Math 760
Chapter 4 HW
Gabrielle Salamanca
March 31, 2024

1. Consider a bivariate normal distribution with $\mu_1 = 1, \mu_2 = 3, \sigma_{11} = 2, \sigma_{22} = 1, \rho_{12} = -.8$.
(a) Write out the bivariate normal density.

The multivariate normal density function is: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left(\frac{x-\mu}{\sigma}\right)^2}{2}}, -\infty < x < \infty$

For the bivariate normal density, let's find our needed variables.
Let's start with σ^2:
```
## [1] 0.72
```

Then, let's find the components of $\frac{x-\mu}{\sigma}$. We already have our μ, so we need to find the inverse of our Σ.
```
##            [,1]      [,2]
## [1,] 1.388889 1.571348
## [2,] 1.571348 2.777778
```

Now, we have everything. The bivariate normal density function for this problem is:
$$f(x) = \frac{1}{\sqrt{2\pi(0.72)}} exp\left(-\frac{1}{2}\left[\frac{1}{0.72}(x_1 - 1)^2 + 2\left(\frac{\sqrt{2}}{9}\right)(x_1 - 1)(x_2 - 3) + \frac{2}{0.72}(x_2 - 3)^2\right]\right)$$
$$= \frac{1}{\sqrt{1.44\pi}} exp\left(-\frac{1}{2}\left[\frac{1}{0.72}(x_1 - 1)^2 + \frac{2\sqrt{2}}{9}(x_1 - 1)(x_2 - 3) + \frac{2}{0.72}(x_2 - 3)^2\right]\right)$$

(b) Write out the squared statistical distance expression $(x - \mu)'\Sigma^{-1}(x - \mu)$ as a quadratic function of x_1 and x_2

The squared statistical distance expression can be written as a quadratic function of x_1 and x_2 as so:
$$\frac{1}{0.72}(x_1 - 1)^2 + \frac{2\sqrt{2}}{9}(x_1 - 1)(x_2 - 3) + \frac{2}{0.72}(x_2 - 3)^2$$

3. Let X be $N_3(\mu, \Sigma)$ with $\mu' = \begin{bmatrix} -3 & 1 & 4 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ Which of the following random variables are independent? Explain.

(a) X_1 and X_2
Because $\sigma_{12} = -2$, X_1 and X_2 aren't independent.

(b) X_2 and X_3
Because $\sigma_{22} = 0$, X_2 and X_3 are independent.

(c) (X_1, X_2) and X_3

Because $\sigma_{13} = 0 = \sigma_{23}$, (X_1, X_2) and X_3 are independent.

(d) $\frac{X_1+X_2}{2}$ and X_3

We know from **Result 4.3**:

If X is distributed as $N_p(\mu, \Sigma)$, the q linear combinations

$$A_{(qxp)}X_{(px1)} = \begin{bmatrix} a_{11}X_1+\ldots+a_{1p}X_p \\ a_{21}X_1+\ldots+a_{2p}X_p \\ \ldots \\ a_{q1}X_1+\ldots+a_{qp}X_p \end{bmatrix}$$

are distributed as $N_q(A\mu, A\Sigma A')$. Also, $X_{(px1)} + d_{(px1)}$, where d is a vector of constants, is distributed as $N_p(\mu + d, \Sigma)$

Thus, $\frac{X_1+X_2}{2}$ and X_3 are jointly normal and their covariance is:

$$\frac{1}{2}(\sigma_{13}) + \frac{1}{2}(\sigma_{23}) = \$(\{13\}) \; + \; (\{23\}) \; = \; \$$$

```
## [1] 0
```

Therefore, $\frac{X_1+X_2}{2}$ and X_3 are independent.

(e) X_2 and $X_2 - \frac{5}{2}X_1 - X_3$

From **Result 4.3**, we can use $N_q(A\mu, A\Sigma A')$ to see if these random variables are independent.

Forming A with the random variables in mind:

$$A = \begin{bmatrix} 0 & 1 & 0 \\ -\dfrac{5}{2} & 1 & -1 \end{bmatrix}$$

If the covariance of $A\Sigma A'$ is 0, then they're independent.

