(10 points): Multivariate Normality Check

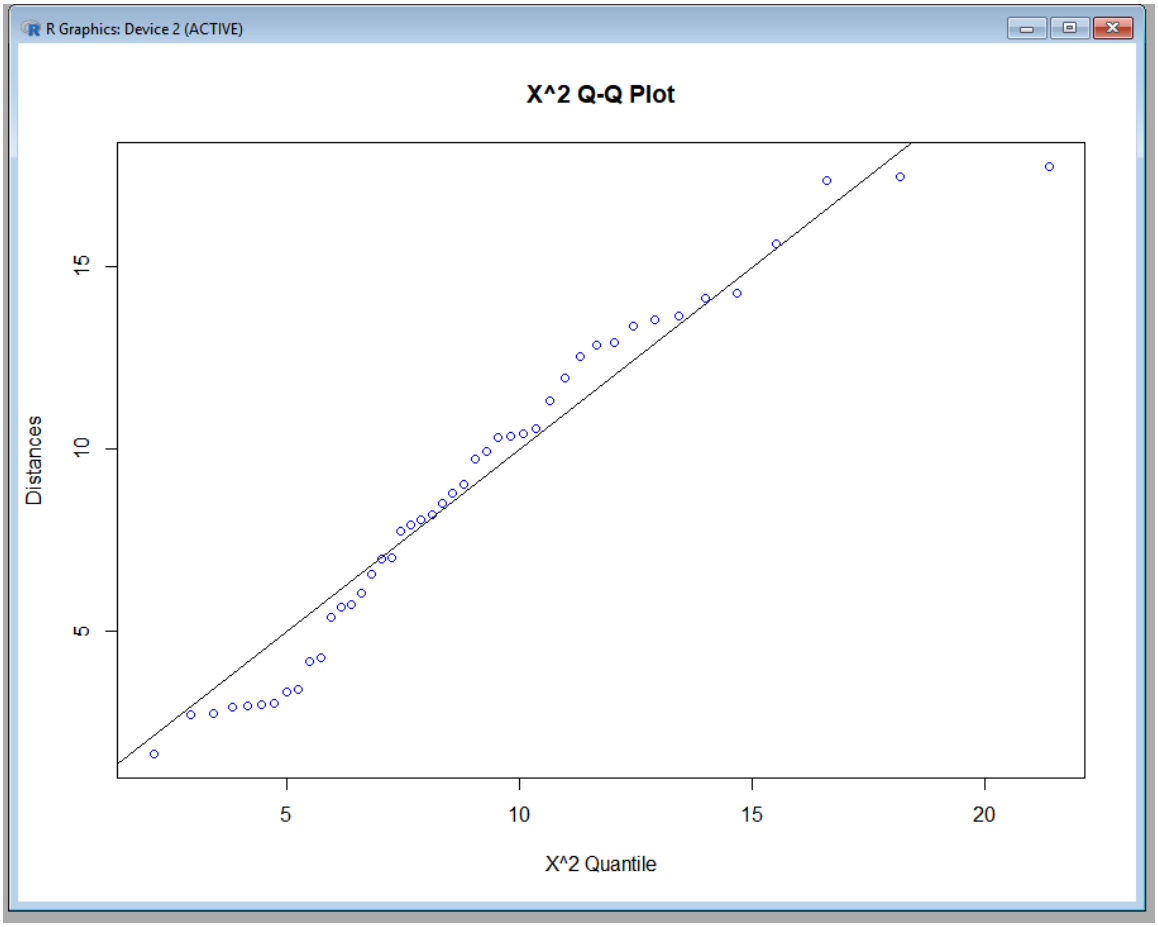
Using the pottery data of Table 1.3 in the text ('pottery.csv'), construct a 'chi-plot' for the 9 oxide columns in the dataset. (8 pts)

