Assignment # 3
Department of Mathematics & Statistics
STA 6746 Multivariate Statistical Analysis
Due: Monday, 10:00 PM of October 19, 2020

This assignment is based on chapter 5. Your hand writing (if not typed) need to be be clear. Some selected questions will be graded. However, to get full credit you must solve all questions and show all necessary work. Pl keeps margin I inch in all sides and empty spaces (1/2 inch) between two answers of the problems. This is your independent work. Your assignment # 3 is due on or before 10:00 PM of October 19 (Monday) and must mail to mmone014@fiu.edu and kibriag@fiu.edu

The **subject of the email and file name should be First name Last Name Sta 6746 Assignment #02 (say, Golam_Kibria_Sta6746_Assignment 02).** ***This is your cover page.***

First & Last Name:---Rachel Prokopius---**Panther ID:**---5274749---

Problem # 1. (Use software) The summary statistics of a bi-variate data with 41 observations are respectively given below

$$\bar{\mathbf{x}} = \begin{bmatrix} 0.550 \\ 0.625 \end{bmatrix} \text{ and } \mathbf{S} = \begin{bmatrix} 0.014 & 0.012 \\ 0.012 & 0.015 \end{bmatrix}$$

(a) Construct a 99% confidence ellipse for $\boldsymbol{\mu}$.
(b) Conduct a test of hypothesis H_0: $\boldsymbol{\mu}'=[0.56, 0.62]$ against H_a: $\boldsymbol{\mu}' \neq [0.56, 0.62]$ at 1% level of significance. Is your result consistent with the 99% confidence ellipse in part (a)? Give a brief expalnation.

Problem # 2. Exercise 5.7, page 261.

Problem # 3. Exercise 5.9, page 262.

Problem # 4. Exercise 5.18, page 267. (For question c use software)

Problem # 5. Exercise 5.20, page 268.

Problem # 6. Exercise 5.23, page 269. (Use software)

Problem # 7. Exercise 5.30, page 270. Consider n=50

Problem # 8.

A physical anthropologist performed a mineral analysis of nine ancient Peruvian hairs. The results for the chromium (x_1) and strontium (x_2) levels, in parts per million (ppm), were as follows:

x_1(Cr)	.48	40.53	2.19	.55	.74	.66	.93	.37	.22
x_2(St)	12.57	73.68	11.13	20.03	20.29	.78	4.64	.43	1.08

Source: Benfer and others, "Mineral Analysis of Ancient Peruvian Hair," *American Journal of Physical Anthropology*, **48**, no. 3 (1978), 277–282.

It is known that low levels (less than or equal to .100 ppm) of chromium suggest the presence of diabetes, while strontium is an indication of animal protein intake.

(a) Construct and plot a 90% joint confidence ellipse for the population mean vector $\boldsymbol{\mu}' = [\mu_1, \mu_2]$, assuming that these nine Peruvian hairs represent a random sample from individuals belonging to a particular ancient Peruvian culture.

(b) Obtain the individual simultaneous 90% confidence intervals for μ_1 and μ_2 by "projecting" the ellipse constructed in Part a on each coordinate axis. (Alternatively, we could use Result 5.3.) Does it appear as if this Peruvian culture has a mean strontium level of 10? That is, are any of the points (μ_1 arbitrary, 10) in the confidence regions? Is $[.30, 10]'$ a plausible value for $\boldsymbol{\mu}$? Discuss.

(c) Do these data appear to be bivariate normal? Discuss their status with reference to Q–Q plots and a scatter diagram. If the data are not bivariate normal, what implications does this have for the results in Parts a and b?

(d) Performs the Shapiro-Wilk test for bivariate nomality assumption.

(e) Repeat the anlaysis with the obvious "outlying" observation removed. Do the inferences change? Comment **PS**: You must give the necessary and concise interpretations of the questions.

Rachel Prokopius

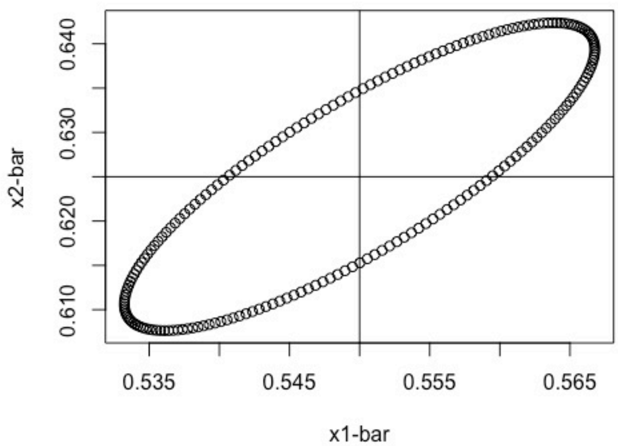
STA 6746

Assignment 3

19th October 2020

Problem 1:

a.



b.

The tsquared value (1.89) is less than the critical value (10.655), so we fail to reject the null hypothesis and at the 1% confidence level, mu is not significantly different from the numbers given for testing. This is consistent with the ellipse from part a because (0.56,0.62) falls within the 99% confidence interval.

Rcode for Problem 1:

```
mean1 = c(0.550,0.625)
> mean1
[1] 0.550 0.625
> cov1 = matrix(c(0.014,0.012,0.012,0.015), nrow = 2, ncol = 2)
> cov1
      [,1]  [,2]
[1,] 0.014 0.012
[2,] 0.012 0.015
> library(mixtools)
> library(matlib)
```

```
> ellipse(mean1,cov1, alpha = .01, npoints = 200, newplot = TRUE, draw = TRUE, xlab
= "x1-bar", ylab = "x2-bar")
> abline(v = 0.550)
> abline(h = 0.625)
> value1 = c(0.56, 0.62)
> dev1 = mean1 - value1
> matrixdev1 = matrix(dev1, nrow = 2, ncol = 1)
> matrixdev1
      [,1]
[1,] -0.010
[2,]  0.005
> transposematrixdev1 = t(matrixdev1)
> transposematrixdev1
     [,1]  [,2]
[1,] -0.01 0.005
> inversecov1 = inv(cov1)
> inversecov1
          [,1]      [,2]
[1,]  227.2727 -181.8182
[2,] -181.8182  212.1212
> tsquared1 = 41 *transposematrixdev1 %*% inversecov1 %*% matrixdev1
> qf(0.99, 2, 39)
[1] 5.194413
> (40*2)/39
[1] 2.051282
> tcrit1 = 2.051282 * qf(0.99,2,39)
> tcrit1
[1] 10.65521
> tsquared1 < tcrit1
     [,1]
[1,] TRUE
> abline(v = 0.56, h = 0.62)
```

Problem 2: Exercise 5.7

The simultaneous 95% T^2 confidence intervals for the data are as follows:

3.398 is less than/equal to mu1 is less than/equal to 5.882
35.052 is less than/equal to mu2 is less than/equal to 55.748
8.571 is less than/equal to mu3 is less than/equal to 11.359

The simultaneous 95% Bonferroni confidence intervals for the data are as follows:

4.071 is less than/equal to mu1 is less than/equal to 5.209
40.659 is less than/equal to mu2 is less than/equal to 50.141
9.326 is less than/equal to mu1 is less than/equal to 10.604

The Bonferroni intervals are slightly smaller than the T^2, so they are less conservative than the T^2

Rcode for Problem 2

```
sweat = table5.7[,2]
sweat
sodium = table5.7[,3]
potassium = table5.7[,4]
table5.7123 = cbind(sweat,sodium,potassium)
table5.7123
matrix5.7 = matrix(c(sweat, sodium, potassium), nrow = 20, ncol = 3)
matrix5.7
xbar2 = matrix(c(mean(sweat), mean(sodium), mean(potassium)), nrow = 3, ncol = 1)
xbar2
cov2 = cov(matrix5.7)
cov2
qf(0.95, 3, 17)
(3*19)/17
tcrit2 = qf(0.95, 3, 17) * (3*19)/17
tcrit2
sqrttcrit = sqrt(tcrit2)
sqrttcrit
sqrtss1n = sqrt(2.879368/20)
sqrtss1n
mu1lower = 4.640 - (sqrttcrit * sqrtss1n)
mu1lower
mu1upper = 4.640 + (sqrttcrit * sqrtss1n)
mu1upper

sqrtss2n = sqrt(199.7884/20)
sqrtss2n
mu2lower = 45.400 - (sqrttcrit * sqrtss2n)
mu2lower
mu2upper = 45.400 + (sqrttcrit * sqrtss2n)
mu2upper

sqrtss3n = sqrt(3.627658/20)
sqrtss3n
mu3lower = 9.965 - (sqrttcrit * sqrtss3n)
mu3lower
mu3upper = 9.965 + (sqrttcrit * sqrtss3n)
mu3upper

conmatrix5.7lower = matrix(c(3.398,35.052,8.571), nrow = 3, ncol = 1)
conmatrix5.7lower
conmatrix5.7upper = matrix(c(5.882,55.748,11.359), nrow = 3, ncol = 1)
```

