Homework 3-1, 3-2

陈旭鹏

2018/4/10

1

density function for the multivariate normal distribution:

$$f(x;\mu,\Sigma)=\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}$$

the likelihood function for the two independent sample:

$$\begin{aligned}L(\mu_1,\mu_2,\Sigma)&=\Pi_{i=1}^{n1}f(X_i;\mu_1,\Sigma)\Pi_{j=1}^{n2}f(X_j;\mu_2,\Sigma)\\&=L(\mu_1,\Sigma)L(\mu_2,\Sigma)\end{aligned}$$

So the likelihood function can be defined as:

$$L(\mu_1,\mu_2,\Sigma)=\frac{1}{(2\pi)^{Np/2}|\Sigma|^{N/2}}exp(-\frac{1}{2}tr[\Sigma^{-1}(\sum_{i=1}^{n1}\Phi_1(X_i)+\sum_{j=1}^{n2}\Phi_2(X_j))])$$

$$\Phi_i(x)=(x-\mu_i)(x-\mu_i)'$$

Using MLE, the maximum is:

$$\hat{\Sigma}=\frac{1}{n_1+n_2}(\sum_{i=1}^{n1}\Phi_1(X_i)+\sum_{j=1}^{n2}\Phi_2(X_j))$$

$$=\frac{1}{n_1+n_2}[(n_1-1)S_1+(n_2-1)S_2]=\frac{n_1+n_2-1}{n_1+n_2}S_{pooled}$$

2

(a)

It can be calculated that:

$$\bar{x}=[4.64,45.4,9.965]^T$$

$$S=\begin{bmatrix}2.879&10.01&-1.81\\10.01&199.788&-5.64\\-1.81&-5.64&3.682\end{bmatrix}$$

we can calculate S's eigen value and its respective eigen vector:

$$\begin{aligned}value&=[200.46209264,1.31804876,4.56885859]\\e_1'&=[-0.05084165,-0.82194341,-0.56729548]\\e_2'&=[-0.99828327,0.02528569,0.0528313]\\e_3'&=[0.02907988,-0.56900761,0.82181792]\end{aligned}$$

The axes of the region are:

$$\sqrt{\lambda_i}\sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)}$$

$$\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha) = \frac{19\times 3}{17\times 20}\times F_{3,17}(0.1) = 0.167\times 2.44 = 0.409$$

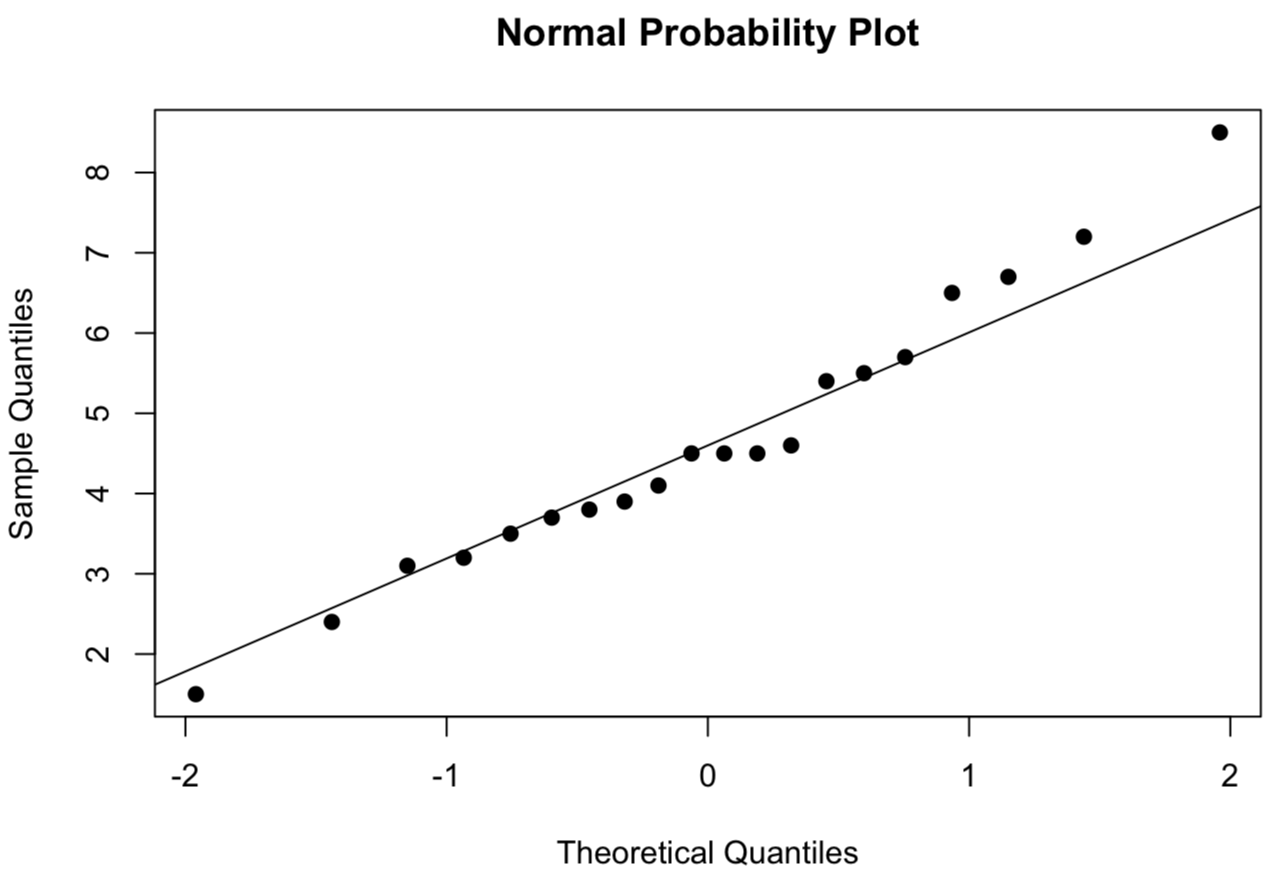
so the axes' lengths are:

