Tutorial 2

Exercise 1:

Commented [MKA1]: Exercise 3.1 page 165pdf

Given the data matrix

$$X = \begin{bmatrix} 9 & 1 \\ 5 & 3 \\ 1 & 2 \end{bmatrix}$$

a) **Graph the scatter plot in p = 2 dimensions, and locate the sample on your diagram.**
b) **Sketch the n = 3-dimentional representation of the data, and plot the deviation vectors $y_1 - \bar{x}_1 1$ and $y_2 - \bar{x}_2 1$**
c) **Sketch the deviation vectors in (b) ementaining from the origin. Calculate the length of these vectors and cosine the angle between them. Relate these quantities to S_n and R.**

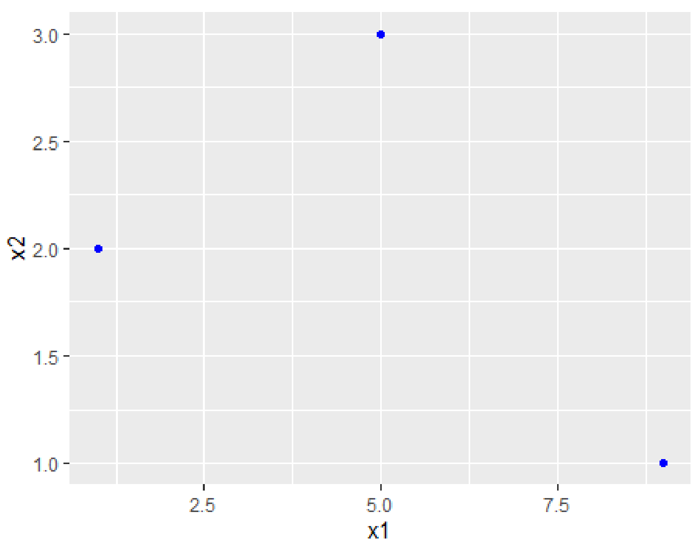
a) $\bar{x} = \begin{bmatrix} \frac{9+5+1}{3} \\ \frac{1+3+2}{3} \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$

R code

```
## Exercise 1
rm(list=ls())
data1 <- read.table("C:/Desktop/stat438/data1_ch3",header =
TRUE, row.names=NULL)

# MeanOfx
meanofx <- matrix(c(mean(x1),mean(x2)),nrow = 2, ncol = 1,
byrow = TRUE)
meanofx
     [,1]
[1,]    5
[2,]    2
#graph
```

```r
library(ggplot2)
ggplot(data1)+aes(x = x1, y = x2)+geom_point(colour = "blue")
```



Exercise 2:

Commented [MKA2]: Exercise 3.5 page 166pdf

Calculate the generalized sample variances |S| for (a) the data matrix X in Exercise 3.1 and (b) the data matrix X in Exercise 3.2

a) $\mathbf{X} = \begin{bmatrix} 9 & 1 \\ 5 & 3 \\ 1 & 2 \end{bmatrix}$

$S_{ik} = \frac{1}{n-1} \sum_{j=1}^{n} (x_{ji} - \bar{x})(x_{jk} - \bar{x})$

R code

```r
## Exercise 2
det(cov(data1))
[1] 12
```

b) $\mathbf{X} = \begin{bmatrix} 3 & 4 \\ 6 & -2 \\ 3 & 1 \end{bmatrix}$

```
data2 <- read.table("C:/Desktop/stat438/data2_ch3",header = TRUE,row.names = 1)

   det(cov(data2))
[1] 6.75
```

Exercise 3: Given

Commented [MKA3]: Exercise 3.8 page 166pdf

$$s = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad s = \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & 1 \end{bmatrix}$$

a) Calculate the total sample variances for each **S**. compare the results.

b) Calculate the generalizes sample variances for each **S**, and compare the results. Comment on the discrepancies, if any, found between Part a and Part b.

a) Total sample variances(S_1) = 1+1+1 = 3
Total sample variances(S_2) = 1+1+1 = 3
They have equal total sample variances

