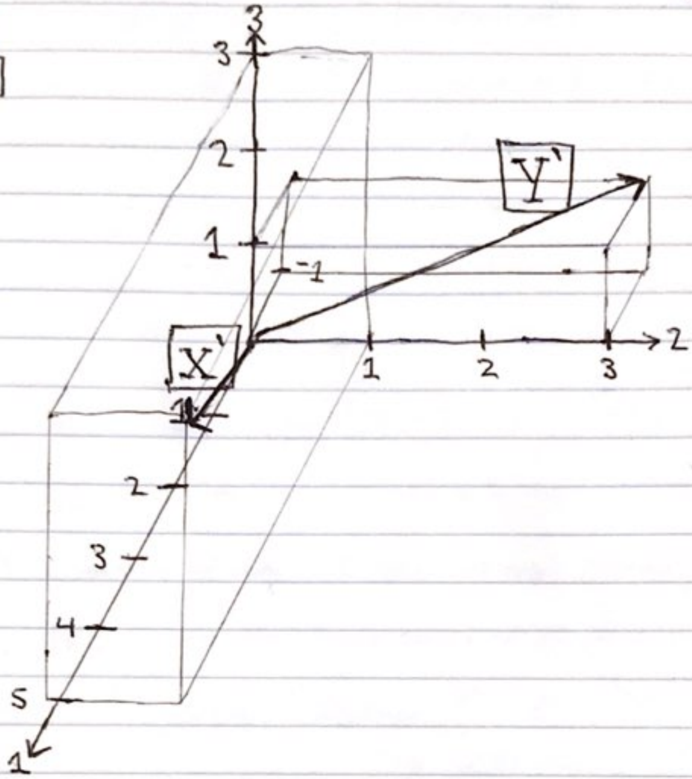
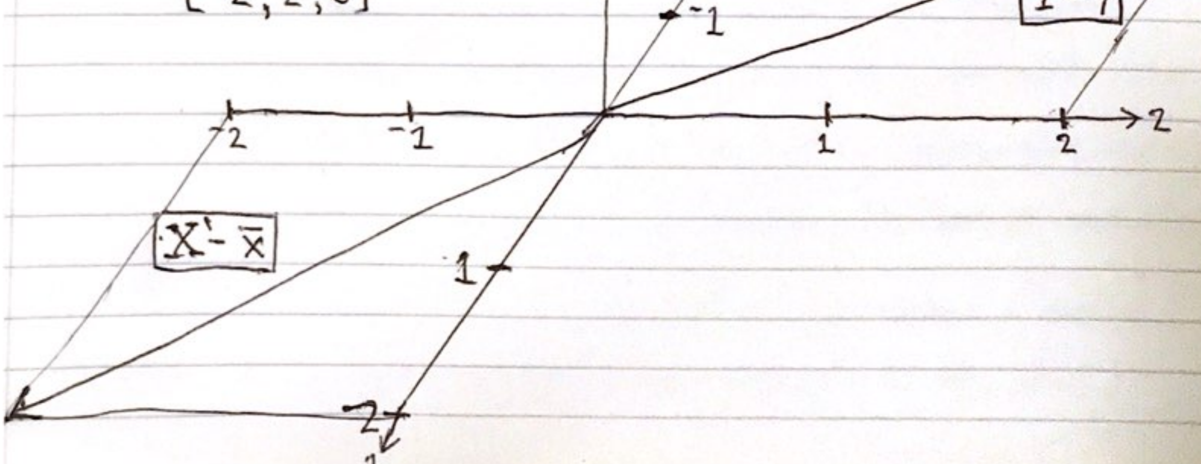
Charles Hwang
Dr. Whalen
STAT 488-001
21 February 2022

2.1A) $X' = [5, 1, 3]$
$Y' = [-1, 3, 1]$



2.1C) $\bar{x} = 3, \; \bar{y} = 1$
$X' - \bar{x} = [5-3, 1-3, 3-3] =$
$[2, -2, 0]$
$Y - \bar{y} = [-1-1, 3-1, 1-1] =$
$[-2, 2, 0]$



STAT 488: Multivariate Statistical Analysis — Homework 2

Charles Hwang

2/21/2022

Problem 2.1

See previous page for Problem 2.1(a).

```
rm(list=ls())
x<-matrix(c(5,1,3))
y<-matrix(c(-1,3,1))
sqrt(x[1]^2+x[2]^2+x[3]^2)                                                    # Problem 2.1(b)(i)
```

```
## [1] 5.91608
```

```
acos((x[1]*y[1]+x[2]*y[2]+x[3]*y[3])/(sqrt(sum(x^2))*sqrt(sum(y^2))))*180/pi # Problem 2.1(b)(ii)
```

```
## [1] 87.07867
```

```
as.numeric(t(y)%*%x/(x[1]^2+x[2]^2+x[3]^2))*x                                 # Problem 2.1(b)(iii)
```

```
##            [,1]
## [1,] 0.14285714
## [2,] 0.02857143
## [3,] 0.08571429
```

See previous page for Problem 2.1(c).

Problem 2.6

Problem 2.6(a)

```
matrix(c(9,-2,-2,6),nrow=2,ncol=2)==t(matrix(c(9,-2,-2,6),nrow=2,ncol=2))
```

```
##      [,1] [,2]
## [1,] TRUE TRUE
## [2,] TRUE TRUE
```

Yes, the matrix $\mathbf{A}$ is symmetric.

