Homework 03
Joseph Blubaugh
jblubau1@tamu.edu
STAT 636-720

4.2) Consider Bivariate Normal $N_2(\mu, \Sigma); \mu = [0, 2]; \sigma_{11} = 2; \sigma_{22} = 1; \rho_{12} = .5$

a) Write out the bivariate normal density:

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{1.5}} * exp\left[-\left[\frac{x_1}{\sqrt{2}}^2 + (x_2 - 2)^2 - \frac{x_1}{\sqrt{2}}^2 (x_2 - 2)^2\right]\right]$$

b) Write out the generalized distance expression for $(x - \mu)'\Sigma^{-1}(x - \mu)$

$$\sigma_{12} = 1$$

$$\Sigma^{-1} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$$

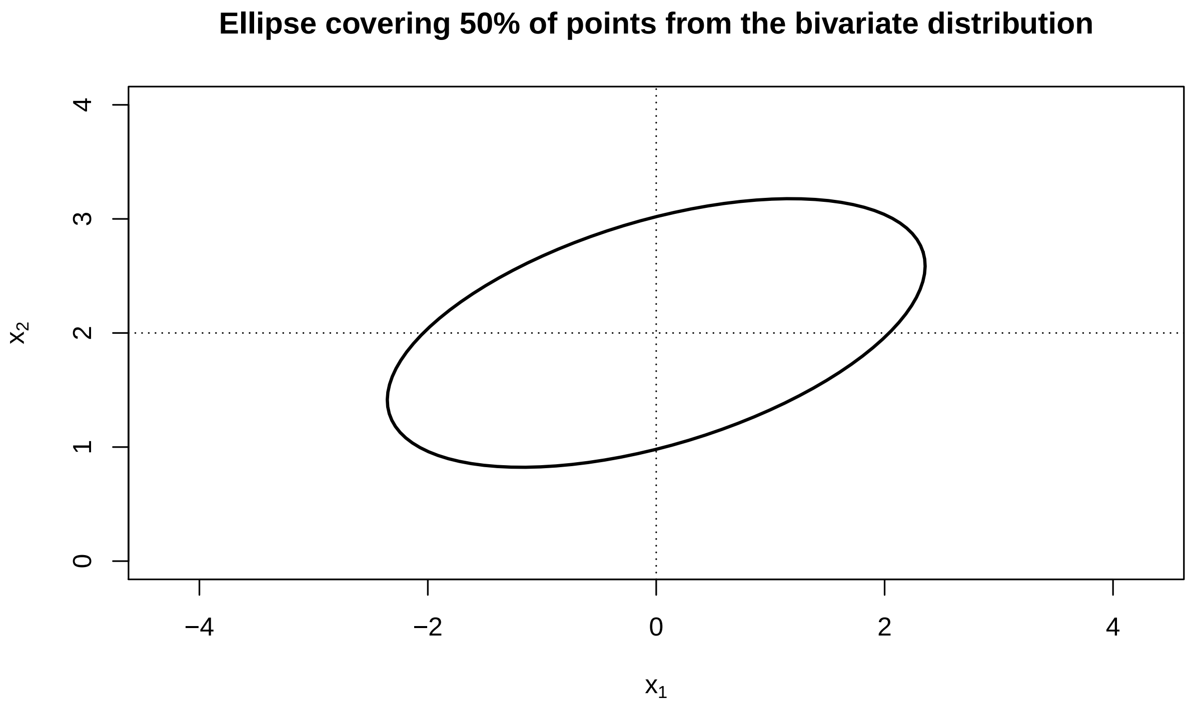
Generalized Distance:

$$x = (x - \mu_i)' \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} (x - \mu_i)$$