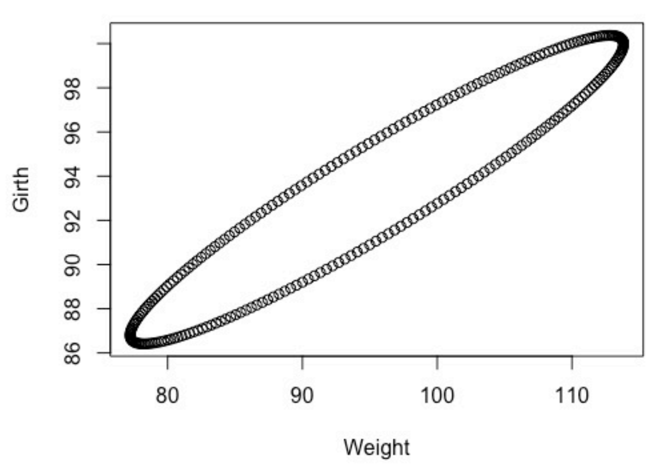
```
conmatrix5.7upper

tcrit0.05 = qt((1-0.05/2*3),19)
bonmu1lower = 4.640 - (tcrit0.05 * sqrtss1n)
bonmu1lower
bonmu1upper = 4.640 + (tcrit0.05 * sqrtss1n)
bonmu1upper

bonmu2lower = 45.400 - (tcrit0.05 * sqrtss2n)
bonmu2lower
bonmu2upper = 45.400 + (tcrit0.05 * sqrtss2n)
bonmu2upper

bonmu3lower = 9.965 - (tcrit0.05 * sqrtss3n)
bonmu3lower
bonmu3upper = 9.965 + (tcrit0.05 * sqrtss3n)
bonmu3upper
```

Problem 3: Exercise 5.9

a. The simultaneous 95% T^2 confidence intervals for the data are as follows:

69.553 is less than/equal to mu1 is less than/equal to 121.487
152.173 is less than/equal to mu2 is less than/equal to 176.587
49.607 is less than/equal to mu3 is less than/equal to 61.773
83.488 is less than/equal to mu4 is less than/equal to 103.292
16.547 is less than/equal to mu5 is less than/equal to 19.413
29.035 is less than/equal to mu6 is less than/equal to 33.225

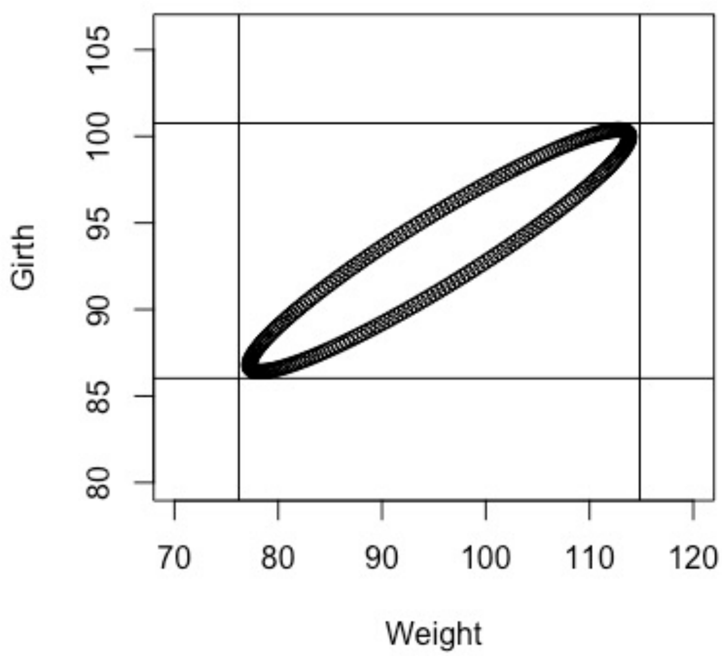
b. The 95% confidence ellipse for weight and girth is as follows:



c. **The simultaneous 95% Bonferroni confidence intervals for the data are as follows:**

76.201 is less than/equal to mu1 is less than/equal to 114.839
155.298 is less than/equal to mu2 is less than/equal to 173.462
51.164 is less than/equal to mu3 is less than/equal to 60.216
86.023 is less than/equal to mu4 is less than/equal to 100.757
16.914 is less than/equal to mu5 is less than/equal to 19.046
29.571 is less than/equal to mu6 is less than/equal to 32.689

d.



The Bonferroni minima and maxima for weight and girth are slightly higher and lower than the confidence ellipse from part b.

e. **The 95% Bonferroni confidence interval for mean head width minus mean head length is as follows:**

12.523 is less than/equal to (mu6-mu5) is less than/equal to 13.777

Rcode for Problem 3:

```
> sample5.9 = matrix(c(95.52,164.38,55.69,93.39,17.98,31.13), nrow = 6, ncol = 1)
> sample5.9
       [,1]
[1,]  95.52
[2,] 164.38
[3,]  55.69
[4,]  93.39
[5,]  17.98
[6,]  31.13
> cov5.9 =
matrix(c(3266.46,1343.97,731.54,1175.50,162.68,238.37,1343.97,721.91,324.25,537.
35,80.17,117.73,731.54,324.25,179.28,281.17,39.15,56.80,1175.50,537.35,281.17,47
4.98,63.73,94.85,162.68,80.17,39.15,63.73,9.95,13.88,
+                238.37,117.73,56.80,94.85,13.88,21.26), nrow = 6, ncol = 6)
> cov5.9
        [,1]    [,2]   [,3]    [,4]   [,5]   [,6]
[1,] 3266.46 1343.97 731.54 1175.50 162.68 238.37
[2,] 1343.97  721.91 324.25  537.35  80.17 117.73
[3,]  731.54  324.25 179.28  281.17  39.15  56.80
[4,] 1175.50  537.35 281.17  474.98  63.73  94.85
[5,]  162.68   80.17  39.15   63.73   9.95  13.88
[6,]  238.37  117.73  56.80   94.85  13.88  21.26
> sqrtchisquare5.9 = sqrt(qchisq(0.95,6))
> sqrtchisquare5.9
[1] 3.548463
> sqrts11n = sqrt(3266.46/61)
> sqrts22n = sqrt(721.91/61)
> sqrts33n = sqrt(179.28/61)
> sqrts44n = sqrt(474.98/61)
> sqrts55n = sqrt(9.95/61)
> sqrts66n = sqrt(21.26/61)
> simmatrix5.9 =
matrix(c(sqrtchisquare5.9*sqrts11n,sqrtchisquare5.9*sqrts22n,sqrtchisquare5.9*sqrts
33n,sqrtchisquare5.9*sqrts44n,sqrtchisquare5.9*sqrts55n,sqrtchisquare5.9*sqrts66n),
nrow = 6, ncol = 1)
> simmatrix5.9
          [,1]
[1,] 25.966535
[2,] 12.207222
[3,]  6.083328
[4,]  9.901773
[5,]  1.433134
[6,]  2.094869
> lower5.9 = sample5.9 - simmatrix5.9
```

```
> lower5.9
          [,1]
[1,]  69.55347
[2,] 152.17278
[3,]  49.60667
[4,]  83.48823
[5,]  16.54687
[6,]  29.03513
> upper5.9 = sample5.9 + simmatrix5.9
> upper5.9
          [,1]
[1,] 121.48653
[2,] 176.58722
[3,]  61.77333
[4,] 103.29177
[5,]  19.41313
[6,]  33.22487
> ## parb b
> sample5.9b = c(95.52,93.39)
> cov5.9b = matrix(c(3266.46,1175.50,1175.50,474.98), nrow = 2, ncol = 2)
> ellipse(sample5.9b,cov5.9b, alpha = 0.95, newplot = TRUE, xlab = "Weight", ylab
= "Girth")
> ## part c
> qnorm((1-0.95)/12)
[1] -2.638257
> zscorecrit5.9 = 2.64
> zscorecrit5.9
[1] 2.64
> bonmatrix5.9 =
matrix(c(zscorecrit5.9*sqrts11n,zscorecrit5.9*sqrts22n,zscorecrit5.9*sqrts33n,zscore
crit5.9*sqrts44n,zscorecrit5.9*sqrts55n,zscorecrit5.9*sqrts66n), nrow = 6, ncol = 1)
> bonmatrix5.9
          [,1]
[1,] 19.318690
[2,]  9.081980
[3,]  4.525900
[4,]  7.366763
[5,]  1.066229
[6,]  1.558550
> bonlower5.9 = sample5.9 - bonmatrix5.9
> bonlower5.9
          [,1]
[1,]  76.20131
[2,] 155.29802
[3,]  51.16410
[4,]  86.02324
```

```
[5,]  16.91377
[6,]  29.57145
> bonupper5.9 = sample5.9 + bonmatrix5.9
> bonupper5.9
          [,1]
[1,] 114.83869
[2,] 173.46198
[3,]  60.21590
[4,] 100.75676
[5,]  19.04623
[6,]  32.68855
> qnorm((1-0.95)/12)
[1] -2.638257
> ## part d
> sample5.9b = c(95.52,93.39)
> cov5.9b = matrix(c(3266.46,1175.50,1175.50,474.98), nrow = 2, ncol = 2)
> ellipse(sample5.9b,cov5.9b, alpha = 0.95, newplot = TRUE, xlab = "Weight", ylab
= "Girth", xlim = c(70,120), ylim = c(80,106))
> abline(v = 76.201,  h = 86.023)
> abline( v = 114.839,  h = 100.757)
> ## part e
> widthminuslength = 31.13 - 17.98
> zscore5.9e = -qnorm((1-0.95)/12)
> sqrt12n = sqrt((9.95-13.88-13.88+21.26)/61)
> sqrt12n
[1] 0.2378179
> coninterval5.9e = zscore5.9e * sqrt12n
> lower5.9e = widthminuslength - coninterval5.9e
> lower5.9e
[1] 12.52258
> upper5.9e = widthminuslength + coninterval5.9e
> upper5.9e
[1] 13.77742
```

