Assignment04 P02

Dawu Liu

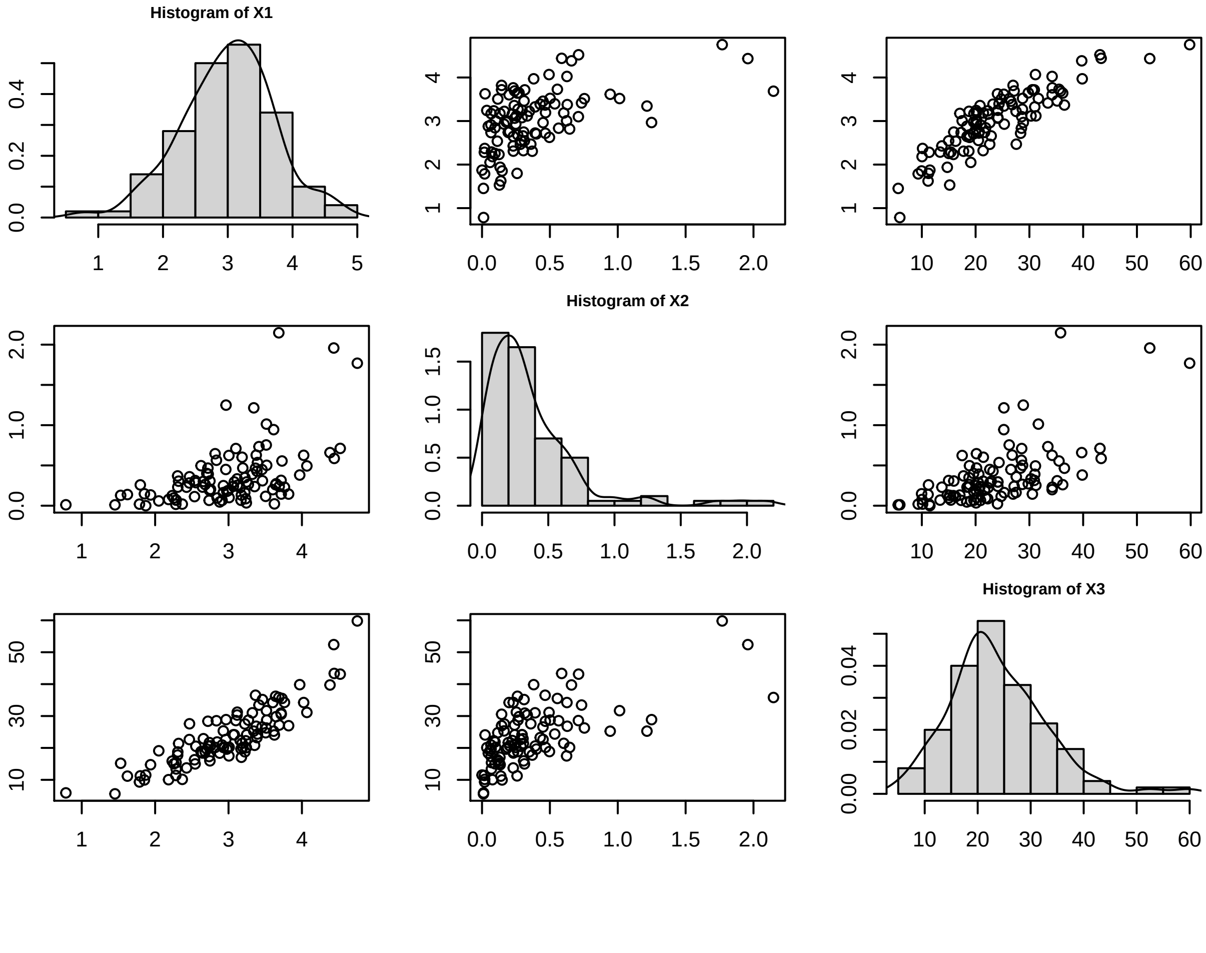
(a) Sample covariance matrix:

$$S_X = \begin{bmatrix} 0.506179723407075 & 0.139961753019714 & 5.72122771503319 \\ 0.139961753019714 & 0.143192401019117 & 2.26270794442784 \\ 5.72122771503319 & 2.26270794442784 & 86.0487659907118 \end{bmatrix} \tag{1}$$

