Discussion 2 (STA 135)

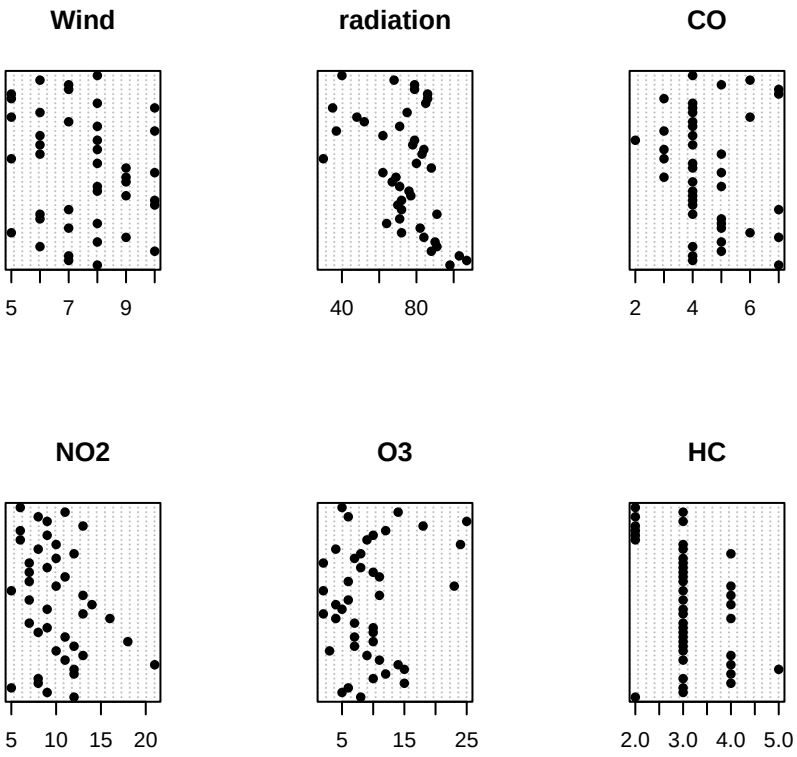
4/5/2022

The air-pollution data set.

```
# Read the data
air_pollution = read.table('T1-5.dat') # Please set your working directory as your file location
# Check the data
colnames(air_pollution) = c('Wind', 'radiation', 'CO', 'NO', 'NO2','O3', 'HC')
head(air_pollution)
```

```
##   Wind radiation CO NO NO2 O3 HC
## 1    8        98  7  2  12  8  2
## 2    7       107  4  3   9  5  3
## 3    7       103  4  3   5  6  3
## 4   10        88  5  2   8 15  4
## 5    6        91  4  2   8 10  3
## 6    8        90  5  2  12 12  4
```

```
# Dot plot
par(mfrow = c(2,4))
for (i in 1:ncol(air_pollution)){
  dotchart(air_pollution[,i],
           pch  = 16,
           main = colnames(air_pollution)[i])
}
```




```
# Calculate the quantities Sn, xbar and R
Sn = matrix(NA, nrow = ncol(air_pollution), ncol= ncol(air_pollution),dimnames = list(colnames(air_pollu
n  = nrow(air_pollution)
xbar = colMeans(air_pollution)
for (i in 1:ncol(air_pollution)){
  for (j in 1:ncol(air_pollution)){
    Sn[i,j] = sum((air_pollution[,i] - xbar[i])*(air_pollution[,j] - xbar[j]))/n
  }
}
var = diag(Sn)
R   = t(Sn / sqrt(var)) / sqrt(var)
cat('The covariance matrix Sn is \n')
```

```
## The covariance matrix Sn is
```

```
print(round(Sn,2))
```

```
##             Wind radiation    CO    NO   NO2    O3   HC
## Wind        2.44     -2.71 -0.37 -0.45 -0.57 -2.18 0.17
## radiation -2.71    293.36  3.82 -1.35  6.60 30.06 0.61
## CO        -0.37      3.82  1.49  0.66  2.26  2.75 0.14
## NO        -0.45     -1.35  0.66  1.15  1.06 -0.79 0.17
## NO2       -0.57      6.60  2.26  1.06 11.09  3.05 1.02
## O3        -2.18     30.06  2.75 -0.79  3.05 30.24 0.58
## HC         0.17      0.61  0.14  0.17  1.02  0.58 0.47
```

```
cat('The mean vector xbar is \n')
```

```
## The mean vector xbar is
```

```
print(round(xbar,2))
```

```
##      Wind radiation        CO        NO       NO2        O3        HC
##      7.50     73.86      4.55      2.19     10.05      9.40      3.10
```

```
cat('The correlation matrix R is \n')
```

```
## The correlation matrix R is
```

```
print(round(R,2))
```

