第九章 主成分与因子分析

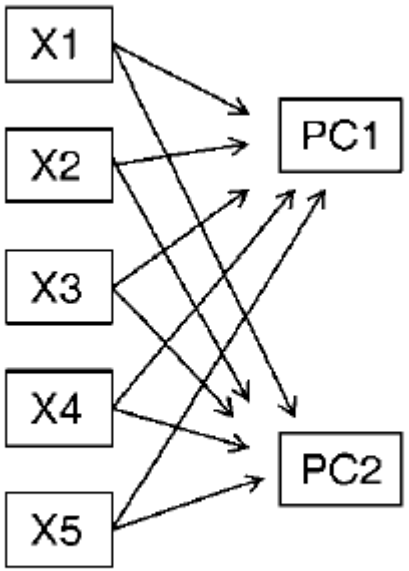
复旦大学附属肿瘤医院 周支瑞

9.1 主成分分析

主成分分析的基本思想

- 主成分分析的基本思想就是将彼此相关的一组指标变量转化为彼此独立的一组新的指标变量，并用其中较少的几个新指标变量，综合反映原多个指标变量中所包含的主要信息，符合专业含义。
- 何为主成分？简而言之，主成分实际上就是由原变量X1~Xm 线性组合出来的m个互不相关、且未丢失任何信息的新变量，也称为综合变量。多指标的主成分分析常被用来寻找判断某种事物或现象的综合指标，并给综合指标所蕴藏的信息以恰当的解释，以便更深刻地揭示事物内在的规律。

主成分分析模型示意图



(a) 主成分分析模型

- 主成分分析（PCA）是一种数据降维技巧，它能将大量相关变量转化为一组少的不相关变量，这些无关变量称为主成分。使用PCA可将30个相关（很可能冗余）的环境变量转化为5个无关的成分变量，并且尽可能保留原始数据集的信息。

主成分分析基本思想应用

- 比如描述儿童生长发育的指标中，身高、腿长和臂长这3 个指标可能是相关的，而胸围、大腿围和臂围这3个围度指标也会有一定的相关性。如果分别用每一个指标对儿童的生长发育做出评价，那么这种评价就是孤立的、片面的，而不是综合的。仅选用几个"重要的"或"有代表性"的指标来评价，就可能失去许多有用的信息， 容易得出片面的结论。我们需要一种综合性的分析方法，既可减少指标变量个数，又尽量不损失原指标变量所包含的信息，对资料进行综合分析。

主成分分析的数学模型

- 假设原变量指标为$x_1, x_2, \ldots, x_k$, 经过标准化后得到标准指标变量$X_1, X_2, \ldots, X_K$:

$$X_j = \frac{x_j - \overline{x}_j}{s_j}, j = 1, 2, \cdots, k$$

- 其中，$\overline{x_j}$是第j个指标变量的均值, sj是第j个指标变量的标准差。它们的综合指标(新变量指标)为$z_1, z_2, \ldots, z_m$, (m<k) 则进行线性变换:

$$\begin{cases} z_1 = l_{11}X_1 + l_{12}X_2 + \cdots + l_{1k}X_k \\ z_2 = l_{21}X_1 + l_{22}X_2 + \cdots + l_{2k}X_k \\ \cdots \\ z_k = l_{k1}X_1 + l_{k2}X_2 + \cdots + l_{kk}X_k \end{cases}$$

- 将k 个标准指标变量$x_1, x_2, \ldots, x_k$,转换成了k 个新变量$z_1, z_2, \ldots, z_k$

线性变换应满足以下3个条件:

- z_i和z_j独立，$i \neq j, i, j = 1, 2, \cdots, k$；
- $\text{var}(z_1) \geqslant \text{var}(z_2) \geqslant \cdots \geqslant \text{var}(z_k)$；
- $l_{i1}^2 + l_{i2}^2 + \cdots + l_{ik}^2 = 1, i = 1, 2, \cdots, k$。

