Homework 3

Luo Bingjun 2017013573 Software 71

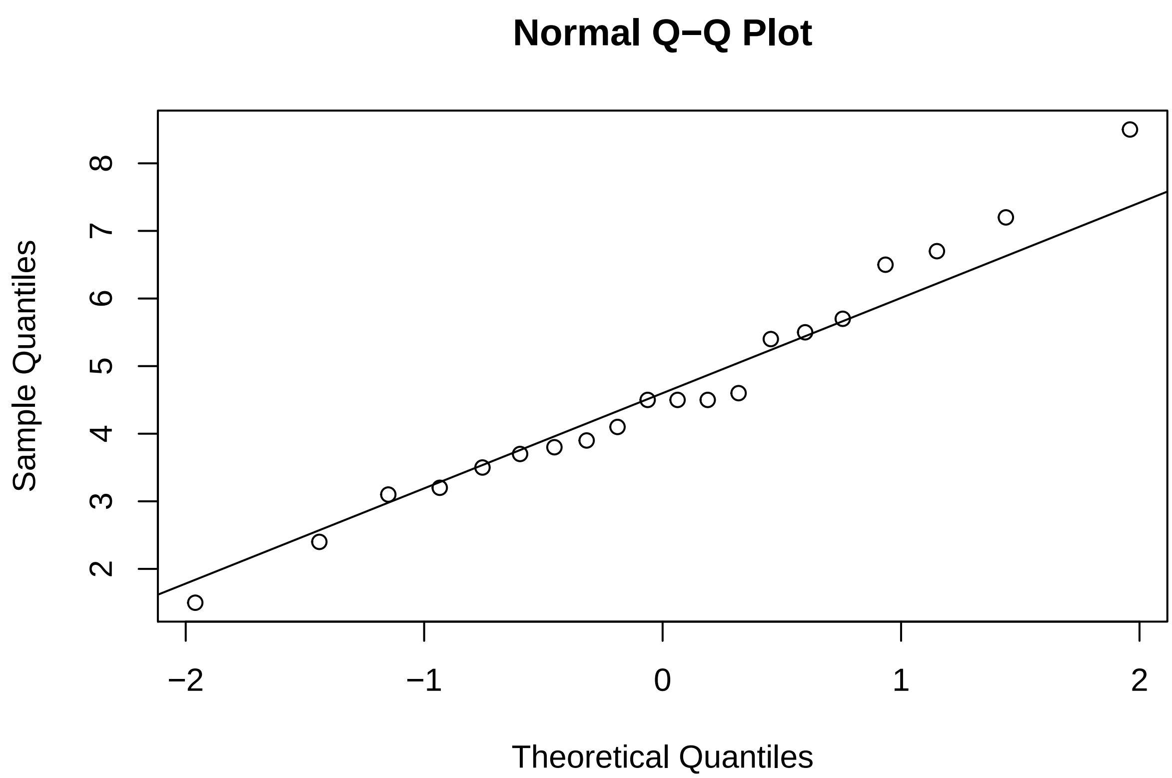
2019/4/12

Chapter 5

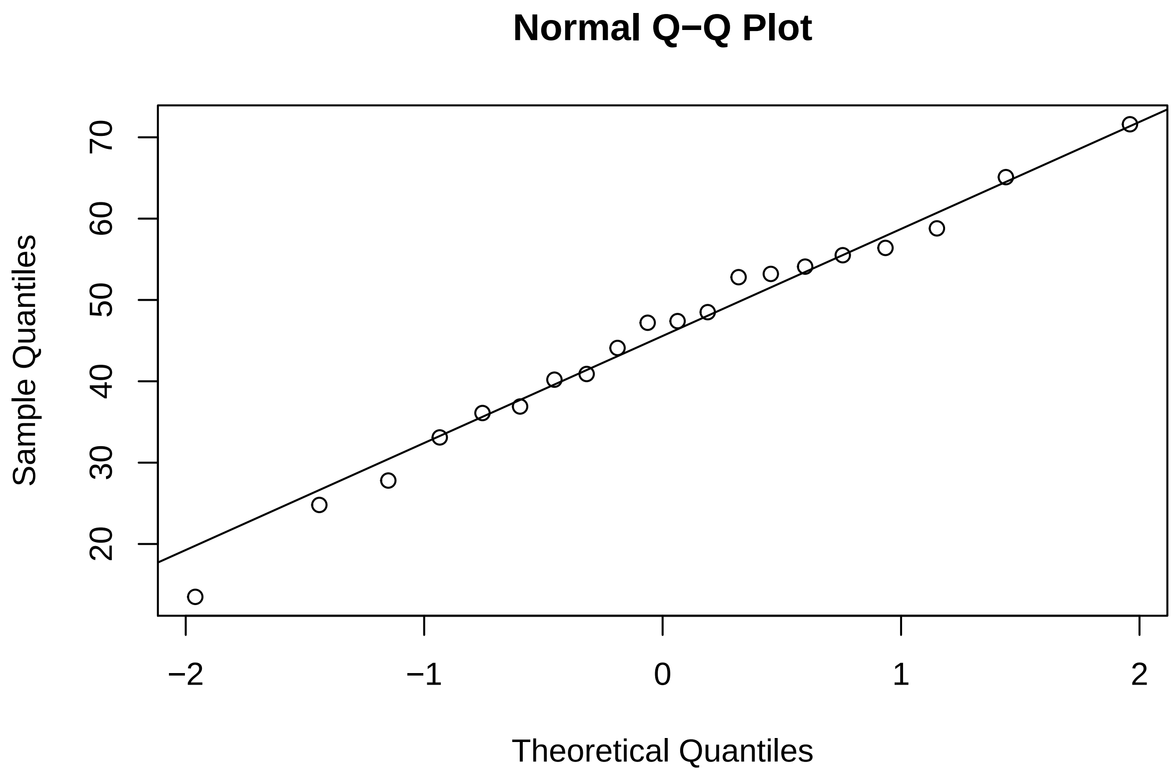
```
data=read.table("T5-1.DAT")
n = dim(data)[1]
p = dim(data)[2]
colnames(data) <- c("X1","X2","X3")
attach(data)
```