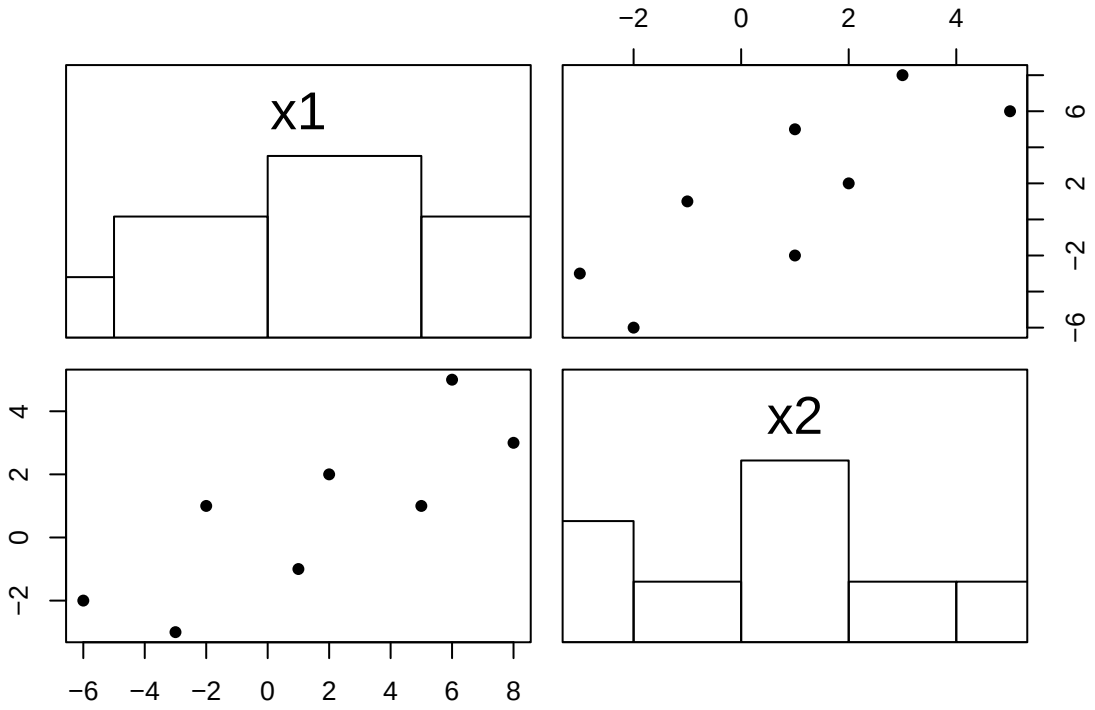
```
##             Wind radiation    CO    NO   NO2    O3   HC
## Wind        1.00     -0.10 -0.19 -0.27 -0.11 -0.25 0.16
## radiation -0.10      1.00  0.18 -0.07  0.12  0.32 0.05
## CO        -0.19      0.18  1.00  0.50  0.56  0.41 0.17
## NO        -0.27     -0.07  0.50  1.00  0.30 -0.13 0.23
## NO2       -0.11      0.12  0.56  0.30  1.00  0.17 0.45
## O3        -0.25      0.32  0.41 -0.13  0.17  1.00 0.15
## HC         0.16      0.05  0.17  0.23  0.45  0.15 1.00
```

