编号 姓名 专业 第*次作业

一、对应分析和典型相关分析

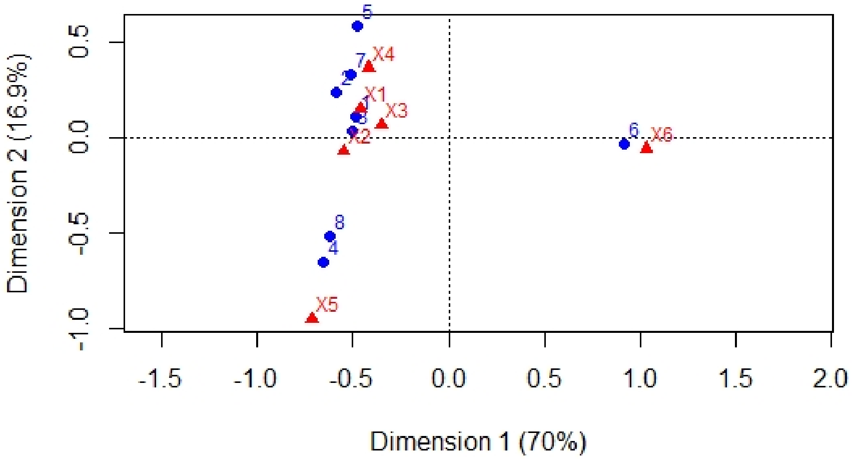
1.

```
Principal inertias (eigenvalues):
           1        2        3        4        5
Value      0.506676 0.122134 0.056584 0.028993 0.009446
Percentage 70%      16.87%   7.82%    4.01%    1.3%
```

由以上可知，有70%可用第一维度表示。

相应的坐标和散布图如下

```
Row scores:
      [,1]     [,2]
1 -0.6781  0.30671
2 -0.8206  0.67031
3 -0.7078  0.08213
4 -0.9228 -1.87861
5 -0.6675  1.65574
6  1.2901 -0.08889
7 -0.7137  0.93068
8 -0.8776 -1.48437

Column scores:
      [,1]    [,2]
x1 -0.6508  0.4202
x2 -0.7725 -0.2051
x3 -0.4937  0.1815
x4 -0.5928  1.0371
x5 -1.0072 -2.7230
x6  1.4506 -0.1662
```



由散布图可知，取样点和污染物可分为三类。

污染物X1,X2,X3,X4主要分布在取样点1，2，3，5，7附近。

污染物X5主要分布在取样点4,8附近。

污染物X6主要分布在取样点6附近。

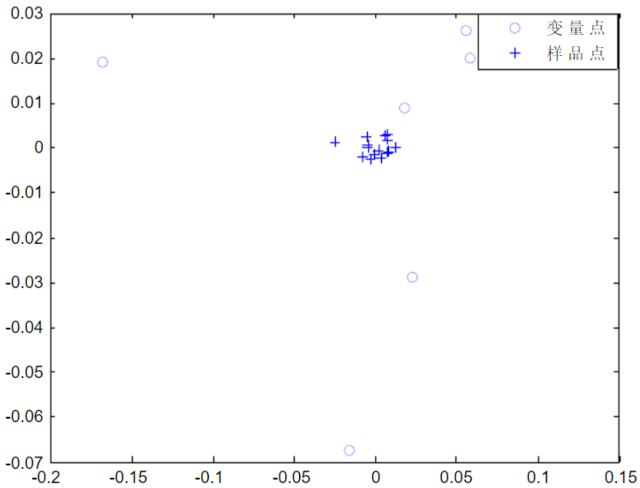
```
data=data.frame(X1=c
    (0.056,0.049,0.038,0.034,0.084,0.064,0.048,0.069),
X2=c(0.084,0.055,0.130,0.095,0.066,0.072,0.089,0.087),
X3=c(0.031,0.100,0.079,0.058,0.029,0.100,0.062,0.027),
X4=c(0.038,0.110,0.170,0.160,0.320,0.210,0.260,0.050),
X5=c(0.0081,0.0220,0.0580,0.2000,0.0120,0.0280,0.0380,0.0890)
    ,
X6=c(0.0220,0.0073,0.0430,0.0290,0.0410,1.3800,0.0360,0.0210)
    )
rownames(data)=c("1","2","3","4","5","6","7","8")
library(ca)
library(MASS)
data.ca=ca(data)
data.ca
```

```
plot(data.ca)
```

2.

