

## 530 陈斯杰 电子信息工程 第15次作业

### 一、主成分分析

1、首先我们将数据进行标准化，并求出相关系数矩阵的特征值与特征向量。由累计贡献率 $\geq 85\%$ 可得，主成分个数为3，相应的特征值为

$$\lambda_1 = 3.1046, \lambda_2 = 2.8975, \lambda_3 = 0.9302$$

相应的特征向量为

	1	2	3
1	-0.4767	-0.2959	-0.1043
2	-0.4729	-0.2778	-0.1629
3	-0.4239	-0.3779	-0.1562
4	0.2127	-0.4515	0.0085
5	0.3885	-0.3310	-0.3211
6	0.3524	-0.4028	-0.1451
7	-0.2148	0.3774	-0.1404
8	-0.0550	-0.2727	0.8912

求解出相应的主成分为 $Z_1, Z_2, Z_3$

$$Y = 0.3881 \times Z_1 + 0.3622 \times Z_2 + 0.1163 \times Z_3$$

	1 X1	2 X2	3 X3	4 X4	5 X5	6 X6	7 X7	8 X8	9 Y
1	90432	52455	101091	19272	82	16.1000	197435	0.1720	-0.9134
2	4903	1973	2035	10313	34.2000	7.1000	592077	0.0030	0.7292
3	6735	21139	3767	1780	36.1000	8.2000	726396	0.0030	0.7237
4	49454	36241	81557	22504	98.1000	25.9000	348226	0.9850	-0.4959
5	139190	203505	215898	10609	93.2000	12.6000	139572	0.6280	-2.6400
6	12215	16219	10351	6328	62.5000	8.7000	145818	0.0660	0.5186
7	2372	6572	8103	12329	184.4000	22.2000	20921	0.1520	0.5114
8	11062	23078	54935	23804	370.4000	41	65486	0.2630	-0.0980
9	17111	23907	52108	21796	221.5000	21.5000	63806	0.2760	-0.0427
10	1206	3930	6126	15586	330.4000	29.5000	1840	0.4370	0.4593
11	2150	5704	6200	10870	184.2000	12	8913	0.2740	0.5678
12	5251	6155	10383	16875	146.4000	27.5000	78796	0.1510	0.4217
13	14341	13203	19396	14691	94.6000	17.8000	6354	1.5740	0.2584

求解出相应的主成分为 $Z_1, Z_2, Z_3$

此时，综合得分越小越好。

2.

将数据输入MATLAB求解，得到综合变量Z1与Z2

```
1 %q1_2
2 clear;clc;
3 A=[ 843 4000 920 1083 1166 1536 1692 2743 2741
4     942 5215 1216 1280 1549 1924 2717 3743 4160
5     895 6379 1836 2287 2715 3218 3240 4572 5439
6     884 3770 978 1010 1227 1558 1755 2552 2632
7     1002 3329 757 684 718 931 1422 1827 2222
8     792 3039 759 765 1253 1296 1902 1822 2233
9     930 2476 559 616 611 863 891 1539 1768
10    758 2127 633 661 762 769 864 1241 1482
11    886 2721 703 836 813 914 895 1272 1431
12    859 1664 427 451 507 551 581 839 1107
13    808 2435 772 790 839 990 771 1132 1316
14    845 2661 557 599 670 855 1128 1639 1861
15    877 2242 549 471 480 643 664 1173 1530
16    787 3088 1061 1021 978 1102 1301 2373 2452
17    943 3322 567 557 630 836 984 1994 2704
18    866 3507 675 678 847 1187 1476 2107 2415
19    924 3412 545 729 821 1017 1079 1878 2325
20    1806 5869 1268 1227 1607 1765 2318 2959 3157
21    1027 3605 651 835 924 1125 1194 2176 2411];%原始数据
22 for i=1:9
23     SA(:,i)=(A(:,i)-mean(A(:,i)))/std(A(:,i)); %标准化
24 end
25 CM=corrcoef(SA);%系数矩阵
26 [V,D]=eig(CM);%特征值与特征向量
27 for i=1:9
28     DS(i,1)=D(10-i,10-i);
29 end
```

```

30 %计算贡献率
31 for i=1:9
32     DS(i,2)=DS(i,1)/sum(DS(:,1));%单个贡献率
33     DS(i,3)=sum(DS(1:i,1))/sum(DS(:,1));%累计贡献率
34 end
35 %假定主成分的信息保留率为T
36 T=0.95;
37 for k=1:9
38     if DS(k,3) >= T
39         com_num=k;
40         break;
41     end
42 end
43 %提取主成分的特征向量
44 for j=1:com_num
45     PV(:,j)=V(:,10-j);
46 end
47 %计算主成分得分
48 new_score=SA*PV;
49 for i=1:19
50     total_score(i,1)=sum(new_score(i,:));
51     total_score(i,2)=i;
52 end
53 %强主成分得分与总分放到同一个矩阵中
54 result_report=[new_score,total_score];
55 %按总分降序排列
56 result_report=sortrows(result_report,-com_num-1);
57 %输出结果
58 disp('特征值、贡献率、累计贡献率:')
59 DS
60 disp('信息保留率对应的主成分数与特征向量: T')

```

```

61 com_num
62 PV
63 disp('主成分得分及排序（最后一列为编号，其次为总得分）')
64 result_report

```

得到排序结果：[3,2,1,4,14,6,18,16,19,17,,5,11,9,15,12,8,7,13,10]

### 3.(1)导入数据，求出相关系数矩阵

```

65 setwd('F:\\XMU_files 数学竞赛数学建模\\\\\\\\CH15')
66 tt <- read.table('3.txt',header = F,sep = '_')
67 rownames(tt) = c("煤矿", '石油', '黑色', '有色', '非金
    属', '其他')
68 colnames(tt)=c('x1', 'x2', 'x3', 'x4', 'x5', 'x6')#资产
    总计固定资产净值从业人员("","'",'/万人工业产值主营业务
    收入利润总额','','','','')
69 a <- cor(tt)