$z_1, z_2, \cdots, z_k$是$X_1, X_2, \cdots, X_k$的k个主成分，其中，z_1为第一主成分，z_2为第二主成分，…，z_k

psych包中的函数

- R基础安装包提供了PCA和EFA的函数，分别是princomp()和factanal()。本章将重点介绍psych包中提供的函数。它们提供了比基础函数更丰富和有用的选项。

表14-1　psych包中有用的因子分析函数

函　数	描　述
principal()	含多种可选的方差旋转方法的主成分分析
fa()	可用主轴、最小残差、加权最小平方或最大似然法估计的因子分析
fa.parallel()	含平行分析的碎石图
factor.plot()	绘制因子分析或主成分分析的结果
fa.diagram()	绘制因子分析或主成分的载荷矩阵
scree()	因子分析和主成分分析的碎石图

主成分与因子分析的一般步骤

1. 数据预处理。PCA和EFA都根据观测变量间的相关性来推导结果。用户可以输入原始数据矩阵或者相关系数矩阵到principal()和fa()函数中。若输入初始数据，相关系数矩阵将会被自动计算，在计算前请确保数据中没有缺失值。
2. 选择因子模型。判断是PCA（数据降维）还是EFA（发现潜在结构）更符合你的研究目标。如果选择EFA方法，你还需要选择一种估计因子模型的方法（如最大似然估计）。
3. 判断要选择的主成分/因子数目。
4. 选择主成分/因子。
5. 旋转主成分/因子。
6. 解释结果。
7. 计算主成分或因子得分。

主成分分析案例 1

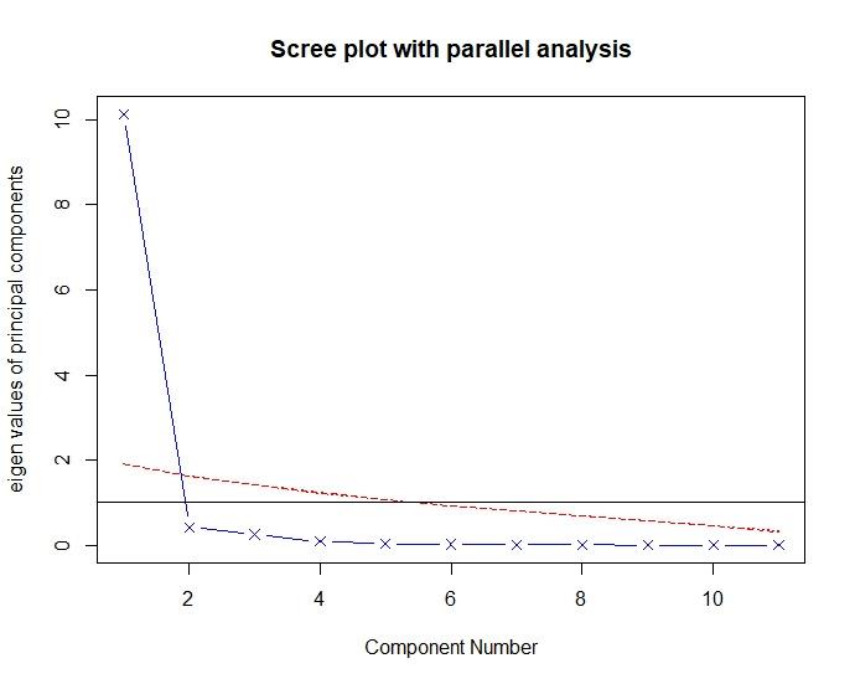
- 数据集USJudgeRatings包含了律师对美国高等法院法官的评分。数据框包含43个观测，12个变量。表14-2列出了所有的变量。

表14-2　USJudgeRatings数据集中的变量

变　量	描　述	变　量	描　述
CONT	律师与法官的接触次数	PREP	审理前的准备工作
INTG	法官正直程度	FAMI	对法律的熟稔程度
DMNR	风度	ORAL	口头裁决的可靠度
DILG	勤勉度	WRIT	书面裁决的可靠度
CFMG	案例流程管理水平	PHYS	体能
DECI	决策效率	RTEN	是否值得保留