(1).对应分析：

求出协方差，特征值及特征向量，得到前两个方差的累计贡献率为87%，可以继续进行，计算载荷矩阵并化为图片可得：



(2).R型因子分析法；

使用pcacov函数（标准化后）得到R因子结果：

F_1主要是生活用品及其他、食品、文化生活服务支出，占贡献率59%

F_2主要是食品和生活用品及其他，占贡献率22%

F_3主要是衣着，占贡献率10%

总贡献率91%,主要因素为生活用品及其他和食品.

总分排行为：上海，北京，浙江，天津，辽宁，江苏，吉林，山东，内蒙古，安徽，福建，黑龙江，山西，河南，江西，河北.

(3).聚类分析：

使用clusterdata函数（标准化后）得到聚类结果：

第一类：北京

第二类：上海

第三类:河北、山西、内蒙、山东、河南

第四类：天津、辽宁、吉林、黑龙江、江苏、浙江、安徽、福建、江西

结果相差不大。

3.先进行卡方检验

```
setwd('F:\\XMU files数学竞赛数学建模\\\\\\CH17')
B <- data.frame(jinhuang=c(326,688,343,98),hong=c
    (38,116,84,48),he=c(241,584,909,403),
                  shenhong=c(110,188,412,681),hei=c(3,4,26,85))
row.names(B)=c('lan','danlan','qianlan','shenlan')
B
chisq.test(B)
```

输出结果为：

```
> chisq.test(B)

        Pearson's Chi-squared test

data:  B
X-squared = 1200, df = 12, p-value <2e-16
```

p值远小于0.05，可见其因素之间存在着较强的相关性。

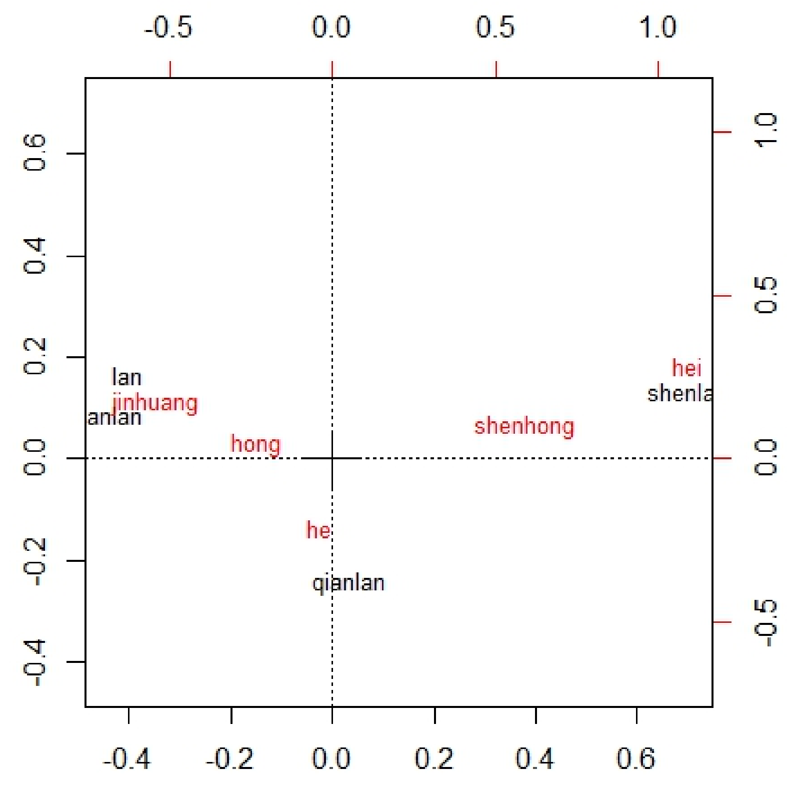
(2)画散点图。

```
B.ca <- corresp(B,nf=2)
B.ca
biplot(B.ca,cex=0.8)
abline(v=0,h=0,lty=3)
```

