Module 3: Multivariate Normal and Related Distribution

Weixing Song

August 15, 2018

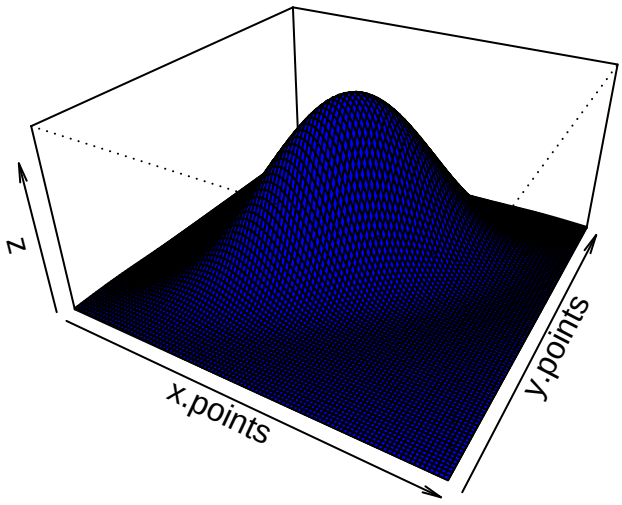
First, we use the following R-codes to install all necessary packages.

```
list.of.packages=c("mvtnorm","MVN")
if(length(which(!list.of.packages %in% installed.packages()))){
  install.packages(list.of.packages[!list.of.packages %in% installed.packages()])}
```

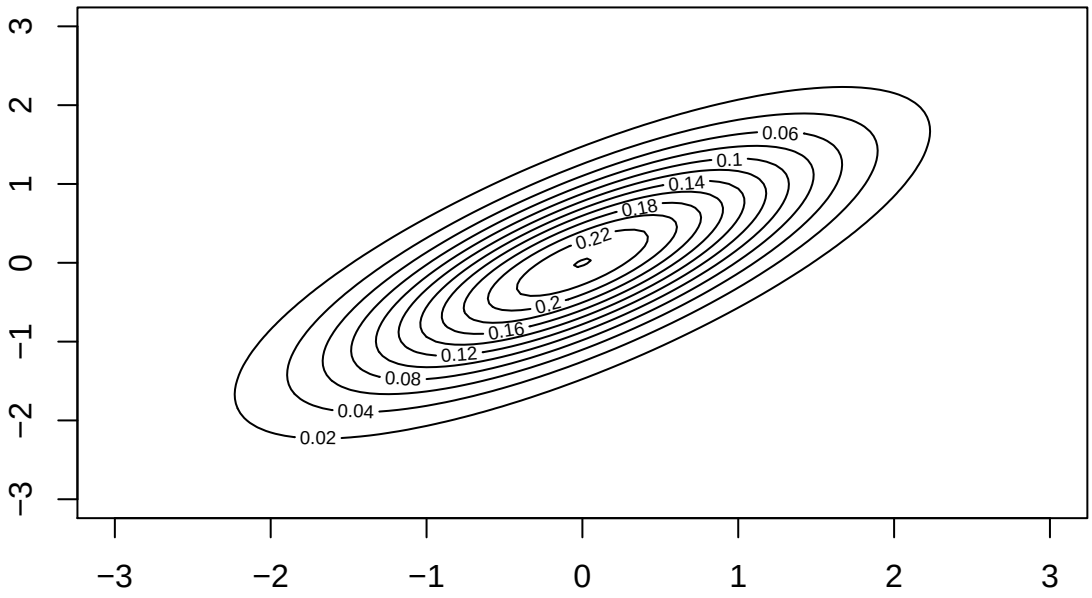
1. Bivariate Normal Density Function

The following R codes generate the surface/contour plots for a bivariate normal density function.

```
library(mvtnorm)

x.points=seq(-3,3,length.out=100)
y.points=x.points
z=matrix(0,nrow=100,ncol=100)
mu=c(0,0)
sigma=matrix(c(1,0.75,0.75,1),nrow=2)
for(i in 1:100)
  {
   for (j in 1:100)
     {
       z[i,j]=dmvnorm(c(x.points[i],y.points[j]),
                mean=mu,sigma=sigma)
     }
  }

# Surface plot of bivariate normal distribution.

persp(x.points, y.points, z,theta=30,phi=30,expand=0.5,col="blue")
```



