REPORT

수강과목	:	다변량통계학(I)
담당교수	:	최용석
학 과	:	통계학과
학 번	:	201611531
이 름	:	정호재
제출일자	:	2020.05.08

Multivariate Statistics (I)

Homeworks for Middle Examination

201611531 통계학과 정호재

Deadline : 2020.05.08.

1. Solve the problems (1) ~ (5) except for (6), in Exercise 1.7.

1.7 [자료 1.3.2](klpga.txt)에는 KLAGA 선수성적은 기술요인변수군(평균퍼팅수, 그린적중률, 파세이브율, 파브레이크율)과 경기성적요인변수군(상금률, 평균타수)으로 나눌 수 있다.

(1) 이 자료를 통해 분석자가 가지는 관심은 무엇인지를 설명하라.

[자료 1.3.2](klpga.txt)에는 KLAGA 선수성적은 기술요인변수군(평균퍼팅수, 그린적중률, 파세이브율, 파브레이크율)과 경기성적요인변수군(상금률, 평균타수)간의 관련성에 대하여 관심을 가진다.

데이터를 살펴보면 아래와 같다.

```
> setwd("D:/2020 1학기 정호재/다변량통계학(1)/200402 다변량 실습1/Rdata")
> library(rgl)
> library(MVN)
> data1.3.2<-read.table("klpga.txt", header=T)
> str(data1.3.2)
'data.frame':   50 obs. of  6 variables:
 $ 평균퍼팅수  : num  30.4 30.9 31.3 30.6 31 ...
 $ 그린적중율  : num  82.7 76.9 79.6 79.3 68.9 ...
 $ 파세이브율  : num  90.1 85.3 88.5 87.7 80 ...
 $ 파브레이크율: num  23.8 23.7 19.3 21.3 17.2 ...
 $ 평균타수    : num  69.6 70.8 70.9 70.5 72.8 ...
 $ 상금율      : num  100 63.7 59.4 50.2 44 38 30.5 21.8 21.1 19.2 ...
> dim(data1.3.2)
[1] 50  6
> colSums(is.na(data1.3.2))
  평균퍼팅수   그린적중율   파세이브율 파브레이크율     평균타수 
           0            0            0            0            0 
      상금율 
           0 
```

```
> summary(data1.3.2)
   평균퍼팅수        그린적중율        파세이브율       파브레이크율
 Min.   :29.81   Min.   :61.30   Min.   :76.26   Min.   :11.76
 1st Qu.:30.87   1st Qu.:68.26   1st Qu.:78.78   1st Qu.:13.24
 Median :31.36   Median :70.19   Median :80.33   Median :14.54
 Mean   :31.36   Mean   :70.92   Mean   :81.23   Mean   :15.25
 3rd Qu.:31.80   3rd Qu.:73.27   3rd Qu.:82.99   3rd Qu.:16.78
 Max.   :32.78   Max.   :82.72   Max.   :90.12   Max.   :23.77
    평균타수           상금율
 Min.   :69.58   Min.   :  5.100
 1st Qu.:72.26   1st Qu.:  7.525
 Median :73.41   Median :  9.000
 Mean   :73.01   Mean   : 16.722
 3rd Qu.:73.79   3rd Qu.: 16.300
 Max.   :74.73   Max.   :100.000
```