$$\sqrt{200.46209264}\times\sqrt{0.409} = 9.055$$
$$\sqrt{1.31804876}\times\sqrt{0.409} = 0.734$$
$$\sqrt{4.56885859}\times\sqrt{0.409} = 1.367$$

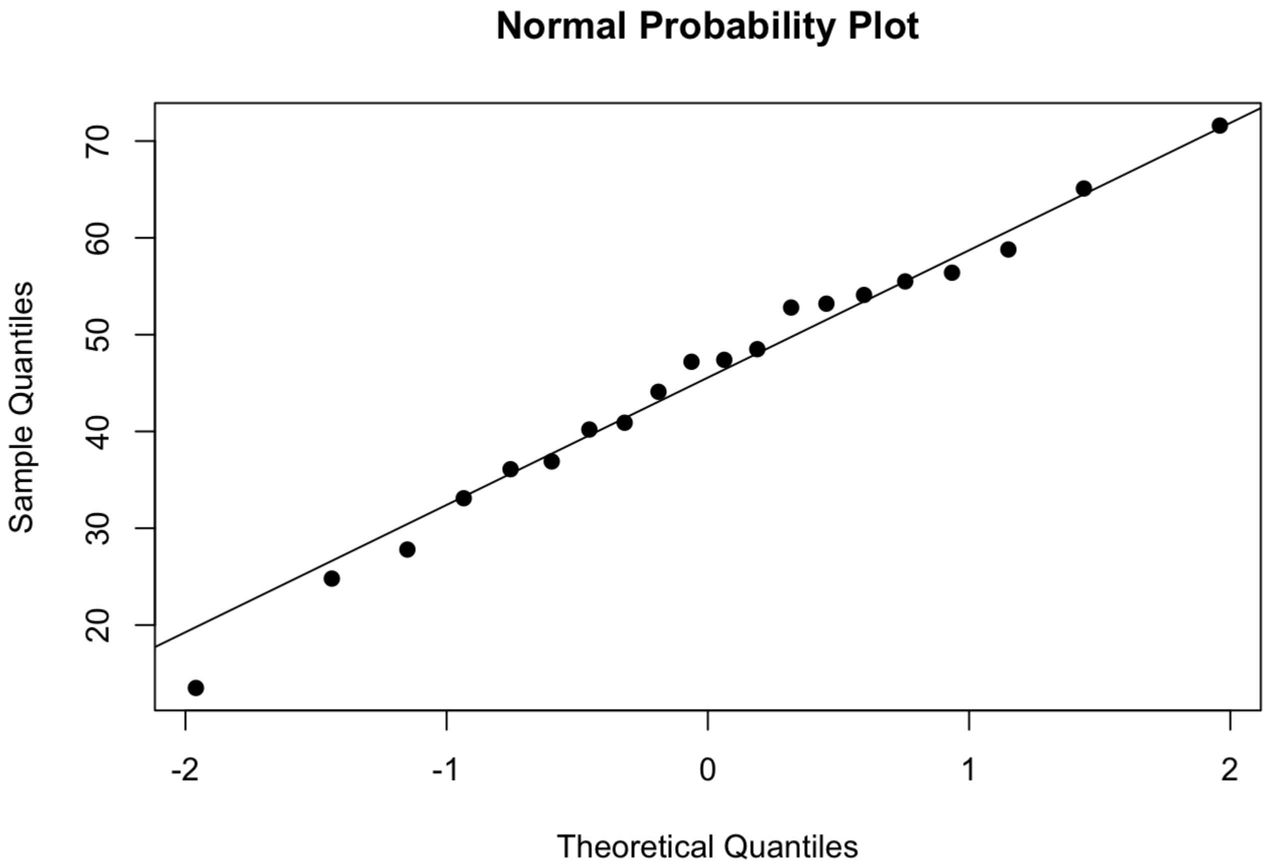
The directions of each aixs is determined by its corresponding eigen vector shown above.