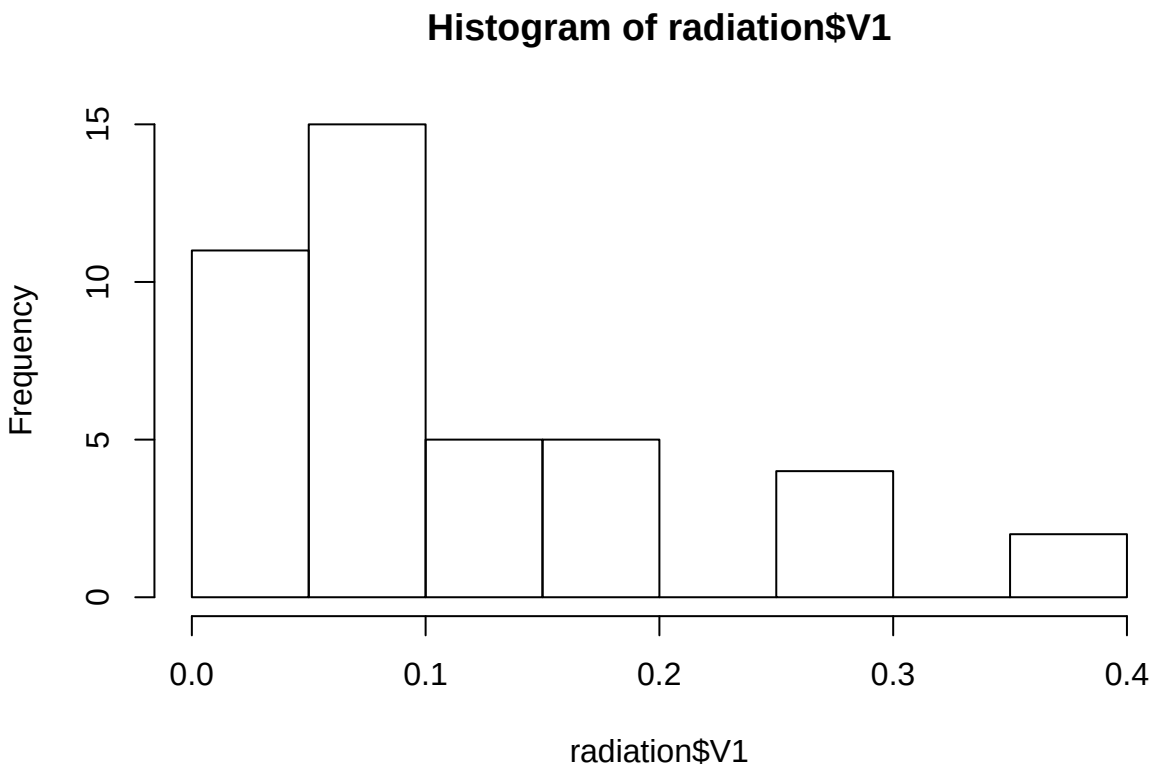

```
# Contour plot of bivariate normal distribution
contour(x=x.points, y=y.points, z=z)
```