```

输出结果为：

```

> a
      x1      x2      x3      x4      x5      x6
x1 1.000000 0.950353 0.896758 0.995527 0.996825 0.776617
x2 0.950353 1.000000 0.719069 0.956663 0.958405 0.929895
x3 0.896758 0.719069 1.000000 0.872742 0.874521 0.418738
x4 0.995527 0.956663 0.872742 1.000000 0.999843 0.802866
x5 0.996825 0.958405 0.874521 0.999843 1.000000 0.802917
x6 0.776617 0.929895 0.418738 0.802866 0.802917 1.000000

```

(2)求出相关系数矩阵的特征值和特征向量。

```

70 eigen(a)

```

输出结果为：

```
> eigen(a)
eigen() decomposition
$values
[1] 5.347190e+00 6.406453e-01 1.134120e-02 7.382813e-04 8.483548e-05 1.267798e-16

$vectors
      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
[1,] -0.4300814 -0.12555696 0.16122259 0.84941959 0.194817966 0.1174948
[2,] -0.4219583 0.26017923 0.62520159 -0.41051924 0.409965268 0.1636562
[3,] -0.3666387 -0.66137708 0.28566780 -0.19655458 -0.541234085 -0.1224023
[4,] -0.4309522 -0.07008813 -0.57433187 -0.21123204 0.002012997 0.6594602
[5,] -0.4313195 -0.07141703 -0.41293643 -0.13912990 0.369153166 -0.6947772
[6,] -0.3616641 0.68491594 -0.03391986 0.08573508 -0.603955290 -0.1637869
```

(3)求解方差贡献率和荷载矩阵。

```
71 test.pr <- princomp(tt, cor = T)
72 summary(test.pr, loading = T)
```

输出结果为：

```
> summary(test.pr, loading = T)
Importance of components:
      Comp.1      Comp.2      Comp.3      Comp.4      Comp.5      Comp.6
Standard deviation 2.3123993 0.8004032 0.1064951 0.0271713314 9.210618e-03 1.485924e-08
Proportion of Variance 0.8911984 0.1067742 0.0018902 0.0001230469 1.413925e-05 3.679952e-17
Cumulative Proportion 0.8911984 0.9979726 0.9998628 0.9999858608 1.000000e+00 1.000000e+00

Loadings:
      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
x1 -0.430 -0.126 0.161 0.849 0.195 0.117
x2 -0.422 0.260 0.625 -0.411 0.410 0.164
x3 -0.367 -0.661 0.286 -0.197 -0.541 -0.122
x4 -0.431      -0.574 -0.211      0.659
x5 -0.431      -0.413 -0.139 0.369 -0.695
x6 -0.362 0.685      -0.604 -0.164
```

(4)数据标准化，求主成分表达式，求综合得分。

```
73 zu <- as.data.frame(scale(tt))
74 attach(zu)
75 pc1 <- -0.430*x1 - 0.422*x2 - 0.367*x3 - 0.431*x4 - 0.431*
      x5 - 0.362*x6
76 y <- 0.8911984*pc1
77 detach(zu)行业
78 <- c("煤矿", "石油", "黑色", "有色", "非金属", "其他")
```



```

79 zu1 <- data.frame(行业(zu,pc1,y,row.names = NULL)
80 zu1

```

输出结果为:

```

> zu1
  行业      x1      x2      x3      x4      x5      x6      pc1      y
1  煤矿  1.6566659  1.2241862  2.00313100  1.5784883  1.5854435  0.5447797 -3.5249868 -3.1414626
2  石油  0.8055535  1.3461216 -0.09715668  0.8770785  0.8784608  1.8005437 -2.2872291 -2.0383749
3  黑色 -0.4263107 -0.5435844 -0.37337130 -0.3220310 -0.3518744 -0.3739073  0.9755412  0.8694007
4  有色 -0.5401569 -0.5842576 -0.41644224 -0.5026870 -0.5060820 -0.5372674  1.2609287  1.1237376
5  非金属 -0.6630386 -0.6433550 -0.41266881 -0.6528269 -0.6526295 -0.6701528  1.5132989  1.3486496
6  其他 -0.8327132 -0.7991108 -0.70349197 -0.9780219 -0.9533184 -0.7639959  2.0624471  1.8380496

```

(5)对得分进行排名:

```

81 Y=arrange(zu1,-y)
82 Y

```

输出结果为:

```

> Y
  行业      x1      x2      x3      x4      x5      x6      pc1      y
1  其他 -0.8327132 -0.7991108 -0.70349197 -0.9780219 -0.9533184 -0.7639959  2.0624471  1.8380496
2  非金属 -0.6630386 -0.6433550 -0.41266881 -0.6528269 -0.6526295 -0.6701528  1.5132989  1.3486496
3  有色 -0.5401569 -0.5842576 -0.41644224 -0.5026870 -0.5060820 -0.5372674  1.2609287  1.1237376
4  黑色 -0.4263107 -0.5435844 -0.37337130 -0.3220310 -0.3518744 -0.3739073  0.9755412  0.8694007
5  石油  0.8055535  1.3461216 -0.09715668  0.8770785  0.8784608  1.8005437 -2.2872291 -2.0383749
6  煤矿  1.6566659  1.2241862  2.00313100  1.5784883  1.5854435  0.5447797 -3.5249868 -3.1414626

```

(6)结果分析:

综合得分超过平均水平的依次是其他采矿业、非金属矿采选业、有色金属矿采选业和黑色金属矿采选业。这几个行业的经济效益较高，经济效益最差的为煤矿开采和洗选业、石油和天然气开采业。