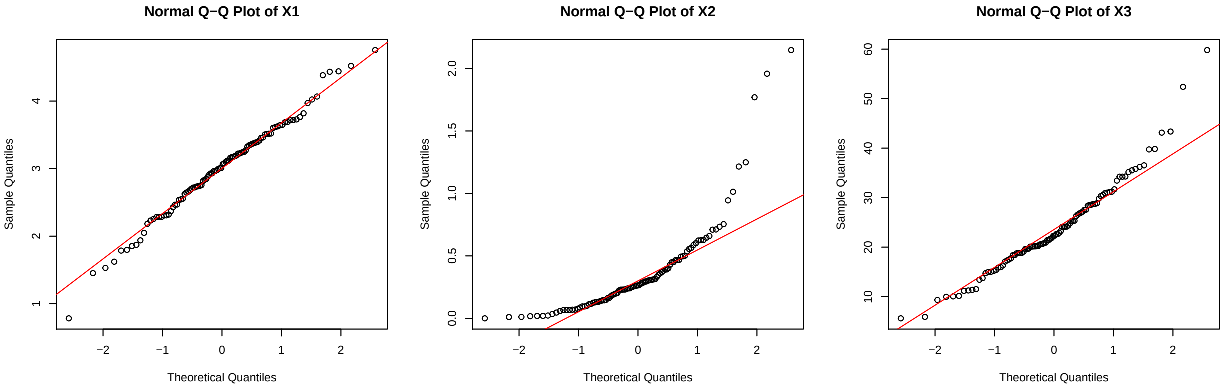
mean vector :

$$\tilde{\tilde{x}} = \begin{pmatrix} 2.9989239 \\ 0.3676117 \\ 23.7540326 \end{pmatrix} \tag{2}$$

(b) Matrix scatter plot with histograms being the diagonals:



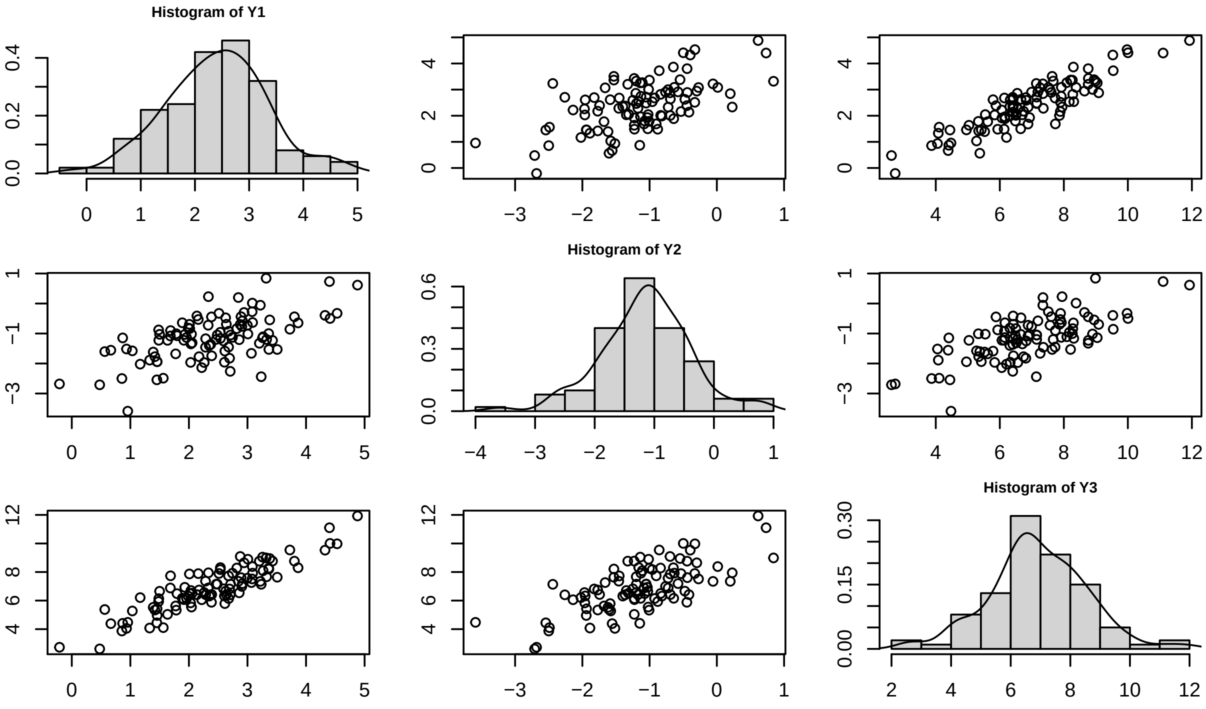
Univariate q-q plots:

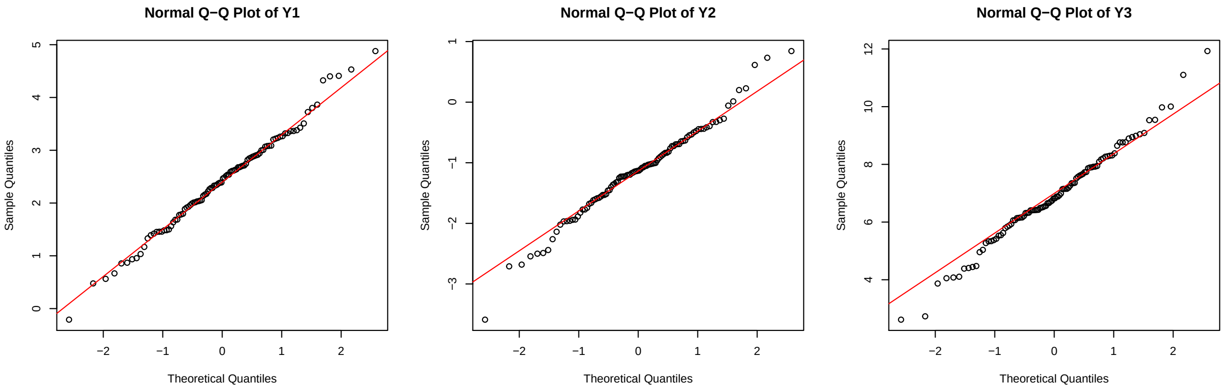


We can see in the scatter plots associated with the 2nd variable, data points appear to fan out from the origin, suggesting that those pairs of data are not bivariate normal. The histograms indicate the the 1st variable appears to be very close to normal, the 2nd variable is heavily right skewed, and the 3rd variable is a bit right screwed. The q-q plots also indicate the 1st variable appears to be close to normal as it fits the straight line closely, but the 2nd and 3rd variables do not as they don't fit the straight line well. The data are not consistent with a normal distribution.

(c) The parameter $\lambda's$ used in the Box-Cox transformations are **1.2626263, 0.2525253, 0.4545455.** As expected, the λ(power) for the 1st variable is close to one because the variable is close to normal. The 2nd variable has the strongest power change as it is strongly skewed. Scroll down to the code section if checking new data set Y is needed, since the matrix is very long.

(d) Matrix scatter plot and univariate q-q plots for the new data set Y:





The Box-Cox transformation improved the data set and made it consistent with a normal distribution or at least pretty close. The histograms of all three variables appear to be normal. The data points in matrix scatter plots for each of all three variables appear to fit in an ellipse, suggesting each pair of the the data is bivariate normal. Also in the q-q plots, the points in all three variables appear to fit the straight lines well, with only light-tailed for the 2nd and 3rd variables, which is probably due to the sample size.

(e) New sample covariance matrix:

$$S_Y = \begin{bmatrix} 0.878946352721069 & 0.416576516113662 & 1.36819113251324 \\ 0.416576516113662 & 0.561684889548861 & 0.851594673734675 \\ 1.36819113251324 & 0.851594673734675 & 2.6931318559889 \end{bmatrix} \tag{3}$$

New mean vector:

$$\tilde{\tilde{y}} = \begin{pmatrix} 2.407611 \\ -1.144121 \\ 6.911003 \end{pmatrix} \tag{4}$$

Subtract the new covariance matrix from the old covariance matrix, we get:

$$S_X - S_Y = \begin{bmatrix} -0.3728 & -0.2766 & 4.353 \\ -0.2766 & -0.4185 & 1.4111 \\ 4.353 & 1.4111 & 83.3556 \end{bmatrix} \tag{5}$$

Transforming the data affects the covariances between all three variables. The biggest change is the covariance between the 1st and 3rd variables, followed by the 2nd and 3rd. The variance of the 3rd variable itself has a massive decrease of 83.36 from 86.05.

Subtract the new mean vector from the old mean vector, we get:

$$\tilde{\tilde{x}} - \tilde{\tilde{y}} = \begin{pmatrix} 0.5913 \\ 1.5117 \\ 16.8430 \end{pmatrix} \tag{6}$$

The means of all three variables have decreased, the biggest change is the mean for the 3rd variable, which has a massive decrease of 16.84 from 23.75.

