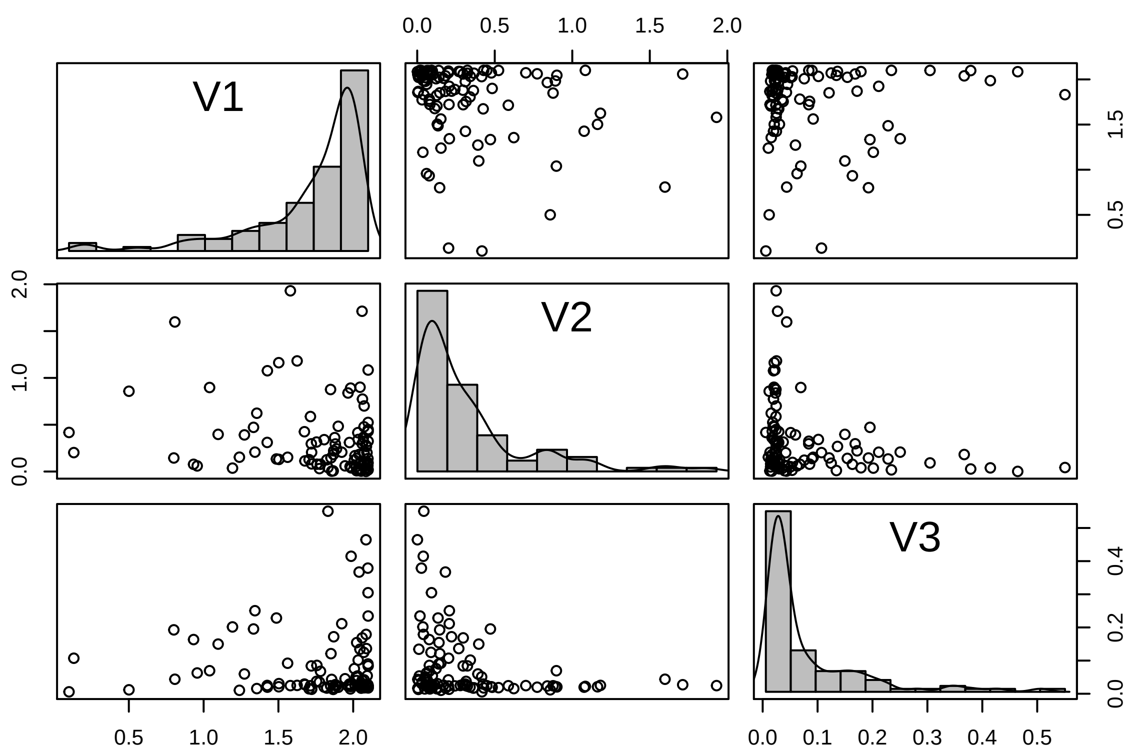
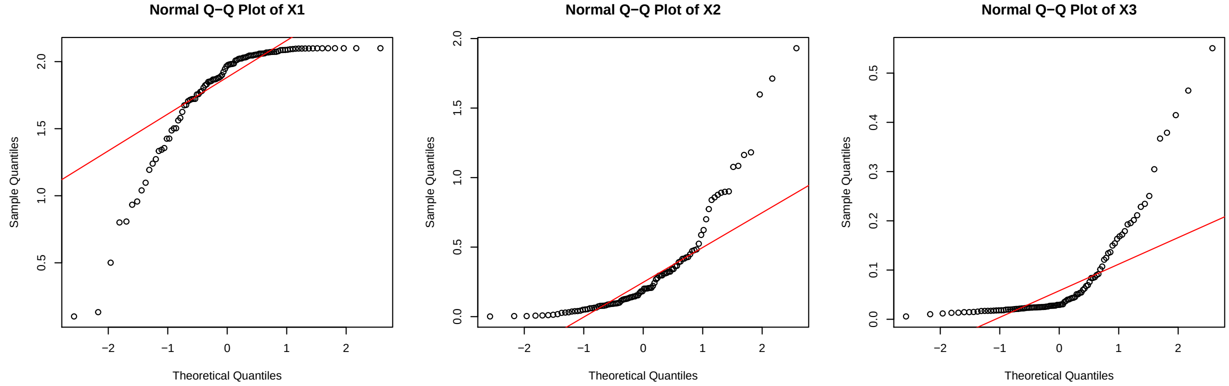
Assignment04 P03