(b)

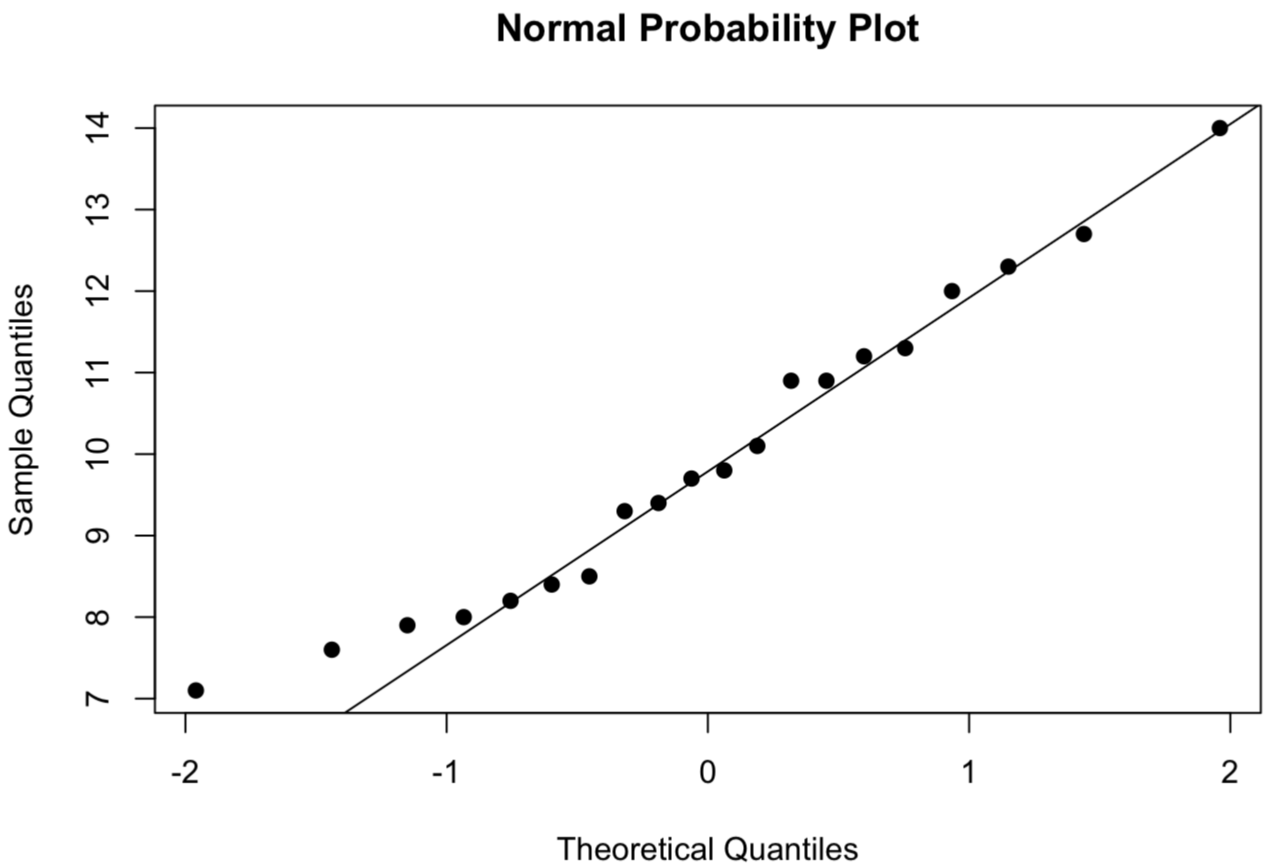
```
# read the data
setwd('~/Desktop/三春/3多元统计分析/作业/作业3-1,3-2/')
dat<-read.csv("data.csv")
x1<-dat$x1
x2<-dat$x2
x3<-dat$x3
qqnorm(x1, main="Normal Probability Plot", pch=19)
qqline(x1)
```



```
qqnorm(x2, main="Normal Probability Plot", pch=19)
qqline(x2)
```