(2) 평균벡터, 공분산행렬S, 상관행렬 R을 구하라.

```
> X<-data1.3.2
> class(X)
[1] "data.frame“
```

```
> X<-as.matrix(X)# 자료행렬
> n<-nrow(X)# 행 개수
> xbar<-t(X)%*%matrix(1,n,1)/n # 평균벡터
> I<-diag(n)
> J<-matrix(1,n,n)
> H<-I-1/n*J# 중심화행렬
> Y<-H%*%X# 중심화 자료행렬
> S<-t(Y)%*%Y/(n-1)# 공분산행렬
> D<-diag(1/sqrt(diag(S)))# 표준편차행렬의 역
> Z<-H%*%X%*%D# 표준화자료행렬
> colnames(Z)<-colnames(X)
> R<-t(Z)%*%Z/(n-1)# 상관행렬
> colnames(xbar)<-c("Mean")
```

```
> xbar; # 평균벡터
                Mean
평균퍼팅수     31.3612
그린적중율     70.9236
파세이브율     81.2336
파브레이크율   15.2524
평균타수       73.0138
상금율         16.7220
> S; # 공분산행렬
             평균퍼팅수  그린적중율  파세이브율  파브레이크율     평균타수
평균퍼팅수    0.4077536   0.3772752  -0.8023922     -0.806554    0.3210647
그린적중율    0.3772752  21.3220766  11.7180174      9.695336   -4.1885915
파세이브율   -0.8023922  11.7180174  11.1798602      6.884448   -3.5515772
파브레이크율 -0.8065540   9.6953361   6.8844483      8.252321   -2.9216460
평균타수      0.3210647  -4.1885915  -3.5515772     -2.921646    1.2846608
상금율       -4.6928637  53.5031437  44.5007559     43.034313  -16.9958608
                 상금율
평균퍼팅수    -4.692864
그린적중율    53.503144
파세이브율    44.500756
파브레이크율  43.034313
평균타수     -16.995861
상금율       326.822976
> R; # 상관행렬
             평균퍼팅수  그린적중율  파세이브율  파브레이크율    평균타수
평균퍼팅수    1.0000000   0.1279513  -0.3758111    -0.4396899   0.4436081
그린적중율    0.1279513   1.0000000   0.7589640     0.7309036  -0.8003114
파세이브율   -0.3758111   0.7589640   1.0000000     0.7167421  -0.9371494
파브레이크율 -0.4396899   0.7309036   0.7167421     1.0000000  -0.8973158
평균타수      0.4436081  -0.8003114  -0.9371494    -0.8973158   1.0000000
상금율       -0.4065208   0.6409265   0.7361951     0.8286489  -0.8294546
                 상금율
평균퍼팅수   -0.4065208
그린적중율    0.6409265
파세이브율    0.7361951
파브레이크율  0.8286489
평균타수     -0.8294546
상금율        1.0000000
```

```
> colMeans(X) # 평균벡터
  평균퍼팅수   그린적중율   파세이브율 파브레이크율     평균타수
     31.3612      70.9236      81.2336      15.2524      73.0138
      상금율
     16.7220
> cov(X) # 공분산행렬
              평균퍼팅수 그린적중율 파세이브율 파브레이크율    평균타수
평균퍼팅수     0.4077536  0.3772752 -0.8023922    -0.806554   0.3210647
그린적중율     0.3772752 21.3220766 11.7180174     9.695336  -4.1885915
파세이브율    -0.8023922 11.7180174 11.1798602     6.884448  -3.5515772
파브레이크율 -0.8065540  9.6953361  6.8844483     8.252321  -2.9216460
평균타수       0.3210647 -4.1885915 -3.5515772    -2.921646   1.2846608
상금율        -4.6928637 53.5031437 44.5007559    43.034313 -16.9958608
                  상금율
평균퍼팅수     -4.692864
그린적중율     53.503144
파세이브율     44.500756
파브레이크율  43.034313
평균타수      -16.995861
상금율        326.822976
> cor(X) # 상관행렬
              평균퍼팅수 그린적중율 파세이브율 파브레이크율   평균타수
평균퍼팅수     1.0000000  0.1279513 -0.3758111   -0.4396899  0.4436081
그린적중율     0.1279513  1.0000000  0.7589640    0.7309036 -0.8003114
파세이브율    -0.3758111  0.7589640  1.0000000    0.7167421 -0.9371494
파브레이크율 -0.4396899  0.7309036  0.7167421    1.0000000 -0.8973158
평균타수       0.4436081 -0.8003114 -0.9371494   -0.8973158  1.0000000
상금율        -0.4065208  0.6409265  0.7361951    0.8286489 -0.8294546
                  상금율
평균퍼팅수    -0.4065208
그린적중율     0.6409265
파세이브율     0.7361951
파브레이크율  0.8286489
평균타수      -0.8294546
상금율         1.0000000
```

공식을 사용하여 xbar, S, R으로 평균백터 공분산행렬, 상관행렬을 구해주었고 R에 내장되어있는 함수를 사용하여 각각 colMeans(X), cov(X), cor(X)와 동일한 값을 가진다.

(3) 다변량 변동량을 구하고 해석하라.

```
> detS <- det(S)
> detR <- det(R)
> trS <- sum(diag(S))
> trR <- sum(diag(R))
> detS; # 데이터의 일반화 분산
[1] 35.01831
> detR; # 상관행렬의 일반화 분산
[1] 0.0001039811
> trS; # 데이터의 총 분산
[1] 369.2696
> trR; # 상관행렬의 총 분산
[1] 6
```

일반화분산 |S|=35.01831으로 평균 또는 중위수를 중심으로 자료가 흩어져있다.
이는 총 분산 tr(S)=369.2696으로 반영이 되어있다.
(|S|=0이면 중심화자료행렬의 열들이 선형종속으로 변수들 간에 공선성이 존재)