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

We see that the covariance is 10, so these random variables aren't independent.

6. Let X be $N_3(\mu, \Sigma)$ with $\mu' = \begin{bmatrix} 1 & -1 & 2 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{bmatrix}$ Which of the following random variables are independent? Explain.

(a) X_1 and X_2

X_1 and X_2 are independent, because σ_{12} is:

```
## [1] 0
```

(b) X_1 and X_3
X_1 and X_3 aren't independent, because σ_{13} is:
```
## [1]  -1
```

(c) X_2 and X_3
X_2 and X_3 are independent, because σ_{23} is:
```
## [1] 0
```

(d) (X_1, X_3) and X_2
(X_1, X_3) and X_2 are independent, because:
```
## sigma12 = 0
## sigma32 = 0
```

(e) X_1 and $X_1 + 3X_2 - 2X_3$
X_1 and $X_1 + 3X_2 - 2X_3$ aren't independent, because when solving for the second random variable:
```
## The covariance is 7
```

19. Let $X_1, X_2, \ldots, X_{20}$ be a random sample of size n = 20 from an $N_6(\mu, \Sigma)$ population. Specify each of the following completely.

(a) The distribution of $(X_1 - \mu)'\Sigma^{-1}(X_1 - \mu)$
From **Result 4.7**:

Let X be distributed as $N_p(\mu, \Sigma)$ with $|\Sigma| > 0$. Then
   (a) $(X - \mu)'\Sigma^{-1}(X - \mu)$ is distributed as χ_p^2, where χ_p^2 denotes the chi-square distribution with p degrees of freedom

Therefore, $(X_1 - \mu)'\Sigma^{-1}(X_1 - \mu) \sim \chi^2$ with 6 degrees of freedom, because $N_6(\mu, \Sigma)$.

(b) The distributions of $\bar{X}$ and $\sqrt{n}(\bar{X} - \mu)$
From (4-23), we know $\bar{X}$ is distributed as $N_p\left(\mu, \frac{1}{n}\Sigma\right)$. Therefore:
$\bar{X} \sim N_6\left(0, \frac{1}{20}\Sigma\right)$
$\sqrt{n}(\bar{X} - \mu) \Rightarrow \sqrt{20}(\bar{X} - \mu) \sim N_6(0, \Sigma)$

(c) The distribution of $(n - 1)S$
From (4-23), we know $(n - 1)S$ is distributed as a Wishart random matrix with $n - 1$ degrees of freedom. Therefore, $(n - 1)S = (20 - 1)S = 19S$ has a Wishart distribution with 19 degrees of freedom.

23. Consider the annual rates of return (including dividend) on the Dow-Jones industrial average for the years 1996-2005. These data, multiplied by 100, are: -0.6, 3.1, 25.3, -16.8, -7.1, -6.2, 25.2, 22.6, 26.0. Use these 10 observations to complete the following.

(a) Construct a Q-Q plot. Do the data seem to be normally distributed? Explain.
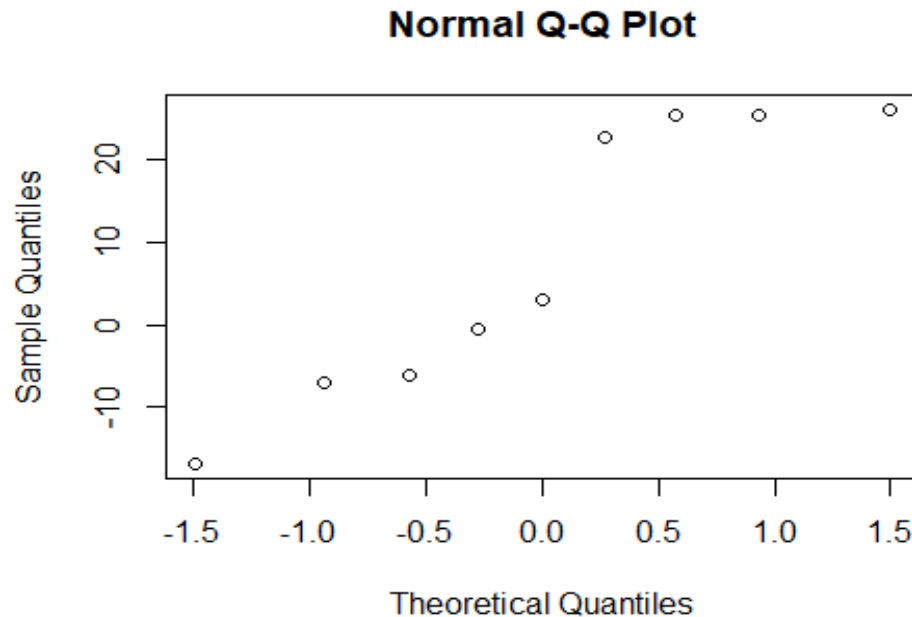
Normal Q-Q Plot