c)

```
mu = c(0, 2)
c2 <- qchisq(1 - .5, 2)
sg = c(2, 1)
rho = .5
Sigma = matrix(c(sg[1] ^ 2, rho * sg[1] * sg[2], rho *
                   sg[1] * sg[2], sg[2] ^ 2), nrow = 2)

lambda = eigen(Sigma)$values
eig.vect = eigen(Sigma)$vectors

theta = acos(abs(eig.vect[1, 1])) * 57.2957795

library(plotrix)

plot(c(-3, 3), c(0, 4),
     xlab = expression(x[1]),
     ylab = expression(x[2]),
     asp = 1, type = "n")
```

```r
abline(h = 2, v = 0, lty = 3)
draw.ellipse(mu[1], mu[2],
             sqrt(c2 * lambda[1]),
             sqrt(c2 * lambda[2]),
             angle = theta,
             border = 1, lwd = 2)
title(main = "Ellipse covering 50% of points from the bivariate distribution")
```



4.3)

a) X_1 and X_2 are not independent because their covariance X_{12} are -2

b) X_2 and X_3 are independent because their covariance X_{23} is 0

c) Since the covariances X_{13} and X_{23} are both 0 (X_1, X_2) and X_3 are independent

d) For the linear combination of $\frac{X_1+X_2}{2}$ and X_3

$$A = \begin{bmatrix} .5 & .5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{aligned} Cov(AX) =& A\Sigma A' \\ =& \begin{bmatrix} .5 & .5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} .5 & 0 \\ .5 & 0 \\ 0 & 1 \end{bmatrix} \\ =& \begin{bmatrix} .5 & 0 \\ 0 & 2 \end{bmatrix} \end{aligned}$$

The linear combination of $\frac{X_1+X_2}{2}$ and X_3 are independent because their covariance are 0

e) The linear combination of X_2 and $X_2 - \frac{5X_1}{2} - X_3$ are not independent because their covariance are not 0

```
(A = matrix(c(0, 1, 0, -2.5, 1, -1), nrow = 2, byrow = TRUE))
```

```
     [,1] [,2] [,3]
[1,]  0.0    1    0
[2,] -2.5    1   -1
```

```
(Cov = matrix(c(1, -2, 0, -2, 5, 0, 0, 0, 2), nrow = 3))
```

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

```
A %*% Cov %*% t(A)
```

```
     [,1]  [,2]
[1,]    5 10.00
[2,]   10 23.25
```

4.18)

```r
(A = matrix(c(3, 4, 5, 4, 6, 4, 7, 7), ncol = 2))
```

```
     [,1] [,2]
[1,]    3    6
[2,]    4    4
[3,]    5    7
[4,]    4    7
```

```r
## MLE mean of A1
(mle.mu.A1 = sum(A[, 1])/nrow(A))
```

```
[1] 4
```

```r
## MLE mean of A2
(mle.mu.A2 = sum(A[, 2])/nrow(A))
```

```
[1] 6
```

```r
Var.A1 = (A[,1] - colMeans(A)[1])^2
Var.A2 = (A[,2] - colMeans(A)[2])^2

## Covariance of A12
(Cov.A = sum(Var.A1 * Var.A2)/ (nrow(A) - 1))
```

```
[1] 0.3333333
```

4.19)
a) $(X_1 - \mu)'\Sigma(X_1 - \mu) = \chi^2_6$
b) $\sqrt{n}(\bar{X} - \mu) = N_6(0, \Sigma)$
c) $(n-1)S = W_{19}(\Sigma)$

4.21)
a) $\bar{X} = N_4(\mu, \frac{\Sigma}{60})$
b) $(X_1 - \mu)'\Sigma(X_1 - \mu) = \chi^2_4$
c) $n(X_1 - \mu)'\Sigma(X_1 - \mu) = n\chi^2_4$
d) $n(X_1 - \mu)'S^{-1}(X_1 - \mu) = \chi^2_4$

4.29)

```
air.polution =
  read.table("C:/Users/Joseph/Projects/learning/Statistics/STAT_636/Homework/T1-5.DAT",
             quote="\"", comment.char="")
colnames(air.polution) = c("Wind", "Solar.Rad", "CO", "NO", "NO2", "O3", "HC")
head(air.polution)
```

```
  Wind Solar.Rad 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
```