Does the plot indicate reasonable MVN or not? (2 pts)

```
> pottery<-read.csv('pottery.csv')

> numvars<-pottery[2:10]

> head(numvars)

  Al2O3 Fe2O3  MgO  CaO Na2O  K2O TiO2   MnO   BaO

1  18.8  9.52 2.00 0.79 0.40 3.20 1.01 0.077 0.015

2  16.9  7.33 1.65 0.84 0.40 3.05 0.99 0.067 0.018

3  18.2  7.64 1.82 0.77 0.40 3.07 0.98 0.087 0.014

4  16.9  7.29 1.56 0.76 0.40 3.05 1.00 0.063 0.019

5  17.8  7.24 1.83 0.92 0.43 3.12 0.93 0.061 0.019

6  18.8  7.45 2.06 0.87 0.25 3.26 0.98 0.072 0.017

> n<-nrow(numvars)

> p<-ncol(numvars)

> c(n,p)

[1] 45  9

> xmeans<-colMeans(numvars)

> xmeans

     Al2O3       Fe2O3        MgO        CaO       Na2O        K2O

15.70888889  5.75622222  2.48844444  0.51355556  0.24288889  3.20000000

      TiO2        MnO        BaO

 0.87666667  0.07051111  0.01651111

> S<-cov(numvars)

> S
```

```
          Al2O3       Fe2O3          MgO          CaO       Na2O
Al2O3  7.306282828 -0.907852020 -3.4490767677  0.2845131313 0.007860101
Fe2O3 -0.907852020  5.787928586  1.6480894444  0.7242160101 0.288911162
MgO   -3.449076768  1.648089444  3.0349497980 -0.1470693434 0.046438687
CaO    0.284513131  0.724216010 -0.1470693434  0.2063688889 0.041789495
Na2O   0.007860101  0.288911162  0.0464386869  0.0417894949 0.031771010
K2O   -1.408931818  1.263231818  1.2985022727  0.0217977273 0.049190909
TiO2   0.341734848 -0.063087879 -0.2151439394  0.0134666667 0.001359848
MnO   -0.071716010  0.075447202  0.0638964949  0.0030065505 0.004451444
BaO    0.002533990  0.001515611 -0.0003453232  0.0003376869 0.000194399
           K2O         TiO2          MnO          BaO
Al2O3 -1.40893182  0.3417348485 -7.171601e-02  2.533990e-03
Fe2O3  1.26323182 -0.0630878788  7.544720e-02  1.515611e-03
MgO    1.29850227 -0.2151439394  6.389649e-02 -3.453232e-04
CaO    0.02179773  0.0134666667  3.006551e-03  3.376869e-04
Na2O   0.04919091  0.0013598485  4.451444e-03  1.943990e-04
K2O    0.72714091 -0.0926318182  3.394068e-02  1.775000e-04
TiO2  -0.09263182  0.0323318182 -4.573712e-03  1.269697e-04
MnO    0.03394068 -0.0045737121  2.190346e-03  2.511919e-05
BaO    0.00017750  0.0001269697  2.511919e-05  8.891919e-06
> invS<-solve(S)
> invS
          Al2O3      Fe2O3        MgO         CaO        Na2O
Al2O3    0.43110342  -0.1592437   0.2703872    0.2967377   -0.1059702
Fe2O3   -0.15924367   1.1495063  -0.1315285   -3.1049820   -2.1688549
```

```
MgO     0.27038723  -0.1315285   2.4678105    1.9247588    1.6314249

CaO     0.29673770  -3.1049820   1.9247588   16.5240210   -3.1121218

Na2O    -0.10597024  -2.1688549   1.6314249   -3.1121218   80.5762121

K2O     0.07207112  -1.2977035  -2.6923939    1.2120849    4.1478309

TiO2    -1.34515823  -2.3928582   2.0382069    7.9879041  -11.4994029

MnO     8.75476442 -18.5924795 -18.7021362   45.5818667 -218.6042886

BaO   -101.12598328 122.2709248   9.9144747 -307.0899722 -481.2207081

            K2O       TiO2         MnO          BaO

Al2O3    0.07207112   -1.345158     8.754764   -101.125983

Fe2O3   -1.29770355   -2.392858   -18.592479    122.270925

MgO     -2.69239393    2.038207   -18.702136     9.914475

CaO      1.21208489    7.987904    45.581867   -307.089972

Na2O     4.14783085  -11.499403  -218.604289   -481.220708

K2O     10.56761566    3.776921   -38.086639   -197.910002

TiO2     3.77692105   80.087398   108.131009   -706.043053

MnO    -38.08663882  108.131009  3205.336320  -6842.638026

BaO   -197.91000168 -706.043053 -6842.638026 176369.849839

> d<-apply(numvars, 1, function(x){t(x-xmeans)%*%invS%*%(x-xmeans)})

> layout(1)

> plot(qchisq((1:n-0.5)/n, df=p), sort(d), xlab='X^2 Quantile',

+ ylab='Distances',main='X^2 Q-Q Plot', col='blue')

> abline(a=0,b=1)

>
```



