FUN = function(x) {

       lam <- powerTransform(x)$lambda;

       (x^lam-1)/lam})


df.ks.p.n <- sapply(1:5, function(x)
lillie.test(df.n[, x])$p.value)


df.norm <- t(cbind(df.ks.p, df.lambda,
df.ks.p.n))

colnames(df.norm) <- names(df)
```