Chapter 6 - Factor Analysis

张笑竹 / 201618070114

2018年11月23日

第六章实验采用课本例6-1，例6-2和例6-3中给出的数据进行因子分析。例6-1是Midwestern银行在1969-1971年之间雇员情况的数据；例6-2是全国重点水泥企业经济效益综合评价；例6-3是中心城市的综合发展评价分析。
将本次的实验任务拆分如下：
1）调用psych包中的函数，利用主成分法，对例6-1进行因子分析；
2）将上述函数进行封装，从而输出更加清晰的结果；
3）利用封装的函数，继续采用主成分法，对例6-2进行因子分析；
4）调用psych包和上述函数，利用主成分法、最大似然法、加权最小二乘法，对例6-3进行分析。

1 例6-1的分析

1.1 相关系数及其检验

首先，在R中读入例6-1的数据集，并对数据集的“行名”等进行处理。

```
X=read.csv("eg6_1.csv",header=T)
rownames(X)=X[,1]
X=X[,-1]
X1=X[,c(3,5:8)]
p=ncol(X1)
head(X1)
```

```
##   educ salary salbegin jobtime prevexp
## 1   15  57000    27000      98     144
## 2   16  40200    18750      98      36
## 3   12  21450    12000      98     381
## 4    8  21900    13200      98     190
## 5   15  45000    21000      98     138
## 6   15  32100    13500      98      67
```

上面输出了数据集的前6行，可以看出，经过处理后，数据集内的各个变量全部为连续变量。
下面，加载psych包，计算变量之间的相关性，从而判断进行因子分析是否合适。

```
library(psych)
corr.test(X1,use = "complete")
```

```
## Call:corr.test(x = X1, use = "complete")
## Correlation matrix
##           educ salary salbegin jobtime prevexp
## educ      1.00   0.66     0.63    0.05   -0.25
## salary    0.66   1.00     0.88    0.08   -0.10
## salbegin  0.63   0.88     1.00   -0.02    0.05
## jobtime   0.05   0.08    -0.02    1.00    0.00
## prevexp  -0.25  -0.10     0.05    0.00    1.00
## Sample Size
## [1] 474
## Probability values (Entries above the diagonal are adjusted for multiple tests.)
##          educ salary salbegin jobtime prevexp
## educ      0.0   0.00     0.00    1.00     0.0
## salary    0.0   0.00     0.00    0.34     0.2
## salbegin  0.0   0.00     0.00    1.00     1.0
## jobtime   0.3   0.07     0.67    0.00     1.0
## prevexp   0.0   0.03     0.33    0.95     0.0
##
##  To see confidence intervals of the correlations, print with the short=FALSE option
```

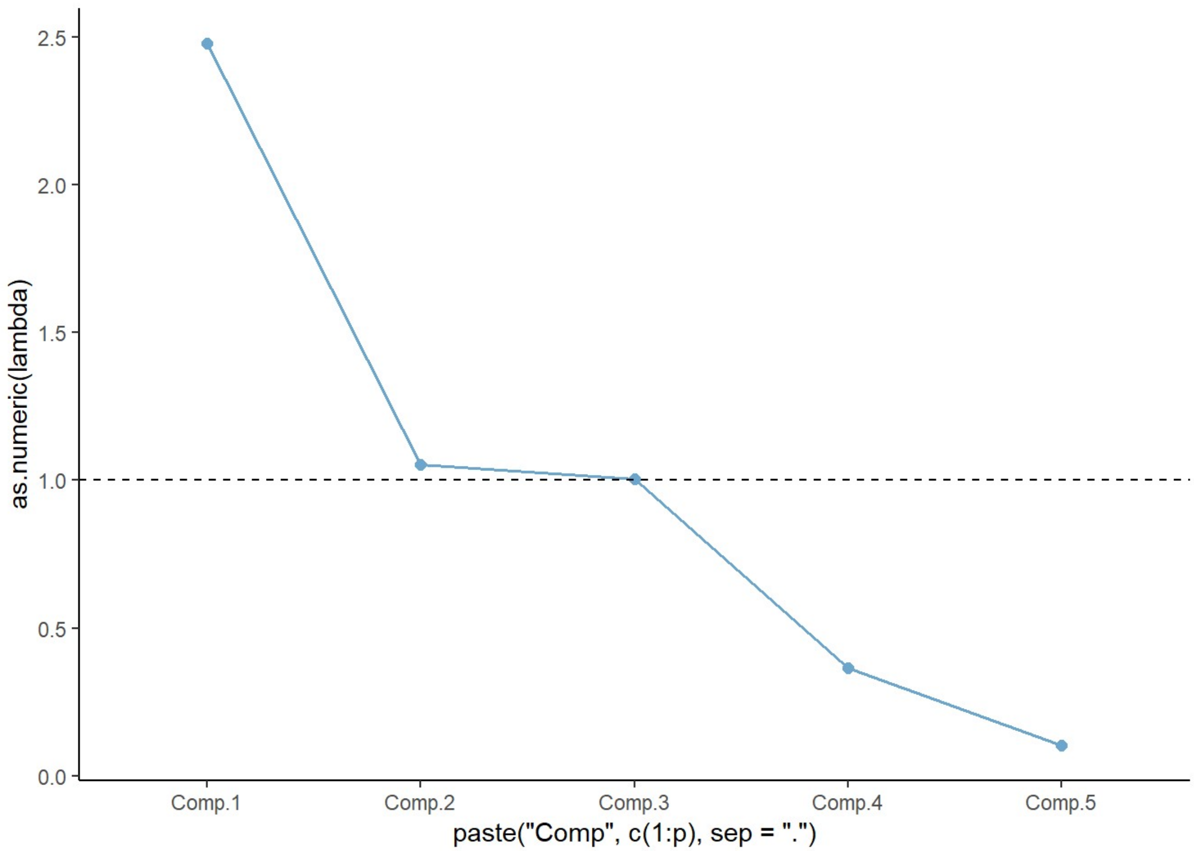
由上面的结果可知，原始变量之间有较强的相关性；对相关系数进行假设检验，得到的结果也大多显著。因此，进行因子分析是合适的。

1.2 主成分法：求解因子载荷

在本节中，为了得到因子载荷，首先对例6-1中的数据进行主成分分析，以便后续利用主成分法提取因子。由于在第五章中，已经编写了主成分分析的函数Pca(),因此直接调用Pca()即可。

```
#先进行主成分分析
source("Pca.R")
X.pca=Pca(X1,cor=T,1)
k=X.pca$`Number of Chosen Components`
X.pca[-7]
```

```
## $`Total Variance Explained`
## Importance of components:
##                           Comp.1    Comp.2    Comp.3     Comp.4     Comp.5
## Standard deviation     1.5738588 1.0258172 1.0017379 0.60410805 0.31975329
## Proportion of Variance 0.4954063 0.2104602 0.2006958 0.07298931 0.02044843
## Cumulative Proportion  0.4954063 0.7058665 0.9065623 0.97955157 1.00000000
##
## $`Variance of Component`
##    Comp.1    Comp.2    Comp.3    Comp.4    Comp.5
## 2.4770314 1.0523010 1.0034789 0.3649465 0.1022422
##
## $`Number of Chosen Components`
## [1] 3
##
## $`Component Matrix`
##                 [,1]        [,2]        [,3]
## educ      0.84618493  0.19385887 -0.01398843
## salary    0.94031332 -0.10446301  0.02857286
## salbegin  0.91669475 -0.26358611 -0.07689996
## jobtime   0.06805872  0.05224622  0.99588719
## prevexp  -0.17847154 -0.96519423  0.06900633
##
## $Communalities
##               [,1]
## educ     0.7538059
## salary   0.8959181
## salbegin 0.9157205
## jobtime  0.9991530
## prevexp  0.9682139
##
## $`Coefficient of Components`
##               Comp.1      Comp.2      Comp.3
## educ      0.53764985  0.18897993 -0.01396416
## salary    0.59745724 -0.10183394  0.02852329
## salbegin  0.58245045 -0.25695231 -0.07676654
## jobtime   0.04324322  0.05093131  0.99415942
## prevexp  -0.11339743 -0.94090273  0.06888661
##
## $`Scree Plot`
```



根据输出的结果，应该提取m=3个主成分，累计贡献率达到了90.66%. 事实上，根据主成分法求解因子载荷的理论，这里就提取m=3个公共因子。从样本相关阵R出发，因子载荷矩阵$\hat{A}$的一个解为

$$\hat{A} = (\sqrt{\lambda_1}\gamma_1, \sqrt{\lambda_2}\gamma_2, \ldots \sqrt{\lambda_m}\gamma_m)$$

其对应的矩阵就是输出的“Component Matrix”，其中每一个元素为主成分(因子)和变量的相关系数$\rho(Y_k, X_i)$.
此外，共同度的估计为

$$\hat{h_i}^2 = \hat{a_{i1}}^2 + \hat{a_{i2}}^2 + \ldots + \hat{a_{im}}^2$$

特殊度的估计为

$$1 - \hat{h_i^2}$$

共同度就是输出的“Communalities”. 特殊度为:

```
1-X.pca$"Communalities"
```

```
##                  [,1]
## educ     0.2461941201
## salary   0.1040819238
## salbegin 0.0842794849
## jobtime  0.0008470436
## prevexp  0.0317861335
```

最后，我们还可以计算出“因子得分系数矩阵”，其含义为，用原始变量表示标准化主成分（公共因子）的系

数矩阵。如果记主成分分析的系数矩阵为$P=(\gamma_1,\gamma_2,\ldots,\gamma_m)$, 且$Y=P^TX$，那么，为了得到标准化主成分（公共因子）$Y_i/\sqrt{\lambda_i}$，系数矩阵$C$有表达式

$$C=(\gamma_1/\sqrt{\lambda_1},\gamma_2/\sqrt{\lambda_2},\gamma_m/\sqrt{\lambda_m})$$