R code

```
   ## Exercise 3
s1 <- diag(c(1,1,1))
sum(diag(s1))
[1] 3
det(s1)
[1] 1

s2 <- matrix(c(1,-0.5,-0.5,-0.5,1,-0.5,-0.5,-0.5,1),nrow = 3)
sum(diag(s2))
[1] 3
det(s2)
[1] 0
```

They have different generalized sample variance

Although the value of the sum of variances was equal, the generalized sample variance values differed

Exercise 4:

Commented [MKA4]: Exercise 3.11 page 167pdf

Use the sample covariance obtained in Example 3.7 to verify (3-29) and (3-30), which state that $R = D^{-1/2}SD^{-1/2}$ and $D^{1/2}RD^{1/2} = S$

Commented [MKA5]: Example 3.7 page 145pdf Employees profit data

R code

```
## Exercise 4
rm(list=ls())
data1 <- read.table("data3_3_Profits per employee",header =
TRUE,row.names = 1)
s <- cov(data1)
R <- cor(data1)
vr <- diag(s)
(D = diag(vr))
       [,1]     [,2]
[1,] 251.434   0.0000
[2,]   0.000 123.6683

D1 <- sqrt(D)
D2 <- sqrt(solve(D))
D2%*%s%*%D2
          [,1]      [,2]
[1,]  1.000000 -0.381439
[2,] -0.381439  1.000000
R
           Employees   Profits
Employees   1.000000 -0.381439
Profits    -0.381439  1.000000

D1%*%R%*%D1
```

Commented [MKA6]: $D^{1/2}$

Commented [MKA7]: $D^{-1/2}$

```
          [,1]      [,2]
[1,] 251.43396 -67.26146
[2,] -67.26146 123.66829
S
          Employees   Profits
Employees 251.43396 -67.26146
Profits   -67.26146 123.66829
```

Exercise 5:

Commented [MKA8]: Exercise 3.14 page 167pdf

Consider the matrix X in Exercise 3.1. We have n = 3 observation on p =2 variables X1 and X2. Form the linear combinations:

$$c'X = \begin{bmatrix}-1 & 2\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} = -x_1 + 2x_2$$

$$b'X = \begin{bmatrix}2 & 3\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} = 2x_1 + 3x_2$$

$$\mathbf{X} = \begin{bmatrix}\mathbf{9} & \mathbf{1}\\\mathbf{5} & \mathbf{3}\\\mathbf{1} & \mathbf{2}\end{bmatrix}$$

a) Evaluate the sample means, variances, and covariance of $b'X$ and $c'X$ from first principles. That is, calculate the observed values of $b'X$ and $c'X$, and then use the sample mean, variance and covariance formulas.
b) Evaluate the sample means, variances, and covariance of $b'X$ and $c'X$ using (3-36). Compare the results in (a) and (b)

a) $\begin{bmatrix}x_{11} & x_{12}\\x_{21} & x_{22}\\x_{31} & x_{32}\end{bmatrix} = \begin{bmatrix}9 & 1\\5 & 3\\1 & 2\end{bmatrix}$

$$b'x_1 = 2x_{11} + 3x_{12} = 18 + 3 = 21$$
$$b'x_2 = 2x_{21} + 3x_{22} = 10 + 9 = 19$$
$$b'x_3 = 2x_{31} + 3x_{32} = 8$$

$$Sample\ mean\ = \frac{21 + 19 + 8}{3} = 16$$

$$sample\ variance = \frac{(21-16)^2 + (19-16)^2 + (8-16)^2}{3-1} = 49$$

$$c'x_1 = -9 + 2 = -7$$

$$c'x_2 = -5+6=1$$
$$c'x_3 = -1+4=3$$
$$sample\ mean = \frac{-7+1+3}{3} = -1$$
$$sample\ variance = \frac{\left(-7-(-1)\right)^2 + \left(1-(-1)\right)^2 + \left(3-(-1)\right)^2}{3-1} = 28$$

$$sample\ covariance = \frac{(21-16)(-7+1)+(19-16)(1+1)+(8-16)(3+1)}{3-1} = -28$$