4.使用以下代码计算:(因数据输入过于冗长选择省略)

```

83 g = [...] %省略
84 g_1=zscore(g) %标准化

```

```

85 g_2=corrcoef(g_1) %相关系数矩阵
86 [x,y,z]=pcacov(g_2) %主成分分析函数

```

得到相关系数矩阵

$$X = \begin{bmatrix} 0.4121 & 0.3023 & 0.1982 & -0.5939 & 0.0951 & -0.4832 & -0.3229 \\ 0.4658 & 0.1514 & -0.0817 & -0.1231 & 0.2533 & 0.1167 & 0.8127 \\ -0.2129 & 0.6000 & 0.3636 & 0.3239 & 0.5756 & 0.1262 & -0.1017 \\ 0.4168 & -0.2932 & -0.3374 & -0.0022 & 0.5033 & 0.4272 & -0.4367 \\ 0.4545 & 0.0193 & -0.1233 & 0.7176 & -0.0805 & -0.5014 & -0.0707 \\ 0.3994 & 0.3508 & 0.2434 & 0.0847 & -0.5681 & 0.5504 & -0.1590 \\ 0.1674 & -0.5624 & 0.7959 & 0.0717 & 0.1139 & 0.0097 & 0.0628 \end{bmatrix},$$

特征值 $y=[4.2825 \ 1.9244 \ 0.4244 \ 0.1812 \ 0.1118 \ 0.0556 \ 0.0201]$ ,

贡献率 $z=[61.1784 \ 27.4918 \ 6.0627 \ 2.5884 \ 1.5967 \ 0.7944 \ 0.2877]$ .

可以得到:

在第一主成分中排行靠前的因素是人均地区生产总值、人均第三产业增加值、第二产业占GDP的比重，排行前三的是上海、北京、天津，

在第二主成分中排行靠前的因素是第三产业固定资产投资比重、人均地区生产总值、第三产业就业人员比重，排行前三的是浙江、黑龙江、江西，

在第三主成分中排行靠前的因素是人均第三产业增加值、第三产业就业人员比重、第二产业占GDP的比重，排行前三的是湖南、云南、江西。

总分排行前三的是上海、北京、天津。

5.根据题意列出matlab代码如下:

```

1      clear
2
3      clc
4
5      %读取数据，并进行标准化变换
6
7      X = [5561.54  1571.74  1286.32  1096.57
8            1563.10  2293.23  2383.52  704.24;
9            5005.09  1153.66  1528.28  817.18  1220.92
10           1567.87  1608.97  520.49;
11           3155.40  1137.22  1097.41  574.84  808.88
12           1062.31  946.38  304.28;
13           2974.76  1137.11  1250.87  471.65  769.79
14           931.33  1041.91  228.53;
15           3553.48  1616.56  1028.19  672.64  869.71
16           1191.70  1383.53  512.81;
17           4378.14  1187.41  1270.95  507.40  913.13
18           1295.70  1145.46  533.29;
19           3307.14  1259.62  1285.28  510.49  914.47
20           954.96  1071.80  425.30;
21           3128.10  1217.04  941.25  494.49  864.89
22           749.05  906.19  321.95;
23           7108.62  1520.61  1646.19  1182.24  755.29
24           3373.19  2874.54  937.21;
25           4544.64  1166.91  1042.10  813.45  794.63
26           1357.96  1799.75  458.10;

```



14

5522.56 1546.46 1333.69 713.31 933.11

2392.63 2195.58 520.95;

15

3905.05 1010.61 988.12 579.59 633.93

920.77 1160.14 325.82;

16

5078.85 1105.31 1300.10 722.17 540.63

1777.06 1453.18 523.83;

17

3633.05 969.58 851.15 623.17 483.96

872.57 945.99 337.91;

18

3699.42 1394.11 1247.04 806.35 799.79

1410.45 1277.43 372.01;

19

3079.82 1141.76 963.59 633.32 790.87

915.12 988.95 324.03;

20

3996.27 1099.16 914.26 604.40 675.32

890.12 1037.24 260.74;

21

3970.42 1090.72 960.82 674.84 790.95

971.05 1110.11 376.62;

22

5866.91 975.06 1748.16 947.54 836.39

262308 1936.38 594.45;

23

4082.99 772.28 891.33 603.84 529.36

1376.03 1081.54 290.04;

24

4226.90 491.84 1106.39 565.51 536.40

1303.50 930.84 247.08;

25

4418.34 1294.30 1096.82 842.09 878.25

```

1044.36 1267.03 305.60;
26 4255.48 1042.45 819.28 590.51 564.93
    1121.45 947.01 338.03;
27 3597.94 851.50 836.54 525.70 471.39
    871.15 934.73 260.27;
28 4272.29 1026.50 739.20 331.94 606.86
    1216.46 732.95 150.42;
29 4262.77 1011.82 634.94 310.22 317.08
    966.74 419.59 400.38;
30 3586.13 1047.61 1007.68 618.16 862.70
    967.52 1281.58 400.68;
31 3183.79 1022.62 846.26 546.23 654.82
    817.17 936.33 301.40;
32 3315.94 945.14 802.73 538.54 610.02
    787.63 880.86 311.72;
33 3352.83 1178.88 1069.15 596.81 816.87
    1096.32 1043.72 403.71;
34 3235.77 1245.02 781.30 535.31 643.48
    1003.89 812.36 411.63];
35 XZ = zscore(X); %数据标准化
36 %主成分分析
37 % 调用函数根据标准化后原始样本观测数据作主成分
    分析princomp
38 %返回主成分表达式的系数矩阵，主成分得分数

```

```

据COEFFSCORE
39 %样本相关系数矩阵的特征值向量和每个观测值的霍特
   林统计量latentT2
40 [COEFF,SCORE,latent ,tsquare] = princomp(
   XZ)
41 % 为了直观，定义元胞数组，用来存放特征值、贡献
   率和累积贡献率等数据result1
42 %函数不返回贡献率，需要用协方差矩阵的特征值向量
   来计算princomlatent
43 explained = 100*latent/sum(latent);%计算
   贡献率
44
45 [m, n] = size(X);%求的行数和列数X
46 result1 = cell(n+1, 4);%定义一个n行、列的元
   胞数组+14
47 %中第一行存放的数据result1
48 result1(1,:) = {'特征值', '差值', '贡献
   率', '累积贡献率'};
49
50 %中第列的第行到最后一行存放的数
   据result112(latent) 特征值
51 result1(2:end,1) = num2cell(latent);
52
53 %中第列的第行到倒数第行存放的数
   据result1222(latent 的方差，特征值的方差)
54 result1(2:end-1,2) = num2cell(-diff(
   latent));
55