Dawu Liu

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



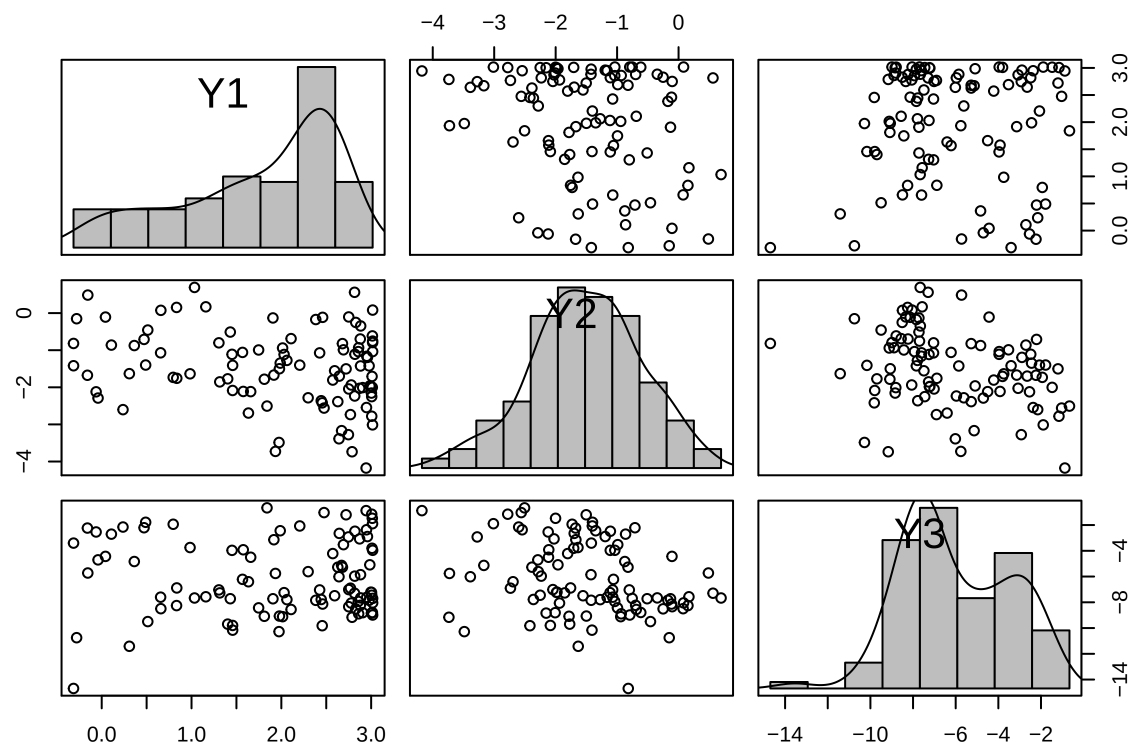
Univariate q-q plots:

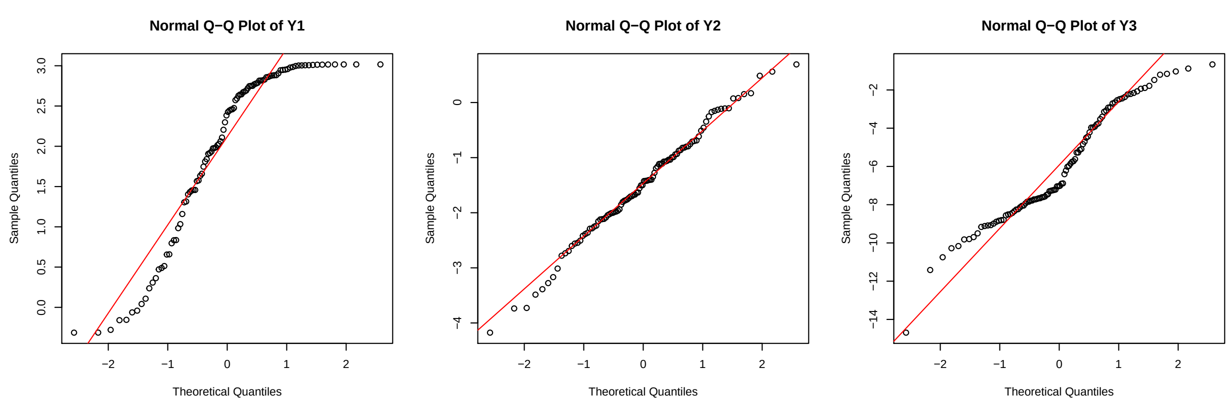


None of the scatter plots appears to be anywhere near an ellipse in shape as they all cluster in a corner and spread out along other two sides, this indicates all the pairs of the data are not bivariate normal. Both the q-q plots and histograms show all three variables are heavily skewed. Data points greatly curve in one direction on q-q plots(don't fit the straight line). The data set is not normal at all.

(b) The parameter $\lambda' s$ used in the Box-Cox transformations are **3.1818182, 0.1515152, -0.3535354.** Scroll down to the code section if checking new data set Y is needed, since the matrix is very long.

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





The 2nd variable appears to be normal as it fits the straight in the q-q plot, and the scatter plots of the 2nd vs 3rd variable appear to be close to ellipse in shape. The q-q plots indicate the 1st and 3rd variables are now less skewed than before as they fit the straight line better. The scatter plots associated with the 1st variable are now be more like ellipse in shape. Overall, the Box-Cox transformations improved the data to be more normal.

(d) No, the transformed data set is still not normal. The q-q plots indicates the data points in the 1st variable are still heavy-skewed. And the scatter plots associated with the 1st variable still show clusters on one side even as they were improved. For the data set to be normal, we need all three variables to be normal and each pair of the variables needs to be bivariate normal.

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

```
rm(list = ls())
library(forecast)
library(MESS)
X <- read.table("C:/Users/John/Desktop/STAT 445/Data/assignment4_data3.txt", sep = ",")
#a
pairs(X,diag.panel = panel.hist)
```