输出结果为：

所以可以分为三类，眼睛蓝色、淡蓝色的和头发金黄色为一类，眼睛浅蓝色和头发为红色、褐色一类，眼睛深蓝与头发深红和黑色一类。

```
> B.ca
First canonical correlation(s): 0.4464 0.1735

 Row scores:
             [,1]    [,2]
lan      -0.89679  0.9536
danlan   -0.98732  0.5100
qianlan   0.07531 -1.4125
shenlan   1.57435  0.7720

 Column scores:
              [,1]    [,2]
jinhuang -1.21871  1.0022
hong     -0.52258  0.2783
he       -0.09415 -1.2009
shenhong  1.31888  0.5993
hei       2.45176  1.6514
```

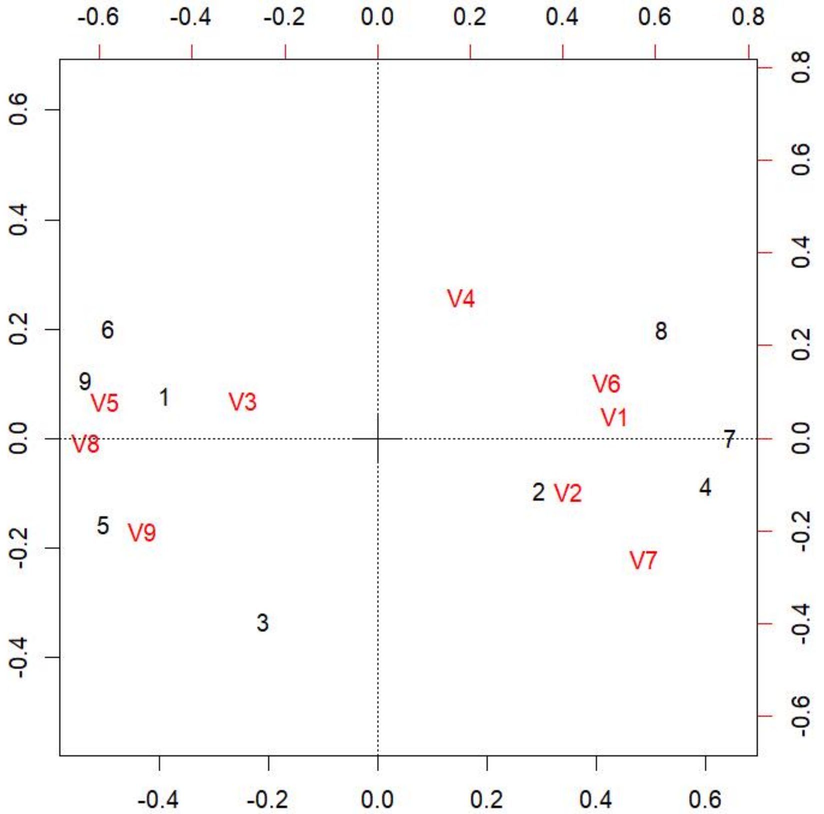


4.利用R语言进行对应分析:

```
> fly <- read.csv("data.csv",header=FALSE)
> library(MASS)
> ca <- corresp(fly,2)
> ca
First canonical correlation(s): 0.4962368 0.1641114

 Row scores:
                [,1]        [,2]
  [1,] -0.7842683  0.47552200
  [2,]  0.5970759 -0.57565401
  [3,] -0.4193079 -2.03182344
  [4,]  1.2136436 -0.53073893
  [5,] -1.0079221 -0.94915578
  [6,] -0.9903145  1.22215357
  [7,]  1.3023041  0.01323004
  [8,]  1.0485474  1.21727682
  [9,] -1.0750760  0.64503980

 Column scores:
            [,1]        [,2]
V1  1.0327867  0.28897091
V2  0.8302158 -0.69834341
V3 -0.5834084  0.50588567
V4  0.3678161  1.85037509
V5 -1.1831337  0.48894732
V6  0.9958246  0.74368787
V7  1.1575836 -1.58884144
V8 -1.2682184 -0.05064382
V9 -1.0208052 -1.22517528
> summary(ca)
          Length Class  Mode
```

```
cor        2      -none-  numeric
rscore  18      -none-  numeric
cscore  18      -none-  numeric
Freq     81      -none-  numeric
> biplot(ca)
```