Q-Q plot:

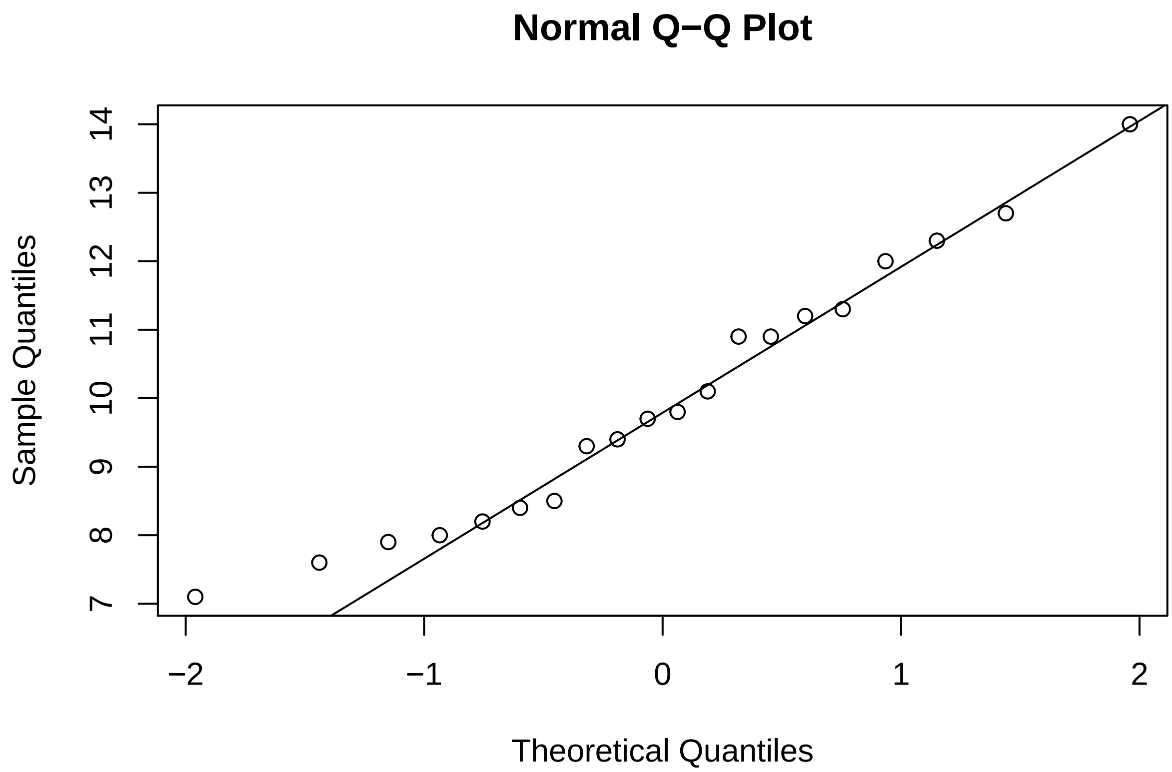
```
qqnorm(X1)
qqline(X1)
```




```
qqnorm(X2)
qqline(X2)
```