Problem 4: Exercise 5.18

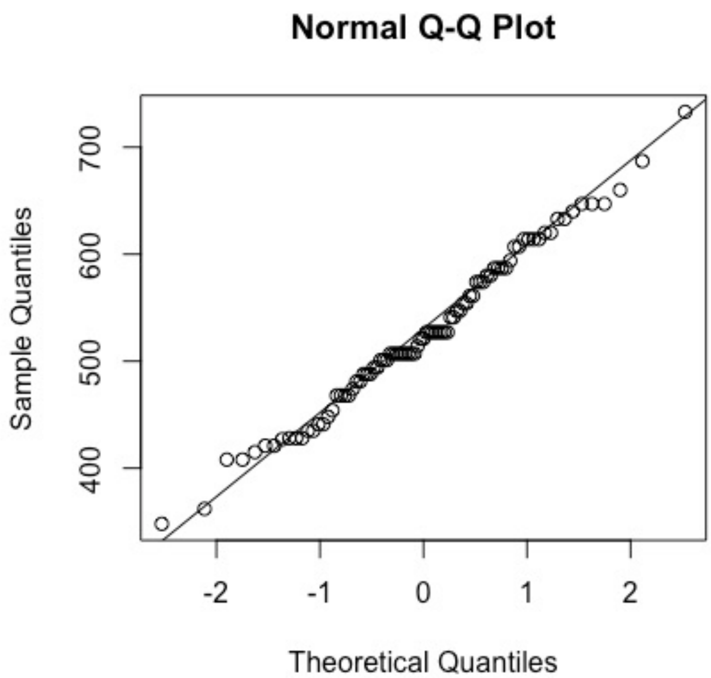
a. **The tsquared value (224.034) is greater than the critical value (83.33), so we reject the null hypothesis and at the 5% confidence level, mu is significantly different from the numbers given for testing. The t-test results show that mu is not equal to these numbers, so the students in the table are scoring differently than the average college students over the past 10 years.**

b. **The lengths and directions for the 95% confidence ellipsoids are as follows:**

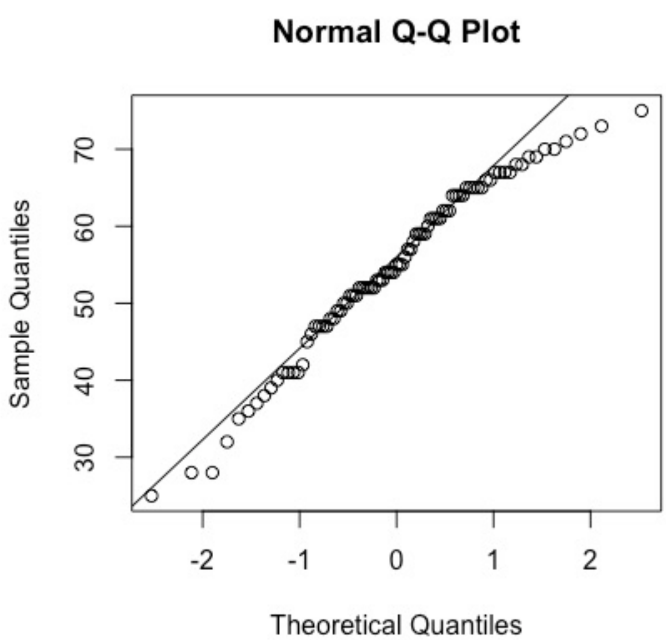
   Social science and history: length = 47.463, direction = 0.994, 0.104, 0.038
   Verbal: length = 4.966, direction = 0.105, -0.994, -0.012

Science: length = 2.365, direction = 0.037, -0.014, 0.999

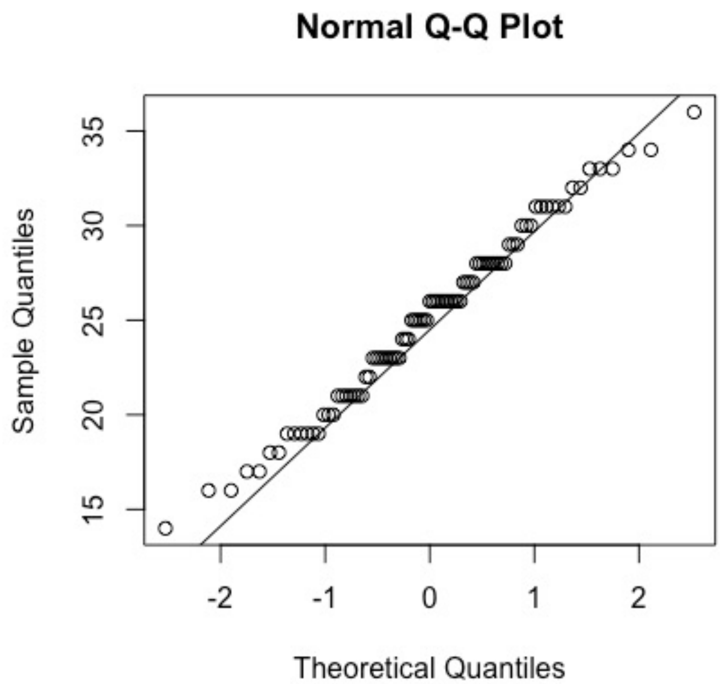
c. The QQ plot for social science and history is as follows:



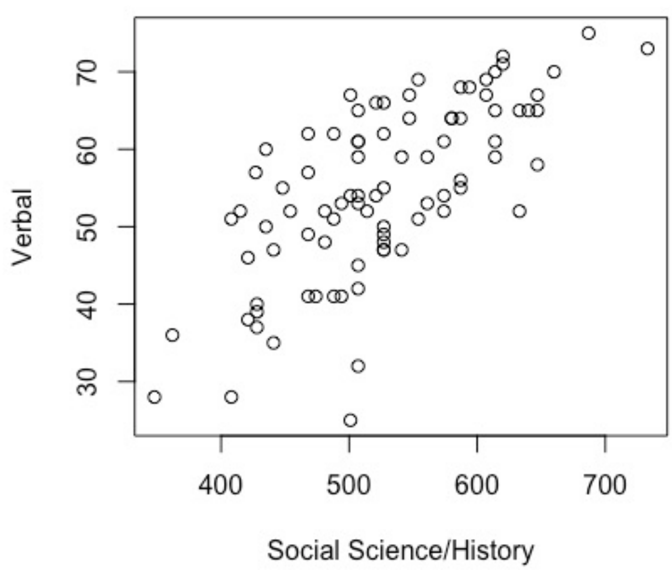
The QQ plot for verbal is as follows:



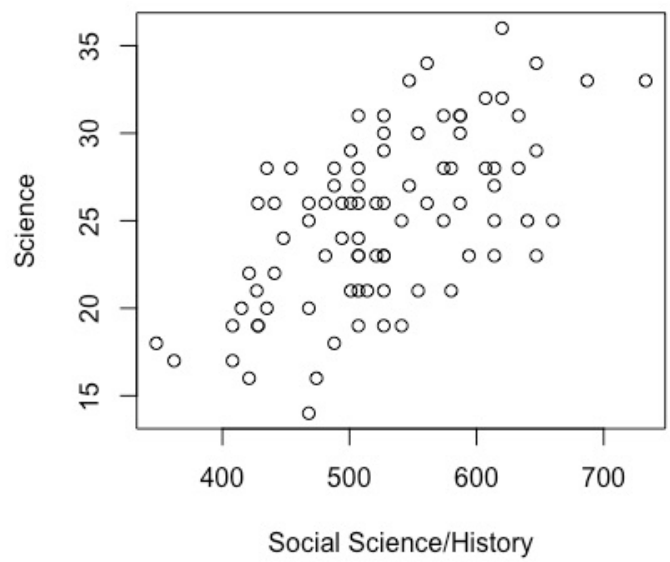
The QQ plot for science is as follows:



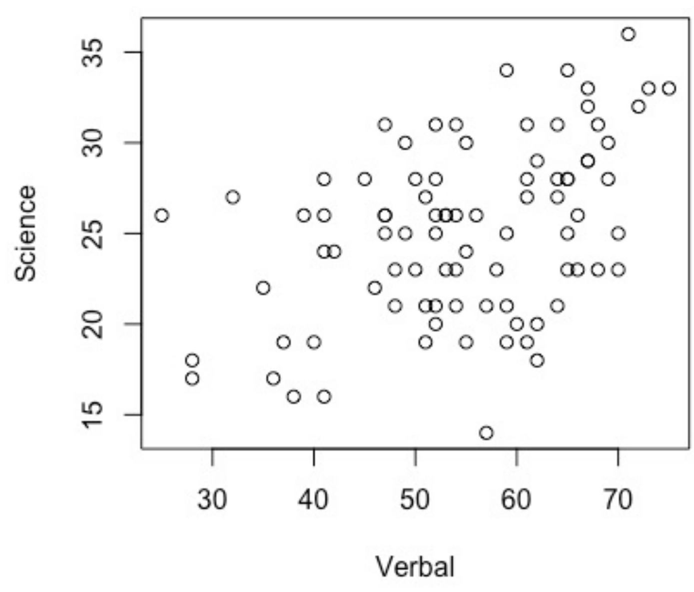
A scatter plot comparing social science and history to verbal is as follows:



A scatter plot comparing social science and history to science is as follows:



A scatter plot comparing verbal to science is as follows:



The data in general seems to be normally distributed, with a clear positive relationship between all of the variables judging by the scatter plots and fairly normal QQ plots, all following a linear progression. If any variable is

not normally distributed it would be verbal, judging by most of the data being on one side of the qqline in the corresponding QQ plot.

