**因子分析**

因子分析，相较于主成分分析而言，通过对因子的旋转处理，使得我们可以更直观的认识到数据内部之间的关系，其目的即用有限个不可观测的因变量来解释原始变量间的相关关系。即用几个少数的综合因子来取代错综复杂关系的变量。

**R中的因子分析函数factanal()，这个函数是基于极大似然方法求解**，其使用方法为：

factanal(X,factors,scores=”none”,rotation=”varimax”)

其中，X为数据（矩阵或者数据框）。factors为因子个数。scores为因子得分的计算方法，可选择”regression”或”Bartlett”等。 rotation为因子旋转方法，常用的正交旋转方法，如:方差最大旋转varimax；另外，还可以指定一些斜交旋转方法，如promax。

**自编因子分析函数factpc()，在mvstats包中，mvstats包的运行环境通常要求为R的3.6.1版本。这个函数是基于主成分方法来求解的**，其使用方法为：

factpc(X,m,scores=”none”,rotation=”varimax”)

其中，X为数据（矩阵或者数据框）。m为因子个数。scores为因子得分的计算方法。 rotation为因子旋转方法，这里常用的正交旋转为方差最大旋转varimax。

**注意：**

**（1）极大似然法要求数据来自多元正态分布，这一点有时是很难满足的。而主成分法没有对正态总体的要求。**

**（2）当原始数据的样本量不太大时，参数估计精度可能受到较大影响。比如，R的factanal函数用极大似然法时，可能会得不到估计结果（即：可能发生参数的可识别性差，得不出因子载荷等的参数估计结果）。这时，建议尝试主成分法确定公因子个数、估计初步的因子载荷矩阵，再考量因子是否旋转及如何旋转。**

例1：对于数据d9.1 水泥行业运营因素做因子分析。

> X=read.table("clipboard",header=T)#读取数据存入X

> cor(X) #计算数据X的相关系数矩阵

输入：

x1 x2 x3 x4 x5

x1 1.00000000 0.9991983 -0.09974689 0.18850763 0.2010041

x2 0.99919830 1.0000000 -0.10420434 0.19672979 0.1903570

x3 -0.09974689 -0.1042043 1.00000000 -0.83715637 -0.4087603

x4 0.18850763 0.1967298 -0.83715637 1.00000000 0.2585103

x5 0.20100410 0.1903570 -0.40876032 0.25851029 1.0000000

x6 0.29778271 0.2874781 0.01518741 -0.02928244 0.5802933

x6

x1 0.29778271

x2 0.28747808

x3 0.01518741

x4 -0.02928244

x5 0.58029333

x6 1.00000000

**#使用极大似然法进行因子分析**

> FA0=factanal(X,3,rotation="none")#X为数据，3为选择的因子个数，rotation="none"为不旋转

> FA0

Call:

factanal(x = X, factors = 3, rotation = "none")

Uniquenesses:

x1 x2 x3 x4 x5 x6

0.005 0.005 0.005 0.271 0.005 0.548

Loadings: #因子载荷矩阵

Factor1 Factor2 Factor3

x1 0.950 -0.307

x2 0.948 -0.310

x3 -0.340 -0.782 0.517

x4 0.363 0.561 -0.531

x5 0.454 0.693 0.556

x6 0.383 0.163 0.527

该分析结果对应的因子模型表达式：

Factor1 Factor2 Factor3

SS loadings 2.402 1.623 1.140

Proportion Var 0.400 0.271 0.190 #方差贡献率

Cumulative Var 0.400 0.671 0.861 #累计方差贡献率

The degrees of freedom for the model is 0 and the fit was 1.1422

**#主成分法进行因子分析**

> library(mvstats) #载入程序包mvstats

> FA1=factpc(X,3) #X为数据，3为选择的成分个数

$Vars

#方差 方差贡献率 累计方差贡献率

Vars Vars.Prop Vars.Cum

Factor1 2.570 0.4283 42.83

Factor2 1.713 0.2855 71.38

Factor3 1.249 0.2082 92.19

$loadings #因子载荷矩阵

Factor1 Factor2 Factor3

x1 0.7829 0.5029 -0.3624

x2 0.7811 0.4964 -0.3756

x3 -0.5786 0.7685 0.0802

x4 0.5951 -0.6990 -0.2415

x5 0.6317 -0.1457 0.6557

x6 0.5084 0.3367 0.6943

$scores #因子得分

Factor1 Factor2 Factor3

冀东水泥 1.10805 0.19287 -0.40233

大同水泥 -1.07195 1.46385 -0.37413

四川双马 -0.58577 -0.49848 0.24193

牡丹江 -1.17442 -0.77791 0.08986

西水股份 -0.05264 -0.46073 2.31615

狮头股份 -1.05007 2.04151 0.25174

太行股份 0.20807 0.48809 -0.23430

海螺水泥 2.20745 0.32524 1.16336

尖峰集团 -1.11541 -1.53235 0.39013

四川金顶 0.09714 -0.60602 -1.45691

祁连山 0.66096 1.03293 0.04173

华新水泥 0.41359 -1.08331 0.19805

福建水泥 0.86840 -0.53255 -1.82104

天鹅股份 -0.51340 -0.05315 -0.40422

**注：关于王斌会教材介绍的根据因子得分进行每一个样本的综合得分、再排名的方法，虽然在mvstats包可以实现，但是：综合得分的含义不明，并且对此方法学界一直有争议，因此不推荐使用。**

**（虽然有些特殊情况下综合得分的含义勉强可用，但因争议故不推荐。详见王学民、何晓群等编著的多元统计的相关书籍。）**

**以下仅仅列出代码，用----包括的部分，仅供参考。**

-----------------------------------------------------------------

$Rank #得分排名

F Ri

冀东水泥 0.48359 3

大同水泥 -0.12910 8

四川双马 -0.37184 11

牡丹江 -0.76615 13

西水股份 0.35587 4

狮头股份 0.20127 5

太行股份 0.19490 6

海螺水泥 1.38882 1

尖峰集团 -0.90457 14

四川金顶 -0.47152 12

祁连山 0.63632 2

华新水泥 -0.09863 7

福建水泥 -0.17273 9

天鹅股份 -0.34622 10

---------------------------------------------------------------

$common #共同度

x1 x2 x3 x4 x5 x6

0.9971 0.9976 0.9318 0.9011 0.8502 0.8539

从结果看出，用极大似然法解释的方差为86%，基本可以全面反映六项财务指标的信息。用主成分法解释的方差为92%，效果要更好。

**注意：若排名中出现小数：例如10.5，则说明有两个样本排名并列第10，且下一名的排名从第12开始。**

因为三个因子的经济含义不明显，需要进行因子旋转。

**#极大似然法**

Call:

> FA0=factanal(X,3,rotation="varimax")#X为数据，3为选择的成分个数，rotation="varimax"为按方差最大法旋转

> FA0

factanal(x = X, factors = 3, rotation = "varimax")

Uniquenesses:

x1 x2 x3 x4 x5 x6

0.005 0.005 0.005 0.271 0.005 0.548

Loadings: #因子载荷矩阵

Factor1 Factor2 Factor3

x1 0.983 0.155

x2 0.985 0.142

x3 -0.990 -0.124

x4 0.127 0.844

x5 0.293 0.953

x6 0.210 0.631

该分析结果对应的因子模型表达式：

Factor1 Factor2 Factor3

SS loadings 1.998 1.800 1.367

Proportion Var 0.333 0.300 0.228#方差贡献率

Cumulative Var 0.333 0.633 0.861#累计方差贡献率

The degrees of freedom for the model is 0 and the fit was 1.1422

#主成分方法

> library(mvstats)

> FA1=factpc(X,3,rotation="varimax")#x为数据，3为选择的成分个数，rotation="varimax"为按方差最大法旋转

> FA1

Factor Analysis for Princomp in Varimax:

$Vars #方差 方差贡献率 累计方差贡献率

Vars Vars.Prop Vars.Cum

Factor1 2.014 33.56 33.56

Factor2 1.938 32.30 65.87

Factor3 1.580 26.33 92.19

$loadings #旋转后载荷矩阵

Factor1 Factor2 Factor3

x1 0.986709 0.07216 0.135305

x2 0.988140 0.07913 0.122314

x3 -0.009491 -0.95685 -0.127000

x4 0.135286 0.93954 0.004538

x5 0.044103 0.32942 0.860082

x6 0.208451 -0.14120 0.889083

该分析结果对应的因子模型表达式：

$scores #旋转后因子得分

Factor1 Factor2 Factor3

冀东水泥 1.0571 0.508465 0.22544

大同水泥 0.2509 -1.704706 -0.68039

四川双马 -0.7922 0.052388 -0.14079

牡丹江 -1.2794 -0.001121 -0.59625

西水股份 -1.3825 -0.096118 1.91289

狮头股份 0.2910 -2.290232 -0.06280

太行股份 0.5235 -0.246292 -0.04099

海螺水泥 1.1476 0.681631 2.13317

尖峰集团 -1.7982 0.594084 -0.39758

四川金顶 0.4175 0.832941 -1.27718

祁连山 1.0061 -0.507764 0.48519

华新水泥 -0.4092 1.074736 0.24757

福建水泥 1.1592 1.253210 -1.19980

天鹅股份 -0.1915 -0.151222 -0.60849

**注：关于王斌会教材介绍的根据因子得分进行每一个样本的综合得分、再排名的方法，虽然在mvstats包可以实现，但是：综合得分的含义不明，并且对此方法学界一直有争议，因此不推荐使用。**

**（虽然有些特殊情况下综合得分的含义勉强可用，但因争议故不推荐。详见王学民、何晓群等编著的多元统计的相关书籍。）**

**以下仅仅列出代码，用----包括的部分，仅供参考。**

-----------------------------------------------------------------

$Rank#得分排名

F Ri

冀东水泥 0.627381 2

大同水泥 -0.700254 13

四川双马 -0.310247 10

牡丹江 -0.636439 12

西水股份 0.009279 8

狮头股份 -0.714476 14

太行股份 0.092589 6

海螺水泥 1.265773 1

尖峰集团 -0.560000 11

四川金顶 0.079124 7

祁连山 0.326924 4

华新水泥 0.298287 5

福建水泥 0.518518 3

天鹅股份 -0.296458 9

-----------------------------------------------------------------

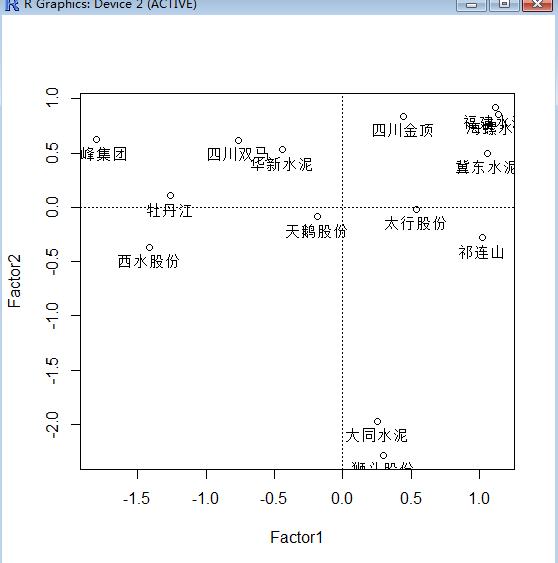
$common#共同度

x1 x2 x3 x4 x5 x6

0.9971 0.9976 0.9318 0.9011 0.8502 0.8539

>library(mvstats)

>factanal.rank(FA1,plot=T)



例2：31个省、市、自治区的消费情况。用**d7.2数据**应用主成分法的因子分析。

> X<-read.table("clipboard",header=T)

> library(mvstats)