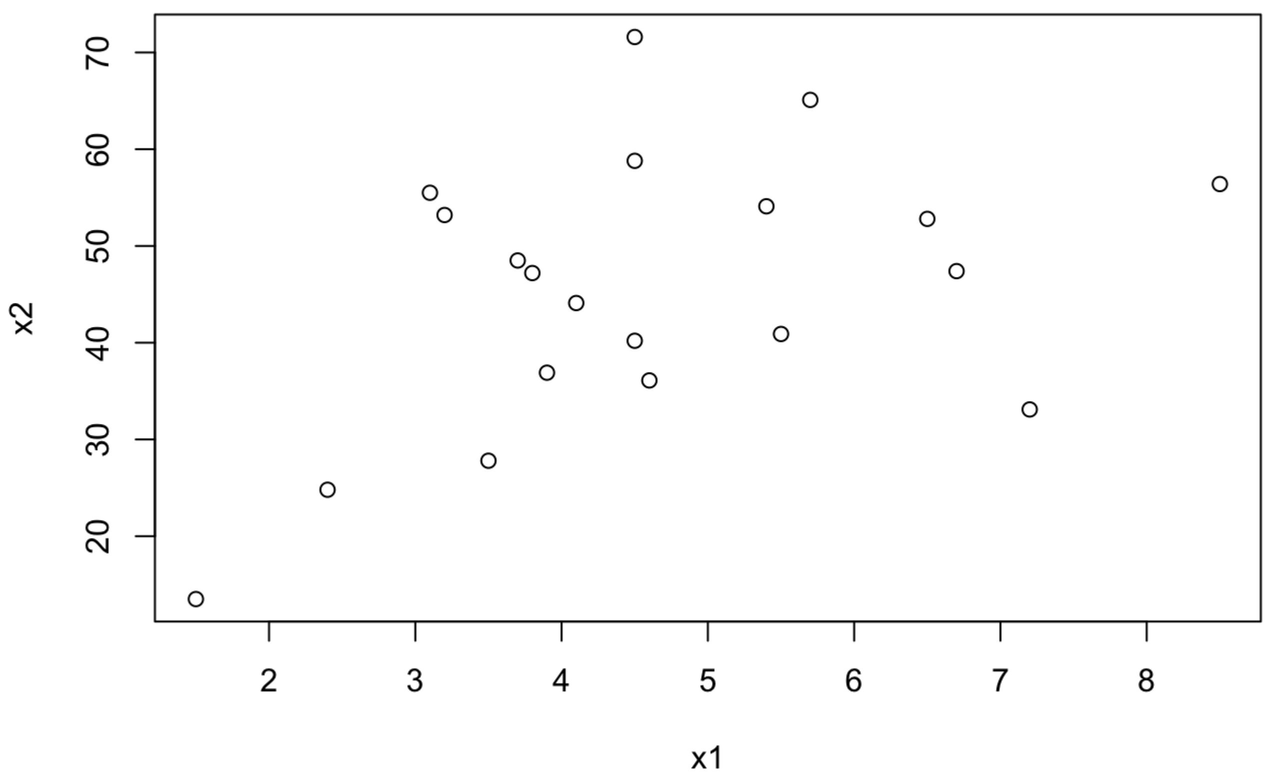



```
qqnorm(x3, main="Normal Probability Plot", pch=19)
qqline(x3)
```