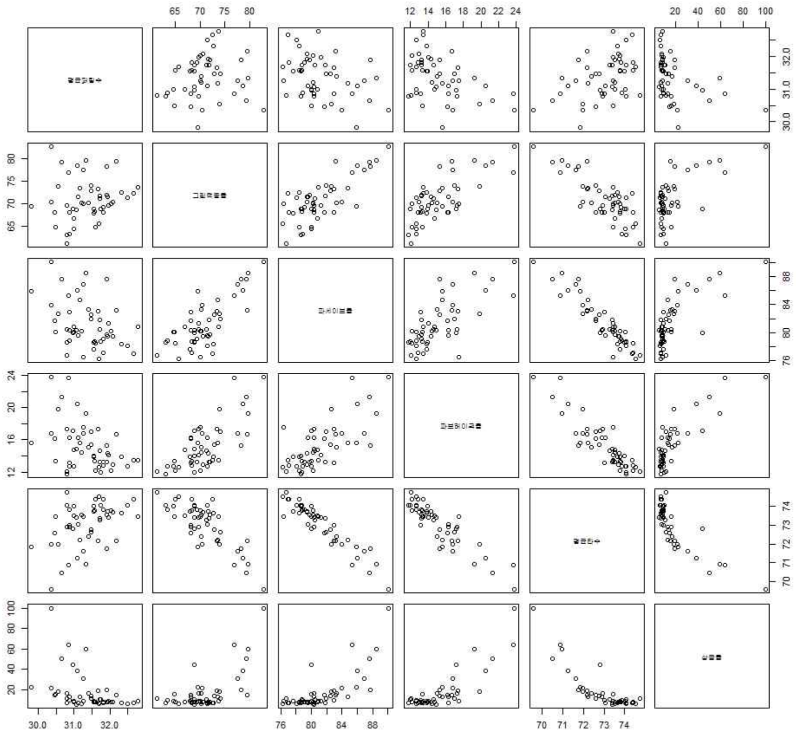
상관행렬의 일반화분산 |R|=0.0001039811이 매우 작으므로 각 변수들의 상관관계가 매우 높다고 할 수 있다. 총 분산 tr(R)=6은 변수가 6개라서 이와 같은 값을 갖는다.
(|S|와 tr(S)는 분산의 크기에 영향을 많이 받으므로 분산 1을 대각원소로 가지는 상관행렬 R을 사용하는 것이 바람직하다.)

(4) 이 자료는 표준화와 중심화 중 어느 것을 선택해야 하는지를 설명하라.

각각의 변수들의 다른 단위를 가지므로 각 변수를 표준화하는 것이 더 적절하다.
특히 상금율의 분산은 326.822976으로 다른 변수들보다 큰 값을 가진다.

(5) 변수들의 연관성을 기술요인변수군과 경기성적용인 변수군으로 나눌 수 있는지에 대하여 살펴보라

```
> plot(data1.3.2)
```




```
> cor(X)
             평균퍼팅수  그린적중율  파세이브율  파브레이크율    평균타수
평균퍼팅수    1.0000000   0.1279513  -0.3758111    -0.4396899   0.4436081
그린적중율    0.1279513   1.0000000   0.7589640     0.7309036  -0.8003114
파세이브율   -0.3758111   0.7589640   1.0000000     0.7167421  -0.9371494
파브레이크율 -0.4396899   0.7309036   0.7167421     1.0000000  -0.8973158
평균타수      0.4436081  -0.8003114  -0.9371494    -0.8973158   1.0000000
상금율       -0.4065208   0.6409265   0.7361951     0.8286489  -0.8294546
                 상금율
평균퍼팅수   -0.4065208
그린적중율    0.6409265
파세이브율    0.7361951
파브레이크율  0.8286489
평균타수     -0.8294546
상금율        1.0000000
```

기술요인변수군의 변수 내에서 상금율과의 상관 계수를 확인하면, 평균퍼팅수는 다른 기술요인변수들과 달리 음수값을 가진다. 평균타수와의 상관계수를 확인하면 앞의 결과와 비슷하게 평균퍼팅수는 다른 기술요인변수들과 달리 양수값을 가진다.
따라서, 변수들의 연관성을 기술요인변수군과 경기성적용인 변수군으로 나눌 수 없다.

2. Solve the problems (1) ~ (3) except for (4) and (5), in Exercise 1.10.

1.10 [자료 1.3.4](irisflower.txt)에는 세 종류의 붓꽃(Setosa, Versicolor, Virginica) 각각 50 포기씩 뽑아서 꽃받침 길이(X1), 꽃받침 폭(X2), 꽃잎 길이(X3), 꽃잎 폭(X4)을 측정한 자료이다.

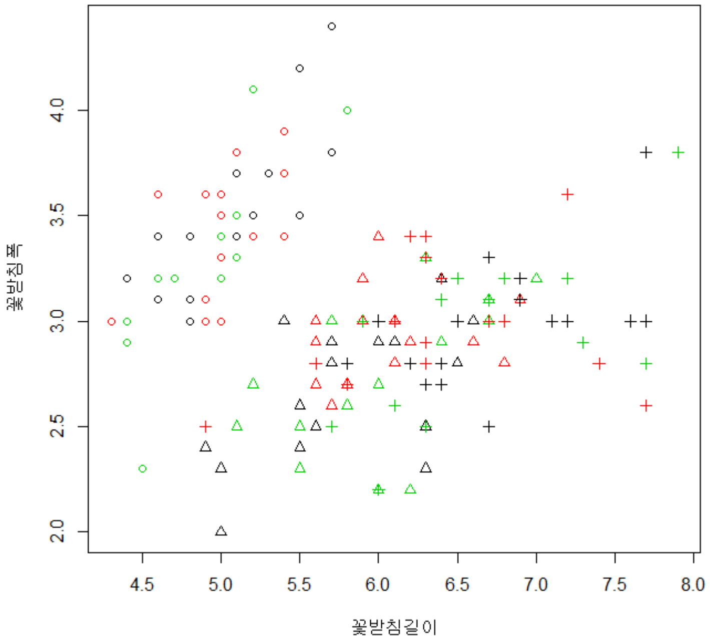
(1) X1과 X2에 의한 산점도를 구하고 두 변수가 연관성이 있는지를 살펴보라.