Problem 2.6(b)

$\mathbf{x}'\mathbf{A}\mathbf{x}$= $9x_1^2 - 4x_1x_2 + 6x_2^2 > 0$ (what we are trying to show)

$9x_1^2 - 4x_1x_2 + \frac{4}{9}x_2^2 + \frac{50}{9}x_2^2 > 0$

$(3x_1 - \frac{2}{3}x_2)^2 + \frac{50}{9}x_2^2 > 0$

$\frac{1}{9}(9x_1 - 2x_2)^2 + \frac{50}{9}x_2^2 > 0$

$(9x_1 - 2x_2)^2 + 50x_2^2 > 0$

$(9x_1 - 2x_2)^2 > 0 \mid 50x_2^2 > 0$. $\square$

Problem 2.7

Problem 2.7(a)

$|\mathbf{A}{-}\lambda\mathbf{I}| = 0$

$(9-\lambda)(6-\lambda)-(-2)(-2) = 0$

$\lambda^2 - 15\lambda + 50 = 0$

$(\lambda-5)(\lambda-10) = 0$

$\lambda = 5, 10$

$\mathbf{Ax}{=}\ \lambda\mathbf{x}$

$9x_1 - 2x_2 = 5x_1$

$-2x_1 + 6x_2 = 5x_2$

$2x_1 = x_2$

We can see there are an infinite number of solutions. If we arbitrarily pick $x_1 = 1$ and $x_2 = 2$, we can see that $\mathbf{e}' = [\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}] =$ [0.4472136, 0.8944272].

$9x_1 - 2x_2 = 10x_1$

$-2x_1 + 6x_2 = 10x_2$

$x_1 = -2x_2$

We can see there are an infinite number of solutions. If we arbitrarily pick $x_1 = 2$ and $x_2 = -1$, we can see that $\mathbf{e}' = [\frac{2}{\sqrt{5}}, \frac{-1}{\sqrt{5}}] =$ [0.8944272, -0.4472136].

Answers: $\lambda_1 = 5, \lambda_2 = 10$, $\mathbf{e}_1' = [\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}]$, $\mathbf{e}_2' = [\frac{2}{\sqrt{5}}, \frac{-1}{\sqrt{5}}]$

Problem 2.7(b)

$\mathbf{A}{=}\ \lambda_1\mathbf{e}_1\mathbf{e}_1' + \lambda_2\mathbf{e}_2\mathbf{e}_2' = 5\mathbf{e}_1\mathbf{e}_1' + 10\mathbf{e}_2\mathbf{e}_2'$

```
e1<-5*matrix(c(1/sqrt(5),2/sqrt(5)))%*%t(matrix(c(1/sqrt(5),2/sqrt(5))))
e1
```

```
##      [,1] [,2]
## [1,]    1    2
## [2,]    2    4
```

```
e2<-10*matrix(c(2/sqrt(5),-1/sqrt(5)))%*%t(matrix(c(2/sqrt(5),-1/sqrt(5))))
e2
```

```
##      [,1] [,2]
## [1,]    8   -4
## [2,]   -4    2
```

```
# We can see that the sum of these matrices is A.
solve(matrix(c(9,-2,-2,6),nrow=2,ncol=2)) # Problem 2.7(c)
```

```
##      [,1] [,2]
## [1,] 0.12 0.04
## [2,] 0.04 0.18
```

Problem 2.7(d)

$|\mathbf{A}^{-1} - \lambda \mathbf{I}| = 0$

$(0.12 - \lambda)(0.18 - \lambda) - (0.04)(0.04) = 0$

$\lambda^2 - 0.3\lambda + 0.02 = 0$

$(\lambda - 0.1)(\lambda - 0.2) = 0$

$\lambda = 0.1, 0.2$

$\mathbf{A}^{-1}\mathbf{x}= \lambda \mathbf{x}$

$0.12x_1 + 0.04x_2 = 0.1x_1$

$0.04x_1 + 0.18x_2 = 0.1x_2$

$0.02x_1 = -0.04x_2$

We can see there are an infinite number of solutions. If we arbitrarily pick $x_1 = 2$ and $x_2 = -1$, we can see that $\mathbf{e}' = [\frac{2}{\sqrt{5}}, \frac{-1}{\sqrt{5}}] =$ 0.8944272, -0.4472136.

$0.12x_1 + 0.04x_2 = 0.2x_1$

$0.04x_1 + 0.18x_2 = 0.2x_2$

$0.04x_1 = 0.02x_2$

We can see there are an infinite number of solutions. If we arbitrarily pick $x_1 = 1$ and $x_2 = 2$, we can see that $\mathbf{e}' = [\frac{2}{\sqrt{5}}, \frac{-1}{\sqrt{5}}] =$ 0.4472136, 0.8944272.

Answers: $\lambda_1 = 0.1, \lambda_2 = 0.2$, $\mathbf{e}'_1 = [\frac{2}{\sqrt{5}}, \frac{-1}{\sqrt{5}}]$, $\mathbf{e}'_2 = [\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}]$

Observe that the eigenvalues for $\mathbf{A}^{-1}$ are the multiplicative inverse/reciprocal ($\lambda^{-1} = \frac{1}{\lambda}$) of the eigenvalues for $\mathbf{A}$ and that the eigenvectors are for $\mathbf{A}^{-1}$ and $\mathbf{A}$ are the same.

Problem 2.15

$\mathbf{x}'\mathbf{A}\mathbf{x}= 3x_1^2 - 2x_1x_2 + 3x_2^2$

$3x_1^2 - 6x_1x_2 + 3x_2^2 + 4x_1x_2$

$3(x_1^2 - 2x_1x_2 + x_2^2) + 4x_1x_2$

$3(x_1 - x_2)^2 + 4x_1x_2$

$3(x_1 - x_2)^2 > 0 \mid 4x_1x_2 > 0.$

We can see that the quadratic form $3x_1^2 - 2x_1x_2 + 3x_2^2$ is definite positive.