The qqplot reveals that this data doesn't follow a normal distribution. However, there are only 10 observations, so we can't truly determine it so if we don't have a larger sample size.

(b) Carry out a test of normality based on the correlation coefficient r_Q. [See (4-31)] Let the significance level be $\alpha = 0.10$

$$H_0: \text{The data is normal}$$
$$H_1: \text{The data is non-normal}$$

The formula for r_Q is:

$$r_Q = \frac{\sum_{j=1}^{n}(x_{(j)} - \bar{x})(q_{(j)} - \bar{q})}{\sqrt{\sum_{j=1}^{n}(x_{(j)} - \bar{x})^2}\sqrt{\sum_{j=1}^{n}(q_{(j)} - \bar{q})^2}}$$

And because we know our sample size is 10 and the $\alpha = 0.10$, our critical point is 0.9351. And our r_Q is:

```
## [1] 0.9383158
```

Because $r_Q = 0.9383158 > 0.9351 = \alpha$, we fail to reject H_0.

26. Exercise 1.2 gives the age x_1, measured in years, as well as the selling price x_2, measured in thousands of dollars, for n = 10 used cars. These data are reproduced as follows:

```
##    x1    x2
## 1   1 18.95
## 2   2 19.00
## 3   3 17.95
## 4   3 15.54
## 5   4 14.00
## 6   5 12.95
## 7   6  8.94
## 8   8  7.49
## 9   9  6.00
## 10 11  3.99
```

(a) Use the results of Exercise 1.2 to calculate the squared statistical distances $(x_j - \bar{x})'S^{-1}(x_j - \bar{x})$ (j = 1, 2, ..., 10), where $x'_j = [x_{j1}], x_{j2}$

Our $\bar{x}$ matrix is:
```
##          [,1]
## [1,]   5.200
## [2,]  12.481
```

Our S matrix is:
```
##            x1         x2
## x1   10.62222 -17.71022
## x2  -17.71022  30.85437
```

And our S^{-1} matrix is:
```
##           x1        x2
## x1 2.189813 1.2569395
## x2 1.256939 0.7538861
```

Then, we find d_j^2.
```
##  [1] 1.8753045 2.0203262 2.9009088 0.7352659 0.3105192 0.0176162 3.7329012
##  [8] 0.8165401 1.3753379 4.2152799
```

(b) (b) Using the distances in Part a determine the proportion of the observations falling within the estimated 50% probability contour of a bivariate normal distribution.

```
## The chi-squared value with 2 degrees of freedom is 1.386294
```

We want at least half of the observations to fall within 1.386294, and we see that 5/10 observations do.

```
##           Xobs   less
## 1   1.8753045 FALSE
## 2   2.0203262 FALSE
## 3   2.9009088 FALSE
## 4   0.7352659  TRUE
## 5   0.3105192  TRUE
## 6   0.0176162  TRUE
## 7   3.7329012 FALSE
## 8   0.8165401  TRUE
## 9   1.3753379  TRUE
## 10  4.2152799 FALSE
```

(c)   (c) Order the distances in Part a and construct a χ^2 plot.