Rcode for Problem 4:

```
> sciencehistory = table4[,2]
> verbal = table4[,3]
> science = table4[,4]
> matrix5.18 = matrix(c(sciencehistory, verbal, science), nrow = 87, ncol = 3)
> matrix5.18
      [,1] [,2] [,3]
 [1,]  468   41   26
 [2,]  428   39   26
 [3,]  514   52   21
 [4,]  547   67   33
 [5,]  614   61   27
 [6,]  501   67   29
 [7,]  421   46   22
 [8,]  527   50   23
 [9,]  527   55   19
[10,]  620   72   32
[11,]  587   68   31
[12,]  541   59   19
[13,]  561   53   26
[14,]  468   62   20
[15,]  614   65   28
[16,]  527   48   21
[17,]  507   32   27
[18,]  580   64   21
[19,]  507   59   21
[20,]  521   54   23
[21,]  574   52   25
[22,]  587   64   31
[23,]  488   51   27
[24,]  488   62   18
[25,]  587   56   26
[26,]  421   38   16
[27,]  481   52   26
[28,]  428   40   19
[29,]  640   65   25
[30,]  574   61   28
[31,]  547   64   27
[32,]  580   64   28
[33,]  494   53   26
[34,]  554   51   21
```

[35,]	647	58	23
[36,]	507	65	23
[37,]	454	52	28
[38,]	427	57	21
[39,]	521	66	26
[40,]	468	57	14
[41,]	587	55	30
[42,]	507	61	31
[43,]	574	54	31
[44,]	507	53	23
[45,]	494	41	24
[46,]	541	47	25
[47,]	362	36	17
[48,]	408	28	17
[49,]	594	68	23
[50,]	501	25	26
[51,]	687	75	33
[52,]	633	52	31
[53,]	647	67	29
[54,]	647	65	34
[55,]	614	59	25
[56,]	633	65	28
[57,]	448	55	24
[58,]	408	51	19
[59,]	441	35	22
[60,]	435	60	20
[61,]	501	54	21
[62,]	507	42	24
[63,]	620	71	36
[64,]	415	52	20
[65,]	554	69	30
[66,]	348	28	18
[67,]	468	49	25
[68,]	507	54	26
[69,]	527	47	31
[70,]	527	47	26
[71,]	435	50	28
[72,]	660	70	25
[73,]	733	73	33
[74,]	507	45	28
[75,]	527	62	29
[76,]	428	37	19
[77,]	481	48	23
[78,]	507	61	19
[79,]	527	66	23
[80,]	488	41	28

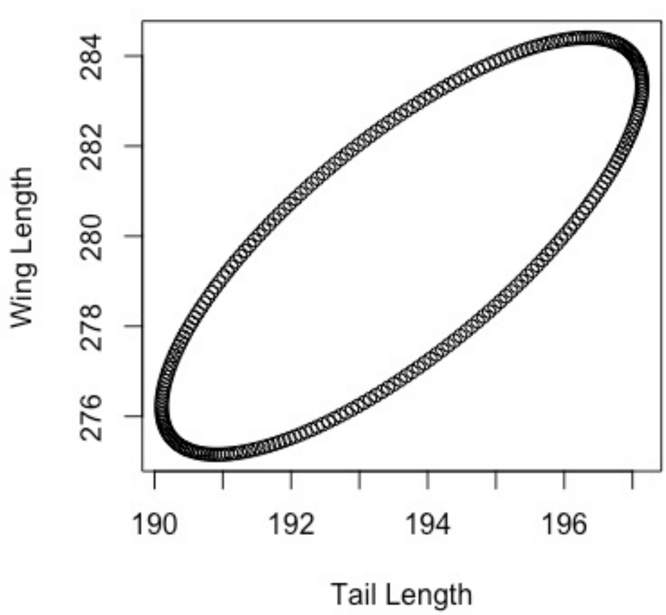
```
[81,]  607   69   28
[82,]  561   59   34
[83,]  614   70   23
[84,]  527   49   30
[85,]  474   41   16
[86,]  441   47   26
[87,]  607   67   32
> meansciencehistory = mean(sciencehistory)
> meanverbal = mean(verbal)
> meanscience = mean(science)
> mean5.18 = matrix(c(meansciencehistory, meanverbal, meanscience), nrow = 3,
ncol = 1)
> mean5.18
          [,1]
[1,] 526.58621
[2,]  54.73563
[3,]  25.12644
> cov5.18 = cov(matrix5.18)
> cov5.18
          [,1]     [,2]      [,3]
[1,] 5808.0593 601.4940 222.02967
[2,]  601.4940 127.3595  23.77800
[3,]  222.0297  23.7780  23.11173
> invcov5.18 = solve(cov5.18)
> invcov5.18
              [,1]         [,2]
[1,]  0.0004305755 -0.001561110
[2,] -0.0015611099  0.015378562
[3,] -0.0025303362 -0.000824634
             [,3]
[1,] -0.002530336
[2,] -0.000824634
[3,]  0.068424888
> dev5.18 = c(meansciencehistory-500, meanverbal-50, meanscience-30)
> dev5.18
[1] 26.586207  4.735632 -4.873563
> dev5.18transpose = t(dev5.18)
> ttest5.18 = 87*dev5.18transpose %*% solve(cov5.18) %*% dev5.18
> ttest5.18
         [,1]
[1,] 224.0341
> library(robustbase)
> library(pcaPP)
> library(rrcov)
> qf(0.95, 3, 84)
[1] 2.713227
```

```
> (86*3)/84
[1] 3.071429
> tcrit15.18 = (86*3)/84 * qf(0.95, 3, 84)
> tcrit15.18
[1] 8.333483
> T2.test(matrix5.18, mu = c(500,50,30), conf.level = 0.95, test = "f")

    One-sample Hotelling test

data:  matrix5.18
T2 = 224.034, F = 72.941, df1 = 3,
df2 = 84, p-value < 2.2e-16
alternative hypothesis: true mean vector is not equal to (500, 50, 30)'

sample estimates:
               [,1]     [,2]     [,3]
mean x-vector 526.5862 54.73563 25.12644

> ## Yes, the t-test results show that mu is not equal to these numbers,
> ## so the students in the table are scoring differently than the average
> ## college students over the past 10 years
> ## part b no f
> eigen5.18 = eigen(cov5.18)
> eigen5.18
eigen() decomposition
$values
[1] 5879.56342   64.37503   14.59216

$vectors
           [,1]        [,2]        [,3]
[1,] 0.99383877  0.10454897 -0.03679688
[2,] 0.10408080 -0.99446427 -0.01442193
[3,] 0.03810098 -0.01050322  0.99921869

> sqrtscihiseigen = sqrt(5879.56342)
> sqrtvereigen = sqrt(64.37503)
> sqrtscieigen = sqrt(14.59216)
> fstat5.18 = ((qf(0.95,3,84))*(3*(87-1)))/(87*(87-3))
> sqrt5.18 = sqrt(fstat5.18)
> lengths5.18 =
matrix(c(2*sqrtscihiseigen*sqrt5.18,2*sqrtvereigen*sqrt5.18,2*sqrtscieigen*sqrt5.18)
, nrow = 3, ncol = 1)
> lengths5.18
          [,1]
[1,] 47.463110
[2,]  4.966408
```

```
[3,]  2.364522
> eigenscihis = eigen5.18$vectors[,1]
> eigenscihis
[1] 0.99383877 0.10408080 0.03810098
> eigenver = eigen5.18$vectors[,2]
> eigenver
[1]  0.10454897 -0.99446427 -0.01050322
> eigensci = eigen5.18$vectors[,3]
> eigensci
[1] -0.03679688 -0.01442193  0.99921869
> ## part c
> scihistoryQQ = qqnorm(sciencehistory)
> qqline(sciencehistory)
> verQQ = qqnorm(verbal)
> qqline(verbal)
> sciQQ = qqnorm(science)
> qqline(science)
> plot(sciencehistory,verbal, xlab = "Social Science/History", ylab = "Verbal")
> plot(sciencehistory,science, xlab = "Social Science/History", ylab = "Science")
> plot(verbal,science, xlab = "Verbal", ylab = "Science")
```

Problem 5: Exercise 5.20

a. The 95% confidence ellipse for tail length and wing length is as follows:



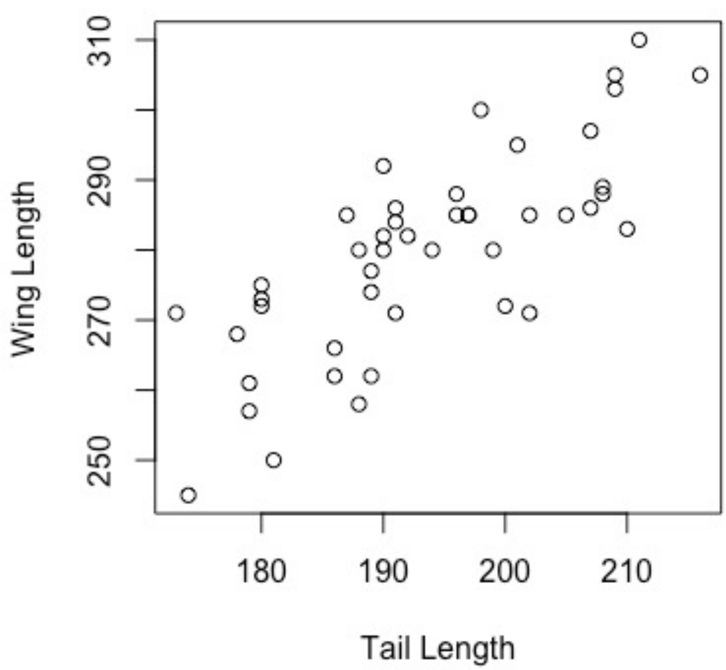
The tsquared value (5.54) is less than the critical value (6.58), so we fail to reject the null hypothesis and at the 5% confidence level, mu is not significantly different from the numbers given for testing. So, statistically male and female values are not different and the male values are plausible for estimating the female values, and vise versa. Though mu values are not in the ellipse pictured above, though it is close. Therefore, the t-test and the confidence ellipse do not exactly match up and further testing is required.