Problem 2.25

```r
E <- matrix(c(25, -2, 4, -2, 4, 1, 4, 1, 9), nrow = 3, ncol = 3)  # Problem 2.25(a)
matrix(c(E[1, 1]/sqrt(E[1, 1] * E[1, 1]), E[1, 2]/sqrt(E[1, 1] * E[2, 2]), E[1, 3]/sqrt(E[1,
    1] * E[3, 3]), E[2, 1]/sqrt(E[2, 2] * E[1, 1]), E[2, 2]/sqrt(E[2, 2] * E[2, 2]),
    E[2, 3]/sqrt(E[2, 2] * E[3, 3]), E[3, 1]/sqrt(E[3, 3] * E[1, 1]), E[3, 2]/sqrt(E[3,
        3] * E[2, 2]), E[3, 3]/sqrt(E[3, 3] * E[3, 3])), nrow = 3, ncol = 3)
```

```
##            [,1]       [,2]      [,3]
## [1,]  1.0000000 -0.2000000 0.2666667
## [2,] -0.2000000  1.0000000 0.1666667
## [3,]  0.2666667  0.1666667 1.0000000
```

```r
matrix(c(sqrt(E[1, 1]), 0, 0, 0, sqrt(E[2, 2]), 0, 0, 0, sqrt(E[3, 3])), nrow = 3,
    ncol = 3)
```

```
##      [,1] [,2] [,3]
## [1,]    5    0    0
## [2,]    0    2    0
## [3,]    0    0    3
```

```r
matrix(c(sqrt(E[1, 1]), 0, 0, 0, sqrt(E[2, 2]), 0, 0, 0, sqrt(E[3, 3])), nrow = 3,
    ncol = 3) %*% matrix(c(E[1, 1]/sqrt(E[1, 1] * E[1, 1]), E[1, 2]/sqrt(E[1, 1] *
    E[2, 2]), E[1, 3]/sqrt(E[1, 1] * E[3, 3]), E[2, 1]/sqrt(E[2, 2] * E[1, 1]), E[2,
    2]/sqrt(E[2, 2] * E[2, 2]), E[2, 3]/sqrt(E[2, 2] * E[3, 3]), E[3, 1]/sqrt(E[3,
    3] * E[1, 1]), E[3, 2]/sqrt(E[3, 3] * E[2, 2]), E[3, 3]/sqrt(E[3, 3] * E[3, 3])),
    nrow = 3, ncol = 3) %*% matrix(c(sqrt(E[1, 1]), 0, 0, 0, sqrt(E[2, 2]), 0, 0, 0,
    sqrt(E[3, 3])), nrow = 3, ncol = 3)  # Problem 2.25(b)
```

```
##      [,1] [,2] [,3]
## [1,]   25   -2    4
## [2,]   -2    4    1
## [3,]    4    1    9
```

We can see that this is the same as the original matrix.

Problem 2.31

```r
matrix(c(4,3))                                                        # Problem 2.31(a)
```

```
##      [,1]
## [1,]    4
## [2,]    3
```

```r
as.numeric(matrix(c(1,-1),ncol=2)%*%matrix(c(4,3)))                   # Problem 2.31(b)
```

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

```r
matrix(c(3,0,0,1),nrow=2,ncol=2)                                      # Problem 2.31(c)
```

```
##      [,1] [,2]
## [1,]    3    0
## [2,]    0    1
```

```r
as.numeric(matrix(c(1,-1),ncol=2)%*%matrix(c(3,0,0,1),nrow=2,ncol=2)%*%t(matrix(c(1,-1),ncol=2)))#d
```

```
## [1] 4
```

```r
matrix(c(2,1))                                                        # Problem 2.31(e)
```

```
##      [,1]
## [1,]    2
## [2,]    1
```

```r
matrix(c(2,0,-1,1),nrow=2,ncol=2)%*%matrix(c(2,1))                    # Problem 2.31(f)
```

```
##      [,1]
## [1,]    3
## [2,]    1
```

```r
matrix(c(9,-2,-2,4),nrow=2,ncol=2)                                    # Problem 2.31(g)
```

```
##      [,1] [,2]
```

```
## [1,]    9   -2
## [2,]   -2    4
```

```
matrix(c(2,0,-1,1),nrow=2,ncol=2)%*%matrix(c(9,-2,-2,4),nrow=2,ncol=2)%*%t(matrix(c(2,0,-1,1),nrow=2,nc
```

```
##      [,1] [,2]
## [1,]   48   -8
## [2,]   -8    4
```

```
matrix(c(2,1,2,0),nrow=2,ncol=2)                                                      # Problem 2.31(i)
```

```
##      [,1] [,2]
## [1,]    2    2
## [2,]    1    0
```

```
matrix(c(1,-1),ncol=2)%*%matrix(c(2,1,2,0),nrow=2,ncol=2)%*%t(matrix(c(2,0,-1,1),nrow=2,ncol=2)) #j
```

```
##      [,1] [,2]
## [1,]    0    2
```

Problem 4.3

We can see that (b) X_2 and X_3, (c) (X_1, X_2) and X_3, and (d) $\frac{X_1+X_2}{2}$ and X_3 are independent because their respective covariances $\text{Cov}(X_2, X_3) = \frac{1}{2}\text{Cov}(X_1, X_3) + \frac{1}{2}\text{Cov}(X_2, X_3) = \text{Cov}(X_1, X_3) = 0$, which is a sufficient condition for independence from Result 4.5(a) on page 159 of the textbook.

We can see that (a) X_1 and X_2 and (e) X_2 and $X_2 - \frac{5}{2}X_1 - X_3$ are not independent because their respective covariances $\text{Cov}(X_1, X_2) = -2 \neq 0$ and $\text{Cov}(X_2, X_2 - \frac{5}{2}X_1 - X_3) = 59 \neq 0$.