데이터를 살펴보면 아래와 같다.

```
> setwd("D:/2020 1학기 정호재/다변량통계학(1)/200421 다변량 실습2/Rdata")
> data1.3.4<-read.table("irisflower.txt", header=T)
> str(data1.3.4)
'data.frame':	150 obs. of  6 variables:
 $ 포기      : int  1 2 3 4 5 6 7 8 9 10 ...
 $ 꽃받침길이: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ 꽃받침폭  : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ 꽃잎길이  : num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ 꽃잎폭    : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ group     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
> dim(data1.3.4)
[1] 150   6
> colSums(is.na(data1.3.4))
      포기 꽃받침길이   꽃받침폭   꽃잎길이     꽃잎폭      group 
         0          0          0          0          0          0 
> summary(data1.3.4)
      포기          꽃받침길이       꽃받침폭        꽃잎길이    
 Min.   :  1.00   Min.   :4.300   Min.   :2.000   Min.   :1.000  
 1st Qu.: 38.25   1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600  
 Median : 75.50   Median :5.800   Median :3.000   Median :4.350  
 Mean   : 75.50   Mean   :5.843   Mean   :3.057   Mean   :3.758  
 3rd Qu.:112.75   3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100  
 Max.   :150.00   Max.   :7.900   Max.   :4.400   Max.   :6.900  
     꽃잎폭              group   
 Min.   :0.100   setosa    :50  
 1st Qu.:0.300   versicolor:50  
 Median :1.300   virginica :50  
 Mean   :1.199                  
 3rd Qu.:1.800                  
 Max.   :2.500                  
```

포기 변수는 분석에 필요 없으므로 제외시킨다.

```
> X <- data1.3.4[,-1]
> head(X)
  꽃받침길이 꽃받침폭 꽃잎길이 꽃잎폭  group
1        5.1      3.5      1.4    0.2 setosa
2        4.9      3.0      1.4    0.2 setosa
3        4.7      3.2      1.3    0.2 setosa
4        4.6      3.1      1.5    0.2 setosa
5        5.0      3.6      1.4    0.2 setosa
6        5.4      3.9      1.7    0.4 setosa
> plot(X[,1:2],pch=unclass(X[,5]), col=1:3)
```

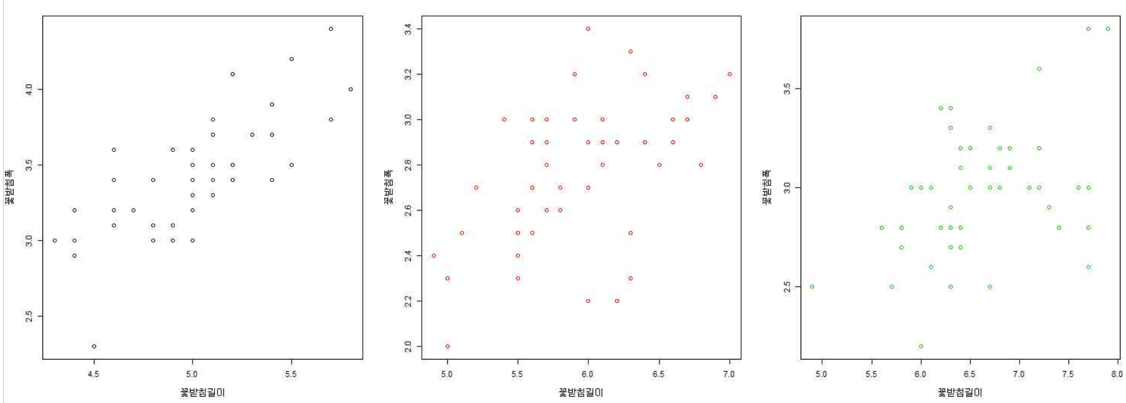


꽃받침 길이(X1)와 꽃받침 폭(X2)에 대한 산점도는 다음과 같고 산점도를 봤을 때 두 변수 사이에 양의 상관을 보이는 듯 하다.

