Module 2-2: Multivariate Data Summarization

Weixing Song

August 15, 2018

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

```
list.of.packages=c("plot3D","plot3Drgl","scatterplot3d","rgl")
if(length(which(!list.of.packages %in% installed.packages()))){
  install.packages(list.of.packages[!list.of.packages %in% installed.packages()])}
```

1. Sample Mean, Covariance, Correlation Coefficient

```
# Read data file into R
myfile="S:/Workshop/Data/T1-3.DAT";
A=read.table(myfile)   #data.frame
B=as.matrix(A)
class(A)
```

```
## [1] "data.frame"
```

```
scan(myfile)  # Read data into a vector or list from the console or file.
```

```
##  [1]   5.526  59.000 113.500  10.401  75.000 142.000   9.213  69.000
##  [9] 124.000   8.953  67.500 125.000   7.063  62.000 129.500   6.610
## [17]  62.000 123.000  11.273  74.000 140.000   2.447  47.000  97.000
## [25]  15.493  86.500 162.000   9.004  69.000 126.500   8.199  70.500
## [33] 136.000   6.601  64.500 116.000   7.622  67.500 135.000  10.067
## [41]  73.000 136.500  10.091  73.000 135.500  10.888  77.000 139.000
## [49]   7.610  61.500 118.000   7.733  66.500 133.500  12.015  79.500
## [57] 150.000  10.049  74.000 137.000   5.149  59.500 116.000   9.158
## [65]  68.000 123.000  12.132  75.000 141.000   6.978  66.500 117.000
## [73]   6.890  63.000 117.000
```

```
matrix(scan(myfile), ncol=3)
```

```
##          [,1]    [,2]    [,3]
##  [1,]   5.526  86.500 118.000
##  [2,]  59.000 162.000   7.733
##  [3,] 113.500   9.004  66.500
##  [4,]  10.401  69.000 133.500
##  [5,]  75.000 126.500  12.015
##  [6,] 142.000   8.199  79.500
##  [7,]   9.213  70.500 150.000
##  [8,]  69.000 136.000  10.049
##  [9,] 124.000   6.601  74.000
## [10,]   8.953  64.500 137.000
## [11,]  67.500 116.000   5.149
## [12,] 125.000   7.622  59.500
## [13,]   7.063  67.500 116.000
## [14,]  62.000 135.000   9.158
## [15,] 129.500  10.067  68.000
```

```
## [16,]   6.610  73.000 123.000
## [17,]  62.000 136.500  12.132
## [18,] 123.000  10.091  75.000
## [19,]  11.273  73.000 141.000
## [20,]  74.000 135.500   6.978
## [21,] 140.000  10.888  66.500
## [22,]   2.447  77.000 117.000
## [23,]  47.000 139.000   6.890
## [24,]  97.000   7.610  63.000
## [25,]  15.493  61.500 117.000
```

```
B=matrix(scan(myfile), ncol=3, byrow=T)
class(B)
```

```
## [1] "matrix"
```

```
# Sample Mean, Covariance, Correlation Matrix
mu=apply(A,2,mean)  # sample mean
mu=colMeans(A);
S=cov(A);    # sample variance with denominator n-1
Sn=cov(A)*24/25; #  sample variance with denominator n
Sn=cov(A)*(dim(A)[1]-1)/(dim(A)[1])
R=cor(A)
```