Problem 4.4(a)

$\mu = E(3X_1 - 2X_2 + X_3) = 3E(X_1) - 2E(X_2) + E(X_3) = 3(2) - 2(-3) + (1) = 6 + 6 + (1) =$

$\mu = 13$

$\Sigma = \text{Var}(3X_1 - 2X_2 + X_3) =$

$\Sigma = a^2\text{Var}(X_1) + b^2\text{Var}(X_2) + c^2\text{Var}(X_3) + 2ab\text{Cov}(X_1, X_2) + 2ac\text{Cov}(X_1, X_3) + 2bc\text{Cov}(X_2, X_3) =$

$\Sigma = 3^2\text{Var}(X_1) + 2^2\text{Var}(X_2) + \text{Var}(X_3) + 2(3)(-2)\text{Cov}(X_1, X_2) + 2(3)(1)\text{Cov}(X_1, X_3) + 2(-2)(1)\text{Cov}(X_2, X_3) =$

$\Sigma = 9(1) + 4(3) + (2) - 12(1) + 6(1) - 4(2) = 9 + 12 + (2) - 12 + 6 - 8 =$

$\Sigma = 9$

We can see the distribution of $3X_1 - 2X_2 + X_3$ is $N_3(13, 9)$.

Problem 4.19

(a) χ^2_6 (from Result 4.7(a) on page 163 of the textbook)

(b) $N_{20}(\mu, \frac{1}{20}\Sigma)$ and $N_{20}(0, \Sigma)$ (from (4-23) 1. on page 174 of the textbook)

(c) $W_p(V, 19)$, where W is the Wishart distribution (from (4-23) 2. on page 174 of the textbook)

Problem 4.29

```
N <- read.table("/Users/newuser/Desktop/Notes/Graduate/STAT 488 - Multivariate Statistical Analysis/T1-5
O <- read.table("/Users/newuser/Desktop/Notes/Graduate/STAT 488 - Multivariate Statistical Analysis/T1-5
m <- matrix(c(mean(N), mean(O)))  # Problem 4.29(a)
S <- solve(matrix(c(cov(N, N), cov(N, O), cov(O, N), cov(O, O)), nrow = 2,
```