b. The simultaneous 95% T^2 confidence intervals for the data are as follows:
189.614 is less than/equal to mu1 is less than/equal to 197.631
274.508 is less than/equal to mu2 is less than/equal to 285.047

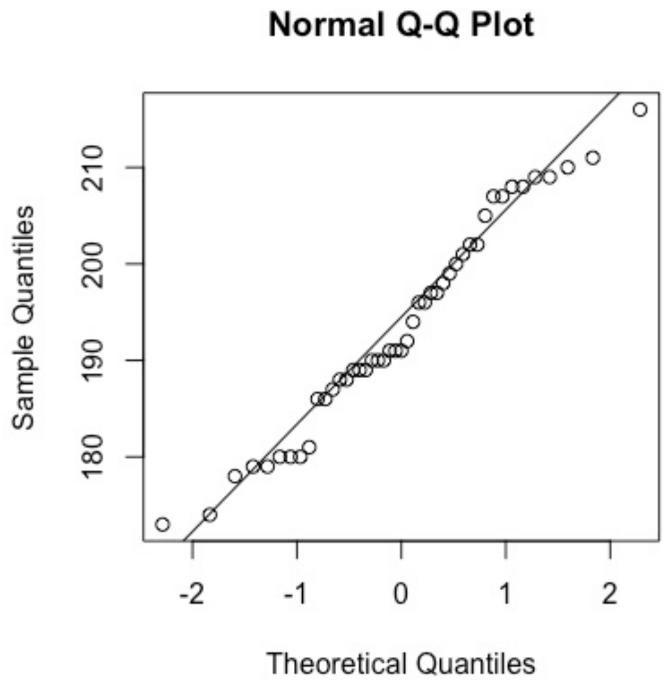
The simultaneous 95% Bonferroni confidence intervals for the data are as follows:
189.951 is less than/equal to mu1 is less than/equal to 197.293
274.953 is less than/equal to mu2 is less than/equal to 284.603

The T^2 intervals are larger and therefore more conservative than Bonferroni.

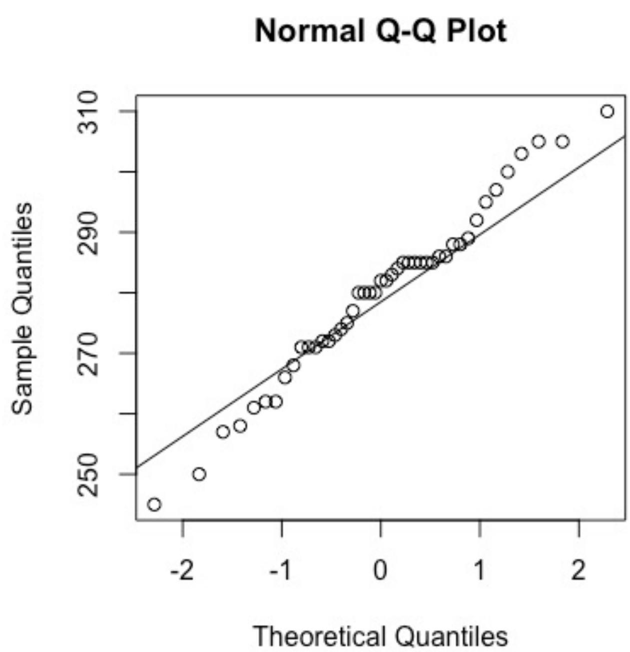
c. The following is a scatter diagram for comparing tail length and wing length



The following is a QQplot for tail length.



The following is a QQplot for wing length.



There seems to be a pretty linear relationship between the two variables in the scatter plot, and both variables follows a pretty linear relationship based

on their respective QQ plots, so a bivariate normal distribution seems to be a viable plot for this data.

Rcode for Problem 5:

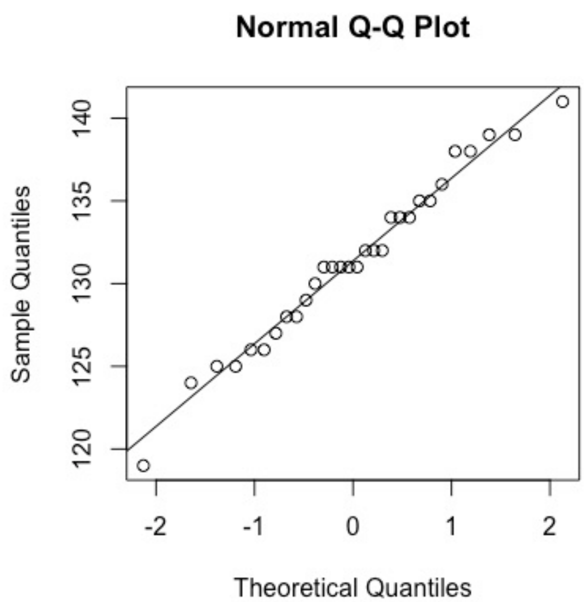
```
> meantaillength = mean(table5[,1])
> meantaillength
[1] 193.6222
> meanwinglength = mean(table5[,2])
> meanwinglength
[1] 279.7778
> mean5.20 = c(meantaillength,meanwinglength)
> mean5.20
[1] 193.6222 279.7778
> matrix5.20 = matrix(c(table5[,1],table5[,2]), nrow = 45, ncol = 2)
> cov5.20 = cov(matrix5.20)
> cov5.20
         [,1]     [,2]
[1,] 120.6949 122.3460
[2,] 122.3460 208.5404
> ellipse(mean5.20,cov5.20, alpha = 0.95, newplot = TRUE, xlab = "Tail Length",
ylab = "Wing Length")
> value5.20 = c(190,275)
> dev5.20 = mean5.20 - value5.20
> dev5.20
[1] 3.622222 4.777778
> matrixdev5.20 = matrix(dev5.20, nrow = 2, ncol = 1)
> matrixdev5.20
         [,1]
[1,] 3.622222
[2,] 4.777778
> transposematrixdev5.20 = t(matrixdev5.20)
> invcov5.20 = solve(cov5.20)
> tsquared5.20 = 45*transposematrixdev5.20 %*% invcov5.20 %*% matrixdev5.20
> tsquared5.20
        [,1]
[1,] 5.54313
> tcritinternal5.20 = qf(0.95, 2, 43)
> tcritinternal5.20
[1] 3.21448
> (2*44)/43
[1] 2.046512
> tcrit5.20 = 2.046512 * tcritinternal5.20
> tcrit5.20
[1] 6.578473
> tsquared5.20 < tcrit5.20
```

```
     [,1]
[1,] TRUE
> ## Fail to reject null hypothesis. So, statistically male and female values
> ## are not different and male values are plausible for female values
> ## part b
> sqrtchisq5.20 = sqrt(qchisq(0.95,2))
> sqrtchisq5.20
[1] 2.447747
> s11n5.20 = sqrt(120.6949/45)
> s22n5.20 = sqrt(208.5404/45)
> tcritmatrix5.20 = matrix(c(sqrtchisq5.20*s11n5.20,sqrtchisq5.20*s22n5.20), nrow
= 2, ncol = 1)
> tcritmatrix5.20
         [,1]
[1,] 4.008711
[2,] 5.269329
> meanmatrix5.20 = matrix(c(meantaillength,meanwinglength), nrow = 2, ncol = 1)
> mu5.20lower = meanmatrix5.20 - tcritmatrix5.20
> mu5.20lower
         [,1]
[1,] 189.6135
[2,] 274.5084
> mu5.20upper = meanmatrix5.20 + tcritmatrix5.20
> mu5.20upper
         [,1]
[1,] 197.6309
[2,] 285.0471
> zscore5.20 = -qnorm((1-0.95)/4)
> zscore5.20
[1] 2.241403
> bonmatrix5.20 = matrix(c(zscore5.20*s11n5.20, zscore5.20*s22n5.20), nrow = 2,
ncol = 1)
> bonlower5.20 = meanmatrix5.20 - bonmatrix5.20
> bonlower5.20
         [,1]
[1,] 189.9514
[2,] 274.9527
> bonupper5.20 = meanmatrix5.20 + bonmatrix5.20
> bonupper5.20
         [,1]
[1,] 197.2930
[2,] 284.6029
> ## Tsquared interval is wider than Bonferroni, so is less conservative
> ## part c
> library(matlib)
> QQ5.20x1 = qqnorm(table5[,1])
```