Code used to solve the questions(graphs are hidden):

```r
rm(list = ls())
library(forecast)
library(MESS)
```

```
X <- read.table("C:/Users/John/Desktop/STAT 445/Data/assignment4_data2.txt", sep = ",")
#a
S_X <- cov(X) ; x_bar <- colMeans(X)
S_X; x_bar
```

```
##           V1        V2        V3
## V1 0.5061797 0.1399618  5.721228
## V2 0.1399618 0.1431924  2.262708
## V3 5.7212277 2.2627079 86.048766
```

```
##         V1         V2         V3
##  2.9989239  0.3676117 23.7540326
```

```
#b
par(mfcol=c(3, 3), mar=c(2,2,2,2))
for(i in 1:3){
  for(j in 1:3) {
    if(i != j){
      plot(X[[i]],X[[j]],type="p", xlab = paste0("X", i), ylab = paste0("X", j))}
    else{
      hist(X[[i]], xlab = paste0("X", i), main = paste0("Histogram of X", i), prob=TRUE,cex.main=0.8)
      lines(density(X[[i]]))}
  }
}
```

```
# c
lam <- vector()
for (i in 1:3) {
  tmp <- MASS::boxcox(X[,i]~1,lambda=seq(-5,5,0.5))
  lam[i] <- tmp$x[which.max(tmp$y)]

}
lam
```

```
## [1] 1.2626263 0.2525253 0.4545455
```

```
Y1 <- BoxCox(X$V1,lam[1])
Y2 <- BoxCox(X$V2,lam[2])
Y3 <- BoxCox(X$V3,lam[3])
Y <- as.data.frame(cbind(Y1,Y2,Y3))
Y
```

```
##             Y1          Y2        Y3
## 1    3.4278072 -1.22897191  8.766276
## 2    2.4836620 -1.05161200  7.153235
## 3    2.0084743 -0.69348000  7.864556
## 4    1.6350794 -1.22802517  5.036414
## 5    0.9344915 -1.51606482  4.053675
## 6    2.5380214 -1.16158178  8.310536
## 7    2.0545752 -1.31153696  6.516047
## 8    2.7118195 -1.05086954  7.151431
## 9   0.9557388 -3.58911255  4.475735
## 10  3.8017466 -0.44231233  8.761314
## 11  2.8903507 -0.43897964  7.607403
## 12  0.8558199 -2.50016471  3.869489
## 13  3.5081207 -1.53136400  7.635394
## 14  1.7838939 -1.00768304  5.330740
## 15  2.6942076 -1.82379077  6.813074
## 16  2.6071855 -1.95765104  6.325405
## 17  1.4996085 -1.02716718  6.653031
## 18 -0.2099175 -2.68065428  2.732601
## 19  2.9419844 -0.29830842  8.645816
## 20  2.0293101 -1.96346767  5.839620
## 21  1.7737204 -1.67773095  5.622118
## 22  2.8176795 -0.83689826  8.279476
## 23  1.6832327 -1.08258282  6.881329
## 24  3.0876191 -0.63308389  7.923299
## 25  2.5354378 -0.95977837  8.171925
## 26  1.3908656 -1.61811108  5.527528
## 27  3.2670209 -1.10832054  8.092414
## 28  2.8768187 -0.69627377  9.088600
## 29  3.7240967 -0.85694795  9.543243
## 30  1.4856838 -1.22553394  6.135862
## 31  1.4831255 -0.87939288  5.933364
## 32  2.7059186 -2.26091671  6.403861
## 33  3.0831375  0.01335347  8.382306
## 34  3.2575416 -1.13931111  9.043472
## 35  2.3853996 -0.44569180  5.880935
## 36  2.1368162 -0.41448301  6.416384
## 37  3.0678273 -1.66096717  7.260870
## 38  4.5306614 -0.32618416  9.975581
## 39  3.3190541  0.84291910  8.988549
## 40  2.4663018 -1.19650843  7.151599
## 41  1.7968331 -1.07298487  6.486315
## 42  2.9037347 -0.76264692  6.991715
## 43  1.8867479 -0.64139652  6.160003
## 44  2.6271061 -0.47610930  6.661702
## 45  3.2033791 -1.32526957  8.761661
## 46  3.3624037 -1.00274316  8.265943
## 47  2.9983057 -0.72489236  7.568471
## 48  3.8644725 -0.64621516  8.298721
## 49  4.4005255  0.73255928 