判断主成分的个数

- library(psych)
- fa.parallel(USJudgeRatings[,-1], fa="pc", n.iter=100, show.legend=FALSE, main="Scree plot with parallel analysis") # 判断主成分个数
- abline(1,0)



- 展示了基于观测特征值的碎石检验（由线段和x符号组成）、根据100个随机数据矩阵推导出来的特征值均值（虚线），以及大于1的特征值准则（y=1的水平线）。
- 左图评价美国法官评分中要保留的主成分个数。碎石图（直线与x符号）、特征值大于1准则（水平线）和100次模拟的平行分析（虚线）都表明保留一个主成分即可

提取主成分

- principal()函数可以根据原始数据矩阵或者相关系数矩阵做主成分分析。格式为：principal(r, nfactors=, rotate=, scores=) 其中：
  - r 是相关系数矩阵或原始数据矩阵；
  - nfactors设定主成分数（默认为1）；
  - rotate指定旋转的方法（默认最大方差旋转（varimax）；
  - scores设定是否需要计算主成分得分（默认不需要）

提取主成分


```
# Listing 14.1 - Principal components analysis of US Judge Ratings
>library(psych)
>pc <- principal(USJudgeRatings[,-1], nfactors=1) # 提取主成分
>pc
```

```
      PC1   h2     u2 com
INTG 0.92 0.84 0.1565   1
DMNR 0.91 0.83 0.1663   1
DILG 0.97 0.94 0.0613   1
CFMG 0.96 0.93 0.0720   1
DECI 0.96 0.92 0.0763   1
PREP 0.98 0.97 0.0299   1
FAMI 0.98 0.95 0.0469   1
ORAL 1.00 0.99 0.0091   1
WRIT 0.99 0.98 0.0196   1
PHYS 0.89 0.80 0.2013   1
RTEN 0.99 0.97 0.0275   1

                 PC1
SS loadings    10.13
Proportion Var  0.92
```

- 第一主成分（PC1）与每个变量都高度相关，也就是说，它是一个可用来进行一般性评价的维度。h2栏指成分公因子方差，即主成分对每个变量的方差解释度。u2栏指成分唯一性，即方差无法被主成分解释的比例（1–h2）。
- 例如，体能（PHYS）80%的方差都可用第一主成分来解释，20%不能。相比而言，PHYS是用第一主成分表示性最差的变量。SS loadings行包含了与主成分相关联的特征值，指的是与特定主成分相关联的标准化后的方差值（本例中，第一主成分的值为10）。
- 最后，Proportion Var行表示的是每个主成分对整个数据集的解释程度。此处可以看到，第一主成分解释了11个变量92%的方差。

主成分分析案例 2

- Harman23.cor数据集包含305个女孩的8个身体测量指标。本例中，数据集由变量的相关系数组成，而不是原始数据集（见表14-3）。

表14-3　305个女孩的身体指标间的相关系数（Harman23.cor）

	身高	指距	前臂	小腿	体重	股骨转子间径	胸围	胸宽
身高	1.00	0.85	0.80	0.86	0.47	0.40	0.30	0.38
指距	0.85	1.00	0.88	0.83	0.38	0.33	0.28	0.41
前臂	0.80	0.88	1.00	0.80	0.38	0.32	0.24	0.34
小腿	0.86	0.83	0.80	1.00	0.44	0.33	0.33	0.36
体重	0.47	0.38	0.38	0.44	1.00	0.76	0.73	0.63
股骨转子间径	0.40	0.33	0.32	0.33	0.76	1.00	0.58	0.58
胸围	0.30	0.28	0.24	0.33	0.73	0.58	1.00	0.54
胸宽	0.38	0.41	0.34	0.36	0.63	0.58	0.54	1.00

来源：Harman, H. H. (1976) *Modern Factor Analysis, Third Edition Revised*, University of Chicago Press, Table 2.3

判断主成分的个数

```
# Principal components analysis Harman23.cor data
>library(psych)
>fa.parallel(Harman23.cor$cov, n.obs=302, fa="pc", n.iter=100,
          show.legend=FALSE, main="Scree plot with parallel analysis") # 判断主成分个数
>abline(a=1,b=0,lwd=1,col="green")
```