Chi-Square Q-Q Plot of cars
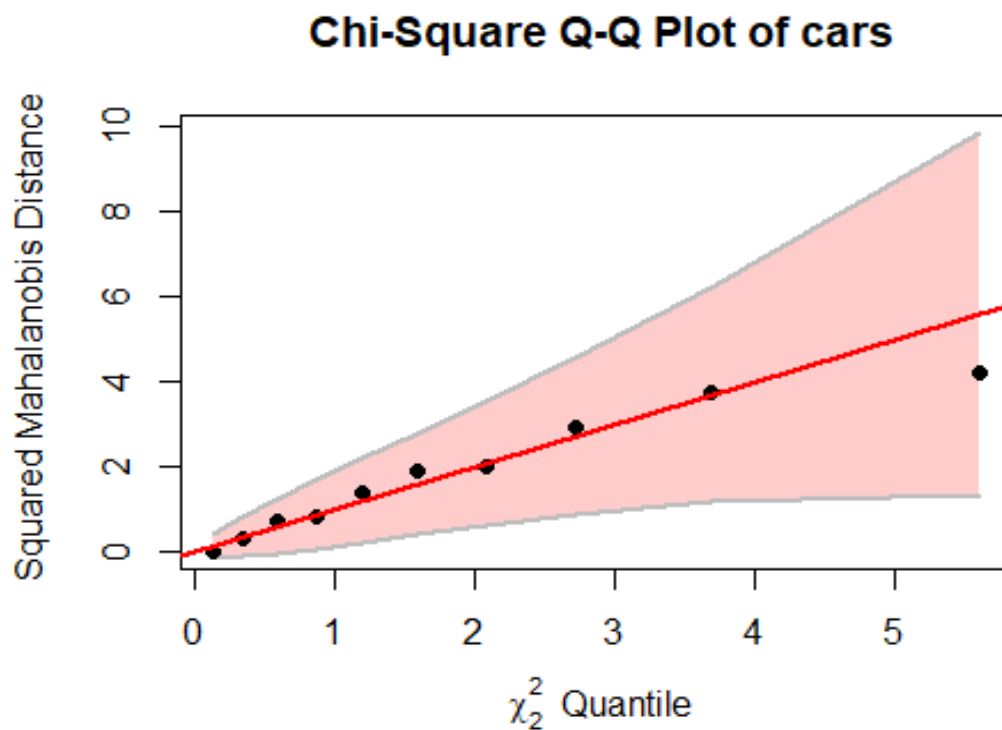
(d) Given the results in Parts b and c, are these data approximately bivariate normal? Explain.
Given our results in (b) and (c), we could say this data is approximately bivariate normal.
However, our sample size is extremely small, only 10 observations. So, we cannot truly say
it is approximately bivariate normal.

37. Refer to Exercise 1.18. Convert the women's track records in Table 1.9 to speeds measured in meters per second. Examine the data on speeds for marginal and multivariate normality.

Let's look into the χ^2 qqplot for the women's track records.

Chi-Square Q-Q Plot of record
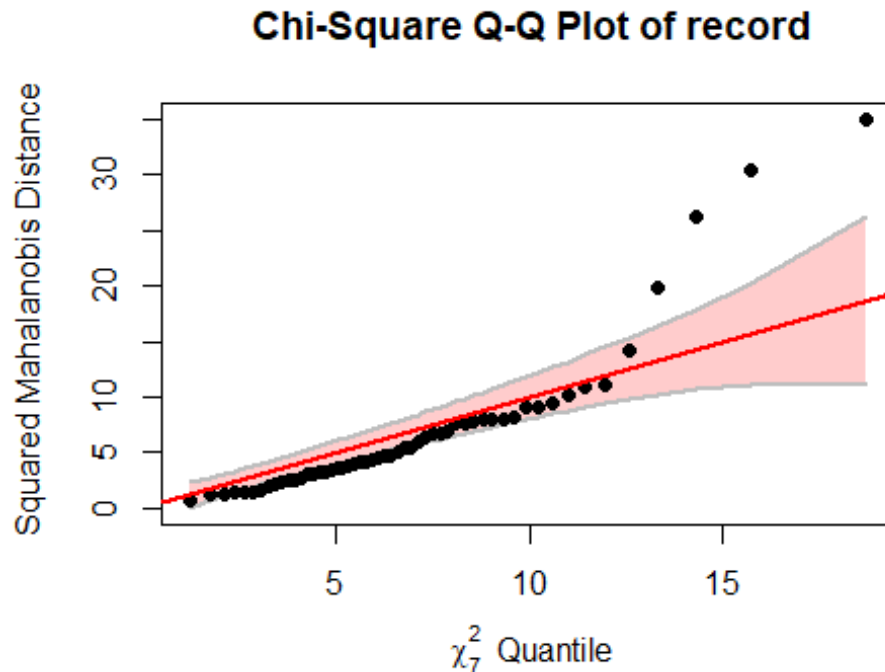
We see a majority fall within the highlighted region, even if a good chunk are toeing the line. Though, we do have four outliers, so we would call into question the normality of this data in terms of multivariate normality.