> FA0=factpc(X,3) #主成分法因子分析 未旋转

> FA0$Vars #方差贡献率

Vars Vars.Prop Vars.Cum

Factor1 5.1925 0.64907 64.91

Factor2 1.2606 0.15757 80.66

Factor3 0.6471 0.08088 88.75

> FA0$loadings #载荷矩阵

Factor1 Factor2 Factor3

X1 0.9114 -0.07120 0.2423

X2 0.3207 0.84400 0.2879

X3 0.8274 -0.01090 -0.3958

X4 0.7803 0.29371 -0.4307

X5 0.9138 -0.15211 0.3033

X6 0.9337 0.03035 -0.1695

X7 0.6569 -0.64671 0.1130

X8 0.9096 0.12046 0.1729

由于公共因子在原始变量上的载荷值不太好解释，故对其进行因子旋转，选用方差最大化正交旋转。

> FA1=factpc(X,3,rotation="varimax")#主成分法因子分析 旋转

> FA1$Vars #旋转后方差贡献率

Vars Vars.Prop Vars.Cum

Factor1 3.229 40.37 40.37

Factor2 2.596 32.45 72.82

Factor3 1.275 15.94 88.75

> FA1$loadings #旋转后载荷矩阵

Factor1 Factor2 Factor3

X1 0.83728 0.4015 0.17946

X2 0.08334 0.1625 0.92987

X3 0.39643 0.8267 -0.02714

X4 0.22838 0.8813 0.22782

X5 0.90324 0.3425 0.13011

X6 0.58608 0.7379 0.11553

X7 0.79127 0.2182 -0.43453

X8 0.72564 0.4887 0.32604

由旋转后的因子载荷矩阵可以看到：

公共因子F1在X1(人均食品支出)、X5（人均交通和通讯支出）、x7（人均居住支出）、x8（人均杂项商品及服务支出）上的载荷值都很大，可视为反映日常必须消费的公共因子。

公共因子F2在X3（人均家庭设备用品及服务支出）、x4（人均医疗保健支出）、x6（人均娱乐教育文化支出）上的载荷值很大，可视为反映相对高档消费的公共因子。

公共因子F3仅在x2（人均衣着支出）上有很大的载荷，可视为衣着因子。

这样就可以对各省、市、自治区的消费情况做评价。

> FA1$scores #因子得分

Factor1 Factor2 Factor3

北京 0.60209 2.93798 1.639329

天津 0.53368 1.39166 -0.946970

河北 -1.00553 0.42878 -0.215804

山西 -0.98315 0.02363 0.005883

内蒙古 -0.40922 -0.75555 0.551076

辽宁 -0.67890 -0.17396 0.484714

吉林 -0.53301 -0.51962 0.137665

黑龙江 -0.85858 -0.22739 0.001374

上海 2.23481 1.27047 0.489859

江苏 -0.11532 0.26186 -0.074807

浙江 0.62223 2.17914 0.168660

安徽 -0.26717 -0.86766 -0.297037

福建 0.94223 -0.67815 -0.337264

江西 -0.12598 -1.25376 -1.226289

山东 -0.52826 0.33116 0.795547

河南 -0.45310 -0.51662 -0.852182

湖北 -0.06496 -0.67388 -0.087642

湖南 0.27740 0.04226 -0.323656

广东 3.52132 -0.59559 -1.409939

广西 0.49514 -0.31264 -1.620601

海南 -0.05670 -0.69585 -1.613855

重庆 0.02695 0.26656 -0.014563

四川 -0.16409 -0.02868 -0.460109

贵州 -0.64279 -0.52856 -0.085301

云南 0.24521 -0.42921 0.308552

西藏 1.41158 -2.38342 3.597715

陕西 -0.96542 0.78188 -0.761364

甘肃 -1.03248 0.24684 0.295909

青海 -0.64835 0.12991 0.646513

宁夏 -1.06282 0.60619 0.411427

新疆 -0.31678 -0.25777 0.793161

**注：关于王斌会教材介绍的根据因子得分进行每一个样本的综合得分、再排名的方法，虽然在mvstats包可以实现，但是：综合得分的含义不明，并且对此方法学界一直有争议，因此不推荐使用。**

**（虽然有些特殊情况下综合得分的含义勉强可用，但因争议故不推荐。详见王学民、何晓群等编著的多元统计的相关书籍。）**

**以下仅仅列出代码，用----包括的部分，仅供参考。**

---------------------------------------------------------------

王斌会教材认为的综合得分公式：；其中，0.40366为F1的方差贡献率，0.32449为F2的方差贡献率，0.15937为F3的方差贡献率，0.8875为前三个因子的方差累积贡献率。