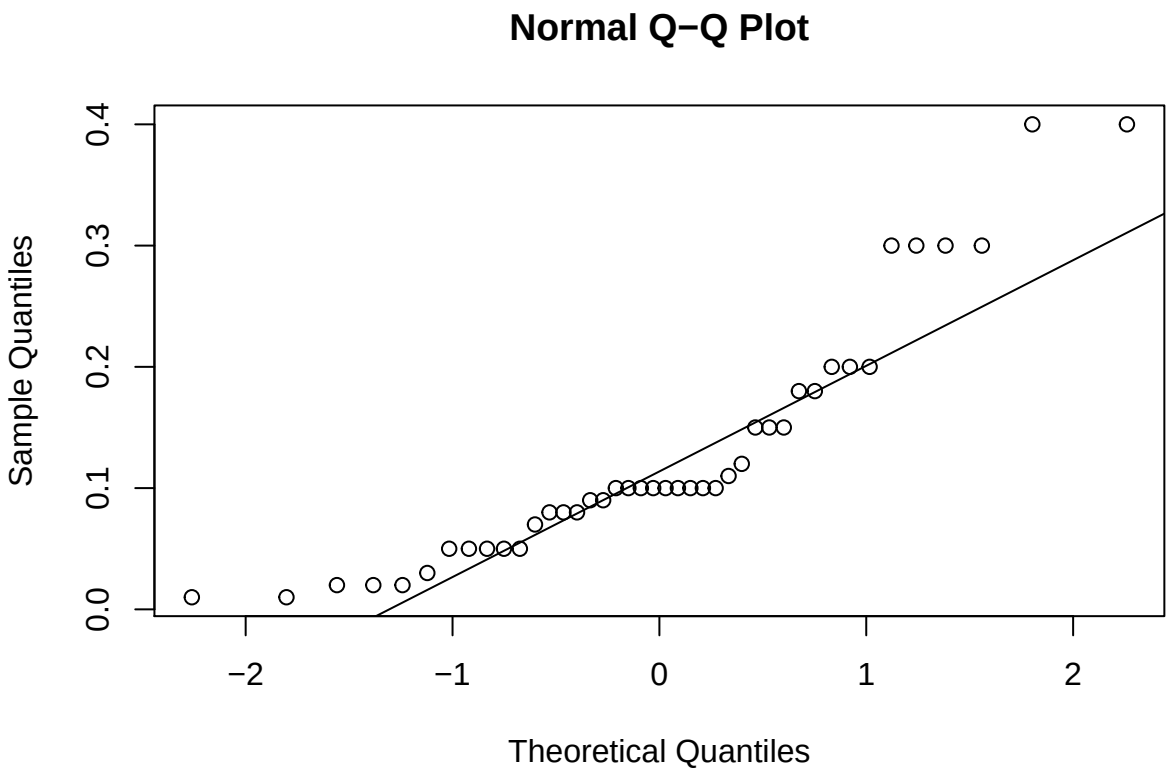
2. Checking Normality

Univariate Case: QQ-plot

```
myfile="S:/Workshop/Data/T4-1.DAT";
radiation=read.table(myfile)
hist(radiation$V1)  # histogram
```




```
qqnorm(radiation$V1) # qq-plot
qqline(radiation$V)  # draw a straight reference line
```



Multivariate Case: Chisquared QQ-plot

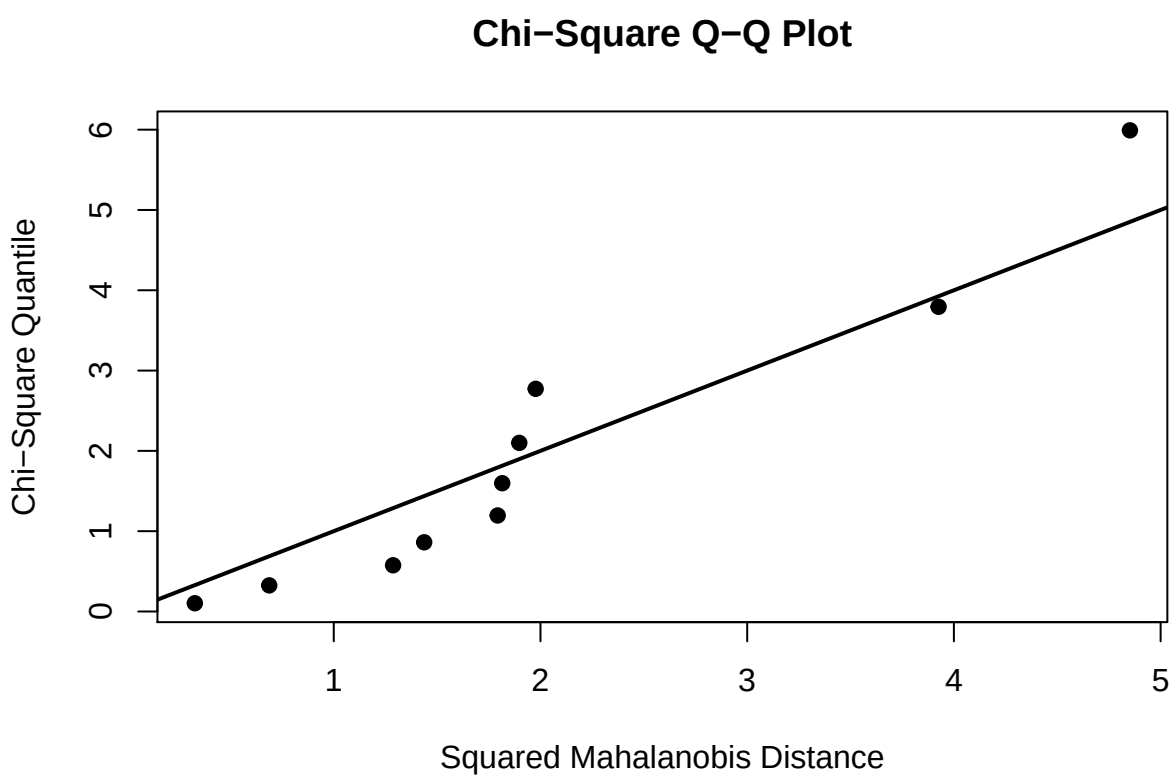
The R-package MVN can be used for this purpose. A reference can be found at https://cran.r-project.org/web/packages/MVN/vignettes/MVN.pdf