最终计算得到：

```
#计算的是"原始变量标准化的系数矩阵"
A=c()
for (i in 1:k){
  a=X.pca$`Coefficient of Components`[,i]/sqrt(X.pca$`Variance of Component`[i])
  A=cbind(A,a)
}
colnames(A)=paste("factor",c(1:k),sep="")
A
```

```
##              factor1     factor2     factor3
## educ      0.34161251  0.18422378 -0.01393993
## salary    0.37961300 -0.09927103  0.02847381
## salbegin  0.37007796 -0.25048547 -0.07663336
## jobtime   0.02747592  0.04964950  0.99243465
## prevexp  -0.07205057 -0.91722258  0.06876710
```

该矩阵给因子得分的计算提供了方便。

从上面的结果不难看出，用主成分法进行因子分析与主成分分析是完全可逆的。然而，我们得到的初始因子解各主因子的典型代表不是很突出，容易使因子的意义含糊不清，不便于对实际问题进行分析。出于这种考虑，就需要对初始公共因子进行线性组合，即进行旋转。

1.3 因子旋转

因子旋转分为正交旋转和斜交旋转，这里我们都进行尝试。

Step 1: 正交旋转

经过正交旋转而得到的新的公共因子仍然保持彼此独立的性质。调用函数principal()进行分析。

```
#正交旋转
X.var=principal((X1),nfactors = 3,rotate="varimax",scores=T)
X.var
```

```
## Principal Components Analysis
## Call: principal(r = (X1), nfactors = 3, rotate = "varimax", scores = T)
## Standardized loadings (pattern matrix) based upon correlation matrix
##            RC1   RC2   RC3   h2      u2 com
## educ      0.81 -0.31  0.04 0.75 0.24619 1.3
## salary    0.94 -0.02  0.07 0.90 0.10408 1.0
## salbegin  0.95  0.13 -0.05 0.92 0.08428 1.0
## jobtime   0.02  0.00  1.00 1.00 0.00085 1.0
## prevexp  -0.05  0.98  0.00 0.97 0.03179 1.0
##
##                        RC1  RC2  RC3
## SS loadings           2.45 1.08 1.01
## Proportion Var        0.49 0.22 0.20
## Cumulative Var        0.49 0.71 0.91
## Proportion Explained  0.54 0.24 0.22
## Cumulative Proportion 0.54 0.78 1.00
##
## Mean item complexity =  1.1
## Test of the hypothesis that 3 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.06
##  with the empirical chi square  30.37  with prob <  NA
##
## Fit based upon off diagonal values = 0.98
```

由结果可以看到，旋转后公共因子解释原始数据的能力并没有提高，共同度$\hat{h_i}^2$和特殊度$1-\hat{h_i}^2$均保持不变。
但是，因子载荷矩阵发生了变化，其中的元素更倾向于0,-1或1.
此时，因子1主要代表变量educ, salary, salbegin；因子2主要代表变量prevexp；而因子3主要代表变量jobtime.
此外，还可以输出因子得分系数矩阵(Component Score Coefficient Matrix).

```
X.var$weights  #Component Score Coefficient Matrix
```

```
##                  RC1         RC2         RC3
## educ      0.31350662 -0.22887048  0.01279603
## salary    0.38824314  0.04910004  0.04034815
## salbegin  0.40350923  0.19318053 -0.07372201
## jobtime  -0.01769493  0.01079848  0.99383943
## prevexp   0.05140954  0.92110594  0.01176316
```

记5各变量分别为$X_1, X_2, \ldots, X_5$，则因子得分为：

$$RC_1 = 0.314X_1 + 0.388X_2 + 0.404X_3 - 0.018X_4 + 0.051X_5$$

$$RC_2 = -0.229X_1 + 0.049X_2 + 0.193X_3 + 0.011X_4 + 0.921X_5$$

$$RC_3 = 0.013X_1 + 0.040X_2 - 0.074X_3 + 0.994X_4 + 0.012X_5$$

各个样本的因子得分计算如下：

```
head(X.var$scores)
```

```
##          RC1        RC2      RC3
## 1  1.1831392  0.6324034 1.640400
## 2  0.4337915 -0.6489026 1.670267
## 3 -0.6036863  2.4873138 1.710249
## 4 -1.0605143  1.1532385 1.660847
## 5  0.5997443  0.3977886 1.667571
## 6 -0.1129660 -0.4486952 1.699354
```

Step 2: 斜交旋转
有时为了使公共因子的实际意义更容易解释，往往需要放弃公共因子之间互不相关的约束而进行斜交旋转，此处采用Promax方法。依旧调用函数principal()进行分析。

```
#斜交旋转
X.pro=principal((X1),nfactors = 3,rotate="promax",scores=T)
X.pro
```

```
## Principal Components Analysis
## Call: principal(r = (X1), nfactors = 3, rotate = "promax", scores = T)
##
##  Warning: A Heywood case was detected.
## Standardized loadings (pattern matrix) based upon correlation matrix
##            RC1   RC2   RC3   h2      u2 com
## educ      0.78 -0.27  0.02 0.75 0.24619 1.2
## salary    0.95  0.03  0.05 0.90 0.10408 1.0
## salbegin  0.97  0.18 -0.06 0.92 0.08428 1.1
## jobtime  -0.01  0.02  1.00 1.00 0.00085 1.0
## prevexp   0.07  0.99  0.02 0.97 0.03179 1.0
##
##                        RC1  RC2  RC3
## SS loadings           2.44 1.08 1.01
## Proportion Var        0.49 0.22 0.20
## Cumulative Var        0.49 0.71 0.91
## Proportion Explained  0.54 0.24 0.22
## Cumulative Proportion 0.54 0.78 1.00
##
##  With component correlations of
##       RC1   RC2   RC3
## RC1  1.00 -0.18  0.05
## RC2 -0.18  1.00 -0.03
## RC3  0.05 -0.03  1.00
##
## Mean item complexity =  1.1
## Test of the hypothesis that 3 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.06
##  with the empirical chi square  30.37  with prob <  NA
##
## Fit based upon off diagonal values = 0.98
```

与正交旋转不同，斜交旋转的输出结果中包括Pattern Matrix和Correlation Matrix. 此外，还可以计算得到Structure Matrix:

$$Structure\ Matrix = Pattern\ Matrix \cdot Correlation\ Matrix$$

其中，Pattern Matrix就是因子载荷矩阵，而Structure Matrix为公共因子与原始变量的相关阵。也就是说，在斜交旋转中，因子载荷系数不再等于公共因子与原始变量的相关系数。
Structure Matrix计算如下：

```
X.pro$loadings[1:p,1:3]%*%X.pro$r.scores #计算Structure
```

```
##                 RC1         RC2         RC3
## educ      0.82721462 -0.40371992  0.06536966
## salary    0.94481643 -0.13738565  0.09471998
## salbegin  0.93699725  0.01640822 -0.02287365
## jobtime   0.04012815 -0.01266422  0.99939894
## prevexp  -0.09910815  0.98108298 -0.01108089
```

可以看出，因子1主要代表变量educ, salary, salbegin; 因子2主要代表变量prevexp; 而因子3主要代表变量jobtime. 这与正交旋转的结论是一致的。
另外，还可以计算出斜交旋转的因子得分系数矩阵、因子得分。

```
X.pro$weights
```

```
##                   RC1           RC2           RC3
## educ      3.253315e-01 -2.657337e-01  0.0253638243
## salary    3.857852e-01  5.935586e-04  0.0511953783
## salbegin  3.913302e-01  1.431813e-01 -0.0643959137
## jobtime  -5.797187e-06  1.323297e-06  0.9926192830
## prevexp   2.928447e-03  9.076197e-01  0.0002599844
```

```
head(X.pro$scores)
```

```
##           RC1        RC2      RC3
## 1  1.17800428  0.4612267 1.665803
## 2  0.49800335 -0.7187985 1.691457
## 3 -0.70265151  2.5201865 1.656158
## 4 -1.08925871  1.2529518 1.612046
## 5  0.60840237  0.2995894 1.678901
## 6 -0.05793339 -0.4534495 1.701411
```

Step 3: 验证
为了验证上一步得到的因子得分是经过标准化的，分别求它们的均值和标准差。

```
#对因子的"标准化"进行验证
#求均值
apply(X.pro$scores,2,mean)
```

```
##          RC1          RC2          RC3
## 6.494028e-17 1.025303e-16 5.318176e-16
```

```
#求标准差
apply(X.pro$scores,2,sd)
```