(2) 세 종류의 붓꽃 군집을 (1)에서 구한 산점도를 통해 살펴보라.

group이 setosa versicolor virginica로 나뉘므로 group별로 산점도를 그려보면

```
> setosa=X[which(X$group=="setosa"),]
> versicolor=X[which(X$group=="versicolor"),]
> virginica=X[which(X$group=="virginica"),]
> par(mfrow=c(1,3))
> plot(setosa[,1:2],col=1)
> plot(versicolor[,1:2],col=2)
> plot(virginica[,1:2],col=3)
```



세 종류의 붓꽃 군집으로 나누었을 때 모두 양의 상관을 가지는 것처럼 보이고 setosa의 경우가 가장 분산이 작아 보인다. 하지만 각 산점도의 범위가 다르므로 정확하게 판단을 할 수가 없다.

(3) 붓꽃 군집별로 꽃받침 길이와 꽃잎 길이의 평균벡터와 공분산행렬, 상관행렬을 구하고 비교하라.

```
> X<-X[,c(1,3,5)]
> head(X)
  꽃받침길이 꽃잎길이  group
1          5.1        1.4 setosa
2          4.9        1.4 setosa
3          4.7        1.3 setosa
4          4.6        1.5 setosa
5          5.0        1.4 setosa
6          5.4        1.7 setosa
> setosa=X[which(X$group=="setosa"),]
> versicolor=X[which(X$group=="versicolor"),]
> virginica=X[which(X$group=="virginica"),]
> setosa_bar<-colMeans(setosa[,-3])
> setosa_S<-cov(setosa[,-3])
> setosa_R<-cor(setosa[,-3])
> versicolor_bar<-colMeans(versicolor[,-3])
> versicolor_S<-cov(versicolor[,-3])
> versicolor_R<-cor(versicolor[,-3])
> virginica_bar<-colMeans(virginica[,-3])
> virginica_S<-cov(virginica[,-3])
> virginica_R<-cor(virginica[,-3])
> cbind(setosa_bar, versicolor_bar, virginica_bar) # 평균벡터
           setosa_bar versicolor_bar virginica_bar
꽃받침길이      5.006          5.936         6.588
꽃잎길이        1.462          4.260         5.552
> cbind(setosa_S, versicolor_S, virginica_S)
           꽃받침길이   꽃잎길이 꽃받침길이  꽃잎길이 꽃받침길이  꽃잎길이
꽃받침길이  0.1242490 0.01635510  0.2664327 0.1828980  0.4043429 0.3032898
꽃잎길이    0.0163551 0.03015918  0.1828980 0.2208163  0.3032898 0.3045878
> cbind(setosa_R, versicolor_R, virginica_R) # 상관행렬
           꽃받침길이  꽃잎길이  꽃받침길이 꽃잎길이  꽃받침길이  꽃잎길이
꽃받침길이  1.0000000 0.2671758   1.000000 0.754049  1.0000000 0.8642247
꽃잎길이    0.2671758 1.0000000   0.754049 1.000000  0.8642247 1.0000000
```

평균벡터값의 크기를 살펴보면

꽃받침길이:setosa<versicolor<virginica, 꽃잎길이:setosa<versicolor<virginica 순서이다.

공분산행렬의 값에서 분산을 살펴보면

꽃받침길이:setosa<versicolor<virginica, 꽃잎길이:setosa<versicolor<virginica 순서이다.

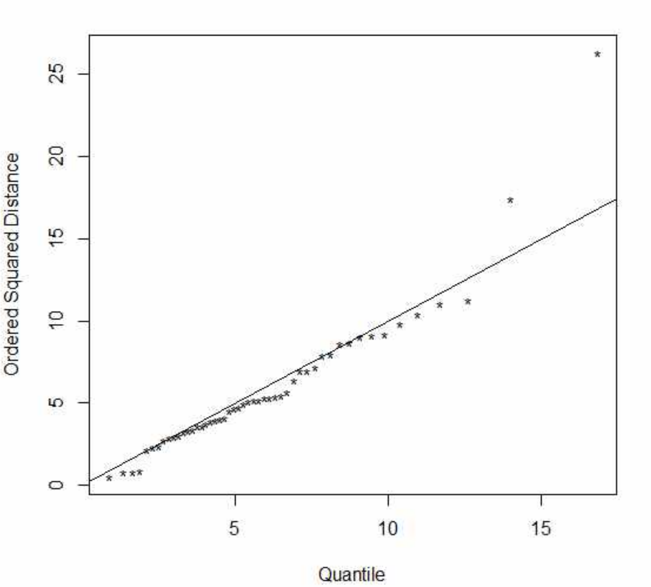
상관행렬의 값에서 상관계수를 살펴보면 모두 양의 상관을 가지고

상관계수의 크기는 setosa<versicolor<virginica 순서이다.

3. Evaluate the multivariate normality of [data 1.3.2](klpga.txt) in Exercise 1.7 by Chi-square Plot and Mardia test based on the skeweness and kurtosis.