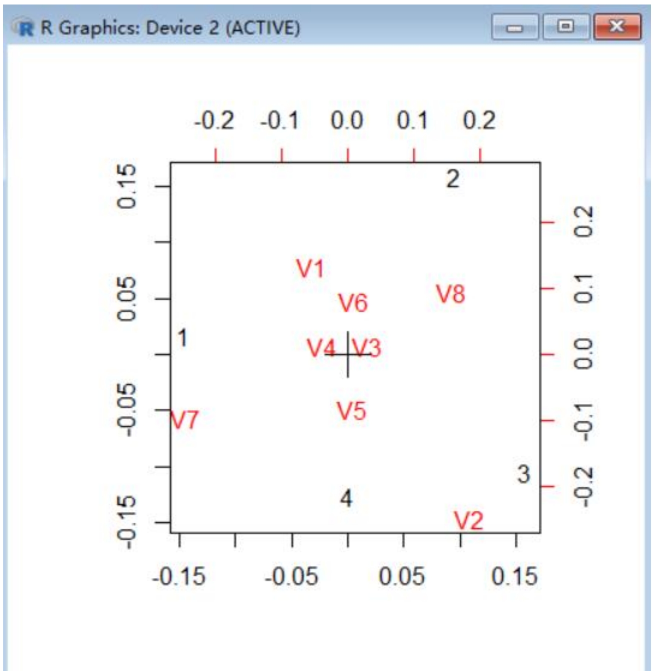
得到图像:



5、利用R语言进行对应分析:

```
data <- read.csv("F:\\R_project\\5.csv",header=FALSE)
library(MASS)
ca <- corresp(data,2)
ca
summary(ca)
biplot(ca)
```

得到图像:



6.

Principal inertias (eigenvalues):

	1	2	3	4	5	6	7	8	9
Value	0.11563	0.090524	0.024469	0.019529	0.012688	0.005121	0.002601	0.001195	0.00054
Percentage	42.46%	33.24%	8.99%	7.17%	4.66%	1.88%	0.96%	0.44%	0.2%

由以上可知，有42.46%可用第一维度表示,33.24%可用第一维度表示

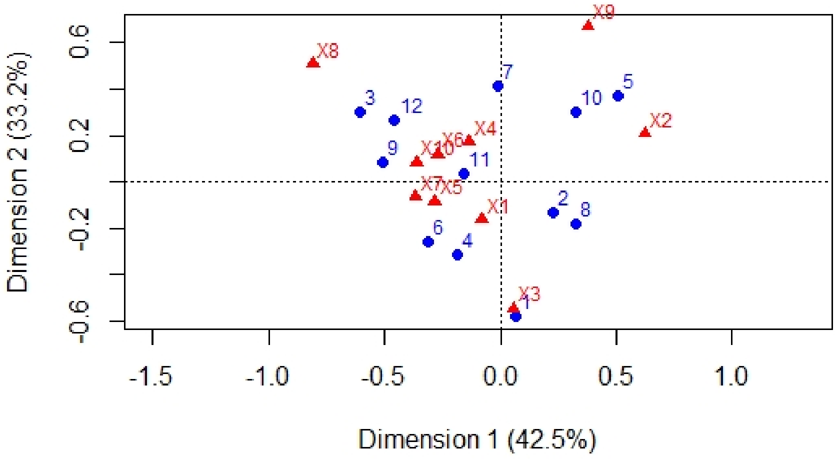
相应的坐标和散布图如下

Rows:

	1	2	3	4	5	6	7	8	9	10	11	12
Mass	0.09535	0.087532	0.05475	0.08532	0.09605	0.04734	0.04413	0.13009	0.06169	0.09879	0.12491	0.07405
ChiDist	0.60240	0.328189	0.73247	0.38555	0.66420	0.50538	0.76681	0.39432	0.67322	0.46223	0.30189	0.58245
Inertia	0.03460	0.009428	0.02937	0.01268	0.04238	0.01209	0.02595	0.02023	0.02796	0.02111	0.01138	0.02512
Dim. 1	0.20329	0.665470	-1.78444	-0.54009	1.49473	-0.91935	-0.03560	0.94633	-1.48549	0.95481	-0.47186	-1.33978
Dim. 2	-1.93501	-0.442746	0.99785	-1.03483	1.24300	-0.86497	1.38073	-0.60538	0.27626	1.00042	0.11432	0.89328

Columns:

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
Mass	0.020371	0.18948	0.21161	0.3059	0.05582	0.13568	0.015506	0.02041	0.004221	0.04095
ChiDist	0.517968	0.66287	0.55375	0.2452	0.42838	0.38191	0.724216	1.30242	1.977959	0.51814
Inertia	0.005465	0.08326	0.06489	0.0184	0.01024	0.01979	0.008132	0.03461	0.016513	0.01099
Dim. 1	-0.234718	1.83390	0.16630	-0.3918	-0.83026	-0.79548	-1.071908	-2.38234	1.111816	-1.05537
Dim. 2	-0.542337	0.68763	-1.81318	0.5731	-0.28224	0.39199	-0.218305	1.68823	2.229085	0.27371



由上图可知，地区11受各种癌症死亡的人数较多，3，9，12，7，6，4，2，8各种癌症死亡人数一般，地区10和5主要受食道癌症的影响。膀胱癌和宫颈癌致死的病例较少，地区一主要死亡因素是胃癌。

```
数据输入省略

library(ca)
library(MASS)
data.ca=ca(data)
data.ca
plot(data.ca)
data.ca1=corresp(data,nf=2)
data.ca1
```

7.

代码如下：

```
xx = [...]
[n p]=size(xx);
T=sum(sum(xx));
xcols=zeros(1,p);
xrows=zeros(1,n);
for k=1:p
        xcols(k)=sum(xx(:,k));
end
for l=1:n
        xrows(l)=sum(xx(l,:));
end
z=zeros(p,p);
for i=1:n
        for j=1:p
        z(i,j)=(xx(i,j)-xrows(i)*xcols(j)/T)/((
            xrows(i)*xcols(j))^(1/2));
        end
end
a=z'*z
[X b]=eig(a)
```

第一个方差的贡献率已达96.1%,接下来计算载荷矩阵：

```
f=zeros(p)
for  t=1:p
          f(t)=(0.0417^0.5)*(X(t,5))
end
f
g=z*f
```

结果如下:

```
f =            g =

   -0.0169       0.0104
   -0.0031      -0.0117
   -0.2029      -0.0079
    0.0106      -0.0041
    0.0109      -0.0091
                -0.0027
                -0.0010
                -0.0011
                -0.0034
                -0.0114
                 0.0014
                -0.0343
                -0.0007
```

8.

```r
R <- data.frame(c(1,0.63,0.24,0.59),c(0.63,1,-0.06,0.07),c(0.24,-0.06,1,0.42),c(0.59,0.07,0.42,1))
colnames(R) <- c("X1",'X2','Y1',"Y2")
row.names(R) <- c("X1",'X2','Y1',"Y2")
rho11 <- matrix(c(1,0.63,0.63,1),nrow = 2)
rho12 <- matrix(c(0.24,-0.06,0.59,0.07),2,2)
rho21 <- matrix(c(0.24,0.59,-0.06,0.07),2,2)
rho22 <- matrix(c(1,0.42,0.42,1),nrow = 2)
A <- solve(rho11)%*%rho12%*%solve(rho22)%*%rho21
A
B <- solve(rho22)%*%rho21%*%solve(rho11)%*%rho12
B
a <- eigen(A)$value
a <- sqrt(a)
a
eigen(A)
eigen(B)
```

```
> a <- eigen(A)$value
> a
[1] 0.50255340 0.01091574
> a <- sqrt(a)
> a
[1] 0.7089100 0.1044784
```

```
> eigen(A)
eigen() decomposition
$values
[1] 0.50255340 0.01091574

$vectors
           [,1]       [,2]
[1,]  0.8715828 -0.1040107
[2,] -0.4902484  0.9945762

> eigen(B)
eigen() decomposition
$values
[1] 0.50255340 0.01091574

$vectors
            [,1]       [,2]
[1,] -0.09706182 -0.8954425
[2,] -0.99527835  0.4451773
```