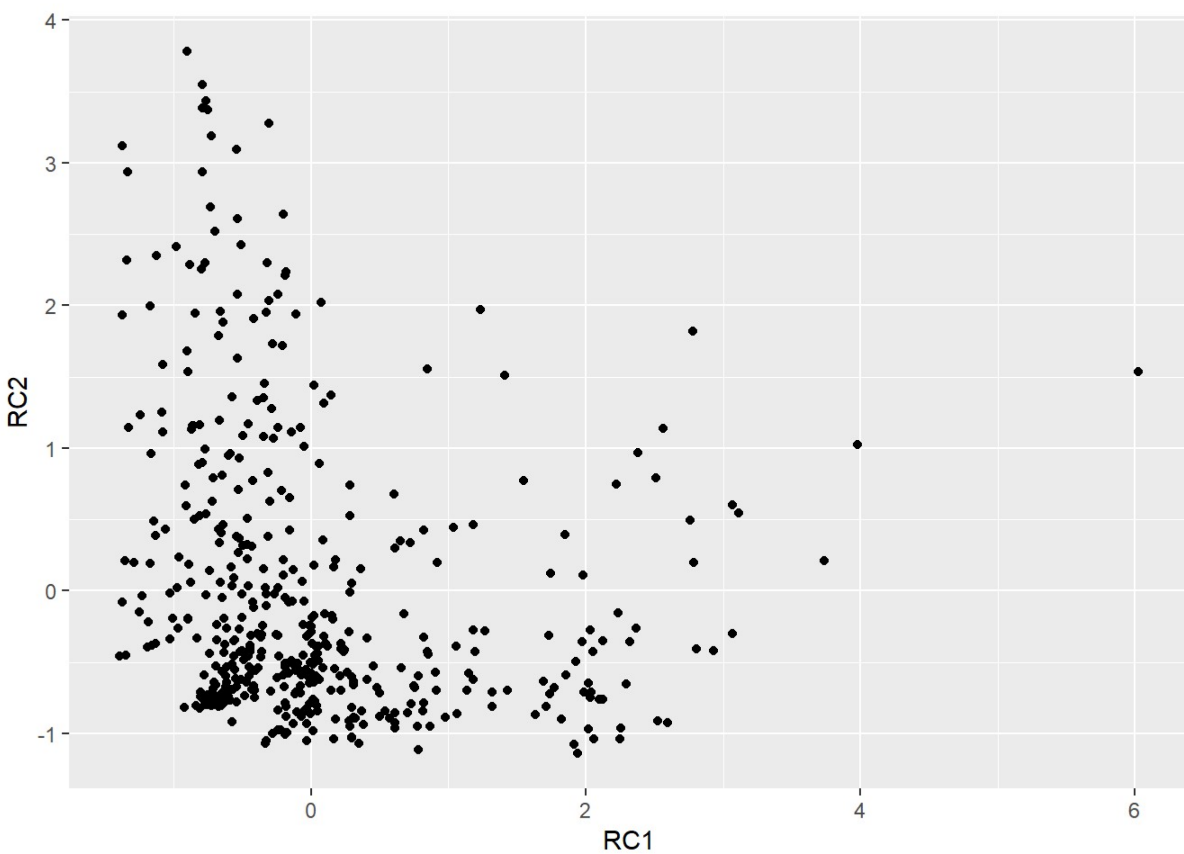
```
## RC1 RC2 RC3
##   1   1   1
```

三个变量的均值均为0，标准差均为1. 因子得分的标准化得到了验证。

1.4 可视化：散点图

最后，以因子1得分为横坐标，因子2得分为纵坐标，将各个样本绘入散点图。

```
#作图
library(ggplot2)
ggplot(X1,aes(x=X.pro$scores[,1], y = X.pro$scores[,2]))+
  geom_point()+
  labs(x = "RC1", y = "RC2")
```



由此可以直观地描述原始数据的散布情况。得分较低的样品占大多数，而得分较高的样品分布比较稀疏。

2 例6-2的分析

2.1 函数的封装

将上一部分中各个函数的输出进行整理，可以得到更为清晰的结果。函数封装如下：

```
Fa.pca<-function(X,cor,z0){
  source("Pca.R")
  #相关性检验
  cor_test=corr.test(X,use = "complete")

  #主成分分析
  pca<-Pca(X,cor,z0)

  #因子个数
  k=pca$"Number of Chosen Components"

  #总方差的解释
  va=pca$"Total Variance Explained"

  #方差
  lambda=pca$"Variance of Component"

  #载荷矩阵
  A=pca$"Component Matrix"

  #共同度&特殊度
  C=data.frame(pca$Communalities,1-pca$Communalities)
  colnames(C)=c("Communalities","Specificities")

  #成分得分系数矩阵
  Coeff=c()
  for (i in 1:k){
    a=pca$`Coefficient of Components`[,i]/sqrt(pca$`Variance of Component`[i])
    Coeff=cbind(Coeff,a)
  }
  colnames(Coeff)=paste("factor",c(1:k),sep="")

  #正交旋转
  X.var=principal((X),nfactors = k,rotate="varimax",scores=T)

  list("Correlation"=cor_test,"Number of Chosen Components"=k,"Total Variance Explained"=va,
       "Variance of Component"=lambda,"Component Matrix"=A,"Communities & Specialities"=C,
       "Component Score Coefficient Matrix"=Coeff,"Varimax"=X.var,"weights of Varimax"=X.var$wei
ghts,
       "scores"=X.var$scores
       )

}
```

函数的输入为：待分析的数据集、是否用相关阵(T&F)、临界特征值；这里输入的参数，与第五章中自行封装的主成分分析函数Pca()一致。 函数的输出为：相关检验结果、选择的因子个数、总方差解释、未标化的因子方差、因子载荷矩阵、自由度和共同度、因子得分系数矩阵、正交旋转结果、正交旋转的因子得分系数矩阵。

2.2 调用封装函数Fa.pca()分析例6-2

首先，在R中读入例6-2的数据集，并对数据集的“行名”等进行处理。由于指标“流动资金周转天数”和“万元产值能耗”为“成本型”变量”，“越小越好”，而其他变量为“收益型”变量，“越大越好”，因此需要将这2个成本型变量取倒数。

```
Y=read.csv("eg6_2.csv",header=T)
rownames(Y)=Y[,1]
Y=Y[,-1]
Y[,6]=1/Y[,6]
Y[,7]=1/Y[,7]
head(Y)
```

```
##          固定资产利税率  资金利税率  销售收入利税率  资金利润率  固定资产产值率
## 琉璃河          16.68       26.75           31.84       18.40           53.25
## 邯郸            19.70       27.56           32.94       19.20           59.82
## 大同            15.20       23.40           32.98       16.24           46.78
## 哈尔滨           7.29        8.97           21.30        4.76           34.39
## 华新            29.45       56.49           40.74       43.68           75.32
## 湘乡            32.93       42.78           47.98       33.87           66.46
##          流动资金周转天数  万元产值能耗  全员劳动生产率
## 琉璃河        0.01818182    0.03468609            1.75
## 邯郸          0.01818182    0.03037667            2.87
## 大同          0.01538462    0.02398657            1.53
## 哈尔滨        0.01612903    0.02545825            1.63
## 华新          0.01449275    0.03748126            2.14
## 湘乡          0.02000000    0.03042288            2.60
```