2. Generalized Variance

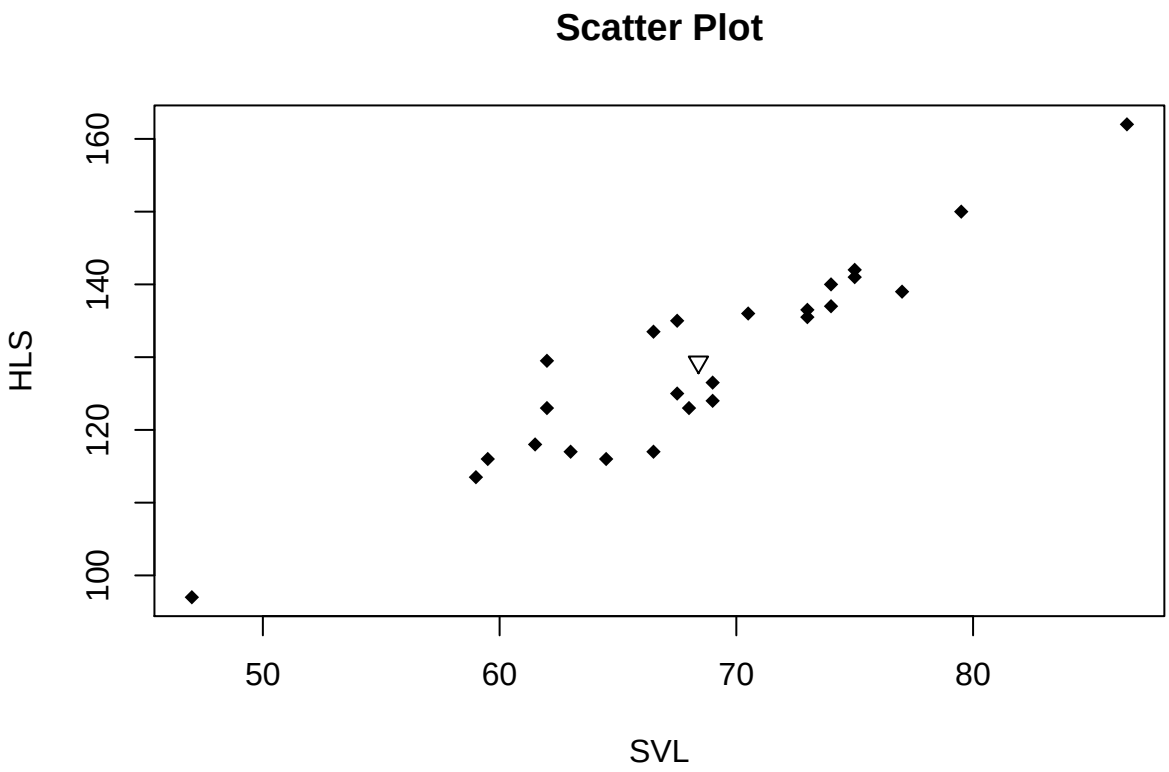
```
gva1=det(S)
gva2=sum(diag(S))
(gva2==sum(eigen(S)$values))
```

```
## [1] FALSE
```

3. Scatter Plot

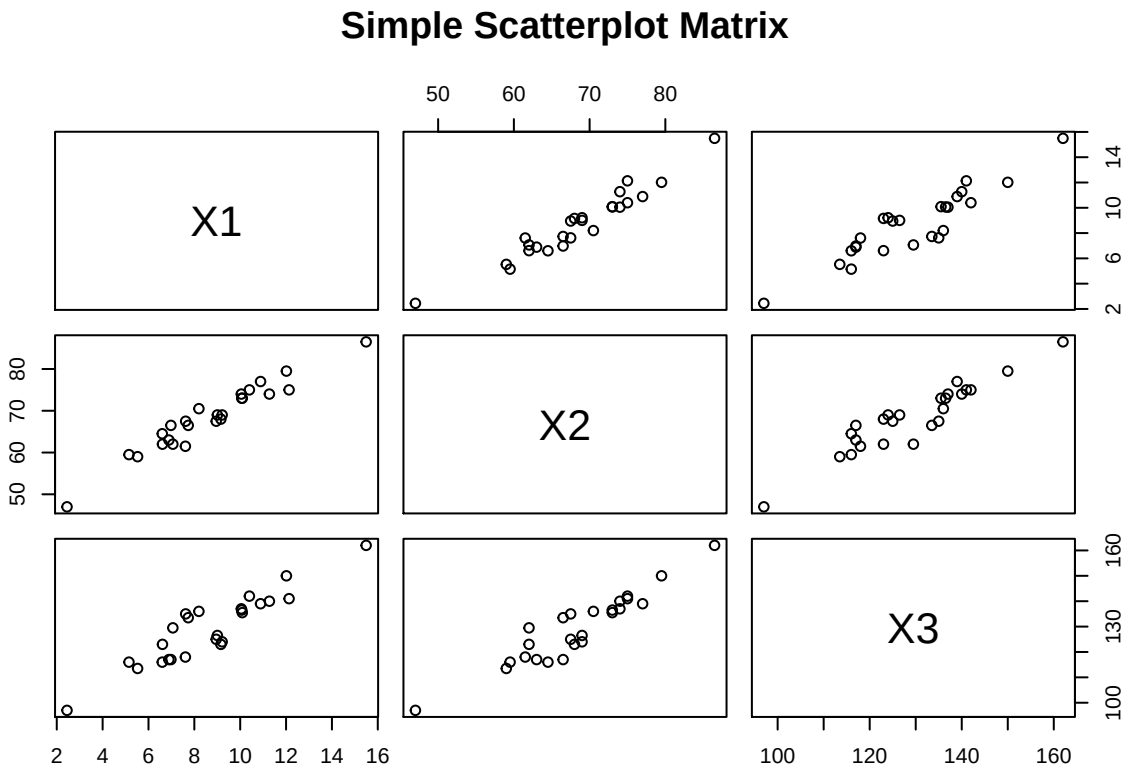
2D scatter plot

```
   X1=A[,1];
   X2=A[,2];
   X3=A[,3];
   plot(X2,X3,main="Scatter Plot", xlab="SVL",ylab="HLS",pch=18);
   points(mean(X2),mean(X3),pch=25)
```