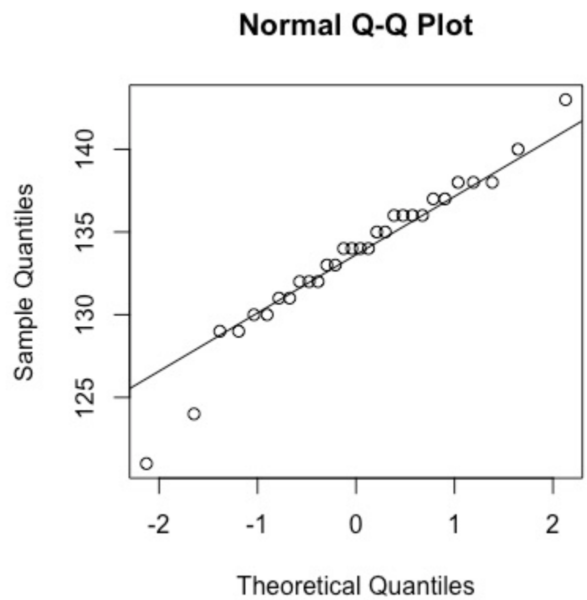
```
> qqline(table5[,1])
> QQ5.20x2 = qqnorm(table5[,2])
> qqline(table5[,2])
> plot(table5[,1], table5[,2], xlab = "Tail Length", ylab = "Wing Length")
> line=abline(0,1)
```

Problem 6: Exercise 5.23

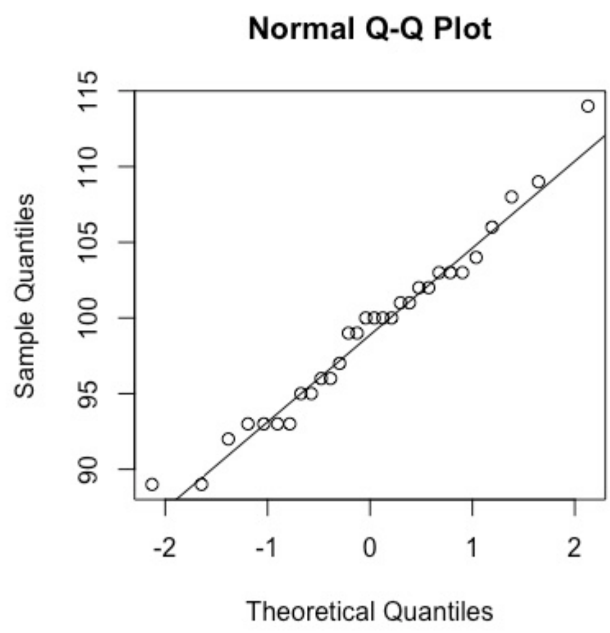
a. The following is a QQplot for maxbreath of Egyptian skulls:



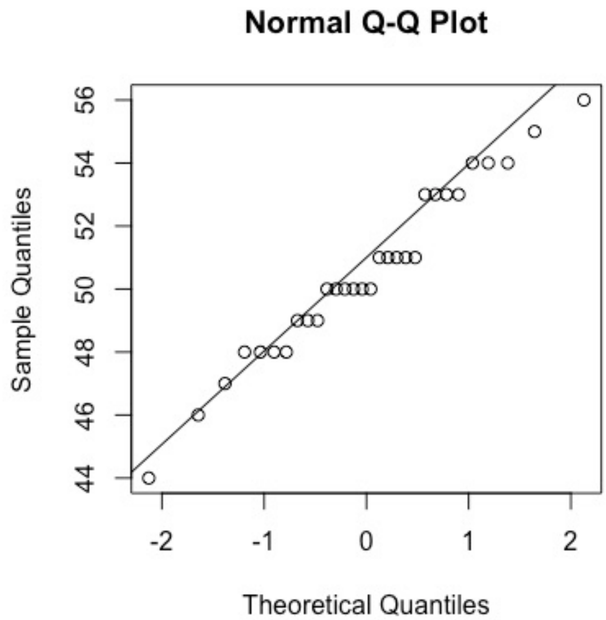
The following is a QQplot for basheight of Egyptian skulls:



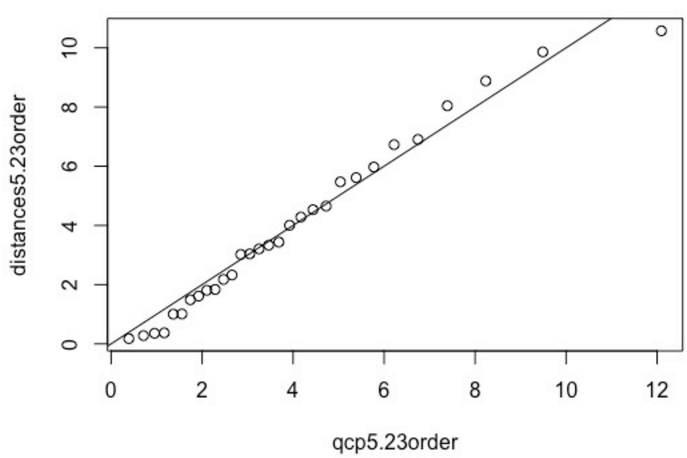
The following is a QQplot for baslength of Egyptian skulls:



The following is a QQplot for nasheight of Egyptian skulls:



The following is a chisquare plot is for data of Egyptian skulls:



All the QQ plots and the chi square plot follow a fairly straight line, so the data appear to be pretty normally distributed.

b. Analysis was conducted using small sample analysis because n-p was less than 30. The simultaneous 95% Bonferroni confidence intervals for the data are as follows:

128.873 is less than/equal to mu1 is less than/equal to 133.801
131.427 is less than/equal to mu2 is less than/equal to 135.773

96.305 is less than/equal to mu3 is less than/equal to 102.028
49.190 is less than/equal to mu4 is less than/equal to 51.877

The simultaneous 95% T^2 confidence intervals for the data are as follows:

128.091 is less than/equal to mu1 is less than/equal to 134.642
130.746 is less than/equal to mu2 is less than/equal to 136.454
95.409 is less than/equal to mu3 is less than/equal to 102.925
48.768 is less than/equal to mu4 is less than/equal to 52.298

As with previous problems, the Bonferroni intervals are slightly smaller than the T^2 intervals, making the T^2 intervals slightly more conservative.

Rcode for Problem 6:

```
maxbreath = table6[,1]
> basheight = table6[,2]
> baslength = table6[,3]
> nasheight = table6[,4]
> matrix5.23 = matrix(c(maxbreath,basheight,baslength,nasheight), nrow = 30, ncol =
> qqmaxbreath = qqnorm(table6[,1])
> qqline(table6[,1])
> qqbasheight = qqnorm(table6[,2])
> qqline(table6[,2])
> qqbaslength = qqnorm(table6[,3])
> qqline(table6[,3])
> qqnasheight = qqnorm(table6[,4])
> qqline(table6[,4])
> XX5.23 = cbind(maxbreath - mean(maxbreath), basheight - mean(basheight),
baslength-mean(baslength), nasheight-mean(nasheight))
> KK5.23 = (as.matrix(XX5.23) %*% solve(cov(matrix5.23)) %*%
t(as.matrix(XX5.23)))
> mKK5.23 = round(diag(KK5.23),4)
> mKK5.23
 [1]  4.5321  3.4359  0.1764  8.0418  5.9720  6.7272  8.8815  3.3321
 [9]  0.3571  0.2784  1.8359  9.8617  5.6144  0.3716  5.4713  2.3262
[17]  1.6127  1.8078  4.0060  1.0095  2.1748  4.2841  6.9045  1.0040
[25]  1.4885  3.0208  3.2070  3.0395 10.5731  4.6524
> J5.23 = seq(1:30)
> qcp5.23 = qchisq((30-J5.23+.5)/30,4)
> qcp5.23order = sort(qcp5.23)
> distances5.23order = sort(mKK5.23)
> distances5.23order
 [1]  0.1764  0.2784  0.3571  0.3716  1.0040  1.0095  1.4885  1.6127
 [9]  1.8078  1.8359  2.1748  2.3262  3.0208  3.0395  3.2070  3.3321
[17]  3.4359  4.0060  4.2841  4.5321  4.6524  5.4713  5.6144  5.9720
```

```
[25]  6.7272  6.9045  8.0418  8.8815  9.8617 10.5731
> plot(qcp5.23order,distances5.23order)
> line = abline(0,1)
> ## part b
> mean5.23 = matrix(c(mean(maxbreath), mean(basheight), mean(baslength), mean(nasheight)), nrow = 4, ncol = 1)
> mean5.23
          [,1]
[1,] 131.36667
[2,] 133.60000
[3,]  99.16667
[4,]  50.53333
> matrix5.23 = matrix(c(maxbreath,basheight,baslength,nasheight), nrow = 30, ncol = 4)
> cov5.23 = cov(matrix5.23)
> cov5.23
          [,1]       [,2]       [,3]       [,4]
[1,] 26.309195  4.1517241  0.4540230  7.2459770
[2,]  4.151724 19.9724138 -0.7931034  0.3931034
[3,]  0.454023 -0.7931034 34.6264368 -1.9195402
[4,]  7.245977  0.3931034 -1.9195402  7.6367816
> tcrit0.055.23 = qt((1-0.05/8), 29)
> bons11n5.23 = sqrt(26.309195/30)
> bons22n5.23 = sqrt(19.9724138/30)
> bons33n5.23 = sqrt(34.6264368/30)
> bons44n5.23 = sqrt(7.6367816/30)
> bonmatrix5.23 = matrix(c(tcrit0.055.23*bons11n5.23, tcrit0.055.23*bons22n5.23, tcrit0.055.23*bons33n5.23, tcrit0.055.23*bons44n5.23), nrow = 4, ncol = 1)
> bonlower5.23 = mean5.23 - bonmatrix5.23
> bonlower5.23
          [,1]
[1,] 128.87267
[2,] 131.42701
[3,]  96.30548
[4,]  49.18965
> bonupper5.23 = mean5.23 + bonmatrix5.23
> bonupper5.23
          [,1]
[1,] 133.86067
[2,] 135.77299
[3,] 102.02785
[4,]  51.87702
> tcrit6 = qf(0.95, 4, 26) * (4*29)/26
> sqrttcrit6 = sqrt(tcrit6)
> sqrttcrit6
[1] 3.498026
```

```
> tmatrix5.23 = matrix(c(sqrttcrit6*bons11n5.23, sqrttcrit6*bons22n5.23,
sqrttcrit6*bons33n5.23, sqrttcrit6*bons44n5.23), nrow = 4, ncol = 1)
> tlower5.23 = mean5.23 - tmatrix5.23
> tupper5.23 = mean5.23 + tmatrix5.23
> bonlower5.23
          [,1]
[1,] 128.87267
[2,] 131.42701
[3,]  96.30548
[4,]  49.18965
> bonupper5.23
          [,1]
[1,] 133.86067
[2,] 135.77299
[3,] 102.02785
[4,]  51.87702
> tlower5.23
          [,1]
[1,] 128.09088
[2,] 130.74584
[3,]  95.40858
[4,]  48.76844
> tupper5.23
          [,1]
[1,] 134.64246
[2,] 136.45416
[3,] 102.92475
[4,]  52.29822
```