(6) 다변량 정규성을 만족하는 지를 카이제곱그림과 왜도와 첨도에 의한 검정을 통해 검토하라.

```
> x<-X
> n<-dim(x)[1]
> p<-dim(x)[2]
> S<-cov(x)
> xbar<-colMeans(x)
> m<-mahalanobis(x, xbar, S)
> m<-sort(m)
> id<-seq(1, n)
> pt<-(id-0.5)/n
> q<-qchisq(pt, p)
> plot(q, m, pch="*", xlab="Quantile", ylab="Ordered Squared Distance")
> abline(0, 1)
```

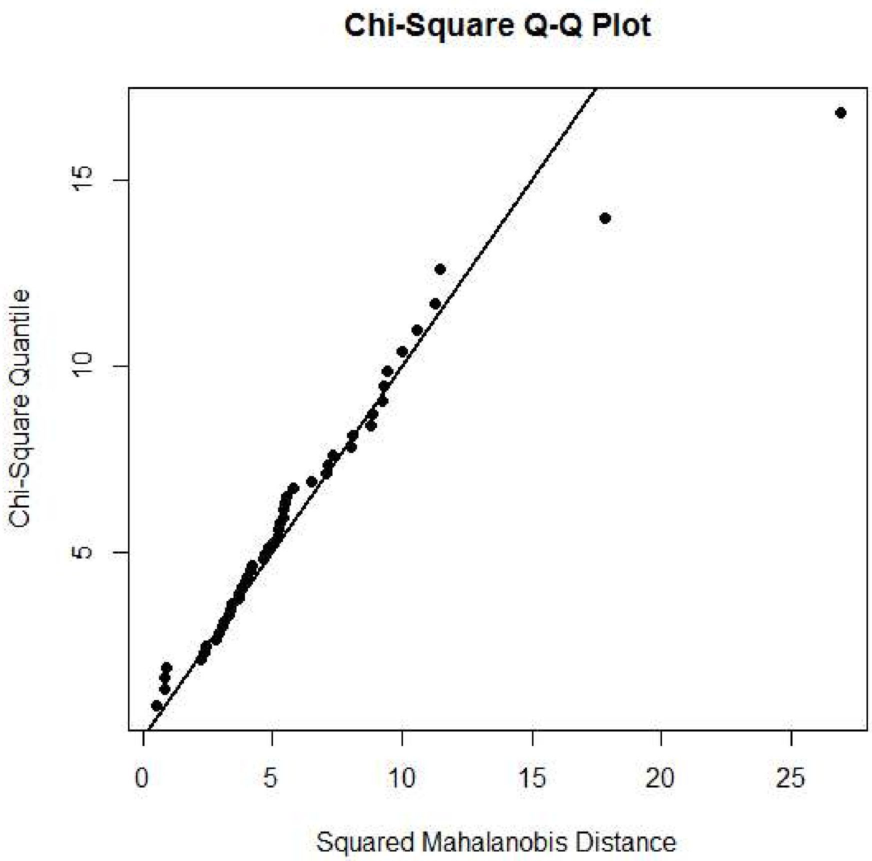


이 그림의 수평 축은 카이제곱분포의 분위수에 해당하고 수직 축은 마할라노비스 거리를 나타낸다. 원점에서 멀어질수록 좌표점이 직선성을 크게 벗어나려고 하고 있다.

```
> rq<-cor(cbind(q, m))[1,2]
> rq
[1] 0.9523092
```

마할라노비스 거리의 상관계수를 구하면 0.9523092으로 1에 가까운값을 가지나 크게 가까운 값은 아니므로 다변량 정규성을 만족하는지 알 수 없다. 따라서 왜도와 첨도에 의한 검정을 통하여 더 정확한 결과를 얻고자한다.

```
> result<-mvn(X, multivariatePlot = "qq")
```



이 그림의 세로축은 카이제곱분포의 분위수를 나타내며 가로축은 마할라노비스 거리를 나타낸다.

```
> result
$multivariateNormality
             Test        Statistic               p value Result
1 Mardia Skewness 134.895403452291  1.85539752395709e-08     NO
2 Mardia Kurtosis 2.74004656643207   0.00614304801363086     NO
3             MVN             <NA>                  <NA>     NO

$univariateNormality
          Test   Variable Statistic   p value Normality
1 Shapiro-Wilk  평균퍼팅수    0.9909  0.9647     YES
2 Shapiro-Wilk  그린적중율    0.9657  0.1542     YES
3 Shapiro-Wilk  파세이브율    0.9359  0.0093     NO
4 Shapiro-Wilk 파브레이크율   0.8796   1e-04     NO
5 Shapiro-Wilk   평균타수     0.9293  0.0052     NO
6 Shapiro-Wilk    상금율      0.6049  <0.001     NO
```

$Descriptives

	n	Mean	Std.Dev	Median	Min	Max	25th	75th
평균퍼팅수	50	31.3612	0.6385559	31.365	29.81	32.78	30.8725	31.7950
그린적중율	50	70.9236	4.6175834	70.190	61.30	82.72	68.2600	73.2700
파세이브율	50	81.2336	3.3436298	80.325	76.26	90.12	78.7825	82.9900
파브레이크율	50	15.2524	2.8726853	14.545	11.76	23.77	13.2400	16.7825
평균타수	50	73.0138	1.1334288	73.410	69.58	74.73	72.2625	73.7875
상금율	50	16.7220	18.0782459	9.000	5.10	100.00	7.5250	16.3000