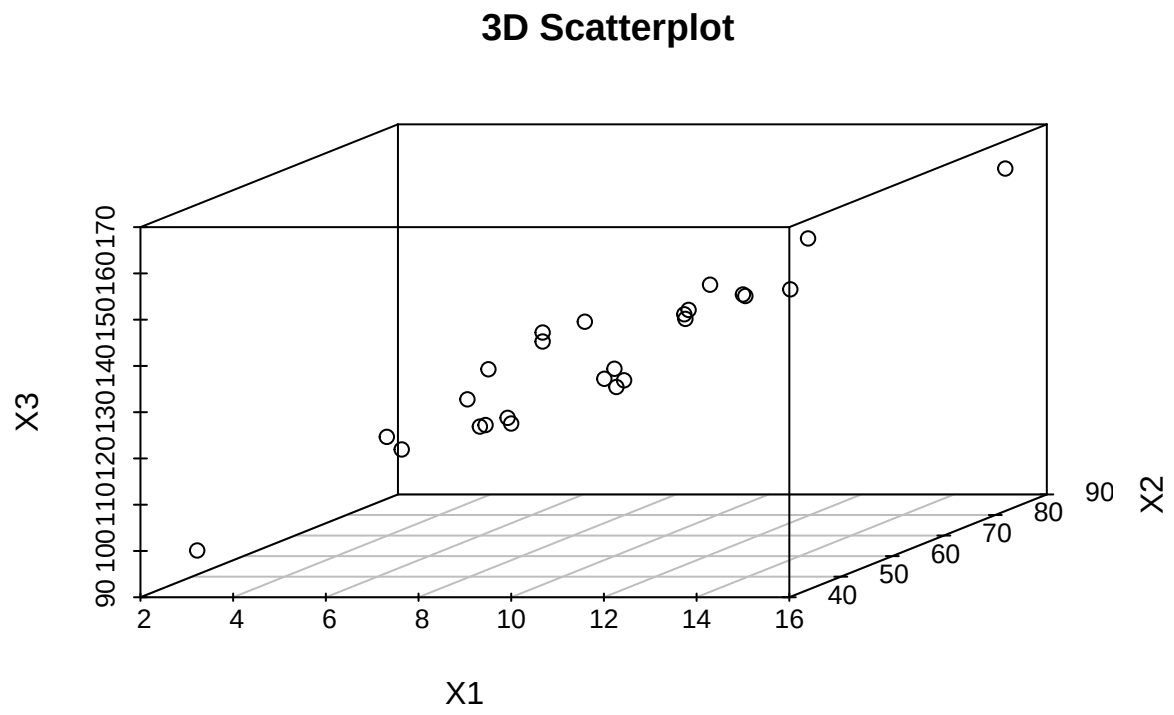
2D Pairwise scatter plot / Scatter plots matrix

```
pairs(~X1+X2+X3, main="Simple Scatterplot Matrix")
```



3D Scatterplots

```
library(scatterplot3d)
scatterplot3d(X1,X2,X3, main="3D Scatterplot")
```



Dynamic 3D Scatter Plot

```
library(rgl)
plot3d(X1,X2,X3, col="red", size=3)
rgl.postscript("myplot.pdf","pdf")    # Save the plot as pdf file
browseURL(paste("file://", writeWebGL(dir=file.path(tempdir(), "webGL"), width=500),
                sep=""))  # Interactive plot
```

4. 3D Vector Plot

3D vector from origin

```
library(plot3D)
library(plot3Drgl)

x0=c(0,0,1);
y0=c(0,0,2);
z0=c(0,0,3);
x1=c(1,1,1);
y1=c(2,4.5,4.5);
z1=c(3,5,5);
```

```
arrows3D(x0,y0,z0,x1,y1,z1, theta=20, phi=20, colvar=1:3, main ="Vector Plot",
           colkey=FALSE, ticktype="detailed");
  # Add starting point of arrow
points3D(x0, y0, z0, add = TRUE, col="darkred", colkey = FALSE, pch = 19,
         cex = 1,xlim=c(0,1.2),ylim=c(0,5),zlim=c(0,5.5));
  # Add labels to the arrows
text3D(x1, y1, z1, c("y1", "y2","y2-y1"), add=TRUE, colkey = FALSE)
```