> FA1$Rank

F Ri

北京 1.642392 1

天津 0.581519 5

河北 -0.339312 22

山西 -0.437458 25

内蒙古 -0.363417 23

辽宁 -0.285347 18

吉林 -0.407689 24

黑龙江 -0.473391 26

上海 1.568902 2

江苏 0.029861 10

浙江 1.110031 4

安徽 -0.492088 27

福建 0.120034 7

江西 -0.735896 31

山东 0.023665 11

河南 -0.547985 29

湖北 -0.291669 20

湖南 0.083501 9

广东 1.130630 3

广西 -0.180103 16

海南 -0.569989 30

重庆 0.107103 8

四川 -0.167735 15

贵州 -0.500925 28

云南 0.009999 12

西藏 0.416570 6

陕西 -0.289922 19

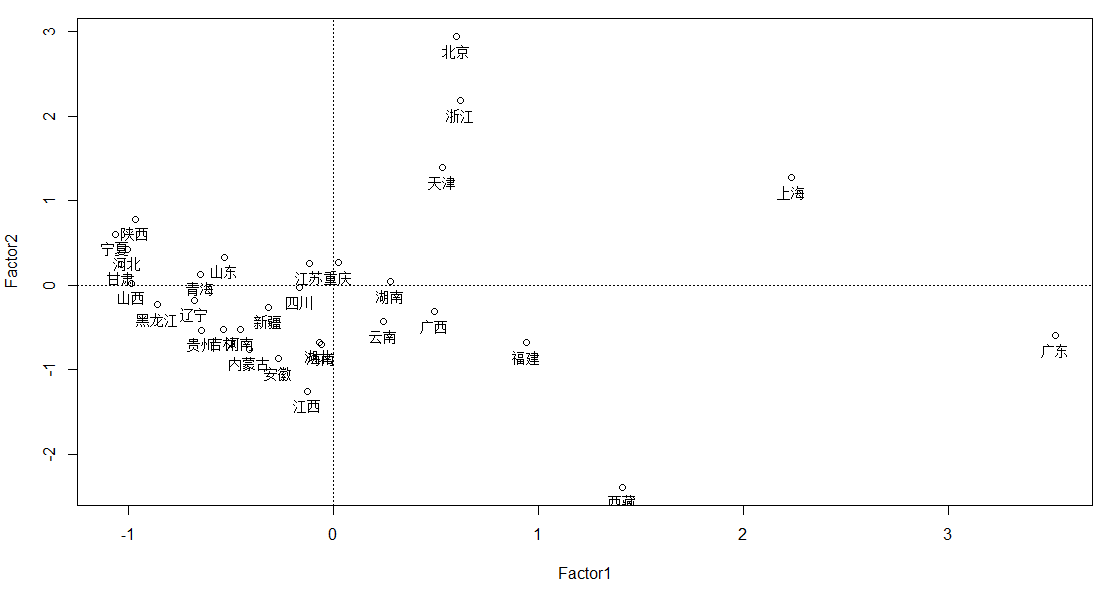
甘肃 -0.326206 21

青海 -0.131298 14

宁夏 -0.187872 17

新疆 -0.095905 13

----------------------------------------------------------------



> plot.text(FA1$scores)

【本章作业】：

1. 教材P271-2（4）：数据集名称为E9.4,详见excel数据文件.

具体要求：使用**主成分法**，分别进行无旋转和varimax正交旋转的因子分析，要求选择累积方差贡献率大于80%的最少个数的因子进行分析，

并根据两种结果选择一种稀疏性较好的结果：计算得出因子得分，作出双信息图，给出每个变量的共同度和各个因子的方差贡献率。

1. 用d7.2数据(31个省、市、自治区的消费情况):

具体要求：应用**极大似然方法，**要求选择累积方差贡献率大于80%的最少个数的因子进行分析，考虑varimax旋转：给出每个变量的共同度和各个因子的方差贡献率，并给出回归法的因子得分，作出双信息图。写出因子模型的表达式。