```

```

56 %中第列和第列的第行到最后一行分别存放主成分的贡
    献率和累积贡献率result1342
57 result1(2:end,3:4) = num2cell([explained
    , cumsum(explained)])
58
59 % 为了直观，定义元胞数组，用来存放前个主成分表
    达式的系数数据result22
60 result2 = cell(n+1, 3); %定义一个n行，列的
    元胞数组+13
61 result2(1,:) = {'标准化变量', '特征向
    量t1', '特征向量t2'};%的第一行数据result2
62 %result2(2:end, 1) = 食品衣着居住家庭设备及
    服务交通和通讯文教娱乐医疗保健其
    他['','','','','','','','','',''];%第
    列result21
63 result2(2:end, 2:end) = num2cell(COEFF
    (:,1:2)) %存放前个主成表达式的系数矩
    阵2
64
65 % 为了直观，定义元胞数组，用来存放每一个地区总
    的消费性支出，以及前个主成分的得分数
    据result32
66 cityname = ['北京','天津','河北','山西','内
    蒙古','辽宁','吉林','黑龙江','上海','江
    苏','浙江','安徽','福建','江西','山东','河
    南','湖北','湖南','广东','广西','海南','重
    庆','四川','贵州','云南','西藏','陕西','甘
    肃','青海','宁夏','新疆'];
67 sumXZ = sum(XZ,2); %按行求和，提取每个
    地区总的消费性支出
68 [s1, id] = sortrows(SCORE,1);%将主成得分数
    据按第一主成分得分SCORE第一列()从小到大排序

```

```

69     result3 = cell(m+1, 4); %定义一个m行，列的元
        胞数组+14
70     result3(1,:) = {'地区', '总支出', '第一主成
        分得分y1', '第二主成分得分y2'}; %第一行的
        数据
71     %result3(2:end, 1) = cityname(id); %的第一
        列的数据，排序后的城市名 result3
72     %第列为按排序的，第列为第一主成分得分，第列为第
        二主成分得分 result32idsumXZ3y14y2
73     result3(2:end, 2:end) = num2cell([sumXZ(
        id), s1(:,1:2)])

```

最后得出的结果为：从结果result1来看，前5个主成分的累积贡献率达到了百

```

COEFF =

    0.3665    0.2619   -0.4652   -0.0708    0.0428    0.5789   -0.3718   -0.3169
    0.2834   -0.5516    0.0853    0.6641   -0.0219   -0.0949   -0.3702   -0.1436
    0.3843    0.1855    0.2731   -0.2070    0.7066   -0.3826   -0.1737   -0.1566
    0.4091    0.0301   -0.0913   -0.2595   -0.5983   -0.4871    0.0934   -0.3900
    0.3033   -0.3785    0.5883   -0.3763   -0.1045    0.4948    0.1443   -0.0213
    0.1622    0.6690    0.4992    0.4423   -0.2555    0.1052   -0.0034    0.0705
    0.4328   -0.0118   -0.1725   -0.1290   -0.1225   -0.0949   -0.2143    0.8344
    0.4058   -0.0293   -0.2612    0.2999    0.2217    0.0862    0.7868    0.0259

```

图 1: 特征矩阵

分之94.9589，所以只用前两个主成分就可以了。有结果result2 写出前5个主成分的表达式如下：

$y_1 = 0.3431x_1 + 0.3384x_2 + 0.3552x_3 + 0.3692x_4 + 0.3752x_5 + 0.3587x_6 + 0.3427x_7 + 0.3427x_8$   
 $y_2 = 0.5035x_1 - 0.4866x_2 + 0.1968x_3 + 0.1088x_4 - 0.0547x_5 - 0.2208x_6 - 0.4783x_7 + 0.4225x_8$

## 一、因子分析

1. 计算原始数据的相关矩阵R及其特征值和特征向量，matlab代码如下：

```

1  clc; clear;
2  data = [1.611, 10.59, 0.69, 1.67;

```

```
result1 =
```

'特征值'	'差值'	'贡献率'	'累积贡献率'
[4.9004]	[3.5694]	[61.2551]	[ 61.2551]
[1.3310]	[0.6084]	[16.6376]	[ 77.8927]
[0.7226]	[0.3315]	[ 9.0321]	[ 86.9248]
[0.3910]	[0.1393]	[ 4.8877]	[ 91.8125]
[0.2517]	[0.0587]	[ 3.1464]	[ 94.9589]
[0.1930]	[0.0419]	[ 2.4121]	[ 97.3710]
[0.1511]	[0.0919]	[ 1.8888]	[ 99.2598]
[0.0592]	[ ]	[ 0.7402]	[ 100]

图 2: result1

```
result2 =
```

'标准化变量'	'特征向量t1'	'特征向量t2'
[ ]	[ 0.3665]	[ 0.2619]
[ ]	[ 0.2834]	[ -0.5516]
[ ]	[ 0.3843]	[ 0.1855]
[ ]	[ 0.4091]	[ 0.0301]
[ ]	[ 0.3033]	[ -0.3785]
[ ]	[ 0.1622]	[ 0.6690]
[ ]	[ 0.4328]	[ -0.0118]
[ ]	[ 0.4058]	[ -0.0293]