然后，调用Fa.pca()进行分析。从相关阵出发进行分析，保留特征根大于0.8的主成分（公共因子）。

```
Y.fa.pca=Fa.pca(Y,cor=T,0.8)
Y.fa.pca
```

```
## $Correlation
## Call:corr.test(x = X, use = "complete")
## Correlation matrix
##              固定资产利税率 资金利税率 销售收入利税率 资金利润率
## 固定资产利税率          1.00       0.85           0.92       0.90
## 资金利税率              0.85       1.00           0.69       0.99
## 销售收入利税率          0.92       0.69           1.00       0.77
## 资金利润率              0.90       0.99           0.77       1.00
## 固定资产产值率          0.65       0.72           0.54       0.69
## 流动资金周转天数        0.31       0.11           0.37       0.12
## 万元产值能耗            0.49       0.60           0.34       0.60
## 全员劳动生产率          0.60       0.26           0.53       0.33
##              固定资产产值率 流动资金周转天数 万元产值能耗
## 固定资产利税率          0.65             0.31         0.49
## 资金利税率              0.72             0.11         0.60
## 销售收入利税率          0.54             0.37         0.34
## 资金利润率              0.69             0.12         0.60
## 固定资产产值率          1.00             0.40         0.44
## 流动资金周转天数        0.40             1.00         0.34
## 万元产值能耗            0.44             0.34         1.00
## 全员劳动生产率          0.36             0.48         0.23
##              全员劳动生产率
## 固定资产利税率          0.60
## 资金利税率              0.26
## 销售收入利税率          0.53
## 资金利润率              0.33
## 固定资产产值率          0.36
## 流动资金周转天数        0.48
## 万元产值能耗            0.23
## 全员劳动生产率          1.00
## Sample Size
## [1] 15
## Probability values (Entries above the diagonal are adjusted for multiple tests.)
##              固定资产利税率 资金利税率 销售收入利税率 资金利润率
## 固定资产利税率          0.00       0.00           0.00       0.00
## 资金利税率              0.00       0.00           0.10       0.00
## 销售收入利税率          0.00       0.00           0.00       0.02
## 资金利润率              0.00       0.00           0.00       0.00
## 固定资产产值率          0.01       0.00           0.04       0.00
## 流动资金周转天数        0.26       0.70           0.18       0.67
## 万元产值能耗            0.06       0.02           0.21       0.02
## 全员劳动生产率          0.02       0.34           0.04       0.23
##              固定资产产值率 流动资金周转天数 万元产值能耗
## 固定资产利税率          0.17             1.00         0.90
## 资金利税率              0.05             1.00         0.35
## 销售收入利税率          0.57             1.00         1.00
## 资金利润率              0.10             1.00         0.35
## 固定资产产值率          0.00             1.00         1.00
## 流动资金周转天数        0.14             0.00         1.00
## 万元产值能耗            0.10             0.21         0.00
## 全员劳动生产率          0.19             0.07         0.42
##              全员劳动生产率
## 固定资产利税率          0.35
## 资金利税率              1.00
## 销售收入利税率          0.63
## 资金利润率              1.00
```

```
## 固定资产产值率             1.00
## 流动资金周转天数           0.91
## 万元产值能耗               1.00
## 全员劳动生产率             0.00
##
##  To see confidence intervals of the correlations, print with the short=FALSE option
##
## $`Number of Chosen Components`
## [1] 3
##
## $`Total Variance Explained`
## Importance of components:
##                           Comp.1    Comp.2    Comp.3     Comp.4     Comp.5
## Standard deviation     2.2046848 1.1265763 0.9148794 0.71909488 0.61495939
## Proportion of Variance 0.6075794 0.1586468 0.1046255 0.06463718 0.04727188
## Cumulative Proportion  0.6075794 0.7662262 0.8708517 0.93548888 0.98276076
##                            Comp.6      Comp.7      Comp.8
## Standard deviation     0.33980626 0.145353408 0.036303945
## Proportion of Variance 0.01443354 0.002640952 0.000164747
## Cumulative Proportion  0.99719430 0.999835253 1.000000000
##
## $`Variance of Component`
##      Comp.1      Comp.2      Comp.3      Comp.4      Comp.5      Comp.6
## 4.860635113 1.269174208 0.837004302 0.517097444 0.378175048 0.115468294
##      Comp.7      Comp.8
## 0.021127613 0.001317976
##
## $`Component Matrix`
##                        [,1]          [,2]        [,3]
## 固定资产利税率   -0.9569816  0.0185733910 -0.23930221
## 资金利税率       -0.8989949  0.3955672959  0.03720129
## 销售收入利税率   -0.8618345 -0.0813848424 -0.33812955
## 资金利润率       -0.9275701  0.3504529503 -0.03762397
## 固定资产产值率   -0.7867504 -0.0002493985  0.18197264
## 流动资金周转天数 -0.4224862 -0.7733416983  0.34535298
## 万元产值能耗     -0.6404981  0.0778927259  0.64218500
## 全员劳动生产率   -0.5707680 -0.6154593954 -0.31276728
##
## $`Communities & Specialities`
##                  Communalities Specificities
## 固定资产利税率       0.9734242    0.02657576
## 资金利税率           0.9660492    0.03395076
## 销售收入利税率       0.8637137    0.13628626
## 资金利润率           0.9846192    0.01538085
## 固定资产产值率       0.6520902    0.34790978
## 流动资金周转天数     0.8958206    0.10417936
## 万元产值能耗         0.8287066    0.17129336
## 全员劳动生产率       0.8023898    0.19761024
##
## $`Component Score Coefficient Matrix`
##                      factor1       factor2     factor3
## 固定资产利税率   -0.19688406  0.0146342329 -0.28590320
## 资金利税率       -0.18495420  0.3116729708  0.04444575
## 销售收入利税率   -0.17730902 -0.0641242486 -0.40397588
## 资金利润率       -0.19083311  0.2761267509 -0.04495075
## 固定资产产值率   -0.16186164 -0.0001965046  0.21740944
```

```
## 流动资金周转天数 -0.08691996 -0.6093266735  0.41260598
## 万元产值能耗      -0.13177251  0.0613727614  0.76724217
## 全员劳动生产率    -0.11742663 -0.4849290125 -0.37367463
##
## $Varimax
## Principal Components Analysis
## Call: principal(r = (X), nfactors = k, rotate = "varimax", scores = T)
## Standardized loadings (pattern matrix) based upon correlation matrix
##                     RC1   RC2   RC3   h2    u2 com
## 固定资产利税率     0.91  0.33  0.17 0.97 0.027 1.3
## 资金利税率         0.88 -0.06  0.43 0.97 0.034 1.5
## 销售收入利税率     0.84  0.40  0.03 0.86 0.136 1.4
## 资金利润率         0.92  0.00  0.37 0.98 0.015 1.3
## 固定资产产值率     0.59  0.27  0.48 0.65 0.348 2.3
## 流动资金周转天数 -0.05  0.85  0.41 0.90 0.104 1.5
## 万元产值能耗       0.31  0.11  0.85 0.83 0.171 1.3
## 全员劳动生产率     0.40  0.79 -0.11 0.80 0.198 1.5
##
##                        RC1  RC2  RC3
## SS loadings           3.77 1.71 1.48
## Proportion Var        0.47 0.21 0.19
## Cumulative Var        0.47 0.69 0.87
## Proportion Explained  0.54 0.25 0.21
## Cumulative Proportion 0.54 0.79 1.00
##
## Mean item complexity =  1.5
## Test of the hypothesis that 3 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.06
##  with the empirical chi square  3.05  with prob <  0.88
##
## Fit based upon off diagonal values = 0.99
## $`weights of Varimax`
##                          RC1         RC2         RC3
## 固定资产利税率     0.28789730  0.07225905 -0.18058767
## 资金利税率         0.24562917 -0.22893886  0.14344626
## 销售收入利税率     0.29224049  0.14593011 -0.30339218
## 资金利润率         0.27471885 -0.18839974  0.06098294
## 固定资产产值率     0.04862226  0.04422374  0.26295671
## 流动资金周转天数 -0.30346618  0.57566559  0.35439410
## 万元产值能耗      -0.17912625 -0.05621109  0.75798764
## 全员劳动生产率     0.08462470  0.51636635 -0.33879543
##
## $scores
##               RC1         RC2        RC3
## 琉璃河 -0.9733702  0.56919157  1.6168952
## 邯郸   -0.3168183  1.98510235 -0.1187031
## 大同   -0.1926769 -0.33827840 -1.0750665
## 哈尔滨 -1.8597453  0.09682695 -0.3889817
## 华新    1.8819412 -0.82303624  1.2263491
## 湘乡    1.2784065  2.19859561 -0.2251075
## 柳州    0.9772931  0.59034084 -0.7598938
## 峨嵋   -0.4689905 -0.98539427 -1.3228722
## 耀县    0.4830541 -0.50774065 -1.1146725
## 永登    0.5640397 -0.12202659 -0.3348595
## 工源   -0.8149367 -0.07307509  1.4835505
```

```
## 抚顺   -0.7553196 -0.23278274  0.9502709
## 大连    0.0832414 -1.10442866 -0.4162849
## 江南    0.8649615 -1.09755665  1.0869464
## 江油   -0.7510800 -0.15573803 -0.6075702
```

根据输出的结果，数据集中的各个变量有较强的相关性且假设检验结果大多显著，因此进行因子分析是合适的。利用主成分法，提取了3个因子。旋转前的载荷矩阵为“Component Matrix”,其对应的共同度和特殊度为“Communities & Specialities”. 进行最大方差旋转后，因子载荷和共同度通过“Varimax”给出。
若记8个变量分别为$X_1, X_2, ..., X_8$,则因子1主要由$X_1, X_2, ..., X_5$解释，因子2主要由X_6, X_8解释，因子3由X_7解释。企业经济效益指标体系的因子分析模型为：

$$X_1 = 0.91RC_1 + 0.33RC_2 + 0.17RC_3 + \epsilon_1$$

$$X_2 = 0.88RC_1 - 0.06RC_2 + 0.43RC_3 + \epsilon_2$$

$$X_3 = 0.84RC_1 + 0.40RC_2 + 0.03RC_3 + \epsilon_3$$

$$X_4 = 0.92RC_1 + 0.00RC_2 + 0.37RC_3 + \epsilon_4$$

$$X_5 = 0.59RC_1 + 0.27RC_2 + 0.48RC_3 + \epsilon_5$$

$$X_6 = -0.05RC_1 + 0.85RC_2 + 0.41RC_3 + \epsilon_6$$

$$X_7 = 0.31RC_1 + 0.11RC_2 + 0.85RC_3 + \epsilon_7$$

$$X_8 = 0.40RC_1 + 0.79RC_2 - 0.11RC_3 + \epsilon_8$$

根据旋转后的因子得分系数矩阵“weights of Varimax”， 因子得分为

$$RC_1 = 0.288X_1 + 0.246X_2 + 0.292X_3 + 0.275X_4 + 0.049X_5 - 0.303X_6 - 0.179X_7 + 0.085X_8$$

$$RC_2 = 0.072X_1 - 0.229X_2 + 0.146X_3 - 0.188X_4 + 0.044X_5 + 0.576X_6 - 0.056X_7 + 0.516X_8$$

$$RC_3 = -0.181X_1 + 0.143X_2 - 0.303X_3 + 0.061X_4 + 0.263X_5 + 0.354X_6 + 0.758X_7 - 0.339X_8$$

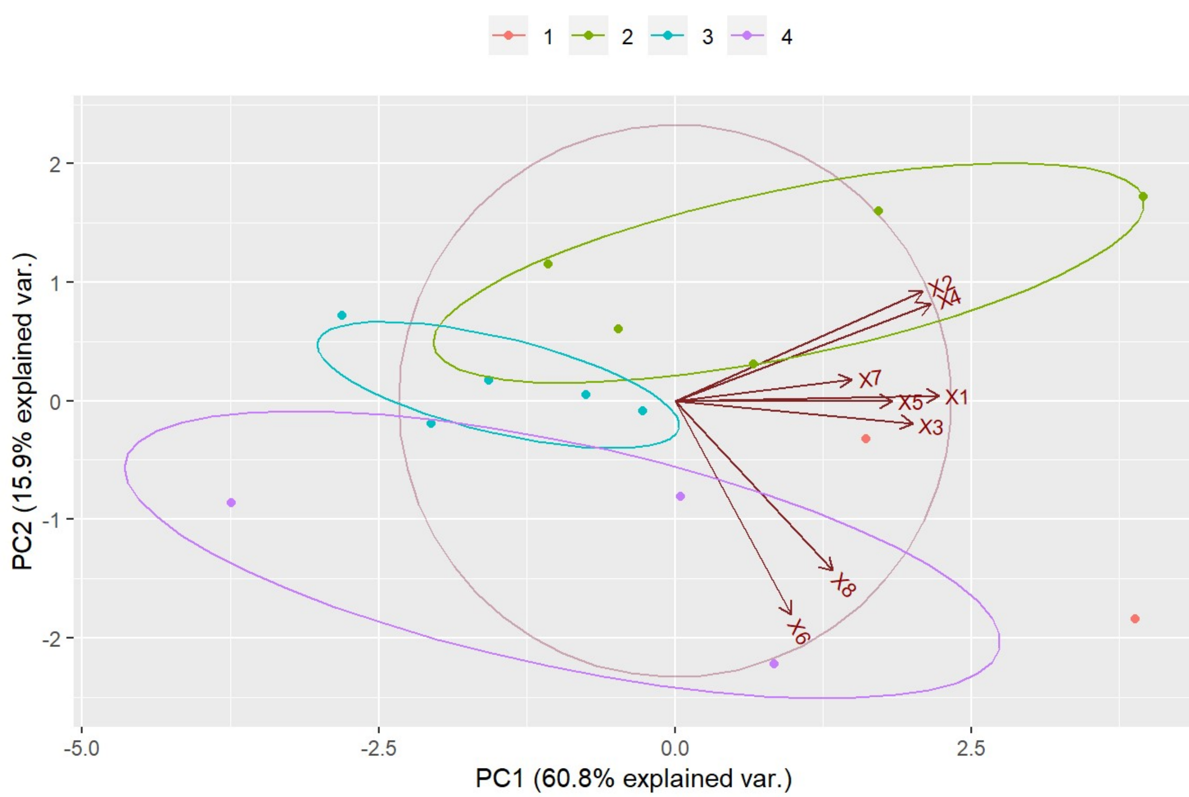
此外，在“Scores”中，得到了所有样本的因子得分。

2.3 可视化

最后，将各个样品的前2个因子得分在二维空间中进行描绘。

```
#先分组
type=c()
for(i in 1:nrow(Y)){
  if(Y.fa.pca$scores[i,1]>0 && Y.fa.pca$scores[i,2]>0){
    type=c(type,1)
  }
  if(Y.fa.pca$scores[i,1]>0 && Y.fa.pca$scores[i,2]<0){
    type=c(type,2)
  }
  if(Y.fa.pca$scores[i,1]<0 && Y.fa.pca$scores[i,2]<0){
    type=c(type,3)
  }
  if(Y.fa.pca$scores[i,1]<0 && Y.fa.pca$scores[i,2]>0){
    type=c(type,4)
  }
}

#再作图
Y=cbind(Y,type)
colnames(Y)=c("X1","X2","X3","X4","X5","X6","X7","X8","type")
Y.var<-prcomp(Y[,1:8], scale. = TRUE)
library(ggbiplot)
ggbiplot(Y.var, obs.scale = 1, var.scale = 1,
         groups = as.factor(type), ellipse = TRUE, circle = TRUE,size=2) +
  scale_color_discrete(name = '') +
  theme(legend.direction = 'horizontal', legend.position = 'top')
```