$U_1 = 0.8715828x_1 - 0.4902484x_2$

$V_1 = -0.09706182y_1 - 0.99527835y_2$

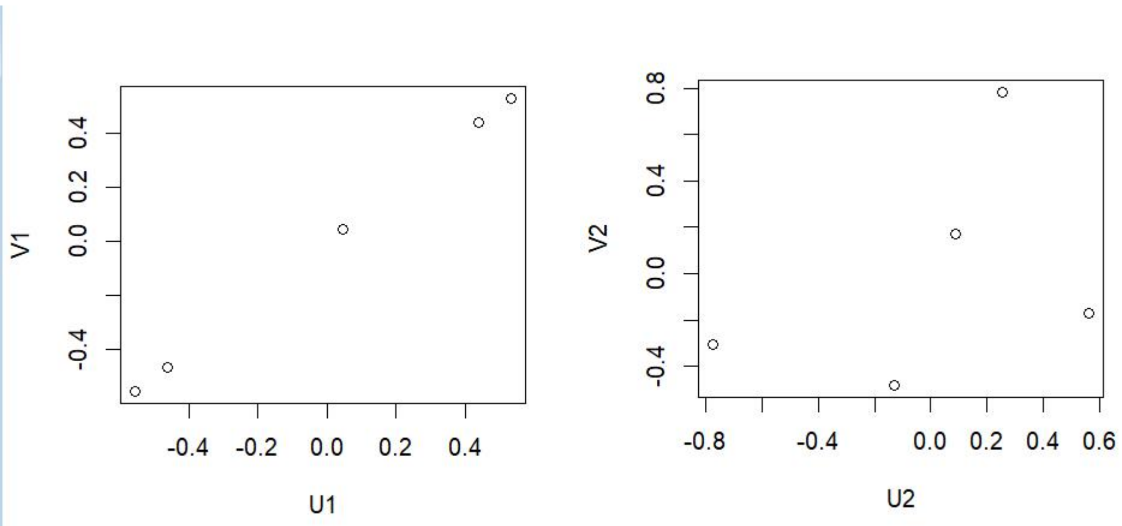
第一典型相关系数为0.7089100，即阅读能力和运算能力之间的相关程度较高。

9.

利用R语言进行典型相关分析:

```
> data = read.table("9.txt",header=T,row.names=1)
> data <- scale(data)
> ca <- cancor(data[,1:2],data[,3:5])
> ca
$cor
[1] 1.000000 0.419282

$xcoef
        [,1]       [,2]
x1 0.3341978  0.9825477
x2 0.1805112 -1.0220097

```

```
$ycoef
                [,1]         [,2]         [,3]
y1 -0.6790076 0.2956356 -0.3848459
y2 -0.1867460 0.6685178 -0.3023967
y3 -0.6302582 0.5688273  0.1119758

$xcenter
               x1              x2
1.332268e-16 9.436896e-17

$ycenter
               y1              y2              y3
7.771561e-17 6.453171e-17 4.440892e-17
```