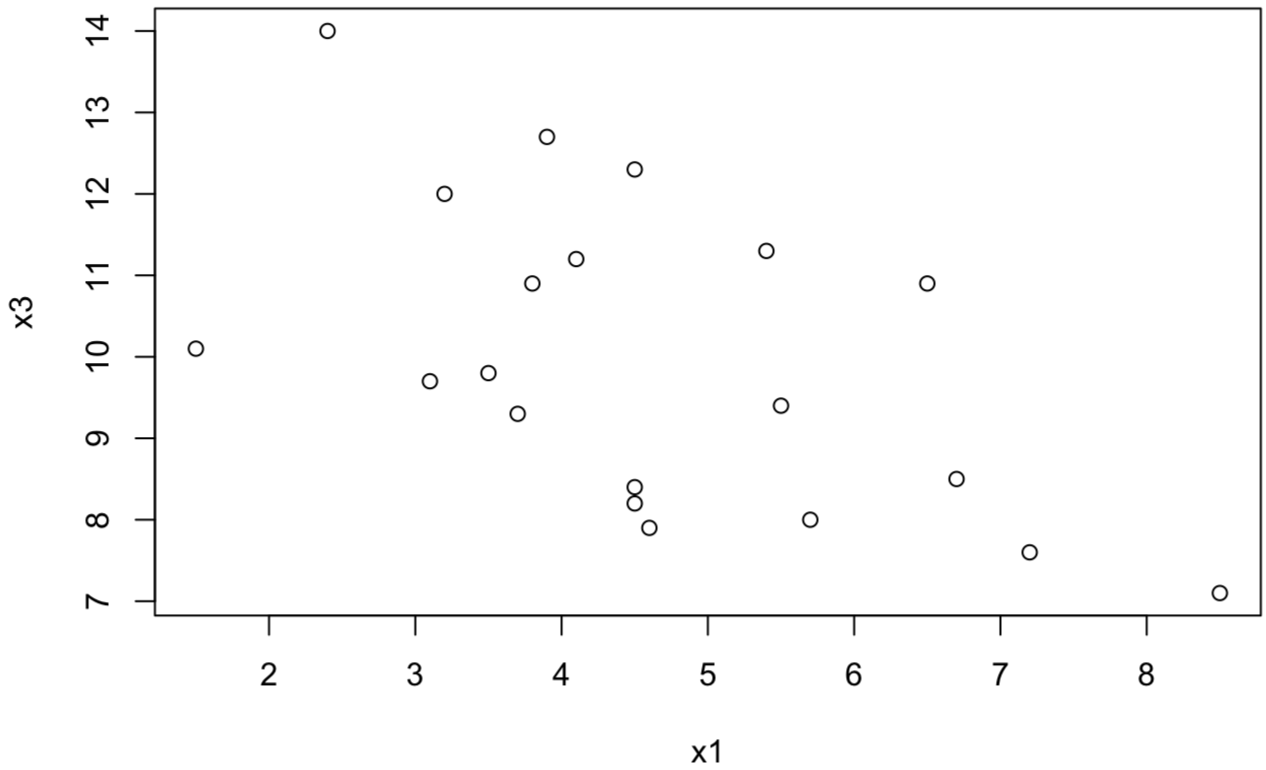
Normal Probability Plot
Sample Quantiles
Theoretical Quantiles
7
8
9
10
11
12
13
14
-2
-1
0
1
2


```
plot(x1,x2)
```