3 例6-3的分析

3.1 加权最小二乘法

在例6-3中，除了利用主成分法计算载荷矩阵，还尝试使用极大似然法、加权最小二乘法等进行求解。

Step 1: 求解初始因子

首先，在R中读入例6-3的数据集，并对数据集的“行名”等进行处理。

```
Z=read.csv("eg6_3.csv",header=T)
rownames(Z)=Z[,1]
Z=Z[,-1]
p=ncol(Z)
```

然后，以主成分个数为原则，确定因子的个数。

```
#选择因子个数
k=Pca(Z,cor=T,1)$"Number of Chosen Components"
k
```

```
## [1] 3
```

选择3个公共因子。

调用函数fa(),参数fm选择“wls”，先不进行旋转。

```
#加权最小二乘法
library(GPArotation)
Z.wls<-fa(cor(Z),nfactors = k,n.obs = nrow(Z),rotate = "none",scores = T,fm="wls")
Z.wls
```

```
## Factor Analysis using method =  wls
## Call: fa(r = cor(Z), nfactors = k, n.obs = nrow(Z), rotate = "none", 
##     scores = T, fm = "wls")
## Standardized loadings (pattern matrix) based upon correlation matrix
##     WLS1  WLS2  WLS3   h2    u2 com
## x1  0.85 -0.30  0.11 0.82 0.182 1.3
## x2  0.82  0.23  0.15 0.75 0.248 1.2
## x3  0.80 -0.19  0.22 0.73 0.269 1.3
## x4  0.75 -0.18 -0.23 0.64 0.357 1.3
## x5  0.93  0.06  0.08 0.87 0.129 1.0
## x6  0.96 -0.03 -0.04 0.92 0.081 1.0
## x7  0.90 -0.19 -0.09 0.86 0.145 1.1
## x8  0.94 -0.04 -0.11 0.90 0.103 1.0
## x9  0.06  0.47  0.68 0.68 0.319 1.8
## x10 0.20  0.81 -0.12 0.70 0.296 1.2
## x11 0.24  0.85 -0.08 0.79 0.213 1.2
## x12 0.23  0.62 -0.29 0.53 0.474 1.7
## 
##                        WLS1 WLS2 WLS3
## SS loadings           6.22 2.23 0.73
## Proportion Var        0.52 0.19 0.06
## Cumulative Var        0.52 0.70 0.77
## Proportion Explained  0.68 0.24 0.08
## Cumulative Proportion 0.68 0.92 1.00
## 
## Mean item complexity =  1.3
## Test of the hypothesis that 3 factors are sufficient.
## 
## The degrees of freedom for the null model are  66  and the objective function was  23.27 with
Chi Square of  678.8
## The degrees of freedom for the model are 33  and the objective function was  8.89 
## 
## The root mean square of the residuals (RMSR) is  0.06 
## The df corrected root mean square of the residuals is  0.08 
## 
## The harmonic number of observations is  35 with the empirical chi square  16.56  with prob <
0.99 
## The total number of observations was  35  with Likelihood Chi Square =  241.38  with prob <
1.6e-33 
## 
## Tucker Lewis Index of factoring reliability =  0.264
## RMSEA index =  0.49  and the 90 % confidence intervals are  0.381 NA
## BIC =  124.06
## Fit based upon off diagonal values = 0.99
## Measures of factor score adequacy             
##                                                   WLS1 WLS2 WLS3
## Correlation of (regression) scores with factors   0.98 0.92 0.86
## Multiple R square of scores with factors          0.95 0.84 0.75
## Minimum correlation of possible factor scores     0.90 0.68 0.49
```

输出的界面与主成分法十分类似，主要的信息集中于“Pattern matrix”，其中各个元素的含义是主成分与原始变量相关系数，最后也输出了共同度和特殊度。但是，此时的各个初始因子意义并不明确，因此需要进行旋转。

Step 2: 正交旋转
先尝试正交旋转。

```
#进行旋转(Varimax)
Z.wls.var<-fa(cor(Z),nfactors = k,n.obs = nrow(Z),rotate = "varimax",scores = T,fm="wls")
Z.wls.var
```

```
## Factor Analysis using method =  wls
## Call: fa(r = cor(Z), nfactors = k, n.obs = nrow(Z), rotate = "varimax",
##     scores = T, fm = "wls")
## Standardized loadings (pattern matrix) based upon correlation matrix
##      WLS1  WLS2  WLS3   h2    u2 com
## x1   0.89 -0.15  0.02 0.82 0.182 1.1
## x2   0.77  0.32  0.25 0.75 0.248 1.6
## x3   0.83 -0.10  0.17 0.73 0.269 1.1
## x4   0.76  0.06 -0.26 0.64 0.357 1.2
## x5   0.90  0.20  0.13 0.87 0.129 1.1
## x6   0.94  0.17 -0.02 0.92 0.081 1.1
## x7   0.92  0.03 -0.12 0.86 0.145 1.0
## x8   0.92  0.18 -0.09 0.90 0.103 1.1
## x9  -0.01  0.21  0.80 0.68 0.319 1.1
## x10  0.04  0.82  0.16 0.70 0.296 1.1
## x11  0.07  0.86  0.22 0.79 0.213 1.1
## x12  0.10  0.72 -0.06 0.53 0.474 1.0
##
##                       WLS1 WLS2 WLS3
## SS loadings           6.07 2.21 0.91
## Proportion Var        0.51 0.18 0.08
## Cumulative Var        0.51 0.69 0.77
## Proportion Explained  0.66 0.24 0.10
## Cumulative Proportion 0.66 0.90 1.00
##
## Mean item complexity =  1.1
## Test of the hypothesis that 3 factors are sufficient.
##
## The degrees of freedom for the null model are  66  and the objective function was  23.27 with
Chi Square of  678.8
## The degrees of freedom for the model are 33  and the objective function was  8.89
##
## The root mean square of the residuals (RMSR) is  0.06
## The df corrected root mean square of the residuals is  0.08
##
## The harmonic number of observations is  35 with the empirical chi square  16.56  with prob <
0.99
## The total number of observations was  35  with Likelihood Chi Square =  241.38  with prob <
1.6e-33
##
## Tucker Lewis Index of factoring reliability =  0.264
## RMSEA index =  0.49  and the 90 % confidence intervals are  0.381 NA
## BIC =  124.06
## Fit based upon off diagonal values = 0.99
## Measures of factor score adequacy
##                                                   WLS1 WLS2 WLS3
## Correlation of (regression) scores with factors   0.97 0.90 0.88
## Multiple R square of scores with factors          0.95 0.81 0.78
## Minimum correlation of possible factor scores     0.89 0.63 0.56
```

仍然关注pattern matrix，此时新的因子载荷系数要么尽可能地接近于零，要么尽可能地远离零。但是，共同

度和特殊度与旋转前相比并没有发生变化，各个新的因子依然保持独立。可以看到因子1与变量$X_1, X_2, \ldots, X_8$关系密切，因子2与X_{10}, X_{11}, X_{12}关系密切，而因子3与变量X_9关系密切。
若记该pattern matrix为A，$X=(X_1,X_2,\ldots,X_{12})^T, F=(F_1,F_2,F_3)^T, \epsilon=(\epsilon_1,\epsilon_2,\ldots,\epsilon_{12})^T$那么就有

$$X = A \cdot F + \epsilon$$

将数据代入即可得到模型的具体表达式。
此外，还可以通过估计的方法计算出各个因子得分。在最小二乘意义下，有

$$\hat{F} = A^T R^{-1} X$$

计算结果为：

```
Z.wls.var$weights
```

```
##            WLS1        WLS2        WLS3
## x1   0.03513541 -0.18015117  0.04642660
## x2  -0.37862986 -0.10053470  0.80655316
## x3  -0.06922732 -0.21227887  0.63553133
## x4  -0.91822928 -0.37473454  1.76682721
## x5   0.48336628  0.14876803 -0.27588038
## x6  -0.25140014 -0.15419543  1.09875588
## x7   0.39166693  0.16173633 -0.28989484
## x8   1.53297477  0.57242290 -3.34530457
## x9  -0.06376092 -0.07049093  0.84052634
## x10 -0.01037187  0.17744808 -0.03703225
## x11 -0.02894636  0.48545251 -0.01232630
## x12 -0.14139664  0.22401531  0.16560474
```