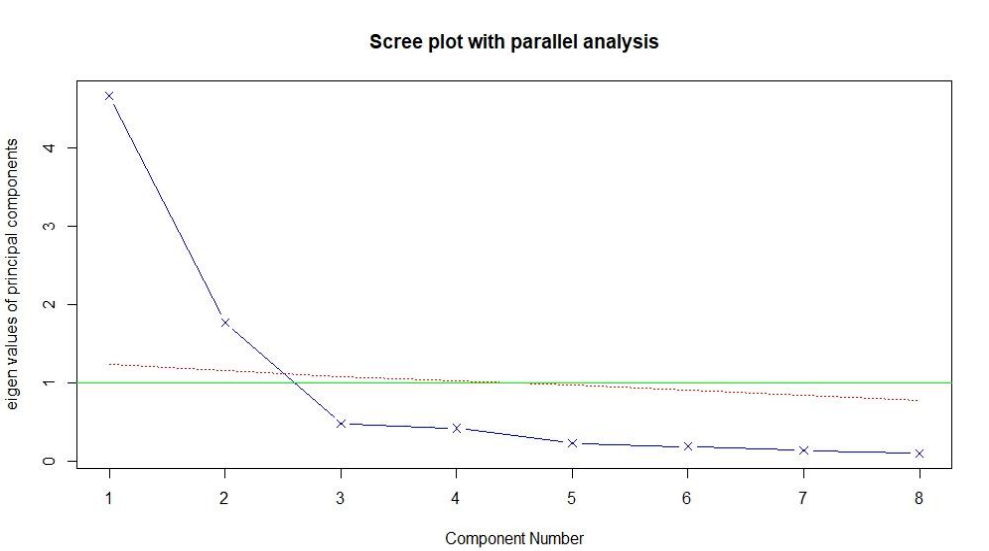


图14-3 判断身体测量数据集所需的主成分数。碎石图（直线和x符号）、特征值大于1准则（水平线）和100次模拟（虚线）的平行分析建议保留两个主成分

提取主成分


```
# Listing 14.2 - Principal components analysis of body measurements
>library(psych)
>PC <- principal(Harman23.cor$cov, nfactors=2, rotate="none") # 提取主成分
>PC
```

	PC1	PC2	h2	u2	com
height	0.86	-0.37	0.88	0.123	1.4
arm.span	0.84	-0.44	0.90	0.097	1.5
forearm	0.81	-0.46	0.87	0.128	1.6
lower.leg	0.84	-0.40	0.86	0.139	1.4
weight	0.76	0.52	0.85	0.150	1.8
bitro.diameter	0.67	0.53	0.74	0.261	1.9
chest.girth	0.62	0.58	0.72	0.283	2.0
chest.width	0.67	0.42	0.62	0.375	1.7

	PC1	PC2
SS loadings	4.67	1.77
Proportion Var	0.58	0.22
Cumulative Var	0.58	0.81
Proportion Explained	0.73	0.27
Cumulative Proportion	0.73	1.00

- 第一主成分解释了身体测量指标58%的方差，而第二主成分解释了22%，两者总共解释了81%的方差。对于高度变量，两者则共解释了其88%的方差。
- 载荷阵解释了成分和因子的含义。第一主成分与每个身体测量指标都正相关，看起来似乎是一个一般性的衡量因子；第二主成分与前四个变量（height、arm.span、forearm和lower.leg）负相关，与后四个变量（weight、bitro.diameter、chest.girth和chest.width）正相关