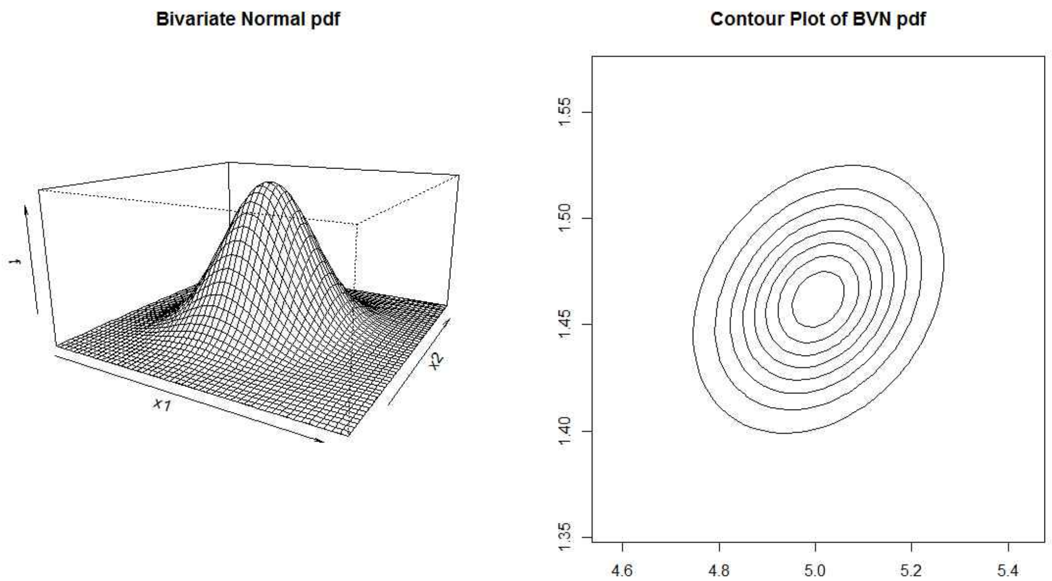
	Skew	Kurtosis
평균퍼팅수	0.04247991	-0.42226125
그린적중율	0.42261179	-0.07071193
파세이브율	0.78060138	-0.12710888
파브레이크율	1.23301980	1.14803846
평균타수	-0.92163471	0.35625579
상금율	2.72993401	7.96346992

multivariateNormality 값을 살펴보면 Mardia Skewness, Mardia Kurtosis의 result가 NO이므로 (alpha=0.05) 귀무가설을 기각한다. 따라서, 다변량정규성을 만족하지 않는다.

4. Solve the problems (4) by using the results of (3) in Exercise 1.10.

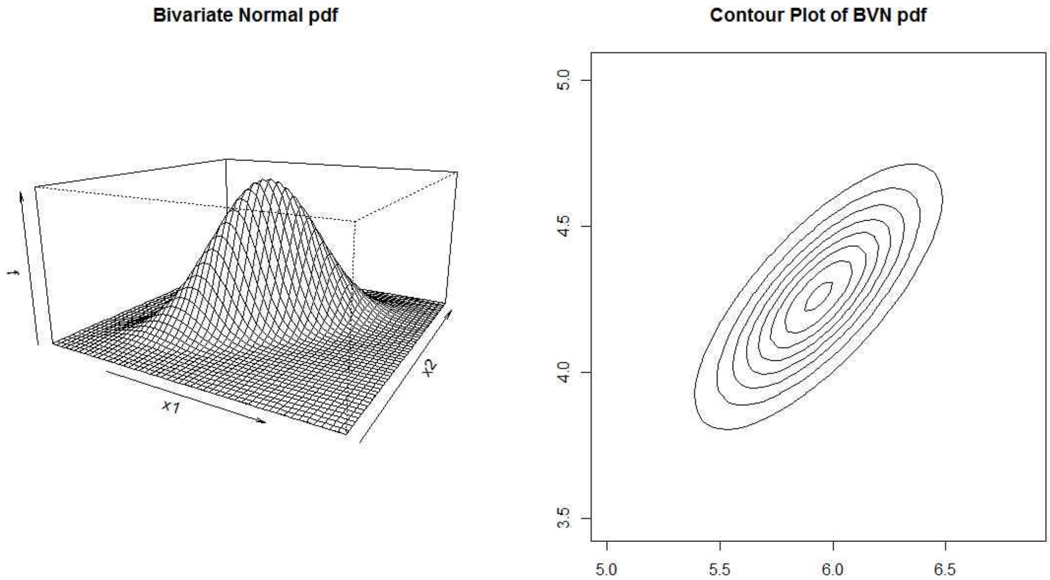
(4) 각 붓꽃 군집별로 (3)에서 구한 정보를 가지고 이변량 정규분포의 밀도함수를 추정하고 그려보라. 단, [R-코드 1.6.1](BVNpdf.R)을 활용하라.

```
> BVNpdf <- function(mu1,mu2,sig1,sig2,rho) {
+    par(mfrow=c(1,2))
+    s12 = sig1*sig2*rho
+    s11 = sig1^2
+    s22 = sig2^2
+    Sig <- matrix(c(s11,s12,s12,s22),ncol=2,nrow=2,byrow=T)
+    Sinv <- solve(Sig)
+    x1 <- seq(mu1 - 3.5*sig1,mu1+3.5*sig1,len=50)
+    fx1 <- seq(-3.5,3.5,len=50)
+    x2 <- seq(mu2 - 3.5*sig2,mu2+3.5*sig2,len=50)
+    fx2 <- seq(-3.5,3.5,len=50)
+    f <- function(x1,x2) {
+       cons <- ((2*pi)*det(Sig)^.5)^{-1}
+       cons*exp(-(.5*(1 - rho^2)^{-1})*(x1^2+x2^2-2*rho*x1*x2))
+    }
+    f <- outer(fx1,fx2,f)
+    persp(x1,x2,f,theta = 30, expand=.50)
+    title(main="Bivariate Normal pdf")
+    contour(x1,x2,f,lty="solid",drawlabels=F)
+    title(main="Contour Plot of BVN pdf")
+ }
> BVNpdf(setosa_bar[1],setosa_bar[2],setosa_S[1,1],setosa_S[2,2],setosa_R[1,2])
```