Step 3: 斜交旋转
再尝试斜交旋转。

```
#进行旋转(Promax)
Z.wls.pro<-fa(cor(Z),nfactors = k,n.obs = nrow(Z),rotate = "promax",scores = T,fm="wls")
Z.wls.pro
```

```
## Factor Analysis using method =  wls
## Call: fa(r = cor(Z), nfactors = k, n.obs = nrow(Z), rotate = "promax", 
##     scores = T, fm = "wls")
## Standardized loadings (pattern matrix) based upon correlation matrix
##      WLS1  WLS2  WLS3   h2    u2 com
## x1   0.92 -0.24  0.04 0.82 0.182 1.1
## x2   0.75  0.20  0.23 0.75 0.248 1.3
## x3   0.86 -0.22  0.19 0.73 0.269 1.2
## x4   0.76  0.05 -0.27 0.64 0.357 1.3
## x5   0.89  0.10  0.11 0.87 0.129 1.1
## x6   0.94  0.09 -0.04 0.92 0.081 1.0
## x7   0.93 -0.03 -0.13 0.86 0.145 1.0
## x8   0.92  0.12 -0.11 0.90 0.103 1.1
## x9  -0.02  0.04  0.81 0.68 0.319 1.0
## x10 -0.05  0.82  0.08 0.70 0.296 1.0
## x11 -0.02  0.84  0.14 0.79 0.213 1.1
## x12  0.02  0.75 -0.14 0.53 0.474 1.1
## 
##                       WLS1 WLS2 WLS3
## SS loadings           6.10 2.14 0.94
## Proportion Var        0.51 0.18 0.08
## Cumulative Var        0.51 0.69 0.77
## Proportion Explained  0.66 0.23 0.10
## Cumulative Proportion 0.66 0.90 1.00
## 
##  With factor correlations of 
##      WLS1 WLS2 WLS3
## WLS1 1.00 0.19 0.04
## WLS2 0.19 1.00 0.31
## WLS3 0.04 0.31 1.00
## 
## Mean item complexity =  1.1
## Test of the hypothesis that 3 factors are sufficient.
## 
## The degrees of freedom for the null model are  66  and the objective function was  23.27 with 
Chi Square of  678.8
## The degrees of freedom for the model are 33  and the objective function was  8.89 
## 
## The root mean square of the residuals (RMSR) is  0.06 
## The df corrected root mean square of the residuals is  0.08 
## 
## The harmonic number of observations is  35 with the empirical chi square  16.56  with prob < 
0.99 
## The total number of observations was  35  with Likelihood Chi Square =  241.38  with prob < 
1.6e-33 
## 
## Tucker Lewis Index of factoring reliability =  0.264
## RMSEA index =  0.49  and the 90 % confidence intervals are  0.381 NA
## BIC =  124.06
## Fit based upon off diagonal values = 0.99
## Measures of factor score adequacy             
##                                                   WLS1 WLS2 WLS3
## Correlation of (regression) scores with factors   0.98 0.91  0.9
## Multiple R square of scores with factors          0.95 0.83  0.8
## Minimum correlation of possible factor scores     0.90 0.66  0.6
```

此时的共同度和特殊度与旋转前相比都发生了变化。由于新的因子不再相互独立，输出结果中不仅包括Pattern Matrix，还有Correlation Matrix，并且因子载荷系数不再等于公共因子与原始变量的相关系数。下面可以计算得到Structure Matrix:

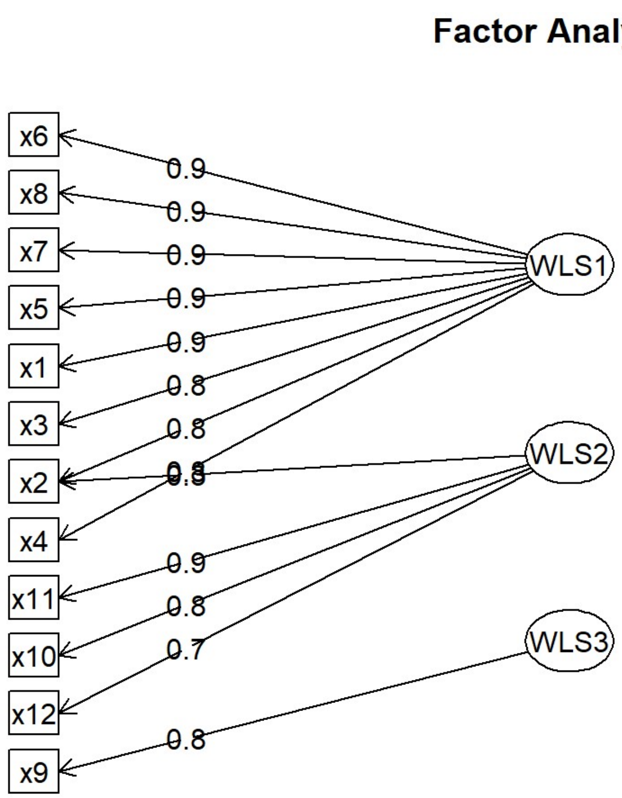
```
Struc=Z.wls.pro$loadings[1:p,1:k]%*%Z.wls.pro$score.cor #计算Structure
colnames(Struc)=c("WLS1","WLS2","WLS3")
Struc
```

```
##            WLS1        WLS2         WLS3
## x1  0.881961475 -0.08079029 -0.002579948
## x2  0.786119318  0.38598198  0.300851081
## x3  0.825103979 -0.02736444  0.146312144
## x4  0.760553523  0.09293125 -0.243020494
## x5  0.913132986  0.27289614  0.162440331
## x6  0.952100946  0.23392611  0.009004842
## x7  0.917304913  0.08363243 -0.112491810
## x8  0.935013187  0.24167991 -0.054728532
## x9  0.004927012  0.26490140  0.823268324
## x10 0.086681790  0.83262824  0.312582482
## x11 0.117622130  0.87238916  0.378177209
## x12 0.141510395  0.71688667  0.075143086
```

将输出的结果与正交对比的结果进行对比，发现变化并不是很大。因此采用正交旋转已经足够。

Step 4: 因子分析图
最后，画出正交旋转的因子分析图。

```
fa.diagram(Z.wls.var,simple=FALSE)
```



3.2 最大似然法

与3.1节步骤完全相似，进行如下分析。

Step 1: 求解初始因子

```
Z.ml<-fa(cor(Z),nfactors = 3,n.obs = nrow(Z),rotate = "none",scores = T,fm="ml")
Z.ml
```

```
## Factor Analysis using method =  ml
## Call: fa(r = cor(Z), nfactors = 3, n.obs = nrow(Z), rotate = "none",
##     scores = T, fm = "ml")
## Standardized loadings (pattern matrix) based upon correlation matrix
##      ML1   ML2   ML3   h2     u2 com
## x1  0.82 -0.12  0.39 0.84 0.1569 1.5
## x2  0.76  0.49  0.34 0.94 0.0565 2.1
## x3  0.73  0.09  0.56 0.86 0.1382 1.9
## x4  0.88 -0.32 -0.30 0.98 0.0215 1.5
## x5  0.92  0.23  0.23 0.95 0.0530 1.3
## x6  0.99  0.06  0.10 0.98 0.0153 1.0
## x7  0.95 -0.17  0.02 0.93 0.0713 1.1
## x8  1.00 -0.01 -0.03 1.00 0.0046 1.0
## x9  0.00  0.44  0.05 0.20 0.8027 1.0
## x10 0.18  0.77 -0.50 0.88 0.1181 1.8
## x11 0.20  0.85 -0.42 0.94 0.0649 1.6
## x12 0.23  0.53 -0.30 0.43 0.5744 2.0
##
##                        ML1  ML2  ML3
## SS loadings           6.42 2.25 1.25
## Proportion Var        0.53 0.19 0.10
## Cumulative Var        0.53 0.72 0.83
## Proportion Explained  0.65 0.23 0.13
## Cumulative Proportion 0.65 0.87 1.00
##
## Mean item complexity =  1.5
## Test of the hypothesis that 3 factors are sufficient.
##
## The degrees of freedom for the null model are  66  and the objective function was  23.27 with
Chi Square of  678.8
## The degrees of freedom for the model are 33  and the objective function was  4
##
## The root mean square of the residuals (RMSR) is  0.03
## The df corrected root mean square of the residuals is  0.04
##
## The harmonic number of observations is  35 with the empirical chi square  3.7  with prob <  1
## The total number of observations was  35  with Likelihood Chi Square =  108.77  with prob <
5.1e-10
##
## Tucker Lewis Index of factoring reliability =  0.732
## RMSEA index =  0.303  and the 90 % confidence intervals are  0.207 0.315
## BIC =  -8.56
## Fit based upon off diagonal values = 1
## Measures of factor score adequacy
##                                                    ML1  ML2  ML3
## Correlation of (regression) scores with factors   1.00 0.98 0.97
## Multiple R square of scores with factors          1.00 0.97 0.94
## Minimum correlation of possible factor scores     0.99 0.93 0.88
```