Problem 7: Exercise 5.30

a. The separate simultaneous 95% Bonferroni confidence intervals for the data are as follows:

0.439 is less than/equal to mu1 is less than/equal to 1.093
0.242 is less than/equal to mu2 is less than/equal to 0.774
0.292 is less than/equal to mu3 is less than/equal to 0.584
0.088 is less than/equal to mu4 is less than/equal to 0.234

The total simultaneous 95% Bonferroni confidence interval for the data is as follows:

1.174 is less than/equal to the sum of mu is less than/equal to 2.572

The simultaneous 95% Bonferroni confidence interval for petroleum minus natural gas is as follows:

0.119 is less than/equal to (mu2-mu1) is less than/equal to 0.397

b. **The separate simultaneous 95% T^2 confidence intervals for the data are as follows:**

0.363 is less than/equal to mu1 is less than/equal to 1.169
0.180 is less than/equal to mu2 is less than/equal to 0.836
0.258 is less than/equal to mu3 is less than/equal to 0.618
0.071 is less than/equal to mu4 is less than/equal to 0.251

The total simultaneous 95% T^2 confidence interval for the data is as follows:

1.011 is less than/equal to the sum of mu is less than/equal to 2.735

The simultaneous 95% T^2 confidence interval for petroleum minus natural gas is as follows:

0.087 is less than/equal to (mu2-mu1) is less than/equal to 0.429

As with previous problems, the Bonferroni intervals are slightly smaller than the T^2 intervals, making the T^2 intervals slightly more conservative.

Rcode for Problem 7:

```
> mean5.30 = matrix(c(0.766,0.508,0.438,0.161), nrow = 4, ncol = 1)
> mean5.30
     [,1]
[1,] 0.766
[2,] 0.508
[3,] 0.438
[4,] 0.161
> cov5.30 =
matrix(c(0.856,0.635,0.173,0.096,0.635,0.568,0.127,0.067,0.173,0.128,0.171,0.039,0.096
,0.067,0.039,0.043), nrow = 4, ncol = 4)
> cov5.30
     [,1]  [,2]  [,3]  [,4]
[1,] 0.856 0.635 0.173 0.096
[2,] 0.635 0.568 0.128 0.067
[3,] 0.173 0.127 0.171 0.039
[4,] 0.096 0.067 0.039 0.043
> s11n5.30 = sqrt(0.856/50)
> s22n5.30 = sqrt(0.568/50)
> s33n5.30 = sqrt(0.171/50)
> s44n5.30 = sqrt(0.043/50)
```

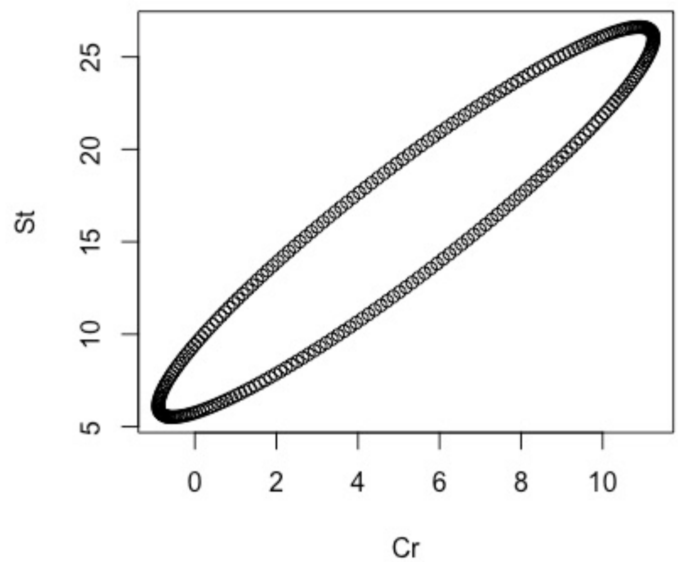
```
> innerzscore5.30 = (1-(0.05/(2*4)) )
> innerzscore5.30
[1] 0.99375
> ## zscore for 0.9938
> zscore5.30 = 2.50
> simmatrix5.30 = matrix(c(zscore5.30*s11n5.30,
zscore5.30*s22n5.30,zscore5.30*s33n5.30,zscore5.30*s44n5.30), nrow = 4, ncol = 1)
> simmatrix5.30
           [,1]
[1,] 0.32710854
[2,] 0.26645825
[3,] 0.14620192
[4,] 0.07331439
> bonlower5.30 = mean5.30 - simmatrix5.30
> bonlower5.30
           [,1]
[1,] 0.43889146
[2,] 0.24154175
[3,] 0.29179808
[4,] 0.08768561
> bonupper5.30 = mean5.30 +simmatrix5.30
> bonupper5.30
          [,1]
[1,] 1.0931085
[2,] 0.7744583
[3,] 0.5842019
[4,] 0.2343144
> ## part a total
> totalmean5.30 = matrix(0.766+0.508+0.438+0.161)
> covlist5.30 = c(cov5.30)
> covlist5.30
 [1] 0.856 0.635 0.173 0.096 0.635 0.568 0.127 0.067 0.173 0.128 0.171
[12] 0.039 0.096 0.067 0.039 0.043
> sumcovlist5.30 = sum(covlist5.30)
> sumcovlist5.30
[1] 3.913
> stotaln5.30 = sqrt(sumcovlist5.30/50)
> simmatrixtotal5.30 = matrix(c(zscore5.30*stotaln5.30), nrow = 1, ncol = 1)
> bontotallower5.30 = totalmean5.30 - simmatrixtotal5.30
> bontotallower5.30
         [,1]
[1,] 1.173625
> bontotalupper5.30 = totalmean5.30 + simmatrixtotal5.30
> bontotalupper5.30
         [,1]
[1,] 2.572375
```

```
> ## part a difference
> pertroleumminusnatural = 0.766 - 0.508
> zscore5.30 = 2.50
> sqrt12n5.30 = sqrt((0.856-0.635-0.635+0.568)/50)
> sqrt12n5.30
[1] 0.05549775
> coninterval5.30a = zscore5.30 * sqrt12n5.30
> bonlower5.30a = pertroleumminusnatural - coninterval5.30a
> bonlower5.30a
[1] 0.1192556
> bonupper5.930a = pertroleumminusnatural + coninterval5.30a
> bonupper5.930a
[1] 0.3967444
> ## part b mean for each
> chsquare5.30 = qchisq(0.95,4)
> chsquare5.30
[1] 9.487729
> sqrtchisq5.30 = sqrt(chsquare5.30)
> sinmatrix5.30b = matrix(c(sqrtchisq5.30*s11n5.30,
sqrtchisq5.30*s22n5.30,sqrtchisq5.30*s33n5.30,sqrtchisq5.30*s44n5.30), nrow = 4, ncol
= 1)
> sinmatrix5.30b
           [,1]
[1,] 0.40302596
[2,] 0.32829956
[3,] 0.18013338
[4,] 0.09032966
> mulower5.30 = mean5.30 - sinmatrix5.30b
> mulower5.30
           [,1]
[1,] 0.36297404
[2,] 0.17970044
[3,] 0.25786662
[4,] 0.07067034
> muupper5.30 = mean5.30 + sinmatrix5.30b
> muupper5.30
          [,1]
[1,] 1.1690260
[2,] 0.8362996
[3,] 0.6181334
[4,] 0.2513297
> ## parb b total
> totalmean5.30 = matrix(0.766+0.508+0.438+0.161)
> covlist5.30 = c(cov5.30)
> covlist5.30
 [1] 0.856 0.635 0.173 0.096 0.635 0.568 0.127 0.067 0.173 0.128 0.171
```

```
[12] 0.039 0.096 0.067 0.039 0.043
> sumcovlist5.30 = sum(covlist5.30)
> sumcovlist5.30
[1] 3.913
> stotaln5.30 = sqrt(sumcovlist5.30/50)
> simmatrixtotal5.30b = matrix(c(sqrtchisq5.30*stotaln5.30), nrow = 1, ncol = 1)
> mutotallower5.30 = totalmean5.30 - simmatrixtotal5.30b
> mutotallower5.30
        [,1]
[1,] 1.01131
> mutotalupper5.30 = totalmean5.30 + simmatrixtotal5.30b
> mutotalupper5.30
        [,1]
[1,] 2.73469
> ## part b difference
> pertroleumminusnatural = 0.766 - 0.508
> chsquare5.30 = qchisq(0.95,4)
> chsquare5.30
[1] 9.487729
> sqrt12n5.30 = sqrt((0.856-0.635-0.635+0.568)/50)
> sqrt12n5.30
[1] 0.05549775
> coninterval5.30b = sqrtchisq5.30 * sqrt12n5.30
> lower5.30b = pertroleumminusnatural - coninterval5.30b
> lower5.30b
[1] 0.08705496
> upper5.930b = pertroleumminusnatural + coninterval5.30b
> upper5.930b
[1] 0.428945
```