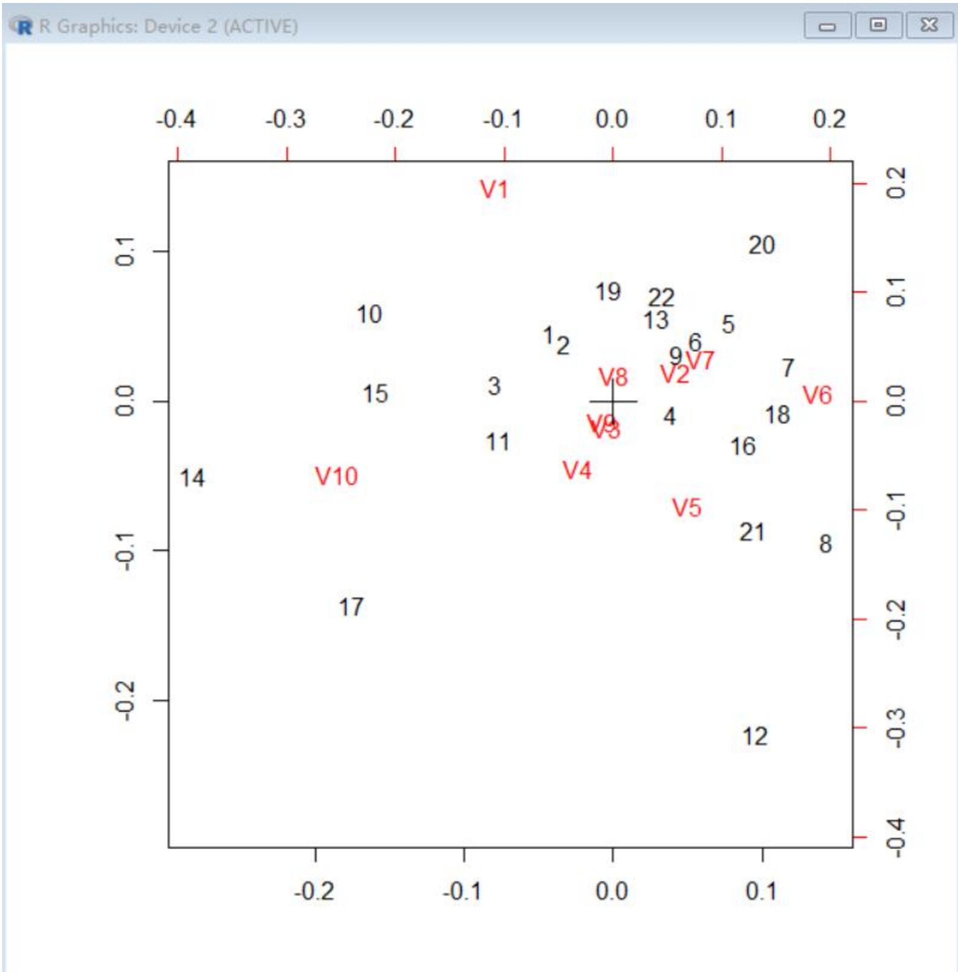


10、利用R语言进行对应分析:

```
data <- read.csv("F:\\R_project\\10.csv",header=FALSE)
library(MASS)
ca <- corresp(data,2)
ca
summary(ca)
```

```
biplot(ca)
```

得到图像:



11.

```
data=load('1.txt');
rho11=data(1:5,1:5);
rho12=data(1:5,6:12);
rho21=data(6:12,1:5);
rho22=data(6:12,6:12)
A=(rho11^(-1))*rho12*(rho22^(-1))*rho21;
```

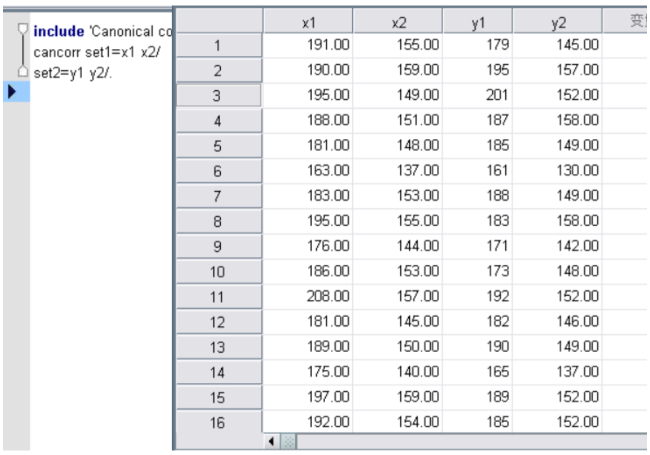
```
B=(rho22^(-1))*rho21*(rho11^(-1))*rho12;
 [X1,B1]=eig(A)
 [X2,B2]=eig(B)
C=sqrt(B1)
```

$$U_1 = -0.6246x_1 - 0.2890x_2 - 0.2483x_3 + 0.0339x_4 - 0.6808x_5$$

$$V_1 = 0.5542y_1 - 0.2723y_2 - 0.0468y_3 + 0.0307y_4 + 0.3784y_5 + 0.6722y_6 - 0.1436$$

第一典型相关系数仅为30.66%，即职业特性与职业满意度之间的关系不是很强烈。

12.



	x1	x2	y1	y2
1	191.00	155.00	179	145.00
2	190.00	159.00	195	157.00
3	195.00	149.00	201	152.00
4	188.00	151.00	187	158.00
5	181.00	148.00	185	149.00
6	163.00	137.00	161	130.00
7	183.00	153.00	188	149.00
8	195.00	155.00	183	158.00
9	176.00	144.00	171	142.00
10	186.00	153.00	173	148.00
11	208.00	157.00	192	152.00
12	181.00	145.00	182	146.00
13	189.00	150.00	190	149.00
14	175.00	140.00	165	137.00
15	197.00	159.00	189	152.00
16	192.00	154.00	185	152.00