图 3: result2



```

result3 =

```

'地区'	'总支出'	'第一主成分得分y1'	'第二主成分得分y2'
[]	[-7.4442]	[-2.7374]	[0.5806]
[]	[-6.6202]	[-2.5314]	[0.1992]
[]	[-6.1778]	[-2.1659]	[0.7123]
[]	[-5.3173]	[-1.9141]	[0.1597]
[]	[-5.6233]	[-1.8320]	[1.8262]
[]	[-4.6887]	[-1.7142]	[-0.0978]
[]	[-4.4989]	[-1.5518]	[0.4357]
[]	[-4.6563]	[-1.5507]	[0.9859]
[]	[-3.5678]	[-1.3744]	[-0.6487]
[]	[-3.4747]	[-1.2345]	[0.2805]
[]	[-3.0538]	[-1.1814]	[-0.3249]
[]	[-2.8620]	[-1.1737]	[-0.8547]
[]	[-2.9286]	[-1.0817]	[-0.0244]
[]	[-2.5603]	[-0.9737]	[-0.5340]
[]	[-2.7723]	[-0.9505]	[0.2584]
[]	[-2.4303]	[-0.9445]	[-0.4404]
[]	[-1.1765]	[-0.4613]	[-0.5336]
[]	[-1.0642]	[-0.3696]	[-0.1776]
[]	[-0.9771]	[-0.3234]	[-0.2828]
[]	[0.0904]	[-0.0471]	[-0.7623]
[]	[1.6723]	[0.5922]	[-0.3250]
[]	[1.8813]	[0.6627]	[-0.5407]
[]	[2.0516]	[0.7297]	[-0.7523]
[]	[2.7202]	[0.8828]	[-1.6333]
[]	[2.2417]	[1.0156]	[0.7190]
[]	[2.7726]	[1.1523]	[-0.1514]
[]	[6.9170]	[2.5844]	[-0.3479]
[]	[7.7675]	[2.8707]	[-0.8143]
[]	[13.5963]	[4.1680]	[4.8030]
[]	[13.9814]	[5.1219]	[-1.8967]
[]	[16.2021]	[6.3329]	[0.1824]

图 4: result3

```

3      1.429,9.44,0.61,1.5;
4      1.447,5.97,0.24,1.25;
5      1.572,10.72,0.75,1.71;
6      1.483,10.99,0.75,1.44;
7      1.371,6.46,0.41,1.31;
8      1.665,10.51,0.53,1.52;
9      1.403,6.11,0.17,1.32;
10     2.620,21.51,1.4,2.59;
11     2.033,24.15,1.8,1.89;
12     2.015,26.86,1.93,2.02;
13     1.501,9.74,0.87,1.48;
14     1.578,14.52,1.12,1.47;
15     1.735,14.64,1.21,1.91;
16     1.453,12.88,0.87,1.52;
17     1.765,17.94,0.89,1.4;
18     1.532,29.42,2.52,1.8;
19     1.488,9.23,0.81,1.45;
20     2.586,16.07,0.82,1.83;
21     1.992,21.63,1.01,1.89];
22     e = mean(data,1);
23     v = std(data,0,1);
24     for i = 1:4
25         data(:,i) = (data(:,i)-e(i))/v(i);
26     end
27     R = cov(data);
28     [ga,lan] = eig(R);

```

得:

$$\lambda = \begin{pmatrix} 0.0369 & 0 & 0 & 0 \\ 0 & 0.1665 & 0 & 0 \\ 0 & 0 & 0.7822 & 0 \\ 0 & 0 & 0 & 3.0145 \end{pmatrix}, \gamma = \begin{pmatrix} -0.3156 & -0.5126 & 0.6590 & 0.4509 \\ 0.6624 & -0.3895 & -0.3516 & 0.5347 \\ -0.6483 & 0.0899 & -0.5758 & 0.4900 \\ 0.2033 & 0.7599 & 0.3325 & 0.5202 \end{pmatrix}.$$

根据提取方差贡献大于1决定因子提取个数的方式，四个变量有一个公因子，因子载荷阵

为:

$$A = \begin{pmatrix} 0.7829 & 0.5828 \\ 0.9284 & -0.3110 \\ 0.8508 & -0.5092 \\ 0.9033 & 0.2941 \end{pmatrix}$$

(2) 调用 $[\lambda, \psi, T, \text{stats}, F] = \text{factoran}(\text{data}, 1)$ ;和 $Y = \text{sort}(F)$ 得因子得分阵F以及排序后的矩阵Y:

$$F = \begin{pmatrix} -0.5524 \\ -0.7215 \\ -1.2269 \\ -0.5307 \\ -0.5001 \\ -1.1492 \\ -0.5750 \\ -1.2104 \\ 1.0274 \\ 1.3930 \\ 1.7774 \\ -0.6651 \\ 0.0101 \\ 0.0429 \\ -0.2309 \\ 0.4704 \\ 2.1540 \\ -0.7397 \\ 0.2265 \\ 1.0001 \end{pmatrix}, Y = \begin{pmatrix} -1.2269 \\ -1.2104 \\ -1.1492 \\ -0.7397 \\ -0.7215 \\ -0.6651 \\ -0.5750 \\ -0.5524 \\ -0.5307 \\ -0.5001 \\ -0.2309 \\ 0.0101 \\ 0.0429 \\ 0.2265 \\ 0.4704 \\ 1.0001 \\ 1.0274 \\ 1.3930 \\ 1.7774 \\ 2.1540 \end{pmatrix}$$

可得以下分类:

公因子小于-1: 企业3、8、6;

公因子大于-1小于0: 企业1、2、4、5、7、12、15、18;