11.101139
## 50  0.5630283 -1.60273342  5.375269
## 51  0.6668850 -1.55662344  4.386843
## 52  1.4567561 -1.93764064  5.422455
## 53  0.4767737 -2.70966211  2.617036
## 54  2.0141223 -0.83500211  6.500498
## 55  2.0349978 -1.34626274  6.705528
## 56  2.6396905 -0.69114262  6.420603
## 57  2.6772849 -1.45153423  7.726613
## 58  1.6866813 -0.90187017  7.734776
## 59  1.5624769 -2.48925271  4.105504
## 60  3.2198223 -0.05683590  7.344907
## 61  1.9891081 -0.81730417  6.312401
## 62  1.0340336 -1.57838118  5.272968
## 63   2.3893279 -1.74375121  6.425073
## 64   2.5911160 -1.24874952  6.824666
## 65   1.9598895 -1.13397295  6.156636
## 66   3.0793961 -0.27249219  7.515870
## 67   2.6138147 -1.58641492  5.789298
## 68   2.1602707 -0.53255159  7.889395
## 69   2.3332209  0.22879611  7.945121
## 70   2.8594304 -1.20525851  6.544592
## 71   1.9294218 -1.03307088  6.930555
## 72   1.3301867 -1.88702779  4.078527
## 73   2.2187045 -2.13805806  6.059809
## 74   2.5127453 -0.32823770  7.897631
## 75   2.9140652 -0.57727745  7.190958
## 76   2.3475639 -1.39865289  6.314372
## 77   1.9134627 -1.22577545  6.067504
## 78   3.0071021 -1.01477145  8.896757
## 79   2.2617743 -1.96810502  6.561717
## 80   3.2346147 -2.43968159  7.140439
## 81   2.2861280 -1.17329938  6.413805
## 82   2.3721672 -1.36341435  6.407423
## 83   2.2840140 -1.45835131  7.359784
## 84   1.4581064 -1.94191460  4.957279
## 85   1.4526088 -2.54408477  4.443531
## 86   1.4196921 -1.77085986  5.339848
## 87   4.3246186 -0.39556946  9.530726
## 88   2.1731640 -1.77197238  6.741339
## 89   3.3800581 -0.54698554  8.946147
## 90   2.7437621 -1.12848644  7.917929
## 91   4.4097763 -0.49868764 10.004113
## 92   3.3654645 -1.53084117  8.210545
## 93   1.1679556 -2.02125902  6.208622
## 94   2.0442477 -1.02158284  5.545978
## 95   3.3267608 -1.19894148  7.668069
## 96   2.6794164 -0.92514708  6.143710
## 97   4.8791164  0.61390290 11.926300
## 98   2.3292478 -0.72373128  6.884544
## 99   0.8691482 -1.14640284  4.407538
## 100  2.8444294  0.19972283  7.348527
```

```r
# d
par(mfcol=c(3, 3), mar=c(2,2,2,2))
```

```r
for(i in 1:3){
  for(j in 1:3) {
    if(i != j){
      plot(Y[[i]],Y[[j]],type="p", xlab = paste0("Y", i), ylab = paste0("Y", j))}
    else{
      hist(Y[[i]], xlab = paste0("Y", i), main = paste0("Histogram of Y", i), prob=TRUE,cex.main=0.8)
      lines(density(Y[[i]]))}
  }
}
```

```
par(mfrow=c(1, 3), mar=c(2,2,2,2))
for (i in 1:3) {
  qqnorm(Y[[i]],main = paste0("Normal Q-Q Plot of Y",i));qqline(Y[[i]],col="red")
}
```

```
# e
S_Y <- cov(Y)
y_bar<- colMeans(Y)

S_Y; S_X-S_Y;
```

```
##           Y1        Y2        Y3
## Y1 0.8789464 0.4165765 1.3681911
## Y2 0.4165765 0.5616849 0.8515947
## Y3 1.3681911 0.8515947 2.6931319
```

```
##            V1         V2        V3
## V1 -0.3727666 -0.2766148  4.353037
## V2 -0.2766148 -0.4184925  1.411113
## V3  4.3530366  1.4111133 83.355634
```

```
y_bar;x_bar-y_bar;
```

```
##        Y1        Y2        Y3
##  2.407611 -1.144121  6.911003
```

```
##         V1         V2         V3
##  0.5913129  1.5117331 16.8430301
```