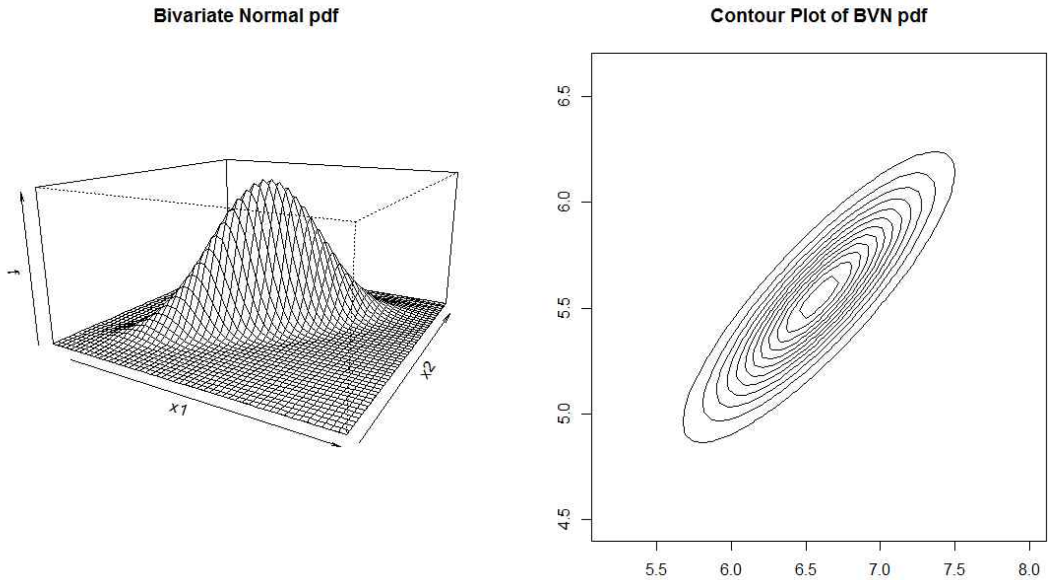
setosa에서의 이변량 정규분포의 확률밀도함수는 왼쪽그림과 같고 오른쪽에 등고선 그래프를 살펴보면 두 변수사이에 약한 양의 상관관계를 가지고 있음을 알 수 있다.

```
> BVNpdf(versicolor_bar[1],versicolor_bar[2],versicolor_S[1,1],
+ versicolor_S[2,2],versicolor_R[1,2])
```



versicolor에서의 이변량 정규분포의 확률밀도함수는 왼쪽그림과 같고 오른쪽에 등고선 그래프를 살펴보면 두 변수사이에 강한 양의 상관관계를 가지고 있음을 알 수 있다.

```
> BVNpdf(virginica_bar[1],virginica_bar[2],virginica_S[1,1],
+ virginica_S[2,2],virginica_R[1,2])
```



virginica에서의 이변량 정규분포의 확률밀도함수는 왼쪽그림과 같고 오른쪽에 등고선 그래프를 살펴보면 두 변수사이에 강한 양의 상관관계를 가지고 있음을 알 수 있다.

군집별로 그래프를 보았을 때 virginica에서 가장 높은 양의 상관관계를 보였고 setosa에서 가장 약한 양의 상관관계를 보였다.

Notice

Send your HWs to your assistant via E-mail: skdltmxogjs@naver.com till noon, May 8