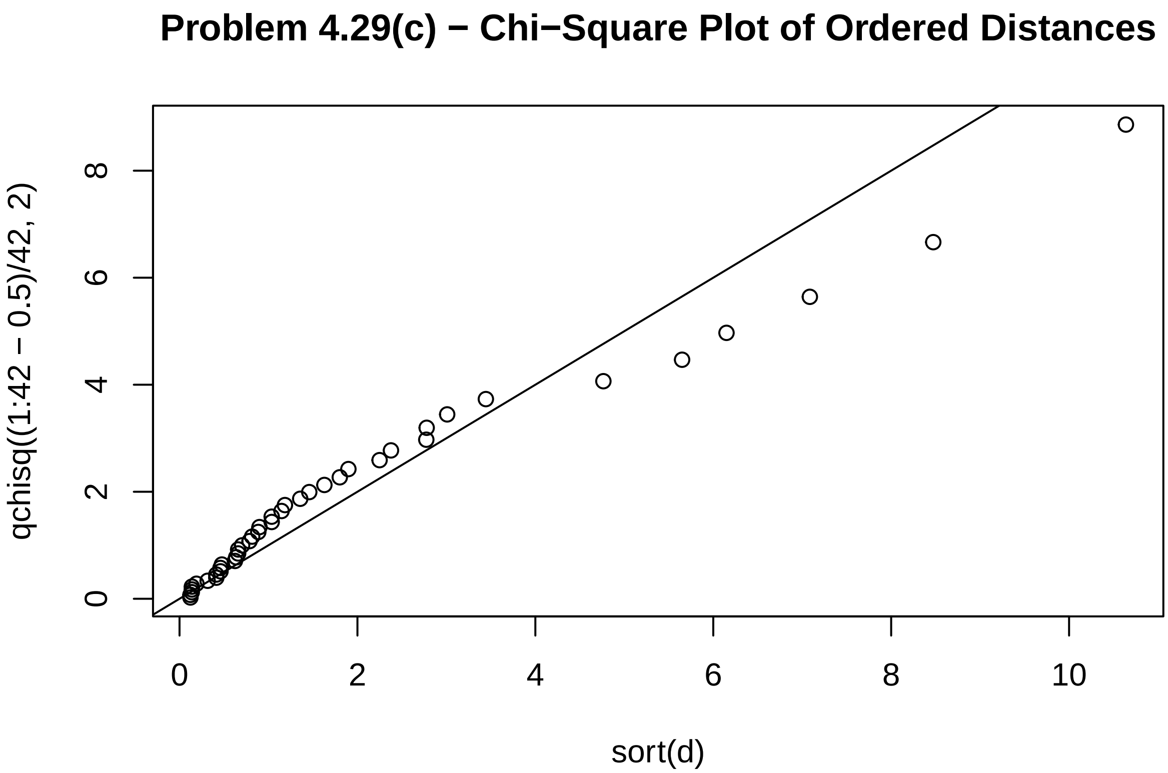
```r
    ncol = 2))
d <- matrix(c(t(matrix(c(N[1], O[1])) - m) %*% S %*% (matrix(c(N[1], O[1])) -
    m), t(matrix(c(N[2], O[2])) - m) %*% S %*% (matrix(c(N[2], O[2])) -
    m), t(matrix(c(N[3], O[3])) - m) %*% S %*% (matrix(c(N[3], O[3])) -
    m), t(matrix(c(N[4], O[4])) - m) %*% S %*% (matrix(c(N[4], O[4])) -
    m), t(matrix(c(N[5], O[5])) - m) %*% S %*% (matrix(c(N[5], O[5])) -
    m), t(matrix(c(N[6], O[6])) - m) %*% S %*% (matrix(c(N[6], O[6])) -
    m), t(matrix(c(N[7], O[7])) - m) %*% S %*% (matrix(c(N[7], O[7])) -
    m), t(matrix(c(N[8], O[8])) - m) %*% S %*% (matrix(c(N[8], O[8])) -
    m), t(matrix(c(N[9], O[9])) - m) %*% S %*% (matrix(c(N[9], O[9])) -
    m), t(matrix(c(N[10], O[10])) - m) %*% S %*% (matrix(c(N[10], O[10])) -
    m), t(matrix(c(N[11], O[11])) - m) %*% S %*% (matrix(c(N[11], O[11])) -
    m), t(matrix(c(N[12], O[12])) - m) %*% S %*% (matrix(c(N[12], O[12])) -
    m), t(matrix(c(N[13], O[13])) - m) %*% S %*% (matrix(c(N[13], O[13])) -
    m), t(matrix(c(N[14], O[14])) - m) %*% S %*% (matrix(c(N[14], O[14])) -
    m), t(matrix(c(N[15], O[15])) - m) %*% S %*% (matrix(c(N[15], O[15])) -
    m), t(matrix(c(N[16], O[16])) - m) %*% S %*% (matrix(c(N[16], O[16])) -
    m), t(matrix(c(N[17], O[17])) - m) %*% S %*% (matrix(c(N[17], O[17])) -
    m), t(matrix(c(N[18], O[18])) - m) %*% S %*% (matrix(c(N[18], O[18])) -
    m), t(matrix(c(N[19], O[19])) - m) %*% S %*% (matrix(c(N[19], O[19])) -
    m), t(matrix(c(N[20], O[20])) - m) %*% S %*% (matrix(c(N[20], O[20])) -
    m), t(matrix(c(N[21], O[21])) - m) %*% S %*% (matrix(c(N[21], O[21])) -
    m), t(matrix(c(N[22], O[22])) - m) %*% S %*% (matrix(c(N[22], O[22])) -
    m), t(matrix(c(N[23], O[23])) - m) %*% S %*% (matrix(c(N[23], O[23])) -
    m), t(matrix(c(N[24], O[24])) - m) %*% S %*% (matrix(c(N[24], O[24])) -
    m), t(matrix(c(N[25], O[25])) - m) %*% S %*% (matrix(c(N[25], O[25])) -
    m), t(matrix(c(N[26], O[26])) - m) %*% S %*% (matrix(c(N[26], O[26])) -
    m), t(matrix(c(N[27], O[27])) - m) %*% S %*% (matrix(c(N[27], O[27])) -
    m), t(matrix(c(N[28], O[28])) - m) %*% S %*% (matrix(c(N[28], O[28])) -
    m), t(matrix(c(N[29], O[29])) - m) %*% S %*% (matrix(c(N[29], O[29])) -
    m), t(matrix(c(N[30], O[30])) - m) %*% S %*% (matrix(c(N[30], O[30])) -
    m), t(matrix(c(N[31], O[31])) - m) %*% S %*% (matrix(c(N[31], O[31])) -
    m), t(matrix(c(N[32], O[32])) - m) %*% S %*% (matrix(c(N[32], O[32])) -
    m), t(matrix(c(N[33], O[33])) - m) %*% S %*% (matrix(c(N[33], O[33])) -
    m), t(matrix(c(N[34], O[34])) - m) %*% S %*% (matrix(c(N[34], O[34])) -
    m), t(matrix(c(N[35], O[35])) - m) %*% S %*% (matrix(c(N[35], O[35])) -
    m), t(matrix(c(N[36], O[36])) - m) %*% S %*% (matrix(c(N[36], O[36])) -
    m), t(matrix(c(N[37], O[37])) - m) %*% S %*% (matrix(c(N[37], O[37])) -
    m), t(matrix(c(N[38], O[38])) - m) %*% S %*% (matrix(c(N[38], O[38])) -
    m), t(matrix(c(N[39], O[39])) - m) %*% S %*% (matrix(c(N[39], O[39])) -
    m), t(matrix(c(N[40], O[40])) - m) %*% S %*% (matrix(c(N[40], O[40])) -
    m), t(matrix(c(N[41], O[41])) - m) %*% S %*% (matrix(c(N[41], O[41])) -
    m), t(matrix(c(N[42], O[42])) - m) %*% S %*% (matrix(c(N[42], O[42])) -
    m)), nrow = 6, ncol = 7)
d
```

```
##           [,1]       [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## [1,] 0.4606524  1.1848895 5.6494392 3.0089122 6.1488606 0.1901188 1.8984708
## [2,] 0.6592206 10.6391792 0.3159498 0.6592206 1.0360061 0.4606524 2.7782596
## [3,] 2.3770610  0.1388339 0.4135364 2.7741416 0.1388339 1.1471939 8.4730649
## [4,] 1.6282902  0.8162468 0.1224973 1.0360061 0.8856041 7.0857237 0.6370218
## [5,] 0.4135364  1.3566301 0.8987982 0.7874152 0.1379719 1.4584229 0.7032485
## [6,] 0.4760726  0.6228096 4.7646873 3.4437748 2.2488867 0.1224973 1.8013611
```

```r
mean(d <= qchisq(0.5, 2))   # Problem 4.29(b)
```

```
## [1] 0.6190476
```

```r
plot(sort(d), qchisq((1:42 - 0.5)/42, 2), main = "Problem 4.29(c) - Chi-Square Plot of Ordered Distances
abline(0, 1)
```