Let's then find the r_Q for each variable and will test it at $\alpha = 0.05$. However, the data has 54 variables, and **Table 42** only has 50 and 55 for n. We will round up to 55, so our critical region is 0.9787.

$$H_0: \text{The data is normal}$$
$$H_1: \text{The data is non-normal}$$

```
## The rq for the 100 m/s is 0.9836334
## The rq for the 200 m/s is 0.9761084
## The rq for the 400 m/s is 0.9698238
## The rq for the 800 m/s is 0.951239
## The rq for the 1500 m/s is 0.9099867
## The rq for the 3000 m/s is 0.8676664
## The rq for the marathon is 0.8605517
```

```
##              rq In_Region
## 1 0.9836334     FALSE
## 2 0.9761084      TRUE
## 3 0.9698238      TRUE
## 4 0.9512390      TRUE
## 5 0.9099867      TRUE
## 6 0.8676664      TRUE
## 7 0.8605517      TRUE
```

We find the values of r_Q decrease as distance increases. We also see that only the 100m/s r_Q is greater than α, so this variable normal; while the rest are less than α, so they're non-normal. So, in terms of marginal normality, the data seems to skew towards the left.

Appendix

```r
knitr::opts_chunk$set(echo = FALSE)
library(heplots)
muMat <- c(1,3)
mu <- matrix(muMat, nrow = 2, ncol = 1, byrow = TRUE)
sigMat <- c(2, -0.8*sqrt(2), -0.8*sqrt(2), 1)
sigma <- matrix(sigMat, nrow = 2, ncol = 2, byrow = TRUE)
sd <- det(sigma)
sd
invSig <- solve(sigma)
invSig
muMat <- c(-3,1,4)
mu <- matrix(muMat, nrow = 1, ncol = 3, byrow = TRUE)
sigMat <- c(1,-2,0,
            -2,5,0,
            0,0,2)
sigma <- matrix(sigMat, nrow = 3, ncol = 3, byrow = TRUE)
sig13 <- sigma[1,3]
sig23 <- sigma[2,3]
0.5*sig13 + 0.5*sig23
aMat <- c(0,1,0,-5/2,1,-1)
A <- matrix(aMat, nrow = 2, ncol = 3, byrow = TRUE)
A %*% sigma %*% t(A)
muMat <- c(1, -1, 2)
mu <- matrix(muMat, nrow = 1, ncol = 3, byrow = TRUE)
sigMat <- c(4, 0, -1,
            0, 5, 0,
            -1, 0, 2)
sigma <- matrix(sigMat, nrow = 3, ncol = 3, byrow = TRUE)
sigma[1,2]
sigma[1,3]
sigma[2,3]
sig12 <- sigma[1,2]
sig32 <- sigma[3,2]

cat("sigma12 =", sig12, "\n")
cat("sigma32 =", sig32)
covar <- sigma[1,1] + 2*sigma[1,2] - 3*sigma[1,3]

cat("The covariance is", covar)
DJ <- c(-0.6, 3.1, 25.3, -16.8, -7.1, -6.2, 25.2, 22.6, 26.0)
qqnorm(DJ)
r <- qqnorm(DJ, plot = FALSE)
rq <- cor(r$x, r$y)
rq
x1 <- c(1,2,3,3,4,5,6,8,9,11)
x2 <- c(18.95,19.00,17.95,15.54,14.00,12.95,8.94,7.49,6.00,3.99)