观察载荷矩阵，各个初始因子意义并不明确，因此需要进行旋转。

Step2: 正交旋转
首先尝试正交旋转。

```
#进行旋转(Varimax)
Z.ml.var<-fa(cor(Z),nfactors = 3,n.obs = nrow(Z),rotate = "varimax",scores = T,fm="ml")
Z.ml.var
```

```
## Factor Analysis using method =  ml
## Call: fa(r = cor(Z), nfactors = 3, n.obs = nrow(Z), rotate = "varimax",
##     scores = T, fm = "ml")
## Standardized loadings (pattern matrix) based upon correlation matrix
##       ML1   ML2   ML3   h2     u2 com
## x1   0.85 -0.15  0.32 0.84 0.1569 1.3
## x2   0.88  0.40  0.04 0.94 0.0565 1.4
## x3   0.92 -0.06  0.06 0.86 0.1382 1.0
## x4   0.43  0.00  0.89 0.98 0.0215 1.4
## x5   0.88  0.25  0.32 0.95 0.0530 1.4
## x6   0.83  0.17  0.52 0.98 0.0153 1.8
## x7   0.71  0.00  0.66 0.93 0.0713 2.0
## x8   0.74  0.16  0.64 1.00 0.0046 2.1
## x9   0.11  0.36 -0.24 0.20 0.8027 1.9
## x10 -0.03  0.93  0.08 0.88 0.1181 1.0
## x11  0.05  0.97  0.01 0.94 0.0649 1.0
## x12  0.09  0.64  0.10 0.43 0.5744 1.1
##
##                        ML1  ML2  ML3
## SS loadings           5.09 2.65 2.19
## Proportion Var        0.42 0.22 0.18
## Cumulative Var        0.42 0.64 0.83
## Proportion Explained  0.51 0.27 0.22
## Cumulative Proportion 0.51 0.78 1.00
##
## Mean item complexity =  1.5
## Test of the hypothesis that 3 factors are sufficient.
##
## The degrees of freedom for the null model are  66  and the objective function was  23.27 with
Chi Square of  678.8
## The degrees of freedom for the model are 33  and the objective function was  4
##
## The root mean square of the residuals (RMSR) is  0.03
## The df corrected root mean square of the residuals is  0.04
##
## The harmonic number of observations is  35 with the empirical chi square  3.7  with prob <  1
## The total number of observations was  35  with Likelihood Chi Square =  108.77  with prob <
5.1e-10
##
## Tucker Lewis Index of factoring reliability =  0.732
## RMSEA index =  0.303  and the 90 % confidence intervals are  0.207 0.315
## BIC =  -8.56
## Fit based upon off diagonal values = 1
## Measures of factor score adequacy
##                                                    ML1  ML2  ML3
## Correlation of (regression) scores with factors   0.99 0.98 0.98
## Multiple R square of scores with factors          0.98 0.96 0.97
## Minimum correlation of possible factor scores     0.95 0.92 0.94
```

观察载荷矩阵，因子1与变量$X_1, X_2, ..., X_8$关系密切，而因子2与变量X_{10}, X_{11}, X_{12}关系密切，因子3与变量X_9关系密切。这与加权最小二乘的结论是一致的。
此外，还可以输出正交旋转的因子得分系数矩阵。

```
Z.ml.var$weights
```

```
##             ML1         ML2          ML3
## x1   0.095196368 -0.08602328 -0.070426609
## x2   0.297528162  0.11198022 -0.333002357
## x3   0.160449484 -0.08477624 -0.150779702
## x4  -0.505455626 -0.06014456  0.828431557
## x5   0.218523149  0.02799039 -0.196394682
## x6   0.400703933 -0.02811916 -0.187966293
## x7   0.022782632 -0.07061099  0.051973970
## x8   0.220491248  0.15568697  0.609146419
## x9   0.005819349  0.01479261 -0.011053589
## x10 -0.107348168  0.31143705  0.057535297
## x11 -0.141052153  0.56760135  0.039083340
## x12 -0.011835742  0.04192288  0.005467763
```

Step3: 斜交旋转

最后，尝试斜交旋转，并计算Structure矩阵。

```
#进行旋转(Promax)
Z.ml.pro<-fa(cor(Z),nfactors = 3,n.obs = nrow(Z),rotate = "promax",scores = T,fm="ml")
Z.ml.pro
```

```
## Factor Analysis using method =  ml
## Call: fa(r = cor(Z), nfactors = 3, n.obs = nrow(Z), rotate = "promax",
##     scores = T, fm = "ml")
##
##  Warning: A Heywood case was detected.
## Standardized loadings (pattern matrix) based upon correlation matrix
##       ML1   ML2   ML3   h2     u2 com
## x1   0.85 -0.27  0.18 0.84 0.1569 1.3
## x2   0.92  0.24 -0.13 0.94 0.0565 1.2
## x3   1.03 -0.24 -0.13 0.86 0.1382 1.1
## x4   0.13  0.07  0.92 0.98 0.0215 1.1
## x5   0.83  0.13  0.19 0.95 0.0530 1.2
## x6   0.70  0.09  0.43 0.98 0.0153 1.7
## x7   0.53 -0.03  0.60 0.93 0.0713 2.0
## x8   0.56  0.13  0.58 1.00 0.0046 2.1
## x9   0.17  0.31 -0.28 0.20 0.8027 2.6
## x10 -0.20  0.99  0.13 0.88 0.1181 1.1
## x11 -0.09  0.99  0.02 0.94 0.0649 1.0
## x12 -0.03  0.66  0.11 0.43 0.5744 1.1
##
##                        ML1  ML2  ML3
## SS loadings           4.90 2.63 2.39
## Proportion Var        0.41 0.22 0.20
## Cumulative Var        0.41 0.63 0.83
## Proportion Explained  0.49 0.27 0.24
## Cumulative Proportion 0.49 0.76 1.00
##
##  With factor correlations of
##      ML1   ML2   ML3
## ML1 1.00  0.28  0.46
## ML2 0.28  1.00 -0.06
## ML3 0.46 -0.06  1.00
##
## Mean item complexity =  1.4
## Test of the hypothesis that 3 factors are sufficient.
##
## The degrees of freedom for the null model are  66  and the objective function was  23.27 with
Chi Square of  678.8
## The degrees of freedom for the model are 33  and the objective function was  4
##
## The root mean square of the residuals (RMSR) is  0.03
## The df corrected root mean square of the residuals is  0.04
##
## The harmonic number of observations is  35 with the empirical chi square  3.7  with prob <  1
## The total number of observations was  35  with Likelihood Chi Square =  108.77  with prob <
5.1e-10
##
## Tucker Lewis Index of factoring reliability =  0.732
## RMSEA index =  0.303  and the 90 % confidence intervals are  0.207 0.315
## BIC =  -8.56
## Fit based upon off diagonal values = 1
## Measures of factor score adequacy
##                                                   ML1  ML2  ML3
## Correlation of (regression) scores with factors   0.99 0.98 0.99
## Multiple R square of scores with factors          0.99 0.96 0.98
## Minimum correlation of possible factor scores     0.97 0.93 0.97
```

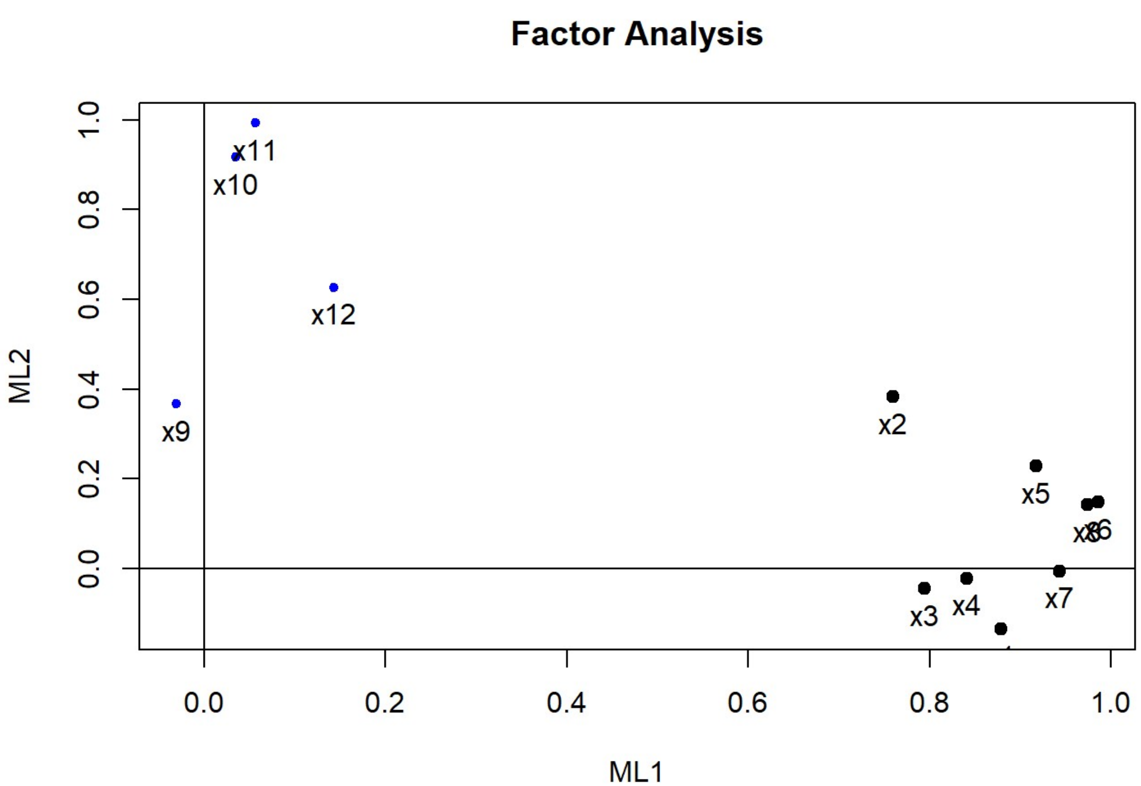
```
Struc2=Z.ml.pro$loadings[1:p,1:3]%*%Z.ml.pro$score.cor #计算Structure
colnames(Struc2)=c("ML1","ML2","ML3")
Struc2
```

```
##              ML1         ML2         ML3
## x1   0.946930518 -0.11156791  0.86893745
## x2   0.859189234  0.40365976  0.63120891
## x3   0.875451914 -0.05553049  0.70496395
## x4   0.899621860  0.11933495  1.03019037
## x5   1.009047099  0.28927779  0.87111132
## x6   1.065401801  0.23243502  1.00091083
## x7   1.018304737  0.08538561  1.03399689
## x8   1.057871645  0.24664588  1.04159285
## x9  -0.005946973  0.33016549 -0.13550832
## x10  0.081221415  0.95459379 -0.01653802
## x11  0.115298227  0.97493460 -0.02217935
## x12  0.182382203  0.65701369  0.10174080
```

同样地，斜交旋转结果与正交旋转结果无太大差别。因此，只考虑正交旋转的结果即可。

Step4: 因子分析图
最后，输出正交旋转的因子分析图。

```
Z.ml.2 <- fa(cor(Z),nfactors = 2,n.obs = nrow(Z),rotate = "varimax",scores = T,fm="ml")
factor.plot(Z.ml.2,labels = rownames(Z.ml.2$loadings))
```