附：d7.2数据，详见excel数据文件：

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 |
| 北京 | 3229.28 | 821.72 | 847.39 | 677.66 | 768.34 | 1429.15 | 587.98 | 561.19 |
| 天津 | 2588.1 | 531.97 | 806.36 | 435.38 | 585.94 | 897.02 | 808.05 | 334.4 |
| 河北 | 1583.68 | 530.02 | 399.04 | 420.15 | 390.24 | 498.06 | 461.18 | 197.39 |
| 山西 | 1412.95 | 518.1 | 316.99 | 347.49 | 317.83 | 567.85 | 391.06 | 250.74 |
| 内蒙古 | 1423.22 | 594.69 | 292.42 | 268.88 | 390.18 | 548.21 | 403.69 | 274.33 |
| 辽宁 | 1846.11 | 592.01 | 272.75 | 378.31 | 347.47 | 575.06 | 412.09 | 230.62 |
| 吉林 | 1650.95 | 547.04 | 257.73 | 325.37 | 344.97 | 528.61 | 453.85 | 228.7 |
| 黑龙江 | 1561 | 531.98 | 259.61 | 353.48 | 318.34 | 534.22 | 432.09 | 201.64 |
| 上海 | 4021.77 | 577.39 | 642.08 | 557.96 | 875.35 | 1359.75 | 732.44 | 569.37 |
| 江苏 | 2194.04 | 525.88 | 603.36 | 297.46 | 483.77 | 691.5 | 438.16 | 298.58 |
| 浙江 | 2888.28 | 669.03 | 926.68 | 532.69 | 689 | 1065.1 | 724.46 | 457.15 |
| 安徽 | 1998.95 | 466.55 | 327.36 | 205.13 | 333.39 | 585.44 | 407.13 | 193.7 |
| 福建 | 2651.11 | 506.94 | 488.37 | 283.24 | 559.69 | 598.95 | 639.83 | 287 |
| 江西 | 1587.55 | 353.44 | 292.1 | 149.98 | 310.94 | 488.24 | 527.16 | 185.11 |
| 山东 | 1801.34 | 700.29 | 522.36 | 327.49 | 411.29 | 777.79 | 441.46 | 270.38 |
| 河南 | 1424.9 | 484.16 | 333.24 | 298.74 | 299.89 | 427.88 | 650.25 | 191.1 |
| 湖北 | 1799.38 | 582.66 | 347.84 | 241.87 | 336.19 | 698.89 | 586.33 | 211.63 |
| 湖南 | 1943.55 | 551.47 | 460.15 | 328.63 | 474.69 | 826.89 | 662.4 | 298.45 |
| 广东 | 3089.63 | 382.98 | 556.14 | 392.41 | 1075.32 | 961.79 | 1126.74 | 514.61 |
| 广西 | 1968.02 | 363.16 | 480.73 | 253.23 | 457.24 | 704.58 | 740.06 | 257.72 |
| 海南 | 2022.19 | 208.83 | 282.49 | 243.85 | 349.44 | 525.92 | 460.06 | 275.06 |
| 重庆 | 2337.65 | 589.28 | 509.82 | 334.05 | 442.5 | 850.15 | 563.72 | 246.51 |
| 四川 | 2082.18 | 489.78 | 460.55 | 300.26 | 381.47 | 674.8 | 530.25 | 256.88 |
| 贵州 | 1748.83 | 486.2 | 361.88 | 249.39 | 371.68 | 522.74 | 333.73 | 199.45 |
| 云南 | 2105.66 | 535.41 | 306.73 | 369.63 | 467.6 | 595.92 | 508.82 | 362.84 |
| 西藏 | 2626.99 | 1001.52 | 258.21 | 220.06 | 628.39 | 494.99 | 369.1 | 395.13 |
| 陕西 | 1589.44 | 443.74 | 529.68 | 361.18 | 366.3 | 642.45 | 452.73 | 252.22 |
| 甘肃 | 1639.17 | 537.94 | 367.32 | 361.4 | 320.88 | 592.72 | 322.94 | 277.94 |
| 青海 | 1790.27 | 532.51 | 350.85 | 374.4 | 361.9 | 594.03 | 295.48 | 399.14 |
| 宁夏 | 1562.57 | 572.03 | 469.18 | 409.96 | 437.72 | 542.39 | 323.16 | 278.39 |
| 新疆 | 1716.8 | 690.14 | 440.37 | 302.82 | 406.72 | 626.58 | 474.61 | 273.37 |