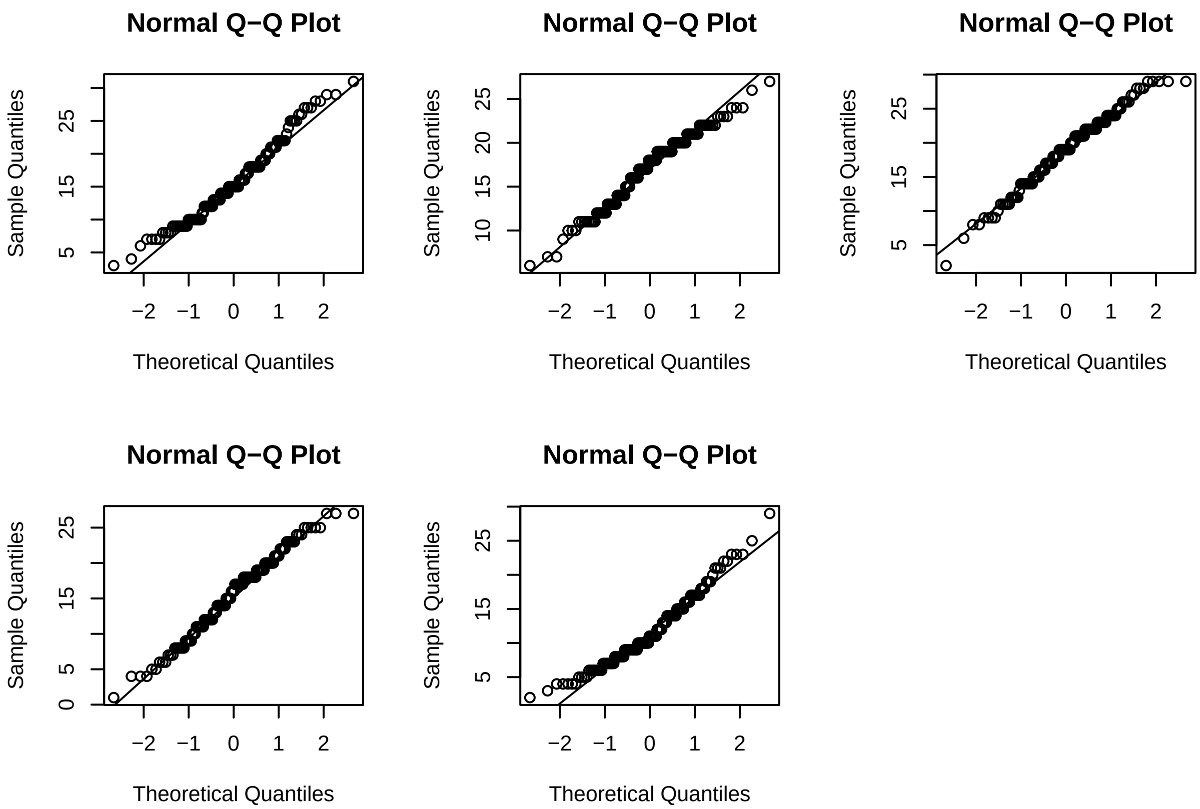


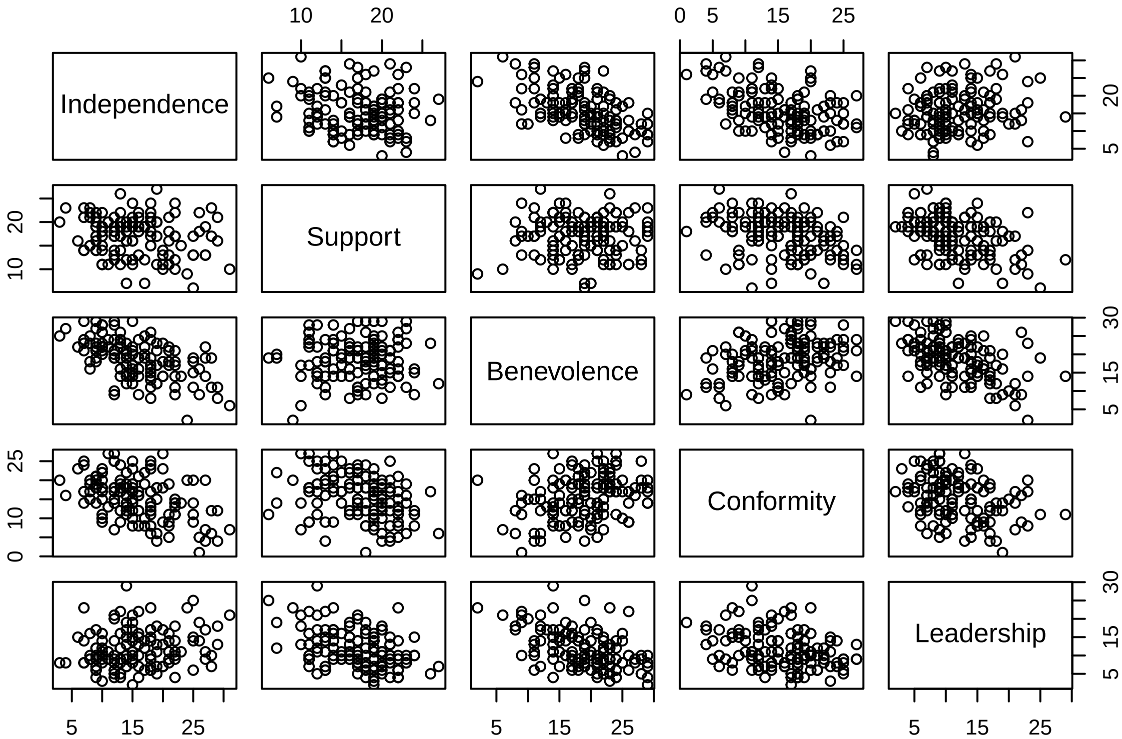
Problem 4.39

```r
T<-read.table("/Users/newuser/Desktop/Notes/Graduate/STAT 488 - Multivariate Statistical Analysis/T4-6.d
library(MVN)
par(mfrow=c(2,3))                                                                  # Problem 4.39(a)
qqnorm(T$I)
qqline(T$I)
qqnorm(T$S)
qqline(T$S)
qqnorm(T$B)
qqline(T$B)
qqnorm(T$C)
qqline(T$C)
qqnorm(T$L)
qqline(T$L)
r<-data.frame(cor(data.frame(qqnorm(T$I,plot.it=F)))[2],cor(data.frame(qqnorm(T$S,plot.it=F)))[2],cor(da
names(r)<-names(T)<-c("Independence","Support","Benevolence","Conformity","Leadership")
row.names(r)<-"r_Q"
r
```

```
##     Independence  Support Benevolence Conformity Leadership
## r_Q    0.9881301 0.989288   0.9925086    0.99338  0.9812888
```

```
# An online table (https://www.itl.nist.gov/div898/handbook/eda/section3/eda3676.htm)
# of critical values of the normal probability plot correlation coefficient shows the
# critical value of r_Q for n = 130 is approximately r = 0.9897. As such, we reject H0
# for the independence, support, and leadership variables at the alpha = 0.05 level.
# There is sufficient evidence these three variables are not normally distributed. We
# fail to reject H0 for the benevolence and conformity variables at the alpha = 0.05
# level. There is insufficient evidence these two variables are not normally distributed.
mvn(T,univariatePlot="scatter")$multivariateNormality                    # Problem 4.39(b)
```