公因子大于0小于1：企业13、14、16、19；

公因子大于1：企业9、10、11、17、20；

而根据最短距离法聚类分析的如下聚类图：

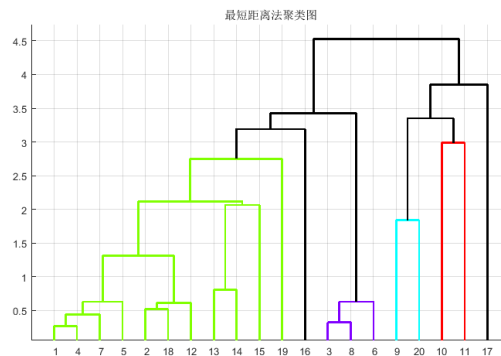


图 5: 聚类图

2.利用matlab进行因子分析求解，代码如下：

```
1 clear;clc;
2 grades=[65 61 72 84 81 79
3         77 77 76 64 70 55
4         67 63 49 65 67 57
5         80 69 75 74 74 63
6         74 70 80 84 81 74
7         78 84 75 64 71 64
8         66 71 67 52 65 57
9         77 71 57 72 86 71
10        83 100 79 41 67 50
11        86 94 97 51 63 55
12        74 80 88 64 73 66
13        67 84 53 58 66 56
14        81 62 69 56 66 52
15        71 64 69 56 66 52
16        78 96 81 80 89 76
17        69 56 67 75 94 80
18        77 90 80 68 66 60
19        84 67 75 60 70 63
20        62 67 83 71 85 77
21        74 65 75 72 90 73
```

```

22         91 74 97 62 71 66
23         72 87 72 79 83 76
24         82 70 83 68 77 85
25         63 70 60 91 85 82
26         74 79 95 59 74 59
27         66 61 77 62 73 64
28         90 82 98 47 71 60
29         77 90 85 68 73 76
30         91 82 84 54 62 60
31         78 84 100 51 60 60];
32     varnames={'数学','物理','化学','语文','历史','英语'}';
33     obsname=1:30;
34     obsname=obsname';
35     [lambda,psi,T,stats,F]=factoran(grades,2)
36     contribut=100*sum(lambda.^2)/6
37     cumcont=cumsum(contribut)
38     result1=num2cell(lambda);
39     result=[varnames,result1]
40     obsF=[num2cell(obsname),num2cell(F)]
41     j=1;k=1;
42     for i=1:30
43         if F(i,1)<F(i,2)
44             li(j)=i;
45             j=j+1;
46         else
47             wen(k)=i;
48             k=k+1;
49         end
50     end
51     fprintf('适合学理科的学生有: ')
52     for i=1:j-1
53         fprintf('%6d',li(i))
54     end
55     fprintf('\适合学文科的学生有: n')
56     for i=1:k-1

```

```

57     fprintf( '%6d' ,wen(i))
58 end

```

运行结果得：

适合学理科的学生有：2 6 7 9 10 11 13 14 17 18 21 25 27 28 29 30

适合学文科的学生有：1 3 4 5 8 12 15 16 19 20 22 23 24 26

首先我们将数据进行标准化，并求出相关系数矩阵的特征值与特征向量。由累计贡献率 $\geq 85\%$ 可得，主成分个数为2。根据主成分分析可求出最后得分。

```

87 data=readtable( '2.3.txt ')
88 data1=table2array( data)
89 data2=zscore( data1)
90 R=corrcoef( data2)
91 [X,B]=eig( R)
92 D=[X(:,4) ,X(:,3) ]
93 Z1=data2*X(:,4)
94 Z2=data2*X(:,3)
95 Y=0.8125*Z1+0.1171*Z2
96 T=[data table( Y) ]

```

4.

根据MATLAB得到的载荷矩阵lambda分析：因子Z1可解释为二三产业总值，Z2解释为二三产业分别占的比重

```

97 %q2_4
98 clear;clc;
99 X=[63029 5186.03 25.7 73.2 72.5 84.9 0.364
100     55473 2383.56 60.1 37.9 44 77.23 0.344
101     23239 853.68 54.2 33.2 26.5 41.9 0.223
102     20398 770.16 61.5 34.2 33 45.11 0.2583
103     32214 1184.36 55 33.3 32.7 51.71 0.2713

```



X1	X2	X3	X4	Y
148	41	72	78	0.129984704927520
139	34	71	76	-1.07456464145637
160	49	77	86	2.32269772178046
149	36	67	79	-0.597122474426747
159	45	80	86	2.25623865516670
142	31	66	67	-2.18911187105665
153	43	76	83	1.23132242731129
150	43	77	68	0.213091849348462
151	42	77	80	0.956714099952568
139	31	68	74	-1.70774258101044
140	29	64	74	-2.18154848446046
161	47	78	84	2.19166715580869
158	49	78	83	2.12244864961014
140	33	67	77	-1.41012536112365
137	31	66	73	-2.06254811279223
152	35	73	79	0.0472259919661948
149	47	82	79	1.63247091664650
145	35	70	77	-0.722326720645010
160	47	74	87	1.95553682319784
156	44	78	85	1.77572350553617
151	42	73	82	0.708048201407704
147	38	73	78	-0.0503339843646239
157	39	68	80	0.194753559948864
147	30	65	75	-1.58974132312199
157	39	68	80	0.194753559948864
151	36	74	80	0.228230252021998
144	36	68	76	-0.951422634578072
141	30	67	76	-1.64573869262286
139	32	68	73	-1.69726925441639
148	38	70	78	-0.281311938504459

```

104      31259 1192.02 55.8 34.5 41.2 60.05 0.255
105      23514 986.40 47.7 38 35.4 53.21 0.2545
106      21727 823.053 52.5 34.4 33 55.4 0.336
107      73124 4361.24 45.5 53.7 55.2 88.6 0.374
108      39622 1669.67 55 38.1 34.7 54.3 0.199
109      42214 1919.64 53.9 41 35.3 57.6 0.261
110      14485 597.99 46.6 37.4 28.3 40.5 0.248
111      30123 1308.21 50 39.3 33.3 49.9 0.336
112      14781 506.04 52.7 30.9 31.9 41.6 0.254
113      33083 1220.10 57 33.4 31 47.6 0.254
114      19593 620.81 56.9 28.6 24.4 36.03 0.201
115      19860 887.24 43.8 40.5 40.1 45.2 0.343
116      17521 731.36 44.2 37.8 30.4 42.15 0.317
117      37589 1787.76 51.6 42.9 38.2 63.37 0.295
118      14966 619.61 42.4 37.4 24.8 38.16 0.324
119      17175 766.11 29.8 40.2 34.9 48 0.484
120      18025 817.38 47.7 41 36.3 49.99 0.321
121      15378 590.05 46.3 34.8 33.1 37.4 0.307
122      8824 403.46 42.3 41.3 36.3 29.11 0.315
123      12587 544.14 43 39.1 25.1 33 0.348
124      13861 852.64 29.2 55.5 33.9 22.61 0.541
125      18246 663.41 56.1 32.9 32.5 42.1 0.407
126      12110 523.051 46.3 39.1 32.8 32.15 0.299
127      17389 652.19 55.1 34 34.2 40.86 0.330
128      17892 718.00 52.9 36.2 30.1 44.98 0.233
129      19893 750.14 49.6 33.9 34.7 39.64 0.273];
130 [lambda,psi,T,stats,F] = factoran(X,2);
131 Contribut = 100*sum(lambda.^2)/7;%计算贡献率
132 CumCont = cumsum(Contribut); %计算累积贡献率
133 disp('载荷, 贡献率和lambdacontributcumcont');
134 lambda,Contribut,CumCont %得到选取个因子载荷合理2