使用spss软件的cancorr函数可得：

两对典型变量中：第一对的典型相关系数达到0.788，属于强相关，而第二对只有0.054，比较弱,

长子的头型变量被自身的第一典型变量解释了86.7%，次子的头型变量被自身的第一典型变量解释了91.8%

13.

```
C <- read.table("13.txt",header = T)
```

```
C
cancor(C[,1:7],C[,8:12])
```

```
> cancor(C[,1:7],C[,8:12])
$cor
[1] 0.8446422 0.6772783 0.5392241 0.3640115 0.2612006

$xcoef
            [,1]          [,2]         [,3]          [,4]          [,5]         [,6]         [,7]
X1 -0.0176900504 -0.0013416603  0.024267009  0.0274619332 -0.0011694343 -0.014166644 -0.040713406
X2 -0.0097306145 -0.0048184242 -0.025647931 -0.0109187810 -0.0038759988  0.010868117 -0.003532008
X3 -0.0057862464  0.0008912145  0.007620476 -0.0004049028  0.0008067791  0.004105284  0.007058277
X4 -0.0025517033  0.0050088988 -0.021522447  0.0141127701  0.0193950276 -0.009032014  0.002444108
X5 -0.0024885265  0.0112180557 -0.003032145 -0.0065216986  0.0003432800  0.001831411  0.002941728
X6 -0.0006364595 -0.0043165567  0.007302303 -0.0113477501  0.0037698698 -0.027892484  0.004026108
X7 -0.0005583015 -0.0026184356  0.006561012  0.0166423584 -0.0178303812 -0.003019923  0.002271563

$ycoef
             [,1]         [,2]          [,3]          [,4]          [,5]
X8   0.1621331150 -0.389885929 -0.0278378038 -0.2699051234  0.1909370496
X9  -0.0011889660 -0.003330845 -0.0001445273  0.0031997318  0.0014109941
X10 -0.0253937298 -0.011295209  0.0326615721 -0.0406126662 -0.0408377643
X11 -0.0043748296  0.003612972 -0.0458265169 -0.0349500428 -0.0001551957
X12  0.0008931099 -0.001221382 -0.0018132491 -0.0004000009 -0.0036974720

$xcenter
       X1        X2        X3        X4        X5        X6        X7
 47.42105  59.68421 114.47368  44.63158  66.90789  15.34211  59.73684

$ycenter
        X8         X9        X10        X11        X12
  7.128947 439.789474  27.815789   7.578947 361.500000
```

所以第一典型相关系数为0.8446422，第二典型相关系数为0.6772783，

$$U_1 = -0.0176900504x_1 - 0.0097306145x_2 - 0.0057862464x_3 - 0.0025517033x_4 - 0.0024$$

$$- 0.0006364595x_6 - 0.0005583015x_7$$

$$U_2 = -0.0013416603x_1 - 0.0048184242x_2 + 0.0008912145x_3 + 0.0050088988x_4 + 0.0112$$

$$- 0.0043165567x_6 - 0.0026184356x_7$$

$$V_1 = 0.1621331150x_8 - 0.0011889660x_9 - 0.0253937298x_10 - 0.0043748296x_11 + 0.00089$$

$$V_2 = -0.389885929x_8 - 0.003330845x_9 - 0.011295209x_10 + 0.003612972x_11 - 0.001221$$

第一组第一典型变量U_1主要代表X_1反复横荡次数，X_2纵跳高度和X_3背力，第二典型变量U_2主要代表X_5塔台升降指数，X_4握力和X_2纵跳高度。

第二组第一典型变量V_1主要代表$X_8$50米跑，第二典型变量V_2主要代表$X_8$50米跑，但呈现负相关。

14.

```
> data = read.table("14.txt",header=T,row.names=1)
> data <- scale(data)
> ca <- cancor(data[,1:3],data[,4:6])
> ca
$cor
[1] 0.79356211 0.19006613 0.02265696

$xcoef
            [,1]         [,2]        [,3]
Dec  -0.01218793  0.211437907  0.04950157
Jan   0.19559525 -0.066110758  0.04986370
Feb   0.03490273  0.001131413 -0.20512044

$ycoef
             [,1]        [,2]        [,3]
high7  0.09130213 -0.02440774  0.17813093
high4 -0.04885330 -0.21147220  0.01508689
high8  0.14233784 -0.12547180 -0.10604916

$xcenter
          Dec           Jan           Feb
-8.807058e-18 -2.775558e-16 -5.898060e-17

$ycenter
        high7         high4         high8
-9.607699e-18  2.435285e-17  3.963176e-17

```

```
> U <- as.matrix(data[,1:3])%*%ca$xcoef
> V <- as.matrix(data[,4:6])%*%ca$ycoef
> plot(U[,1], V[,1], xlab="U1", ylab="V1")
```