Example 1: The data set considered in this example contains some information of the world's 10 largest companies yield.

```
library(MVN)
```

```
## sROC 0.1-2 loaded
```

```
myfile="S:/Workshop/Data/P1-4.DAT";
mydata=read.table(myfile);
mydata=mydata[,1:2];

mvn(mydata,multivariatePlot="qq")
```




```
## $multivariateNormality
##              Test          Statistic           p value Result
## 1 Mardia Skewness   1.92106204524825 0.750274862568088    YES
## 2 Mardia Kurtosis -0.898406734719832 0.368968747315826    YES
## 3             MVN               <NA>              <NA>    YES
##
## $univariateNormality
##           Test  Variable Statistic   p value Normality
## 1 Shapiro-Wilk    V1        0.8591    0.0744    YES
## 2 Shapiro-Wilk    V2        0.9422    0.5778    YES
##
## $Descriptives
##     n    Mean   Std.Dev  Median   Min    Max    25th     75th      Skew
## V1 10 155.603 86.466486 130.320 62.97 285.06 92.7675 239.4125 0.4094674
## V2 10  14.704  5.117647  14.935  8.10  25.33 10.9650  16.9350 0.5589459
##      Kurtosis
## V1 -1.6803972
## V2 -0.6667657
```

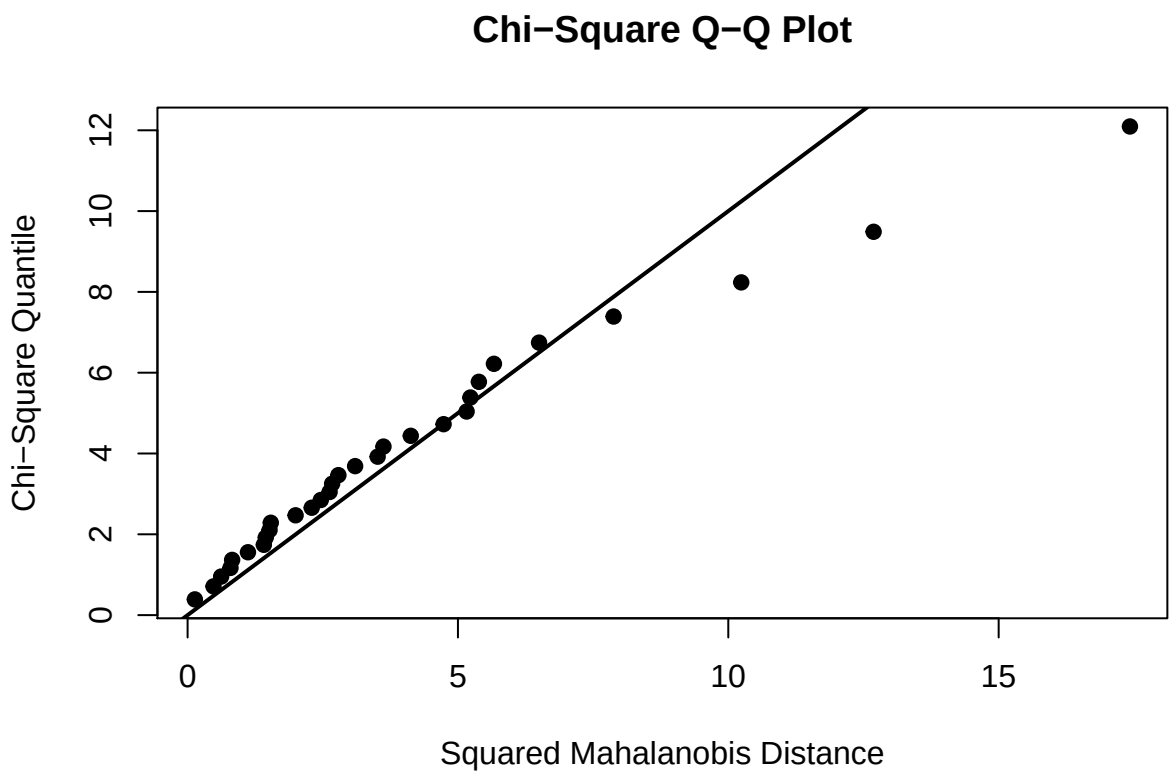
Example 2: The data considered in this example were obtained by taking four different measures of stiffness, x_1, x_2, x_3, and x_4, of each of $n = 30$ boards. The first measurement involves sending a shock wave down the board, the second measurement is determined while vibrating

the board, and the last two measurements are obtained from static tests. Generalized squared distances and Chi-square plot.

```r
library(MVN)
myfile="S:/Workshop/Data/t4-3.DAT";
mydata=as.matrix(read.table(myfile)[,1:4]);