```

```

135 obsnum=[1:31]';
136 obsF=[obsnum,F];
137 disp('按因子排序【第一列为城市序号】Z1');
138 F1=sortrows(obsF,-2)
139 disp('按因子排序【第一列为城市序号】Z2');
140 F2=sortrows(obsF,-3)
141
142 % obsF =
143 %      number      factor1      factor2
144 %      1.0000      3.0048      3.0163
145 %      2.0000      1.6345     -1.4503
146 %      3.0000     -0.1731     -0.7215
147 %      4.0000     -0.2314     -0.6999
148 %      5.0000      0.2748     -1.0122
149 %      6.0000      0.3005     -0.9146
150 %      7.0000     -0.1459     -0.0770
151 %      8.0000     -0.2023     -0.5165
152 %      9.0000      3.0644      0.0471
153 %     10.0000      0.6990     -0.7743
154 %     11.0000      0.8733     -0.5028
155 %     12.0000     -0.6522      0.2160
156 %     13.0000      0.1662     -0.1280
157 %     14.0000     -0.5438     -0.6348
158 %     15.0000      0.3100     -1.0649
159 %     16.0000     -0.3310     -1.1524
160 %     17.0000     -0.4028      0.5400
161 %     18.0000     -0.5275      0.2717
162 %     19.0000      0.6708     -0.1187
163 %     20.0000     -0.6863      0.3601
164 %     21.0000     -0.6504      0.9925
165 %     22.0000     -0.4445      0.4721

```

```

166 %      23.0000    -0.6165     0.0052
167 %      24.0000    -1.0510     1.0840
168 %      25.0000    -0.8375     0.6526
169 %      26.0000    -1.0580     2.9729
170 %      27.0000    -0.4020    -0.5226
171 %      28.0000    -0.8243     0.5953
172 %      29.0000    -0.4436    -0.3811
173 %      30.0000    -0.4116    -0.2147
174 %      31.0000    -0.3625    -0.3396

```

5.(1)计算相关系数矩阵:

```

175 q5 <- read.table('5.txt',header = F,sep = ' ',col.
      names = c('x1','x2','x3','x4','x5','x6','x7','x8'
      ))
176 b <- cor(q5)

```

输出结果为:

```

> b
      x1      x2      x3      x4      x5      x6      x7      x8
x1 1.0000000 0.2822944 0.6463794 0.7244395 0.27185664 0.35236476 0.8193804 0.7655592
x2 0.2822944 1.0000000 0.3745849 0.4833445 0.62118987 -0.12124871 0.5729358 0.5997374
x3 0.6463794 0.3745849 1.0000000 0.7115044 0.56542137 0.47958057 0.7716317 0.6932502
x4 0.7244395 0.4833445 0.7115044 1.0000000 0.56400649 0.30108495 0.8968621 0.7682495
x5 0.2718566 0.6211899 0.5654214 0.5640065 1.00000000 0.06778063 0.5833063 0.4821320
x6 0.3523648 -0.1212487 0.4795806 0.3010850 0.06778063 1.00000000 0.2585385 0.2413172
x7 0.8193804 0.5729358 0.7716317 0.8968621 0.58330635 0.25853852 1.0000000 0.8460551
x8 0.7655592 0.5997374 0.6932502 0.7682495 0.48213200 0.24131719 0.8460551 1.0000000

```

(2)分析提取因子数:

```

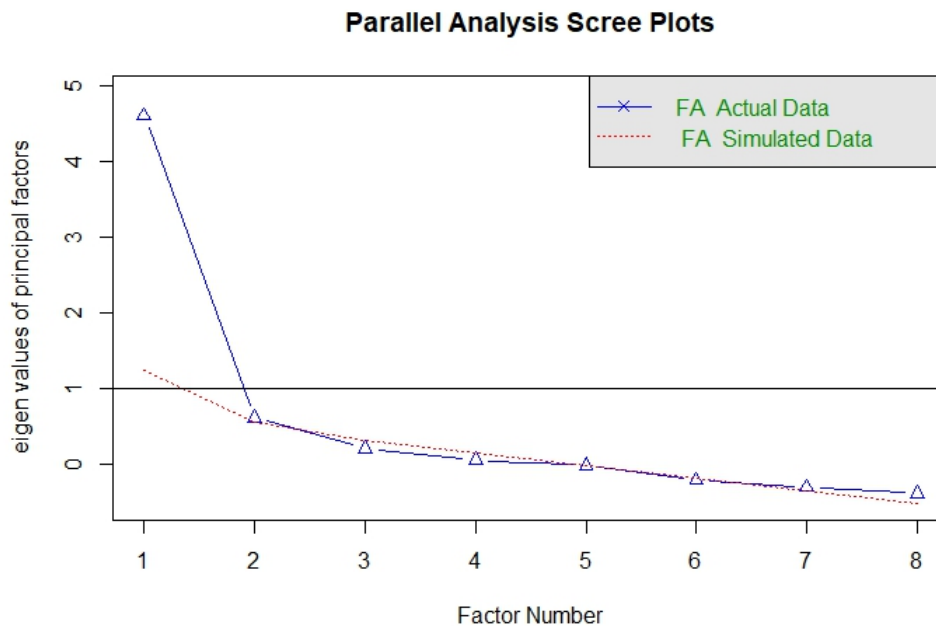
177 library(psych)
178 fa.parallel(b,n.obs = 31,fa = 'fa',n.iter = 100)