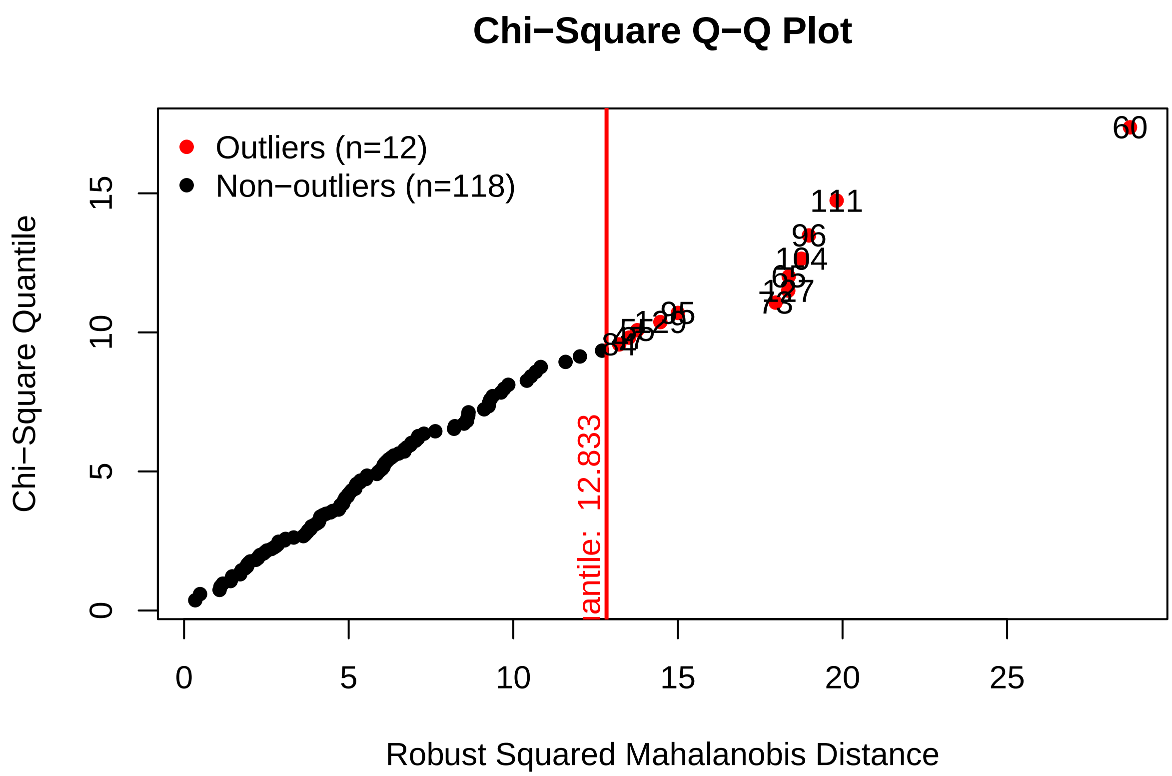





```
##              Test        HZ   p value MVN
## 1 Henze-Zirkler 0.9684355 0.104286 YES
```

```
# We fail to reject H0 at the alpha = 0.05 level. There is insufficient
# evidence (p = 0.104286) that the multivariate normality assumption has been violated.
```

```
mvn(T,multivariateOutlierMethod="quan",showOutliers=TRUE)$multivariateOutliers
```




```
##     Observation Mahalanobis Distance Outlier
## 60           60               28.732    TRUE
## 111         111               19.821    TRUE
## 96           96               18.980    TRUE
## 104         104               18.752    TRUE
## 65           65               18.366    TRUE
## 127         127               18.351    TRUE
## 73           73               17.966    TRUE
## 95           95               14.999    TRUE
## 129         129               14.468    TRUE
## 55           55               13.765    TRUE
## 47           47               13.515    TRUE
## 84           84               13.217    TRUE
```

```
# We can see from the quantile method that there are 12 observed statistical outliers in
# the data, with observation 60 having the greatest Mahalanobis Distance of d = 28.732.
mvn(sqrt(T[c("Independence","Support","Leadership")]))$univariateNormality # Problem 4.39(c)
```

```
##               Test     Variable Statistic   p value Normality
## 1 Anderson-Darling Independence    0.4148    0.3302    YES
## 2 Anderson-Darling   Support       1.8426    0.0001    NO
## 3 Anderson-Darling  Leadership     0.4767    0.2342    YES
```

```
# Since these three variables are count data, (4-33) 1. on page 192 of the textbook says
# the square root scale is the most appropriate transformation in this situation. We can see
# the p-values of the Anderson-Darling test for the independence and leadership variables
# are greater than alpha = 0.05, but the p-value of the Anderson-Darling test for the
# variable for support is not, warranting additional analysis for this support variable.
mvn(T[c("Independence","Support","Leadership")],bc=TRUE)$univariateNormality
```

```
##               Test     Variable Statistic   p value Normality
```

```
## 1 Anderson-Darling Independence    0.4148    0.3302    YES
## 2 Anderson-Darling   Support       1.2096    0.0036    NO
## 3 Anderson-Darling  Leadership     0.4767    0.2342    YES
```

```
mvn(T[c("Independence","Support","Leadership")],bc=TRUE)$BoxCoxPowerTransformation
```

```
## Independence      Support   Leadership
##          0.5          1.0          0.5
```

```
# I first performed a Box-Cox transformation using the function's built-in argument,
# but the p-value of the Anderson-Darling test for the support variable was still
# less than alpha = 0.05. We can see from the output that for whatever reason,
# lambda = 1.0 for the support variable, which meant that no scaling or transformation
# was actually performed. We can also note that lambda = 0.5 for the independence and
# leadership variables, which is the same as the initial square root transformation.
```

```
mvn(data.frame(T$I,(T$S^1.8-1)/1.8,T$L),univariatePlot="histogram",bc=TRUE)$univariateNormality
```

```
##               Test           Variable Statistic   p value Normality
## 1 Anderson-Darling        T.I            0.4148    0.3302    YES
## 2 Anderson-Darling X.T.S.1.8...1..1.8    0.8342    0.0307    NO
## 3 Anderson-Darling        T.L            0.4767    0.2342    YES
```

```
mvn(data.frame(T$I,(T$S^1.5-1)/1.5,T$L),univariateTest="SF",bc=TRUE)$univariateNormality
```