# The following 4 lines codes calcualting d-squares
n=length(mydata[,1]);
Omatrix=matrix(1,nrow=30,ncol=30);
A=(mydata-Omatrix%*%mydata/n);
dsquare=diag(A%*%solve(cov(mydata))%*%t(A));
dsquare
```

```
##  [1]  0.6000129  5.4770196  7.6166439  5.2076098  1.3980776  2.2191409
##  [7]  4.9883498  1.4876570 12.2647550  0.7665400  1.9307771  0.4635159
## [13]  2.6959024  0.1295714  1.0792484 16.8474070  3.5018290  3.9900603
## [19]  1.3632124  1.4649908  9.8980384  5.0557446  0.7962096  2.5385575
## [25]  4.5767867  3.3979804  2.3816050  2.9951752  6.2837628  2.5838186
```

```r
mvn(mydata,multivariatePlot="qq") # Chi-square plot
```




```
## $multivariateNormality
##              Test        Statistic             p value Result
## 1 Mardia Skewness   37.68401850797 0.00967629675522986     NO
## 2 Mardia Kurtosis 2.53811341655708  0.0111451859079121     NO
## 3             MVN             <NA>                 <NA>     NO
##
## $univariateNormality
##           Test  Variable Statistic   p value Normality
## 1 Shapiro-Wilk    V1        0.9307    0.0512    YES
```

```
## 2 Shapiro-Wilk    V2        0.9127    0.0175    NO
## 3 Shapiro-Wilk    V3        0.9326    0.0575    YES
## 4 Shapiro-Wilk    V4        0.9613    0.3337    YES
##
## $Descriptives
##     n     Mean  Std.Dev Median  Min  Max    25th    75th      Skew
## V1 30 1906.100 324.9866 1863.0 1325 2983 1715.25 2057.25 1.0380842
## V2 30 1749.533 318.6065 1680.0 1170 2794 1595.50 1888.75 1.1435912
## V3 30 1509.133 303.1783 1466.0 1002 2412 1295.75 1623.75 0.9800274
## V4 30 1724.967 322.8436 1674.5 1176 2581 1520.25 1880.75 0.5978431
##       Kurtosis
## V1  2.03586397
## V2  1.94986381
## V3  0.99683699
## V4 -0.04626509
```