```

输出结果为:

Parallel analysis suggests that the number of factors = 2 and the number of

components = NA



现在我们确定提取两个因子，利用函数 $fa$ 求解荷载矩阵。

```
179 fa.varimax <- fa(b, nfactors = 2, n.obs = 31, rotate  
    = "varimax", fm = "pa", scores = T )
```

```
> fa.varimax
Factor Analysis using method = pa
Call: fa(r = b, nfactors = 2, n.obs = 31, rotate = "varimax", scores = T,
  fm = "pa")
Standardized loadings (pattern matrix) based upon correlation matrix
  PA1  PA2  h2  u2 com
x1 0.79  0.30 0.71 0.285 1.3
x2 0.00  0.94 0.88 0.124 1.0
x3 0.74  0.40 0.71 0.293 1.6
x4 0.70  0.55 0.80 0.204 1.9
x5 0.24  0.66 0.49 0.514 1.3
x6 0.55 -0.13 0.32 0.684 1.1
x7 0.74  0.64 0.95 0.055 2.0
x8 0.64  0.60 0.77 0.227 2.0

          PA1  PA2
SS loadings      2.97 2.64
Proportion Var    0.37 0.33
Cumulative Var    0.37 0.70
Proportion Explained 0.53 0.47
Cumulative Proportion 0.53 1.00
```

我们可以看到，因子1与X1（食品），X3（居住），X4（家庭设备用品及服务）

, X7（文教娱乐物品及服务）,X8（其他商品和服务）相关性强，关系密切，因子2与X2（衣着）,X5（医疗保健）联系紧密，我们可以把因子1叫做“享受型消费因子”，因子2叫做“一般消费因子”，我们分别对因子1、2进行排序。

```
180 fa$varimax$weights
181 attach(q5)
182 y1 <- 0.14163200*x1-0.53316620*x2+0.22846917*x3
      +0.07609864*x4-0.04213721*x5+0.12310146*x6
      +0.50609429*x7+0.19818503*x8
183 y2 <- -0.19853661*x1+0.75535260*x2-0.05765704*x3
      +0.01931537*x4+0.04482218*x5-0.04933633*x6+
      0.39197565*x7-0.01565477*x8
184 detach(q5)省份
185 <-c('BJ','TJ','HEB','SX1','NMG','LN','JL','HLJ','
      SH','JS','ZJ','AH','FJ','JX','SD','HEN','HUB','
      HUN',
186      'GD','GX','HAN','CQ','SC','GZ','YN','XZ','
      SX2','GS','QH','NX','XJ')
187 q51 <- data.frame省份(,q5,y1,y2)
188 q52 <- arrange(q51,-y1)
189 head(q52)
190 q53 <- arrange(q51,-y2)
191 head(q53)
```



```
> head(q52)
省份      x1      x2      x3      x4      x5      x6      x7      x8      y1      y2
1  GD 5866.91  975.06 1748.16  947.54  836.39 262308.00 1936.38 594.45 34135.637 -12664.8914
2  SH 7108.62 1520.61 1646.19 1182.24  755.29  3373.19 2874.54 937.21  2686.089   644.7072
3  BJ 5561.54 1571.74 1286.32 1096.57 1563.10  2293.23 2383.52 704.24  1889.316   910.2429
4  ZJ 5522.56 1546.46 1333.69  713.31  933.11  2392.63 2195.58 520.95  1786.273   784.8125
5  TJ 5005.09 1153.66 1528.28  817.18 1220.92  1567.87 1608.97 520.49  1564.144   405.2948
6  FJ 5078.85 1105.31 1300.10  722.17  540.63  1777.06 1453.18 523.83  1517.242   263.5195

> head(q53)
省份      x1      x2      x3      x4      x5      x6      x7      x8      y1      y2
1  NMG 3553.48 1616.56 1028.19  672.64  869.71 1191.70 1383.53 512.81  839.3688 983.7572
2  BJ 5561.54 1571.74 1286.32 1096.57 1563.10  2293.23 2383.52 704.24  1889.3163 910.2429
3  ZJ 5522.56 1546.46 1333.69  713.31  933.11  2392.63 2195.58 520.95  1786.2726 784.8125
4  SD 3699.42 1394.11 1247.04  806.35  799.79 1410.45 1277.43 372.01  987.0906 723.4082
5  SH 7108.62 1520.61 1646.19 1182.24  755.29  3373.19 2874.54 937.21  2686.0888 644.7072
6  JL 3307.14 1259.62 1285.28  510.49  914.47  954.96 1071.80 425.30  835.0482 637.9596
```

可见，广东、上海、北京、浙江、天津、福建、江苏等省（市、自治区）在享受型消费排名靠前，内蒙古、北京、浙江、山东、上海等省（市、自治区）在一般型消费排名靠前。北京、上海、浙江对于两种消费类型都比较注重。

使用以下代码计算：（因数据输入过于冗长选择省略）

```
192 X=[...] %省略
193 [ lambda , psi ,T, stats ,F] = factoran (X,3)
194 contribut = 100*sum(lambda .^2) /7
195 cumCont = cumsum( contribut )
```

得到如下结果：

$$\text{lambda} = \begin{bmatrix} 0.9610 & -0.1786 & 0.1406 \\ 0.2565 & -0.2532 & 0.7596 \\ 0.9074 & -0.2720 & -0.0568 \\ -0.0577 & -0.1177 & 0.9889 \\ -0.2794 & 0.9514 & -0.1086 \\ -0.1534 & 0.8566 & -0.2837 \\ 0.9672 & -0.1302 & 0.1646 \end{bmatrix},$$

贡献率为40.76,26.28,24.25，累积为91.29，基本覆盖。

可以解释为主要参考因素为地区生产总值、规定资产投资与工业总产值，基本符合实际情况。