```
# bumpus_sparrows
#########################################################
#HOMEWORK SECTION
#Jia Yue Gao
#Homework for Week 2
#########################################################################################
#Some of the following codes are adopted from sample code in class
#Bumpus_sparrows <- read_excel("Desktop/Multivariate Analysis/
Bumpus_sparrows.xlsx")
Bumpus_sparrows
sparr <- Bumpus_sparrows
sparr
sparr[,-1]
#We will drop column 1 as it does not contain a number
sparr.num <- sparr[,-1]
# Computing the means of each variable in data frame sparr.num
colMeans(sparr.num)
# Covariance matrix
cov(sparr.num)
# Correlation matrix
cor(sparr.num)


# Survivorship == 'S' for survived sparrows, and "NS" for non-survived
sparrows
sparr_s <- sparr[(sparr$Survivorship == "S"),]
sparr_ns <- sparr[(sparr$Survivorship == "NS"),]
# dim(sparr_s)
# dim(sparr_ns)
# > dim(sparr_s)
# [1] 21  6
# > dim(sparr_ns)
# [1] 28  6


#drop column 1 for both datasets as they does not contain a number
sparr_s.num <- sparr_s[,-1]
sparr_ns.num <- sparr_ns[,-1]

# calculate column means for both datasets
colMeans(sparr_s.num)
colMeans(sparr_ns.num)


# > colMeans(sparr_s.num)
#   Total_length    Alar_extent    L_beak_head     L_humerous
L_keel_sternum
#      157.38095      241.00000       31.43333       18.50000
20.80952
# > colMeans(sparr_ns.num)
#   Total_length    Alar_extent    L_beak_head     L_humerous
L_keel_sternum
#      158.42857      241.57143       31.47857       18.44643
20.83929
```

```
# Column means of Total_length and Alar_extent for non-survived sparrows
are greater than that of survived sparrows. Sparrows with longer total
length and greater alar extent may have a higher chance of surviving. Means
of other variables such as L_beak_head, L_humerous, and L_keel_sternum and
comparable for both groups.

# calculate the following for both datasets:Variance
var(sparr_s.num)
var(sparr_ns.num)

# > var(sparr_s.num)
#                Total_length Alar_extent L_beak_head L_humerous
L_keel_sternum
# Total_length      11.047619        9.10   1.5566667     0.8700
1.2861905
# Alar_extent        9.100000       17.50   1.9100000     1.3100
0.8800000
# L_beak_head        1.556667        1.91   0.5313333     0.1890
0.2396667
# L_humerous         0.870000        1.31   0.1890000     0.1760
0.1325000
# L_keel_sternum     1.286190        0.88   0.2396667     0.1325
0.5749048

# > var(sparr_ns.num)
#                Total_length Alar_extent L_beak_head L_humerous
L_keel_sternum
# Total_length      15.068783   17.190476   2.2428571  1.7460317
2.9306878
# Alar_extent       17.190476   32.550265   3.3978836  2.9502646
4.0656085
# L_beak_head        2.242857    3.397884   0.7284127  0.4695503
0.5590212
# L_humerous         1.746032    2.950265   0.4695503  0.4344312
0.5058862
# L_keel_sternum     2.930688    4.065608   0.5590212  0.5058862
1.3209921

# calculate the following for both datasets:Covariance matrix
cov(sparr_s.num)
cov(sparr_ns.num)

# > cov(sparr_s.num)
#                Total_length Alar_extent L_beak_head L_humerous
L_keel_sternum
# Total_length      11.047619        9.10   1.5566667     0.8700
1.2861905
# Alar_extent        9.100000       17.50   1.9100000     1.3100
0.8800000
# L_beak_head        1.556667        1.91   0.5313333     0.1890
0.2396667
# L_humerous         0.870000        1.31   0.1890000     0.1760
0.1325000
```

```
# L_keel_sternum     1.286190        0.88   0.2396667     0.1325
0.5749048
# > cov(sparr_ns.num)
#                Total_length Alar_extent L_beak_head L_humerous
L_keel_sternum
# Total_length      15.068783   17.190476   2.2428571  1.7460317
2.9306878
# Alar_extent       17.190476   32.550265   3.3978836  2.9502646
4.0656085
# L_beak_head        2.242857    3.397884   0.7284127  0.4695503
0.5590212
# L_humerous         1.746032    2.950265   0.4695503  0.4344312
0.5058862
# L_keel_sternum     2.930688    4.065608   0.5590212  0.5058862
1.3209921

# The diagonal values in the covariance matrices show that the non-survived
group has greater variance across all 5 variables. This increase in
variance may also contribute to higher covariance between variables such as
Total_length and Alar_extent.