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

```
#b
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]  3.1818182  0.1515152 -0.3535354
```

```
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    2.87876406 -0.6981760  -8.2347660
## 2    3.01528152 -3.0128305  -1.8950724
## 3    1.74698878 -0.9939226  -8.4315592
## 4    2.62867195 -2.3862159  -5.2729084
## 5    0.65850300  0.0742149  -8.4955431
## 6    0.65574490 -1.0706441  -7.6009321
## 7    0.83559100 -1.7572385  -6.8840988
## 8    0.23684957 -2.6019424  -2.1546593
## 9    1.30466722 -0.8016301  -7.0367997
## 10  -0.15929306 -1.6753381  -2.2343340
## 11  -0.04094266 -2.2898226  -4.7041429
## 12   2.64184962 -3.3880211  -6.0162454
## 13   2.81662977 -2.2342879  -5.9676012
## 14   2.75020666 -0.1038484  -8.3403021
## 15   0.51299669 -0.4568406  -9.4972696
```

```
## 16   2.29825538 -2.2827365  -5.6191295
## 17   1.15917575  0.1689394  -7.5723265
## 18   2.06197690 -1.2747764  -7.7930756
## 19  -0.27947721 -0.1517113 -10.7488215
## 20   2.81920265 -1.1097819  -2.4805278
## 21  -0.15496989  0.4856856  -5.7249188
## 22   2.75034586 -2.0449888  -7.0132829
## 23  -0.06240343 -2.1216495  -2.5378355
## 24   0.10761686 -0.8621279  -2.7063178
## 25   0.36274014 -0.8746793  -4.8332537
## 26   1.45076288 -1.1110725  -3.9654450
## 27   2.99682523 -1.9781934  -7.2108334
## 28   0.48923483 -1.3998360  -1.7859661
## 29   3.00408308 -1.9867748  -7.2308275
## 30   1.43164802 -0.5109666  -7.7221949
## 31   3.01643379 -1.0370675  -3.9608906
## 32   1.90645323 -0.1309552  -7.7253259
## 33   1.93417541 -3.7279805  -5.7587991
## 34   2.68034199 -0.8226299  -5.2749810
## 35   0.30770582 -1.6325623 -11.4156527
## 36   1.65876884 -2.1166023  -4.5010307
## 37   2.44408636 -2.3620906  -7.7808466
## 38   1.56758508 -1.0593057  -6.2167842
## 39   2.03052422 -1.1155331  -7.2506342
## 40   2.88179469 -1.4248030  -5.8549519
## 41   0.04169576 -0.1072335  -4.4346831
## 42   1.45768045 -1.4086994 -10.1629606
## 43   3.00982596 -1.7065114  -3.8028550
## 44   2.90588743 -2.0082355  -8.8076993
## 45   2.88315921 -0.3465255  -7.6579024
## 46   2.20497903 -1.4026653  -2.0719441
## 47   2.95870050 -1.1936324  -2.8939791
## 48   2.76918002 -2.7355709  -6.9027115
## 49   2.82861169 -0.2517183  -8.5092023
## 50   2.47548049 -2.5596466  -1.0326420
## 51   2.94781422 -2.5449369  -2.3679782
## 52   2.42646379 -1.0726314  -7.0395881
## 53   1.91683574 -1.6707630  -3.1409070
## 54   1.03313973  0.6916886  -7.6654225
## 55   2.74686088 -3.2766613  -2.9244805
## 56   3.00620189 -2.2519008  -7.4507700
## 57   2.10820251 -0.6882980  -8.5578850
## 58   0.83342676  0.1525509  -8.2550776
## 59   2.77827337 -1.9372425  -8.0619737
## 60   2.01388100 -0.9410886  -9.1162314
## 61   1.45709365 -2.0858992  -9.7973809
## 62   2.97724672 -1.4223603  -7.8430527
## 63   3.01273277 -2.0017423  -1.4766838
## 64   2.98508597 -1.9625088  -5.0889799
## 65   3.01664831 -0.7545313  -7.6976503
## 66   2.64523295 -1.6984800  -2.6490204
## 67  -0.31378434 -1.4191155  -3.4021785
## 68   2.86076369 -0.9327983  -8.8928946
## 69   2.87160478 -2.0263834  -3.0749422
```

```
## 70   3.01372227 -0.6148030  -8.7912915
## 71   0.79593823 -1.7318692  -1.9388758
## 72   2.72068357 -1.5029342  -1.2027603
## 73   2.85732796 -1.0382481  -7.9330816
## 74   1.84038679 -2.5057915  -0.6644017
## 75   3.00307714 -2.1549963  -8.8390429
## 76   2.95088498 -1.1629508  -7.6111681
## 77   2.57257993 -1.8046461  -4.2158481
## 78   0.98346066 -1.6367613  -3.7469415
## 79   2.94396696 -4.1760769  -0.8807735
## 80   2.45965546 -0.1145105  -8.1355157
## 81   3.01647755  0.0813927  -8.0452464
## 82  -0.31407894 -0.8183412 -14.6913084
## 83   2.45496658 -2.4186782  -9.8179116
## 84   2.69160891 -0.9895653  -3.5230631
## 85   0.47124837 -0.7096255  -2.2095051
## 86   2.67110625 -3.1684538  -5.1356844
## 87   1.97414547 -3.4861936 -10.2775748
## 88   2.78752760 -3.7368822  -9.1624298
## 89   1.57593377 -2.1118969  -3.9204684
## 90   1.31432983 -1.8538581  -7.2704253
## 91   1.97875328 -1.5001943  -9.0677235
## 92   2.59326534 -1.5547599  -7.4905262
## 93   2.38320698 -0.1731790  -7.8456823
## 94   3.01663927 -0.7942624  -8.9875554
## 95   1.63365900 -2.6935954  -6.4037874
## 96   1.98747450 -1.3485015  -2.4425672
## 97   1.40259949 -1.7715308  -9.6951164
## 98   2.81514354  0.5602328  -7.2931781
## 99   3.00653964 -2.7798215  -1.1578442
## 100  1.81049193 -1.7823331  -9.0861055
```

```r
# c
pairs(Y,diag.panel = panel.hist)
```

```r
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")
}
```