主成分旋转


```
# Listing 14.3 - Principal components analysis with varimax rotation
>rc <- principal(Harman23.cor$cov, nfactors=2, rotate="varimax")
>rc
```

```
                  RC1  RC2   h2    u2 com
height           0.90 0.25 0.88 0.123 1.2
arm.span         0.93 0.19 0.90 0.097 1.1
forearm          0.92 0.16 0.87 0.128 1.1
lower.leg        0.90 0.22 0.86 0.139 1.1
weight           0.26 0.88 0.85 0.150 1.2
bitro.diameter   0.19 0.84 0.74 0.261 1.1
chest.girth      0.11 0.84 0.72 0.283 1.0
chest.width      0.26 0.75 0.62 0.375 1.2

                       RC1  RC2
SS loadings           3.52 2.92
Proportion Var        0.44 0.37
Cumulative Var        0.44 0.81
Proportion Explained  0.55 0.45
Cumulative Proportion 0.55 1.00
```

- 列的名字都从PC变成了RC，以表示成分被旋转。观察RC1栏的载荷，你可以发现第一主成分主要由前四个变量来解释（长度变量）。RC2栏的载荷表示第二主成分主要由变量5到变量8来解释（容量变量）。注意两个主成分仍不相关，对变量的解释性不变，这是因为变量的群组没有发生变化。另外，两个主成分旋转后的累积方差解释性没有变化（81%），变的只是各个主成分对方差的解释度（成分1从58%变为44%，成分2从22%变为37%）。

主成分在线性回归中的应用案例

例 16-2　某学校 20 名一年级女大学生体重（公斤）、胸围（厘米）、肩宽（厘米）及肺活量（升）实测值如表 16-3 所示，试对影响女大学生肺活量的有关因素作多元回归分析。

表 16-3　　20 名一年级女大学生肺活量及有关变量测量结果

编　　号	体重（公斤）	胸围（厘米）	肩宽（厘米）	肺活量（升）
1	51.3	73.6	36.4	2.99
2	48.9	83.9	34.0	3.11
3	42.8	78.3	31.0	1.91
4	55.0	77.1	31.0	2.63
5	45.3	81.7	30.0	2.86
6	45.3	74.8	32.0	1.91
7	51.4	73.7	36.5	2.98
8	53.8	79.4	37.0	3.28
9	49.0	72.6	30.1	2.52
10	53.9	79.5	37.1	3.27
11	48.8	83.8	33.9	3.10
12	52.6	88.4	38.0	3.28
13	42.7	78.2	30.9	1.92
14	52.5	88.3	38.1	3.27
15	55.1	77.2	31.1	2.64
16	45.2	81.6	30.2	2.85
17	51.4	78.3	36.5	3.16
18	48.7	72.5	30.0	2.51
19	51.3	78.2	36.4	3.15
20	45.2	74.7	32.1	1.92

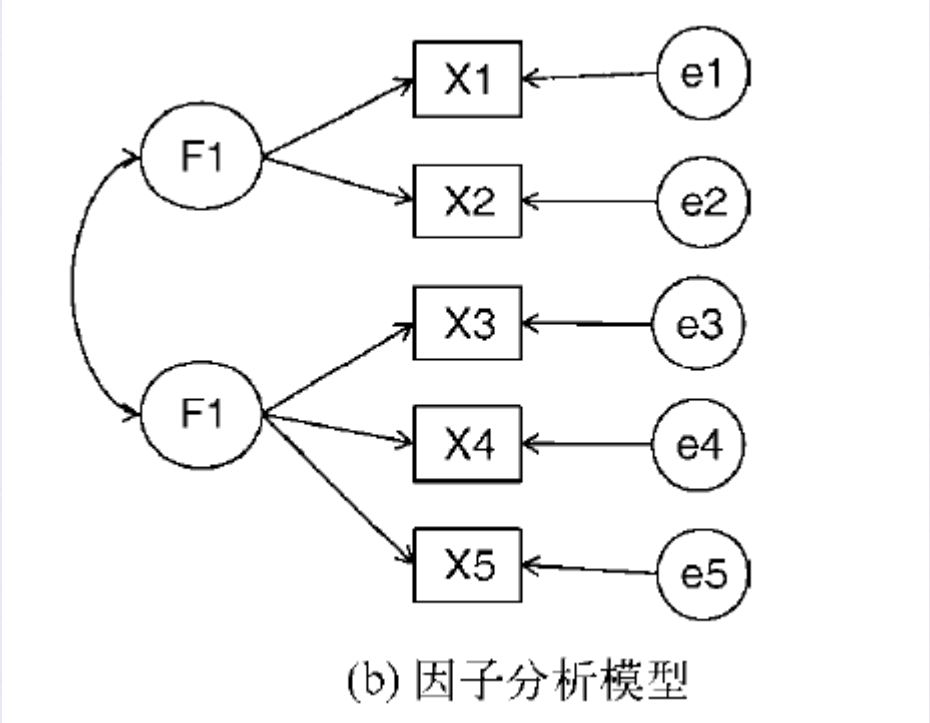
主成分在线性回归中的应用案例 代码


```
> install.packages("car")
> library(car)
> example16_2  <- read.table ("example16_2.csv", header=TRUE, sep=",")
> example16_2
> fit  <- lm(y~x1+x2+x3, data=example16_2)
> summary(fit)
> vif(fit)

> library(psych)
> describe(example16_2)
> fa.parallel(example16_2[-4], fa="pc", n.iter=100, show.legend=FALSE, main="Screen plot with
parallel analysis")
> pc <-  principal(example16_2[-4], nfactors=2, rotate= "varimax", score=TRUE)
> pc
> pc$weights
> pc$scores
> newdata <- data.frame(example16_2,  pc$scores)
> newdata
> fit  <- lm(y~ RC1+RC2, data=newdata)
> summary(fit)
```