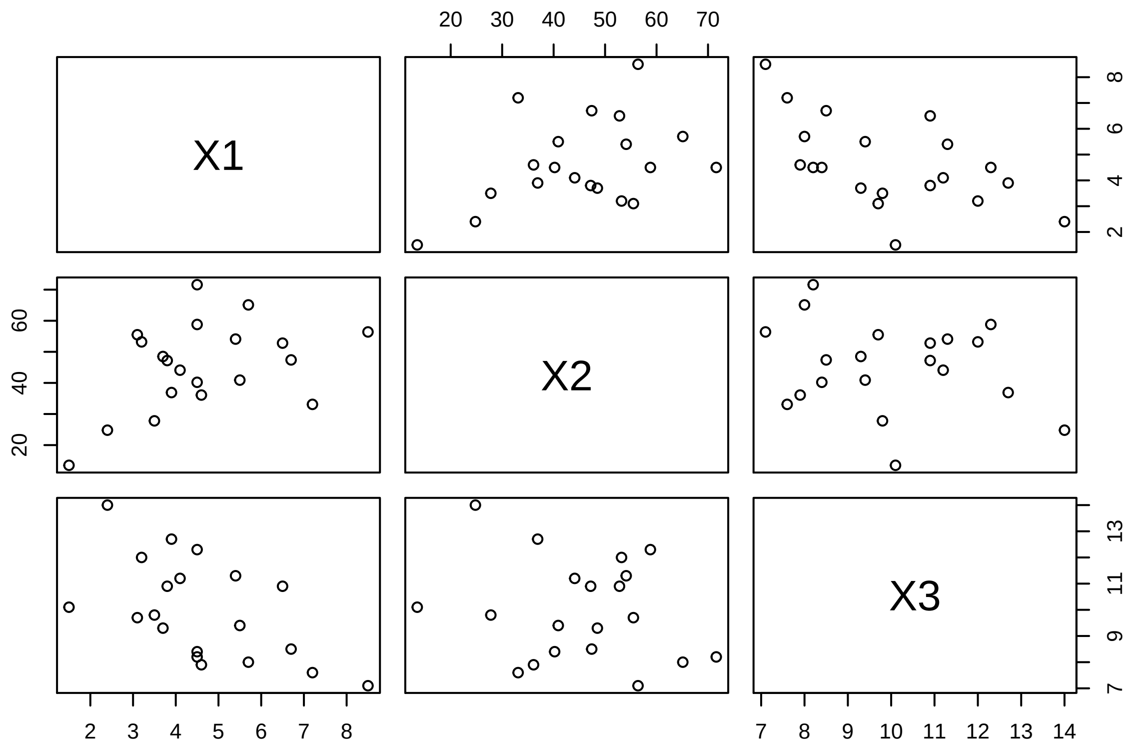



```r
qqnorm(X3)
qqline(X3)
```



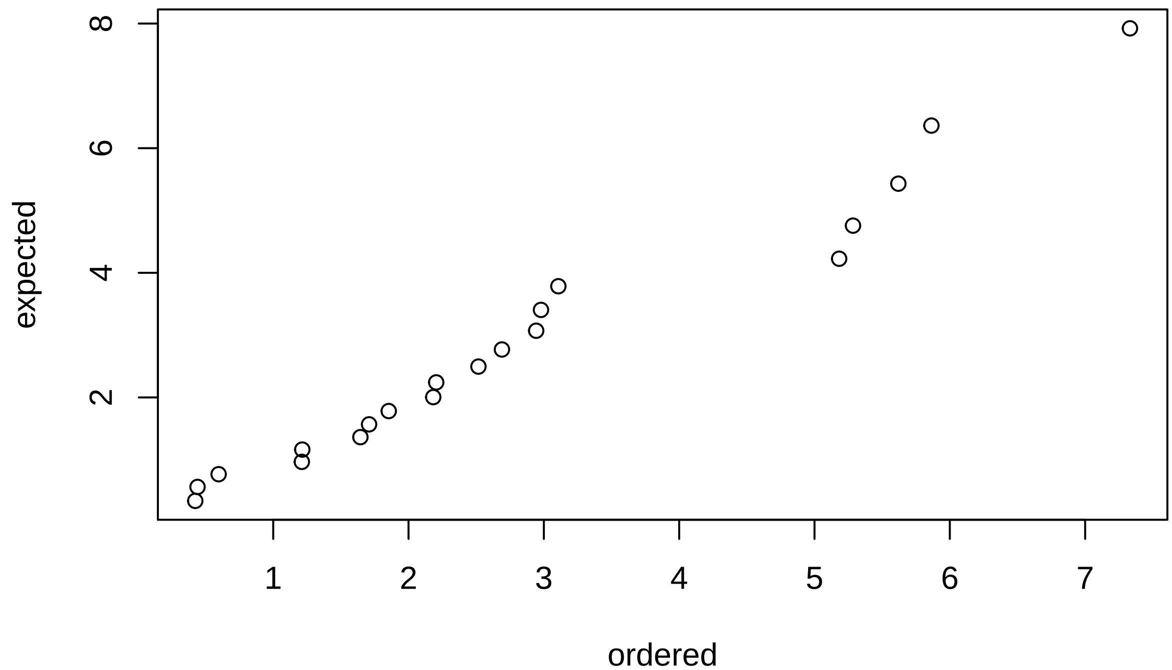
scatter plot matrix:

```r
plot(data)
```