# calculate the following for both datasets: Correlation matrix
cor(sparr_s.num)
cor(sparr_ns.num)

# > cor(sparr_s.num)
#                Total_length Alar_extent L_beak_head L_humerous
L_keel_sternum
# Total_length      1.0000000   0.6544674   0.6425068  0.6239195
0.5103557
# Alar_extent       0.6544674   1.0000000   0.6263698  0.7464418
0.2774378
# L_beak_head       0.6425068   0.6263698   1.0000000  0.6180476
0.4336368
# L_humerous        0.6239195   0.7464418   0.6180476  1.0000000
0.4165447
# L_keel_sternum    0.5103557   0.2774378   0.4336368  0.4165447
1.0000000
# > cor(sparr_ns.num)
#                Total_length Alar_extent L_beak_head L_humerous
L_keel_sternum
# Total_length      1.0000000   0.7761963   0.6769768  0.6824212
0.6568714
# Alar_extent       0.7761963   1.0000000   0.6978185  0.7845546
0.6200093
# L_beak_head       0.6769768   0.6978185   1.0000000  0.8347046
0.5698878
# L_humerous        0.6824212   0.7845546   0.8347046  1.0000000
0.6677936
# L_keel_sternum    0.6568714   0.6200093   0.5698878  0.6677936
1.0000000
```

```
# The correlation matrices are normalized by the variance of each
individual variable. The comparison of two correlation matrices shows that
in the non-survived group, the correlations between Total_length &
Alar_extent and Total_length & L_keel_sternum are much higher than the
survived group. It is possible that sparrows, given the same total length,
may have a lower chance of survival if it has greater Alar_extent and
L_keel_sternum.

# Calculate distance for survived sparrows
distance_s = dist(scale(sparr_s.num,center = FALSE))
distance_s
distance_s_matrix <- as.matrix(distance_s)
distance_s_matrix

x <- sparr_s.num
cm <- colMeans(x)
S <- cov(x)
d_s <- apply(x, MARGIN = 1, function(x)t(x - cm) %*% solve(S) %*% (x - cm))
d_s
# > d_s
#  [1]  3.398261  5.430132  2.761144  4.248201  2.310336  3.959346
2.220920 11.420509  5.062491  7.126981  3.495930  3.217816  4.112069
6.266423  7.583973
# [16]  1.860433  3.509591  3.456548  6.268125  6.168717  6.122055

# Calculate distance for non-survived sparrows
distance_ns = dist(scale(sparr_ns.num,center = FALSE))
distance_ns
distance_ns_matrix <- as.matrix(distance_ns)
distance_ns_matrix

x <- sparr_ns.num
cm <- colMeans(x)
S <- cov(x)
d_ns <- apply(x, MARGIN = 1, function(x)t(x - cm) %*% solve(S) %*% (x -
cm))
d_ns
# > d_ns
#  [1]  3.256952  1.233208  3.801387  4.665528  6.562139  3.536768
6.679629  9.198939  3.959222 16.246076  2.227248  2.205544 10.482729
4.903730  5.065758
# [16]  5.137205  1.821819  2.468433  6.459364  5.832981  2.269359
2.136120  2.147935  1.634513  4.529446  7.868488  4.752499  3.916982

qqnorm(sparr_s$Total_length, main = "Total_length_Survive");
qqline(sparr_s$Total_length)
qqnorm(sparr_s$Alar_extent, main = "Alar_extent_Survive");
qqline(sparr_s$Alar_extent)
qqnorm(sparr_s$L_beak_head, main = "L_beak_head_Survive");
qqline(sparr_s$L_beak_head)
qqnorm(sparr_s$L_humerous, main = "L_humerous_Survive");
qqline(sparr_s$L_humerous)
```

```
qqnorm(sparr_ns$Total_length, main = "Total_length_Nonsurvive");
qqline(sparr_ns$Total_length)
qqnorm(sparr_ns$Alar_extent, main = "Alar_extent_Nonsurvive");
qqline(sparr_ns$Alar_extent)
qqnorm(sparr_ns$L_beak_head, main = "L_beak_head_Nonsurvive");
qqline(sparr_ns$L_beak_head)
qqnorm(sparr_ns$L_humerous, main = "L_humerous_Survive");
qqline(sparr_ns$L_humerous)


##########################################################
#STOPS HERE
```