9.2 探索性因子分析

因子分析



(b) 因子分析模型

- 探索性因子分析（EFA）是一系列用来发现一组变量潜在结构的方法。它通过寻找一组更小的、潜在的或隐藏的结构来解释已观测到的、显式的变量间的关系。
- Harman74.cor包含了24个心理测验间的相互关系，受试对象为145个七年级或八年级的学生。应用EFA探索该数据表明276个测验间的相互关系可用四个学生能力的潜在因子（语言能力、反应速度、推理能力和记忆能力）进行解释。

因子分析模型形式

$$X_i = a_1F_1 + a_2F_2 + \cdots + a_pF_p + U_i$$

- 其中Xi是第i个可观测变量（i=1…k），Fj是公共因子（j=1…p），并且p<k。Ui是Xi变量独有的部分（无法被公共因子解释）。ai可认为是每个因子对复合而成的可观测变量的贡献值。回到本章开头的Harman74.cor的例子，我们认为每个个体在24个心理学测验上的观测得分，是根据四个潜在心理学因素的加权能力值组合而成。

因子分析 案例

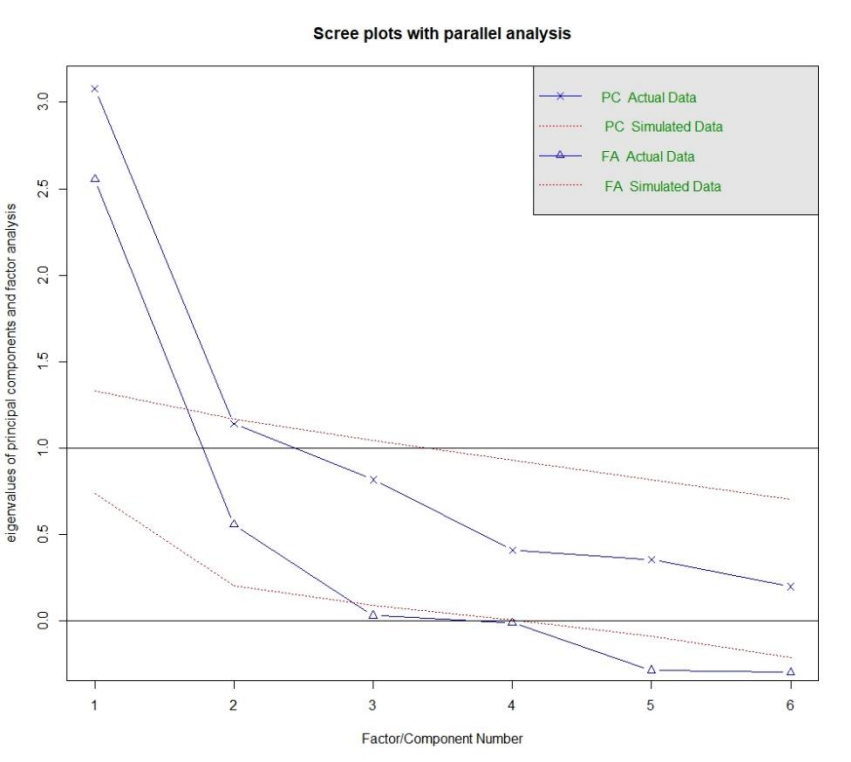
- 虽然PCA和EFA存在差异，但是它们的许多分析步骤都是相似的。为阐述EFA的分析过程，我们用它来对六个心理学测验间的相关性进行分析。112个人参与了六个测验，包括非语言的普通智力测验（general）、画图测验（picture）、积木图案测验（blocks）、迷宫测验（maze）、阅读测验（reading）和词汇测验（vocab）。我们如何用一组较少的、潜在的心理学因素来解释参与者的测验得分呢？数据集ability.cov提供了变量的协方差矩阵，你可用cov2cor()函数将其转化为相关系数矩阵。数据集没有缺失值。

```
## Exploratory factor analysis of ability.cov data
>options(digits=2)
>library(psych)
>covariances <- ability.cov$cov
# convert covariances to correlations
>correlations <- cov2cor(covariances)
>correlations
```

判断需提取的公共因子数

- 用fa.parallel()函数可判断需提取的因子数：

```
# determine number of factors to extract
> fa.parallel(correlations, n.obs=112, fa="both", n.iter=100,
          main="Scree plots with parallel analysis")
> abline(0,0)
```



- 图14-4 判断心理学测验需要保留的因子数。图中同时展示了PCA和EFA的结果。PCA结果建议提取一个或者两个成分，EFA建议提取两个因子。注意，代码中使用了fa="both"，因子图形将会同时展示主成分和公共因子分析的结果。