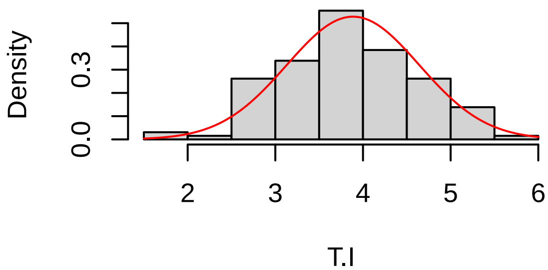
```
##              Test           Variable Statistic   p value Normality
## 1 Shapiro-Francia        T.I            0.9902    0.4143    YES
## 2 Shapiro-Francia X.T.S.1.5...1..1.5    0.9847    0.1344    YES
## 3 Shapiro-Francia        T.L            0.9920    0.5821    YES
```

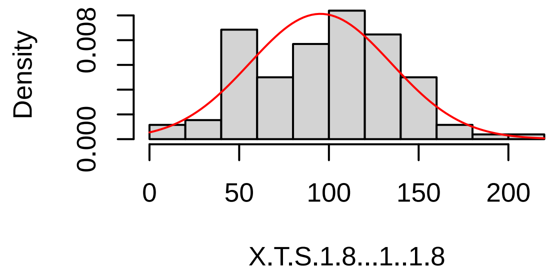
```
mvn(data.frame(T$I,(T$S^1.5-1)/1.5,T$L),univariateTest="SW",bc=TRUE)$univariateNormality
```

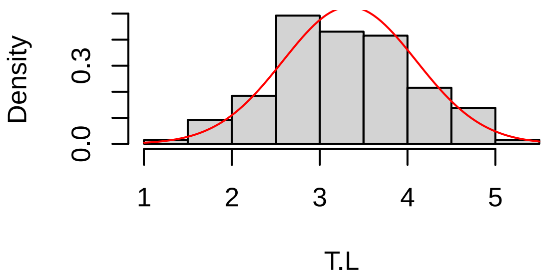
```
##           Test           Variable Statistic   p value Normality
## 1 Shapiro-Wilk        T.I            0.9891    0.3944    YES
## 2 Shapiro-Wilk X.T.S.1.5...1..1.5    0.9832    0.1098    YES
## 3 Shapiro-Wilk        T.L            0.9914    0.6053    YES
```

```
# I then proceeded to perform a Box-Cox transformation manually using (4-34) on page 193.
# After testing various values of lambda, I found that lambda = 1.8 maximized the p-value
# of the Anderson-Darling test for the support variable, but it was still less than
# alpha = 0.05. We can see the histogram for the support variable appears to be bimodal
# with a mode around 50, which the function may consider to be not normally distributed.
# I eventually tried using different arguments for the univariate normality tests (since
# the default argument is the Anderson-Darling test) and additional values of lambda. This
# resulted in p-values greater than alpha = 0.05 when using the Shapiro-Francia test and
# the Shapiro-Wilk test with which I am most familiar, which both had maximized p-values
# when using lambda = 1.5. These results illustrate how different the various univariate
# normality tests can vary in methodology, test statistic, p-value, and conclusion.
rS<-data.frame(0.9897,cor(data.frame(qqnorm(T$S,plot.it=F)))[2],cor(data.frame(qqnorm(sqrt(T$S),plot.it
names(rS)<-c("Critical","Original","Square Root","Natural Log","Box-Cox (l=1.8)","Box-Cox (l=1.5)")
row.names(rS)<-"r_Q"
rS
```

```
##     Critical Original Square Root Natural Log Box-Cox (l=1.8) Box-Cox (l=1.5)
## r_Q   0.9897 0.989288   0.9798977   0.9622772       0.9911138        0.992072
```







I calculated the r_Q values for the various transformations, and after some experimentation, I noticed a Box-Cox transformation using $\lambda = 1.5$, not $\lambda = 1.8$, provided the greatest value for r_Q for the support variable.

In conclusion, it appears a **square root** ($\sqrt{x_j}$) transformation is the most appropriate for the variables for independence and leadership and a **Box-Cox** transformation with $\lambda = 1.8$ is the most appropriate for the variable for support when using the Shapiro-Wilk test for univariate normality.