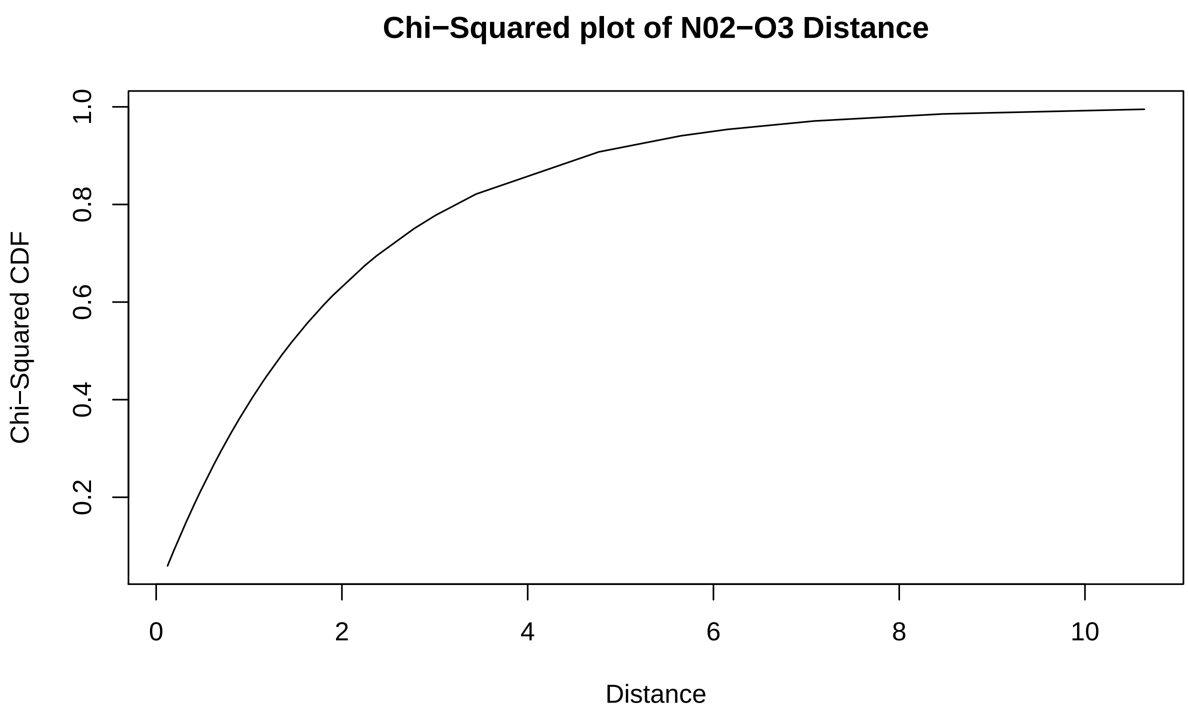
```
x = as.matrix(air.polution[, 5:6], nrow = 2)
cov.x = as.matrix(solve(cov(x)))

## A) Distance calculation
distance = data.frame()
for (i in 1:42) {
  y = t(x[i,] - colMeans(x)) %*% cov.x %*% (x[i,] - colMeans(x))
  distance = rbind(distance, y)
}
colnames(distance) = c("distance")

head(distance)
```

```
   distance
1 0.4606524
2 0.6592206
3 2.3770610
4 1.6282902
5 0.4135364
6 0.4760726
```

```
## B) Proportion of points inside of a 50% probability contour
distance$chi.sq.critical = rep(qchisq(p = .5, df = 2), 42)
distance$diff = with(distance, chi.sq.critical - distance)
nrow(distance[distance$diff > 0, ])/42
```

```
[1] 0.6190476
```

```
## C)
distance$chi.sq = pchisq(distance$distance, df = 2)
distance = distance[order(distance$distance), ]

plot(distance$distance, distance$chi.sq, type = "l",
     xlab = "Distance", ylab = "Chi-Squared CDF",
     main = "Chi-Squared plot of N02-O3 Distance")
```