3.3 主成分法

最后，回到主成分法，对例6-3进行因子分析。

```
Z.fa.pca<-Fa.pca(Z,cor=T,1)
Z.fa.pca[c(8,9)]
```

```
## $Varimax
## Principal Components Analysis
## Call: principal(r = (X), nfactors = k, rotate = "varimax", scores = T)
## Standardized loadings (pattern matrix) based upon correlation matrix
##      RC1   RC2   RC3   h2    u2  com
## x1  0.93 -0.18  0.04 0.90 0.102 1.1
## x2  0.81  0.31  0.34 0.86 0.136 1.7
## x3  0.87 -0.15  0.25 0.84 0.157 1.2
## x4  0.79  0.09 -0.44 0.83 0.175 1.6
## x5  0.93  0.20  0.15 0.93 0.066 1.1
## x6  0.97  0.17 -0.05 0.97 0.026 1.1
## x7  0.95  0.03 -0.19 0.93 0.066 1.1
## x8  0.95  0.20 -0.15 0.97 0.029 1.1
## x9  0.01  0.21  0.84 0.75 0.252 1.1
## x10 0.03  0.91  0.17 0.87 0.133 1.1
## x11 0.07  0.92  0.26 0.92 0.079 1.2
## x12 0.09  0.81 -0.11 0.67 0.325 1.1
##
##                        RC1  RC2  RC3
## SS loadings           6.52 2.65 1.28
## Proportion Var        0.54 0.22 0.11
## Cumulative Var        0.54 0.76 0.87
## Proportion Explained  0.62 0.25 0.12
## Cumulative Proportion 0.62 0.88 1.00
##
## Mean item complexity =  1.2
## Test of the hypothesis that 3 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.06
##  with the empirical chi square  14.6  with prob <  1
##
## Fit based upon off diagonal values = 0.99
## $`weights of Varimax`
##              RC1          RC2           RC3
## x1   0.157913225 -0.129953420  0.0838529262
## x2   0.120456400  0.034162463  0.2588045367
## x3   0.151483208 -0.146640327  0.2575935375
## x4   0.111636763  0.070415369 -0.3671722949
## x5   0.142357876  0.010356540  0.1199356077
## x6   0.144440677  0.033942965 -0.0516697192
## x7   0.144973701 -0.003321596 -0.1453869212
## x8   0.138162446  0.061920660 -0.1421518710
## x9   0.009972953 -0.051067525  0.6775027754
## x10 -0.036566542  0.355514913 -0.0007471679
## x11 -0.029655590  0.343235811  0.0700690673
## x12 -0.028310340  0.354392539 -0.2201717541
```

这里得到的结果，与加权最小二乘法、最大似然法得到的结果并无太大差别。事实上，观察它们的方差贡献率，有如下对比：

方法	方差贡献率
加权最小二乘法	77%
最大似然法	83%
主成分法	87%

因此，在计算因子得分时，不妨采用主成分法的结果进行分析。

3.4 因子得分

以各因子的方差贡献率占三个因子总方差贡献率的比重作为权重进行加权汇总，即

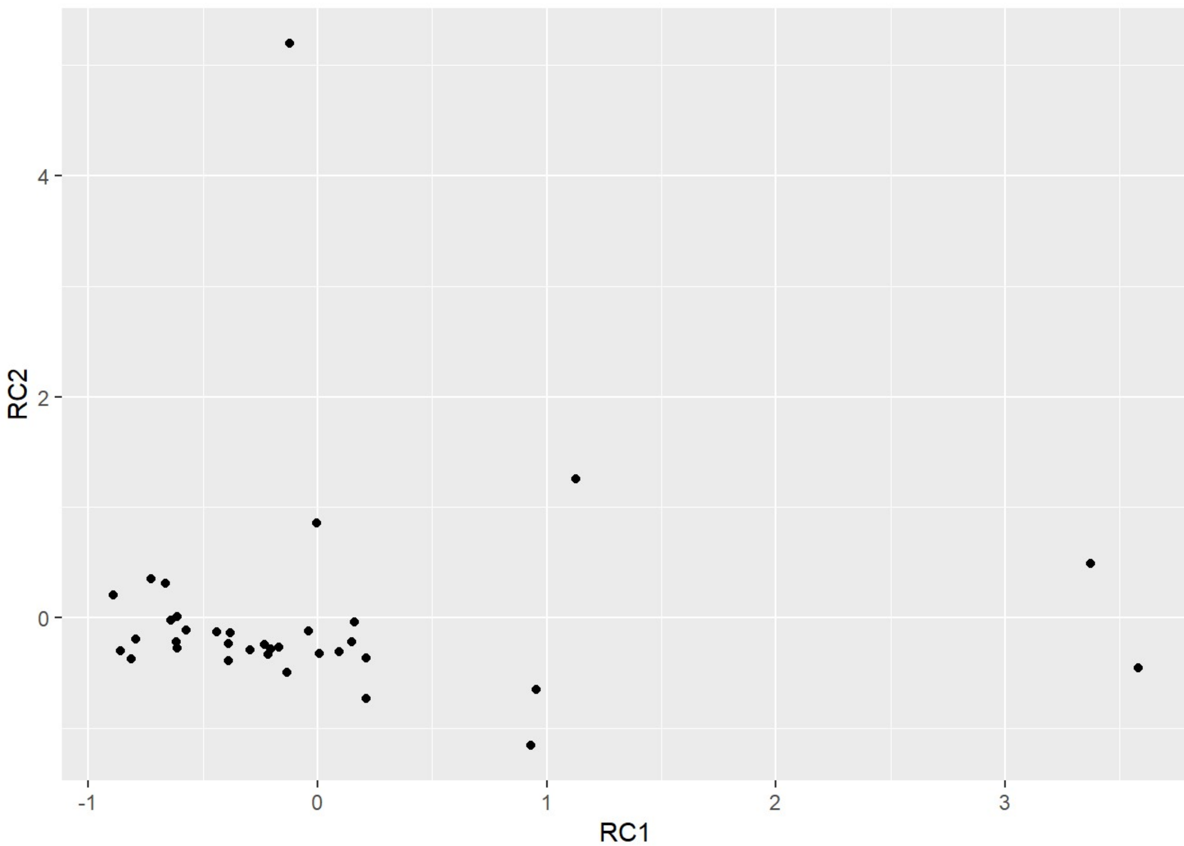
$$F=(54.381F_1+22.077F_2+10.647F_3)/87.105$$

```
F=(54.381*Z.fa.pca$scores[,1]+22.077*Z.fa.pca$scores[,2]+10.647*Z.fa.pca$scores[,3])/87.105
Z.Score=data.frame(Z.fa.pca$scores,F)
Z.Score
```

```
##                RC1         RC2          RC3           F
## 北京      3.374224077  0.49201267 -3.039872368  1.85971325
## 天津      0.955571137 -0.65270781  0.963708422  0.54894308
## 石家庄   -0.216011233 -0.33077426  0.349207748 -0.17601051
## 太原     -0.388031038 -0.38682330 -0.250711293 -0.37094010
## 呼和浩特 -0.792149243 -0.19428297  0.258885051 -0.51214860
## 沈阳      0.008121260 -0.32831530 -0.679844040 -0.16124073
## 长春     -0.231292396 -0.24039703 -0.650337357 -0.28482060
## 哈尔滨    0.151495038 -0.22203747 -1.587062242 -0.15568476
## 上海      3.581540637 -0.45341683  2.453030429  2.42092983
## 南京     -0.005001935  0.85426198 -0.615963232  0.13810196
## 杭州      0.094242861 -0.30554345  0.358685374  0.02523921
## 合肥     -0.725120746  0.35031172 -0.009445654 -0.36507121
## 福州     -0.379114380 -0.13561778  0.326406035 -0.23116248
## 南昌     -0.617184141 -0.21808281 -0.112994402 -0.45440281
## 济南     -0.203395675 -0.28409657  0.883348702 -0.09101483
## 郑州     -0.295071349 -0.28905820  0.605261174 -0.18349805
## 武汉      0.214089872 -0.72831774  0.005271687 -0.05029013
## 长沙     -0.439143602 -0.13128335  0.297879339 -0.27102795
## 广州      1.127864207  1.25646175 -0.577347066  0.95202658
## 南宁     -0.638346870 -0.02502599  0.172714631 -0.38376152
## 海口     -0.812763907 -0.37670688  0.968967701 -0.48445982
## 成都      0.212909098 -0.36447826  0.254519259  0.07165478
## 贵阳     -0.664996784  0.31153815 -0.961635217 -0.45374999
## 昆明     -0.389711388 -0.23140477 -0.776960177 -0.39692226
## 西安     -0.132456887 -0.49580505 -0.922211246 -0.32108156
## 兰州     -0.611627891 -0.27601710 -0.760147658 -0.54472026
## 西宁     -0.859639062 -0.29978654 -0.097469017 -0.62458150
## 银川     -0.892742301  0.20630723 -0.914843044 -0.61688661
## 乌鲁木齐 -0.571560961 -0.11466620  0.736045040 -0.29592872
## 大连     -0.040115352 -0.12417417 -0.883308110 -0.16448525
## 宁波     -0.170301831 -0.27117889  0.345100684 -0.13287083
## 厦门     -0.613379423  0.01009493  0.997995294 -0.25839693
## 青岛      0.160343654 -0.03734866 -0.152727250  0.07197078
## 深圳     -0.121402941  5.19481903  1.266659845  1.39567343
## 重庆      0.930159494 -1.15846008  1.749192959  0.50090509
```

最后，以F_1因子得分为x轴，以F_2因子得分为y轴，画出各城市的因子得分图。

```
ggplot(Z,aes(x=Z.Score[,1], y = Z.Score[,2]))+
  geom_point()+
  labs(x = "RC1", y = "RC2")
```



可见，大部分城市密集地集中于两个分值均为负的第三象限，而北京、天津、上海、南京、杭州、广州、成都、青岛、深圳、重庆则拥有较好的发展水平。