A synthetic data example

```
# Create the data
x1   = c(-6, -3, -2,  1, 2, 5, 6, 8)
x2   = c(-2, -3,  1, -1, 2, 1, 5, 3)
data = cbind(x1, x2)
colnames(data) = c('x1','x2')
data
```

```
##      x1 x2
## [1,] -6 -2
## [2,] -3 -3
## [3,] -2  1
## [4,]  1 -1
## [5,]  2  2
## [6,]  5  1
## [7,]  6  5
## [8,]  8  3
```

```
# scatter plot
pairs(data, pch = 16, diag.panel = panel.hist)
```




```
# Calcuate statistics Sn, R, xbar
# Copied from 'E1-2.tex'
n        = length(x1)
xbar1    = sum(x1)/n
xbar2    = sum(x2)/n
s11      = sum((x1-xbar1)^2)/n
s22      = sum((x2-xbar2)^2)/n
s12      = sum((x1-xbar1)*(x2-xbar2))/n
Sn       = matrix( c(s11,s12,s12,s22),
                   nrow=2,ncol=2,
                   dimnames = list(c('x1','x2'), c('x1','x2')))
var      = diag(Sn)
R        = t(Sn / sqrt(var)) / sqrt(var)
xbar     = c(xbar1, xbar2)
cat('The covariance matrix Sn is \n')
```

```
## The covariance matrix Sn is
```

```
print(round(Sn,2))
```

```
##       x1   x2
## x1 20.48 9.09
## x2  9.09 6.19
```

```
cat('The mean vector xbar is \n')
```

```
## The mean vector xbar is
```

```
print(round(xbar,2))
```

```
## [1] 1.38 0.75
```

```
cat('The correlation matrix R is \n')
```

```
## The correlation matrix R is
```

```
print(round(R,2))
```

```
##      x1   x2
## x1 1.00 0.81
## x2 0.81 1.00
```

Table 1.9: National Track Record for Women Data

```
data = read.table('T1-9.dat', sep = '\t') # The text in the file was separated using '\t'
rownames(data) = data[,1]
data = data[,-1]
data[4:7] = data[4:7] * 60
colnames(data) = c('100m', '200m', '400m', '800m', '1500m', '3000m','Marathon')
head(data)
```

```
##      100m  200m  400m  800m 1500m 3000m Marathon
## ARG 11.57 22.94 52.50 123.0 255.0 551.4   9019.2
## AUS 11.12 22.23 48.63 118.8 241.2 517.8   8610.6
## AUT 11.15 22.70 50.62 116.4 243.0 526.8   9261.0
## BEL 11.14 22.48 51.45 118.2 244.8 529.2   8583.0
## BER 11.46 23.05 53.30 124.2 257.4 588.6  10450.8
## BRA 11.17 22.60 50.62 118.2 250.2 542.4   8844.6
```

```
# convert the data to be speed
data = sweep(1/data, 2,c(100,200,400,800,1500,3000, 42195), '*' )
# Calcuate statistics Sn, R, xbar
Sn   = matrix(NA, nrow = ncol(data), ncol= ncol(data),
              dimnames = list(colnames(data), colnames(data)))
n    = nrow(data)
xbar = colMeans(data)
for (i in 1:ncol(data)){
  for (j in 1:ncol(data)){
    Sn[i,j] = sum((data[,i] - xbar[i])*(data[,j] - xbar[j]))/n
  }
}
#calculate R
var = diag(Sn)
R   = t(Sn / sqrt(var)) / sqrt(var)
cat('The covariance matrix Sn is \n')
```

```
## The covariance matrix Sn is
```

```
print(round(Sn,2))
```

```
##          100m 200m 400m 800m 1500m 3000m Marathon
## 100m     0.09 0.09 0.09 0.06  0.08  0.09     0.08
## 200m     0.09 0.11 0.11 0.07  0.09  0.10     0.09
## 400m     0.09 0.11 0.14 0.08  0.09  0.11     0.10
## 800m     0.06 0.07 0.08 0.07  0.08  0.10     0.09
## 1500m    0.08 0.09 0.09 0.08  0.12  0.14     0.12
## 3000m    0.09 0.10 0.11 0.10  0.14  0.17     0.14
## Marathon 0.08 0.09 0.10 0.09  0.12  0.14     0.16
```

```
cat('The mean vector xbar is \n')
```

```
## The mean vector xbar is
```

```
print(round(xbar,2))
```

```
##     100m     200m     400m     800m    1500m    3000m Marathon
##     8.81     8.66     7.71     6.60     5.99     5.54     4.62
```

```
cat('The correlation matrix R is \n')
```

```
## The correlation matrix R is
```

```
print(round(R,2))
```

```
##          100m 200m 400m 800m 1500m 3000m Marathon
## 100m     1.00 0.94 0.87 0.80  0.78  0.73     0.66
## 200m     0.94 1.00 0.91 0.82  0.81  0.74     0.67
## 400m     0.87 0.91 1.00 0.80  0.73  0.69     0.67
## 800m     0.80 0.82 0.80 1.00  0.91  0.88     0.85
## 1500m    0.78 0.81 0.73 0.91  1.00  0.97     0.82
## 3000m    0.73 0.74 0.69 0.88  0.97  1.00     0.85
## Marathon 0.66 0.67 0.67 0.85  0.82  0.85     1.00
```