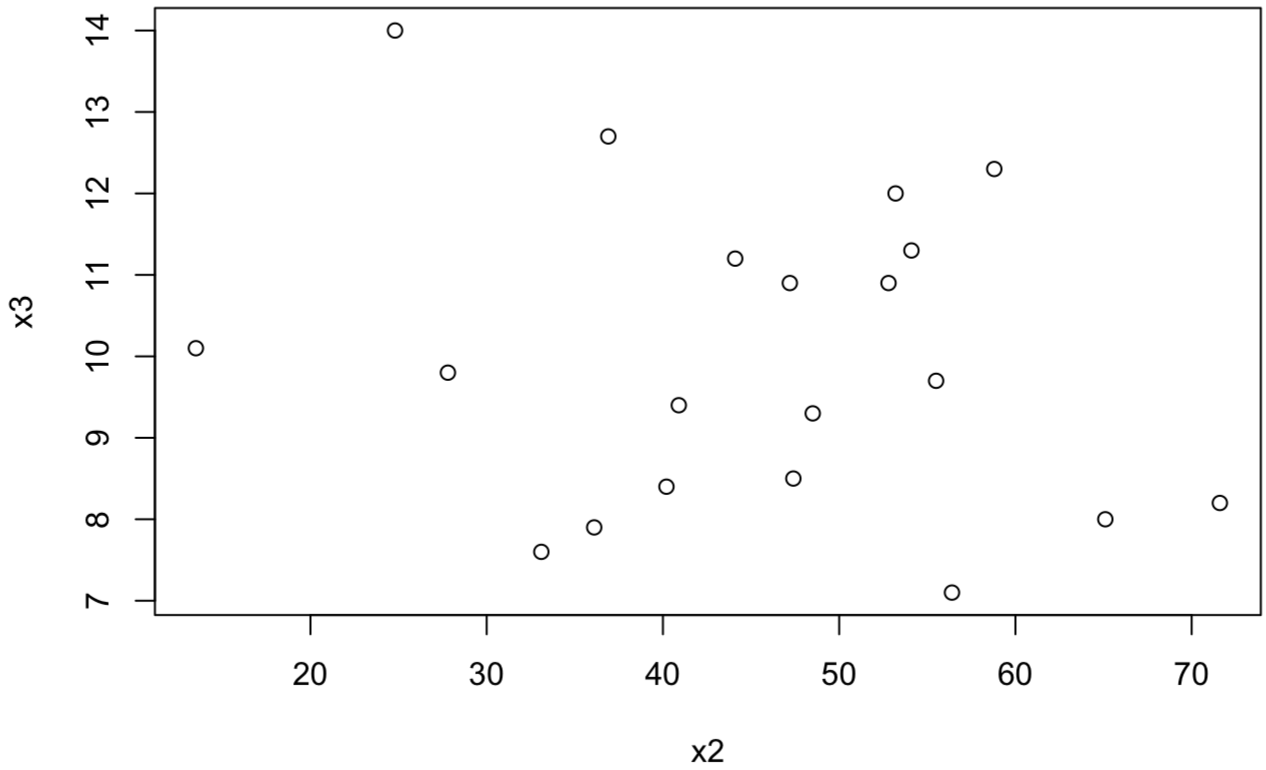


x2
x1
20
30
40
50
60
70
2
3
4
5
6
7
8

```
plot(x1,x3)
```




```
plot(x2,x3)
```



It seems that each variable's normality is fine and they don't have a significant relationship with each other, so the multivariate normal assumption seems justied.

3

$$T^2 = \sqrt{n}(\bar{X} - \mu_0)'\left(\frac{\sum_{j=1}^{n}(X_j - \bar{X})(X_j - \bar{X})'}{n-1}\right)^{-1}\sqrt{n}(\bar{X} - \mu_0)$$

$$\text{So it can be calculated that } T^2 = 9.74$$

$$\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha) = \frac{19 \times 3}{17} \times F_{3,17}(0.05) = \frac{19 \times 3}{17} \times 3.20 = 10.729$$

the confidence region is defined as :

$$\left(\bar{x}_i - \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{ii}}{n}}, \bar{x}_i + \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{ii}}{n}}\right)$$

$$F_{p,n-p}(\alpha) = F_{3,17}(0.05) = 3.20,\ \bar{x}_1 = 4.64,\ \bar{x}_2 = 45.4,\ \bar{x}_3 = 9.965$$

$$s_{11} = 2.879,\ s_{22} = 199.788,\ s_{33} = 3.628$$

so the three regions are:

$$(\bar{x}_1 - \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{11}}{n}}, \bar{x}_1 + \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{11}}{n}})$$
$$= [3.397, 5.882]$$
$$(\bar{x}_2 - \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{22}}{n}}, \bar{x}_2 + \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{22}}{n}})$$
$$= [35.047, 55.752]$$
$$(\bar{x}_3 - \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{22}}{n}}, \bar{x}_3 + \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{33}}{n}})$$
$$= [8.569, 11.360]$$

(b) The Bonferroni region is defined as :

$$[\bar{x}_p - t_{n-1}(\frac{\alpha}{2p})\sqrt{\frac{s_{pp}}{n}},\ \bar{x}_p + t_{n-1}(\frac{\alpha}{2p})\sqrt{\frac{s_{pp}}{n}}]$$
$$t_{19}(0.0083) = 2.625106$$

so the Bonferroni regions are:

$$[3.644, 5.635]$$
$$[37.103, 53.696]$$
$$[8.846, 11.083]$$

which are smaller than T^2 confidence region because it focus on single confidence interval.