cars <- data.frame(x1,x2)
```

```r
print(cars)
n1 <- length(x1)
n2 <- length(x2)
x1bar <- sum(x1)/n1
x2bar <- sum(x2)/n2
# matrix
xbarMat <- c(x1bar, x2bar)
xbar <- matrix(xbarMat, nrow = 2, ncol = 1, byrow = TRUE)
xbar
covar <- cov(cars)
covar
inverse <- solve(covar)
inverse
d <- mahalanobis(cars, xbar, covar)
d
half <- qchisq(p = .5, df = 2, lower.tail = FALSE)
cat("The chi-squared value with 2 degrees of freedom is", half)
Xobs <- c(d[1], d[2], d[3], d[4], d[5],
          d[6], d[7], d[8], d[9], d[10])
less <- c(d[1] <= half, d[2] <= half, d[3] <= half, d[4] <= half, d[5] <=
half,
          d[6] <= half, d[7] <= half, d[8] <= half, d[9] <= half, d[10] <=
half)
x <- data.frame(Xobs, less)
x
distCars <- c(1.8753045, 2.0203262, 2.9009088, 0.7352659, 0.3105192,
              0.0176162, 3.7329012, 0.8165401, 1.3753379, 4.2152799)
cqplot(cars)
track <- read.table("D:/Coding/R Storage/T1-9.dat", header = FALSE, sep =
"\t")

# vars
x1 <- track$V1 # country
x2 <- track$V2 # 100m/s
x3 <- track$V3 # 200m/s
x4 <- track$V4 # 400m/s
x5 <- track$V5# 800m/min
x6 <- track$V6 # 1500m/min
x7 <- track$V7 # 3000m/min
x8 <- track$V8 # marathon/min

# per second
OG <- track[,1:4]
second <- track[,5:8]*60
second$V1 <- track$V1

record <- merge(OG, second, by = "V1")
record <- record[,-1]
cqplot(record)
r1 <- qqnorm(record$V2, plot = FALSE)
```

```
r2 <- qqnorm(record$V3, plot = FALSE)
r3 <- qqnorm(record$V4, plot = FALSE)
r4 <- qqnorm(record$V5, plot = FALSE)
r5 <- qqnorm(record$V6, plot = FALSE)
r6 <- qqnorm(record$V7, plot = FALSE)
r7 <- qqnorm(record$V8, plot = FALSE)
# rq
rq1 <- cor(r1$x, r1$y)
rq2 <- cor(r2$x, r2$y)
rq3 <- cor(r3$x, r3$y)
rq4 <- cor(r4$x, r4$y)
rq5 <- cor(r5$x, r5$y)
rq6 <- cor(r6$x, r6$y)
rq7 <- cor(r7$x, r7$y)
# crit
a1 <- rq1 < 0.9787
a2 <- rq2 < 0.9787
a3 <- rq3 < 0.9787
a4 <- rq4 < 0.9787
a5 <- rq5 < 0.9787
a6 <- rq6 < 0.9787
a7 <- rq7 < 0.9787
# cat
cat("The rq for the 100 m/s is", rq1, "\n")
cat("The rq for the 200 m/s is", rq2, "\n")
cat("The rq for the 400 m/s is", rq3, "\n")
cat("The rq for the 800 m/s is", rq4, "\n")
cat("The rq for the 1500 m/s is", rq5, "\n")
cat("The rq for the 3000 m/s is", rq6, "\n")
cat("The rq for the marathon is", rq7)
# chart
rq <- c(rq1, rq2, rq3, rq4, rq5, rq6, rq7)
In_Region <- c(a1,a2,a3,a4,a5,a6,a7)
x <- data.frame(rq, In_Region)
x
```