Chi-square plot:

```
X=as.matrix(data)
d2=c()
for (i in 1:20){
  d2<-c(d2, t(matrix(X[i,]-colMeans(X)))%*%solve(cov(X))%*%(X[i,]-colMeans(X)))
}
# ordered values
ordered <- sort(d2)
# expected values
expected <- qchisq(rank(ordered)/(length(ordered) + 1),p)
plot(ordered, expected)
```



Since there is a relatively strong linear ralationship in Q-Q plot and Chi-square plot, and the variables seems independent from each other, the multivariate normal assumption seems justified.

Chapter 6

1.

a situation fit for paired comparisons:

- the data of two samples shares a one to one correspondence, and the corresponded data has the same variables.
- the same variable of one sample is independent from each other.

a situation not fit for paired comparisons: - the data of the two samples is not one-to-one corresponded. - the same variable of one sample may be not independent.

2.

a.

$$H_0 : \mu_1 = \mu_2 \leftrightarrow H_1 : \mu_1 \neq \mu_2$$

```
n1 = 45
n2 = 55
p=2
```

```r
alpha = 0.05
X1 = matrix(c(204.4,556.6), nrow = p)
X2 = matrix(c(130.0,355.0), nrow = p)
S1 = matrix(c(13825.3,23823.4,23823.4,73107.4), nrow=p)
S2 = matrix(c(8632.0,19616.7,19616.7,55964.5), nrow=p)
Sp = (n1-1)/(n1+n2-2)*S1+(n2-1)/(n1+n2-2)*S2 # S_pooled
T2 = 1/(1/n1+1/n2)*t(X1-X2) %*% solve(Sp) %*% (X1-X2)
# test statistic
F=(n1+n2-p-1)/(p*(n1+n2-2))*T2
F
```

```
##          [,1]
## [1,] 7.951139
```

```r
df(1-alpha,p,n1+n2-p-1)
```

```
## [1] 0.382811
```

Since $F > F_{p,n_1+n_2-p-1}(\alpha)$, we conclude H_1, that $\mu_1 \neq \mu_2$.

The linear combination is

$$S_{pooled}^{-1}(\bar{X}_1 - \bar{X}_2)$$

b.

```r
S = S1/n1+S2/n2
# test statistic
T2 = t(X1-X2) %*% solve(S) %*% (X1-X2) # H0: \mu1-\mu2=0
T2
```

```
##          [,1]
## [1,] 15.65853
```

```r
qchisq(1-alpha,p)
```

```
## [1] 5.991465
```

Since $T^2 > \chi^2_p(\alpha)$, we conclude H_1, that $\mu_1 \neq \mu_2$.

c.

$$H_1 : \Sigma_1 = \Sigma_2 \leftrightarrow H_1 : \Sigma_1 \neq \Sigma_2$$

```r
Sp = (n1-1)/(n1+n2-2)*S1+(n2-1)/(n1+n2-2)*S2 # S_pooled
u = (2*p^2+3*p-1)/(6*(p+1)*(2-1))*(1/(n1-1)+1/(n2-1)-1/(n1+n2-2))
# test statistic
C = (1-u)*((n1+n2-2)*log(det(Sp))-(n1-1)*log(det(S1))-(n2-1)*log(det(S2)))
C
```

```
## [1] 18.93306
```

```r
v=1/2*p*(p+1)
qchisq(1-alpha,v)
```

```
## [1] 7.814728
```

Since $C > \chi^2_v(\alpha)$, we conclude H_1, that $\Sigma_1 \neq \Sigma_2$, to reject the equal variance matrix assumption.

d.

Prefer (b), because it doesn't require the equal variance matrix assumption.

3: 6.9

a.

$$d_j = (x_{j,1} - x_{j,p+1}, \cdots, x_{j,p} - x_{j,2p}) = Cx_j$$

b.

$$\bar{d} = \frac{1}{n}\sum_{j=1}^{n} d_j = \frac{1}{n}\sum_{j=1}^{n} Cx_j = C\frac{1}{n}\sum_{j=1}^{n} x_j = C\bar{x}$$

c.

$$S_d = \frac{1}{n-1}\sum_{j=1}^{n}(d_j - \bar{d})(d_j - \bar{d})' = C\frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})(x_j - \bar{x})'C' = CSC'$$

4: 6.11