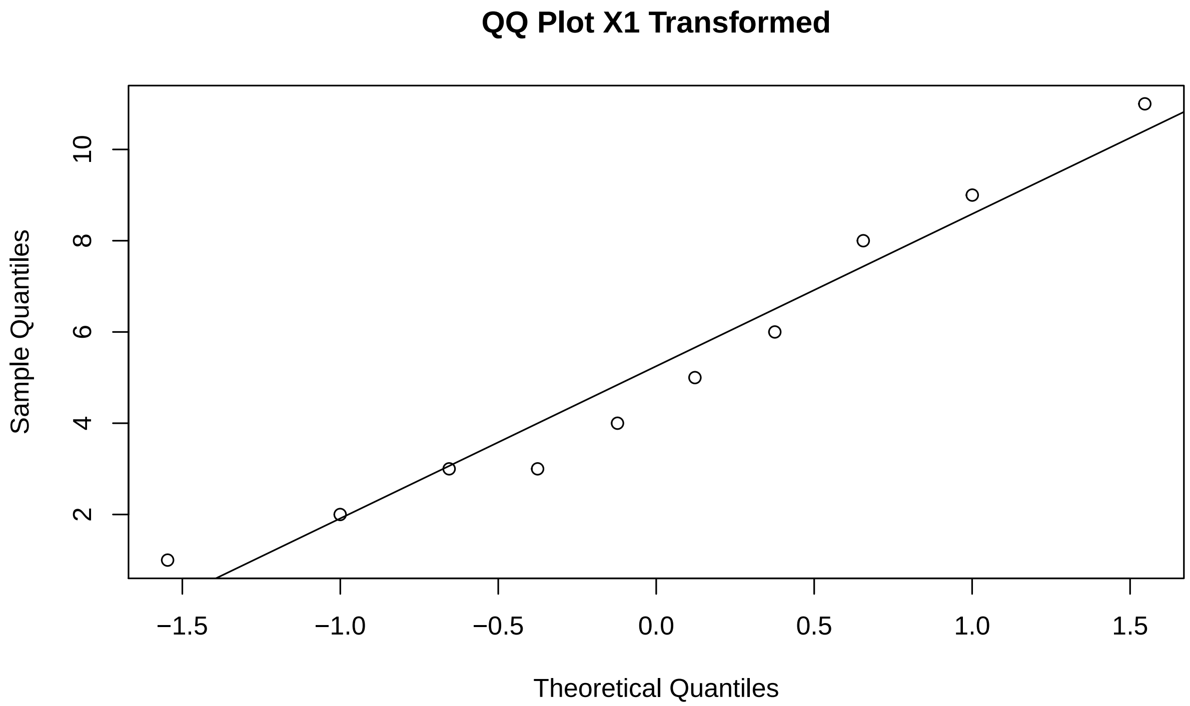


4.30)

```
library(MASS)

cars = data.frame(
  x1 = c(1, 2, 3, 3, 4, 5, 6, 8, 9, 11),
  x2 = c(18.95, 19, 17.95, 15.54, 14, 12.95, 8.94, 7.49, 6, 3.99)
)

## A) Box cox transformation for X1
bc = data.frame(boxcox(x1 ~ 1, data = cars, plotit = FALSE))
bc = bc[order(bc$y), ]
tail(bc, 1)
```

```
     x         y
25 0.4 -7.637662
```