The plot indicate reasonable MVN

(20 points): MANOVA For the 'pottery' data of Table 1.3 in the text:

Perform a MANOVA for the 1-way means model of the 9 oxide response variables vs. the categorical factor 'kiln'. Compute the Wilk's lambda statistic. (10 pts)

Does the 'kiln' factor have a detectable effect on the mean values (i.e., different means for different kilns)? (Answer using the P-value for the Wilk's statistic.) (2 pts)

Show the ANOVA results for each of the 9 response variables vs. kiln. Which of them have detectable differences among the means for different kilns? (Answer using P-values.) (9 pts)

> manova_pottery<- manova(data.matrix(numvars) ~ pottery$kiln)

> summary(manova_pottery, test='Wilks')

Df Wilks approx F num Df den Df Pr(>F)

pottery$kiln 1 0.070027 51.645 9 35 < 2.2e-16 *

Residuals 43**

Signif. codes: 0 ‘*’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

>

> summary.aov(manova_pottery)

Response Al2O3 :

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
pottery$kiln	1	2.74	2.7362	0.3691	0.5467
Residuals	43	318.74	7.4126		

Response Fe2O3 :

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
pottery$kiln	1	224.473	224.473	319.66	< 2.2e-16 ***
Residuals	43	30.196	0.702		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Response MgO :

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
pottery$kiln	1	10.377	10.3767	3.6229	0.0637 .
Residuals	43	123.161	2.8642		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Response CaO :

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
pottery$kiln	1	5.1402	5.1402	56.098	2.531e-09 ***
Residuals	43	3.9400	0.0916		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Response Na2O :

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
pottery$kiln	1	0.62076	0.62076	34.346	5.831e-07	***
Residuals	43	0.77716	0.01807			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Response K2O :

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
pottery$kiln	1	7.1445	7.1445	12.363	0.001046	**
Residuals	43	24.8497	0.5779			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Response TiO2 :

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
pottery$kiln	1	0.02827	0.028267	0.8717	0.3557
Residuals	43	1.39433	0.032426		

Response MnO :

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
pottery$kiln	1	0.030734	0.0307338	20.133	5.324e-05	***
Residuals	43	0.065641	0.0015265			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Response BaO :

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
pottery$kiln	1	0.00001254	1.2539e-05	1.4237	0.2393
Residuals	43	0.00037871	8.8071e-06		

>

The MANOVA for all 9 responses is highly significant (P-value < 0.0001), the 'kiln' factor have a detectable effect on the mean values.

The individual ANOVAs show that Fe2O3 , CaO, Na2O , K2O , MnO have significant differences among the means for different kilns, but Al2O3 (P-value=0.5467), MgO (P-value=0.0637), TiO2 (P-value=0.3557), BaO(P-value=0.2393) don't.