Problem 8:

a. **The 90% confidence ellipse for Cr and St is as follows:**

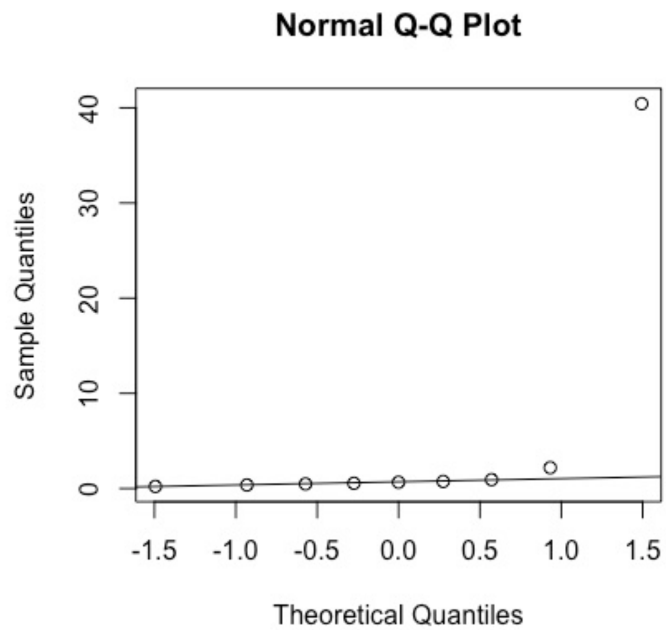


b. **The separate simultaneous 95% T^2 confidence intervals for the data are as follows:**

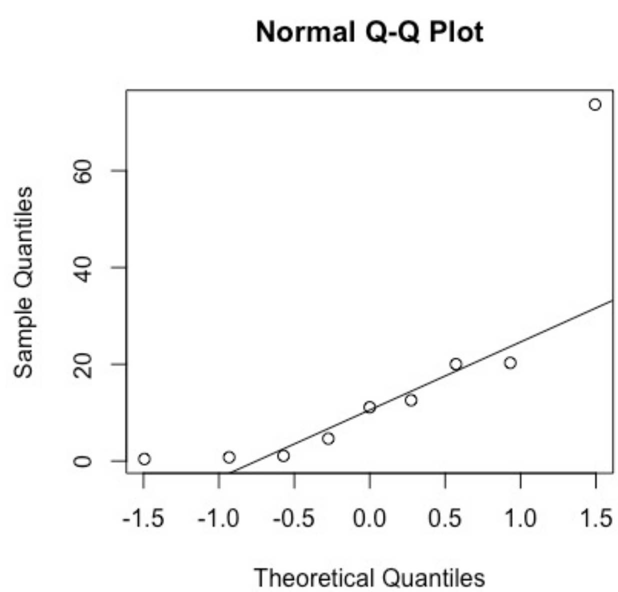
-6.862 is less than/equal to mu1 is less than/equal to 17.210
-4.828 is less than/equal to mu2 is less than/equal to 36.966

The tsquared value (1.775) is less than the critical value (7.44), so we fail to reject the null hypothesis and at the 10% confidence level, mu is not significantly different from the numbers given for testing. Based on the data, 10 for Sr is a plausible number.

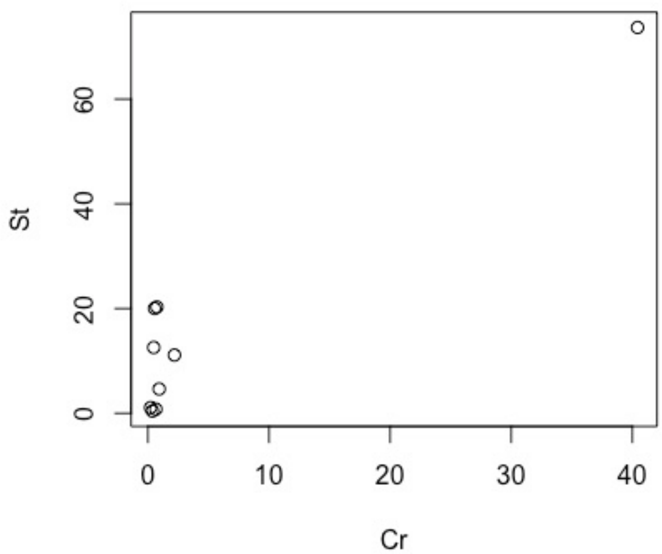
c. The QQ plot for Cr is as follows:



The QQ plot for St is as follows:



The scatterplot relating Cr and St is as follows:



The dataset has a big outlier, and based on the incredible skew of the QQplots and the scatter plot, this dataset is not normally distributed. Therefore, the confidence intervals and tests we performed in parts a and b of this problem are not accurate.

d. The results for the Shapiro-Wilks are as follows:
Shapiro-Wilk normality test

data: Z
W = 0.42294, p-value = 7.91e-07
The data are not normally distributed.

e. The results for the Shapiro-Wilks test without the outlier are as follows:
Shapiro-Wilk normality test

data: Z
W = 0.73601, p-value = 0.005713

The data still are not normally distributed.

Rcode for Problem 8:

```
> x18 = c(.48,40.43,2.19,.55,.74,.66,.93,.37,.22 )
> x28 = c(12.57,73.68,11.12,20.03,20.29,.78,4.64,.43,1.08)
> meanmatrix8 = matrix(c(mean(x18),mean(x28)), nrow = 2, ncol = 1)
> meanmatrix8
```

```
         [,1]
[1,]  5.174444
[2,] 16.068889
> mean8 = c(5.174444, 16.068889)
> matrix8 = matrix(c(x18,x28), nrow = 9, ncol = 2)
> matrix8
      [,1]  [,2]
 [1,]  0.48 12.57
 [2,] 40.43 73.68
 [3,]  2.19 11.12
 [4,]  0.55 20.03
 [5,]  0.74 20.29
 [6,]  0.66  0.78
 [7,]  0.93  4.64
 [8,]  0.37  0.43
 [9,]  0.22  1.08
> cov8 = cov(matrix8)
> cov8
         [,1]     [,2]
[1,] 175.1217 286.5248
[2,] 286.5248 527.8617
> ellipse(mean8,cov8, alpha = 0.90, newplot = TRUE, xlab = "Cr", ylab = "St")
> ##partb
> sqrtfstat8 = sqrt(((2*8)/7)*qf(0.90,2,7))
> sqrtfstat8
[1] 2.728659
> s11n8 = sqrt(175.1217/9)
> s22n8 = sqrt(527.8617/9)
> simmatrix8 = matrix(c(sqrtfstat8*s11n8,sqrtfstat8*s22n8), nrow = 2, ncol = 1)
> simmatrix8
         [,1]
[1,] 12.03644
[2,] 20.89720
> mulower8 = meanmatrix8 - simmatrix8
> mulower8
          [,1]
[1,] -6.861995
[2,] -4.828313
> muupper8 = meanmatrix8 + simmatrix8
> muupper8
         [,1]
[1,] 17.21088
[2,] 36.96609
> tcrit8 = ((2*8)/7) * qf(0.90,2,7)
> tcrit8
[1] 7.445582
```

```
> value8 = matrix(c(.30,10), nrow = 2, ncol = 1)
> dev8 = meanmatrix8-value8
> dev8
        [,1]
[1,] 4.874444
[2,] 6.068889
> tsquared8 = 9 * t(dev8) %*% solve(cov8) %*% dev8
> tsquared8
       [,1]
[1,] 1.7749
> tsquared8 < tcrit8
     [,1]
[1,] TRUE
> ## can't reject normality
> ## part c
> qqnorm(x18)
> qqline(x18)
> qqnorm(x28)
> qqline(x28)
> plot(x18,x28, xlab = "Cr", ylab = "St")
> table8 = cbind(x18,x28)
> table8
        x18   x28
 [1,]  0.48 12.57
 [2,] 40.43 73.68
 [3,]  2.19 11.12
 [4,]  0.55 20.03
 [5,]  0.74 20.29
 [6,]  0.66  0.78
 [7,]  0.93  4.64
 [8,]  0.37  0.43
 [9,]  0.22  1.08
> ## part d
> library(mvnormtest)
> mshapiro.test(t(table8))

      Shapiro-Wilk normality test

data:  Z
W = 0.42294, p-value = 7.91e-07

> ## do reject hypothesis of normality
> ## part e
> x18remove = c(.48,2.19,.55,.74,.66,.93,.37,.22 )
> x28remove = c(12.57,11.12,20.03,20.29,.78,4.64,.43,1.08)
> table8remove = cbind(x18remove,x28remove)
```

```
> mshapiro.test(t(table8remove))

	Shapiro-Wilk normality test

data:  Z
W = 0.73601, p-value = 0.005713
```