b) $\bar{x} = \begin{bmatrix} \frac{9+5+1}{3} \\ \frac{1+3+2}{3} \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$

$$s = \begin{bmatrix} \frac{(9-5)^2+(5-5)^2+(1-5)^2}{3-1} & \frac{(9-5)(1-2)+(5-5)(3-2)+(1-5)(2-2)}{3-1} \\ \frac{(9-5)(1-2)+(5-5)(3-2)+(1-5)(2-2)}{3-1} & \frac{(1-2)^2+(3-2)^2+(2-2)^2}{3-1} \end{bmatrix} = \begin{bmatrix} 16 & -2 \\ -2 & 1 \end{bmatrix}$$

Sample mean of b'x = $b'\bar{x} = [2 \quad 3]\begin{bmatrix} 5 \\ 2 \end{bmatrix} = 10+6=16$

Sample mean of c'x = $c'\bar{x} = [-1 \quad 2]\begin{bmatrix} 5 \\ 2 \end{bmatrix} = -5+4 = -1$

Sample variance of b'x = $b'sb = [2 \quad 3]\begin{bmatrix} 16 & -2 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ 3 \end{bmatrix} =$

$[26 \quad -1]\begin{bmatrix} 2 \\ 3 \end{bmatrix} = 49$

Sample variance of c'x = $c'sc = [-1 \quad 2]\begin{bmatrix} 16 & -2 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} -1 \\ 2 \end{bmatrix} =$

$[-20 \quad 4]\begin{bmatrix} -1 \\ 2 \end{bmatrix} = 20+8=28$

sample covariance of b'x and c'x = $b' s\ c =$

$$[2 \quad 3]\begin{bmatrix}16 & -2\\ -2 & 1\end{bmatrix}\begin{bmatrix}-1\\ 2\end{bmatrix} = [26 \quad -1]\begin{bmatrix}-1\\ 2\end{bmatrix} = -26 - 2 = -28$$

R code

```
##Exercise 5
##a
rm(list=ls())
bt <- matrix(c(2,3),nrow =1,byrow = TRUE)
x1 <- matrix(c(9,1),nrow = 2)
x2 <- matrix(c(5,3),nrow = 2)
x3 <- matrix(c(1,2),nrow = 2)
samplemeanofbt <- (bt %*% x1 + bt %*% x2 + bt %*% x3)/3
samplemeanofbt
     [,1]
[1,]   16
samplevarianceofbt <- var(c(bt %*% x1,bt %*% x2,bt %*% x3))
samplevarianceofbt
[1] 49
 ct <- matrix(c(-1,2), nrow = 1 , byrow = TRUE)
 samplemeanofct <- (ct %*% x1 + ct %*% x2 + ct %*% x3)/3
 samplemeanofct
     [,1]
[1,]   -1
samplevarianceofct <- var(c(ct %*% x1,ct %*% x2,ct %*% x3))
samplevarianceofct
[1] 28
samplecovariance <- cov(c(bt %*% x1,bt %*% x2,bt %*% x3),c(ct
%*% x1,ct %*% x2,ct %*% x3))
samplecovariance
[1] -28

##b
```

```
x <- matrix(c(9,5,1,1,3,2), nrow = 3, ncol = 2, byrow = FALSE)
meanofx <- matrix(c(mean(x[ ,1]),mean(x[ ,2])), nrow = 2)
varx1 <- var(x[ ,1])
varx2 <- var(x[ ,2])
covx1x2 <- cov(x[ ,1],x[ ,2])
s <- matrix(c(varx1,covx1x2,covx1x2,varx2), nrow = 2, byrow =
TRUE)
> s
     [,1] [,2]
[1,]   16   -2
[2,]   -2    1
bt %*% meanofx
     [,1]
[1,]   16
ct %*% meanofx
     [,1]
[1,]   -1
bt %*% s %*% t(bt)
     [,1]
[1,]   49
ct %*% s %*% t(ct)
     [,1]
[1,]   28
bt %*% s %*% t(ct)
     [,1]
[1,]  -28
ct %*% s %*% t(bt)
     [,1]
[1,]  -28
```