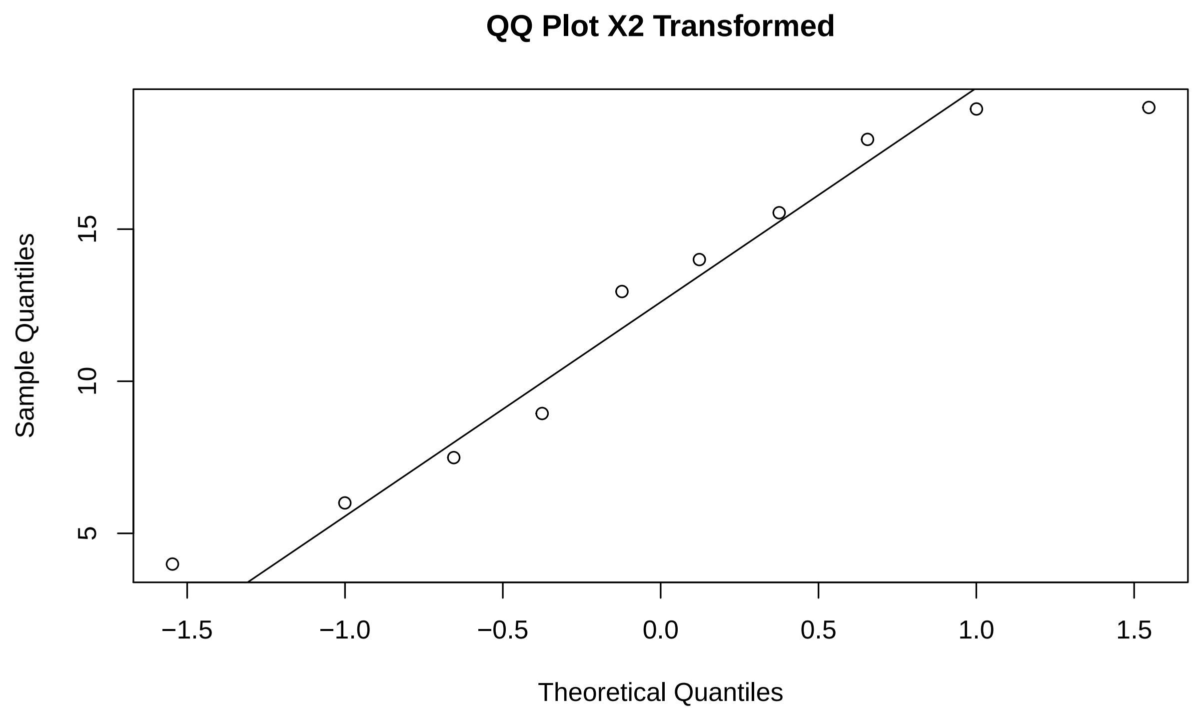
```r
cars$x1.trans = (cars$x1^.4 - 1)/.4
qqnorm(cars$x1, main = "QQ Plot X1 Transformed")
qqline(cars$x1)
```




```r
## B) Box cox transformation for X2
bc = data.frame(boxcox(x2 ~ 1, data = cars, plotit = FALSE))
bc = bc[order(bc$y), ]
tail(bc, 1)
```

```
     x         y
30 0.9 -4.033217
```

```r
cars$x2.trans = (cars$x2^.9 - 1)/.9
qqnorm(cars$x2, main = "QQ Plot X2 Transformed")
qqline(cars$x2)
```




```
## C) Multivariate Box Cox
cars = cars[, -c(3,4)]

results = data.frame()

for (lam.1 in seq(-1, 1, .1)) {
  for (lam.2 in seq(-1, 1, .1)) {

      x = data.frame(lam.1, lam.2)
      results = rbind(results, x)
  }
}

 score = c()

for (j in 1:nrow(results)) {

  cars.trans = cars
  cars.trans$x1 = cars.trans$x1^results[j, 1]
  cars.trans$x2 = cars.trans$x2^results[j, 2]

  x1 = -5 * log(det(cov(cars.trans)))
```

```r
  x2.1 = (results[j, 1] - 1) * sum(log(cars$x1))
  x2.2 = (results[j, 2] - 1) * sum(log(cars$x2))

  x2.sum = sum(x2.1, x2.2)

  score = c(score, x1 + x2.sum)

}

results$score = score
results = subset(results, score < Inf)
results = results[order(results$score), ]
tail(results, 1)
```

```
    lam.1 lam.2    score
243   0.1   0.1 26.17309
```

```r
## Multivariate Box Cox --> X1 (Lambda 1): -.8, X2 (Lambda 2): -1.0
## vs
## Univariate Box Cox of X1: .4, X2: .9
```