提取公共因子

- 现在你决定提取两个因子，可以使用fa()函数获得相应的结果。fa()函数的格式如下：
- fa(r, nfactors=, n.obs=, rotate=, scores=, fm=) 其中：
  - r是相关系数矩阵或者原始数据矩阵；
  - nfactors设定提取的因子数（默认为1）；
  - n.obs是观测数（输入相关系数矩阵时需要填写）；
  - rotate设定旋转的方法（默认互变异数最小法）；
  - scores设定是否计算因子得分（默认不计算）；
  - fm设定因子化方法（默认极小残差法）。

提取公共因子


```
# Listing 14.6 - Principal axis factoring without rotation
> fa <- fa(correlations, nfactors=2, rotate="none", fm="pa")
> fa
```

```
          PA1   PA2   h2    u2 com
general 0.75  0.07 0.57 0.432 1.0
picture 0.52  0.32 0.38 0.623 1.7
blocks  0.75  0.52 0.83 0.166 1.8
maze    0.39  0.22 0.20 0.798 1.6
reading 0.81 -0.51 0.91 0.089 1.7
vocab   0.73 -0.39 0.69 0.313 1.5

                       PA1  PA2
SS loadings           2.75 0.83
Proportion Var        0.46 0.14
Cumulative Var        0.46 0.60
Proportion Explained  0.77 0.23
Cumulative Proportion 0.77 1.00
```

- 可以看到，两个因子解释了六个心理学测验60%的方差。不过因子载荷阵的意义并不太好解释，此时使用因子旋转将有助于因子的解释。

用正交旋转提取因子


```
# Listing 14.7 - Factor extraction with orthogonal rotation
> fa.varimax <- fa(correlations, nfactors=2, rotate="varimax", fm="pa")
> fa.varimax
```

```
          PA1  PA2   h2    u2 com
general  0.49 0.57 0.57 0.432 2.0
picture  0.16 0.59 0.38 0.623 1.1
blocks   0.18 0.89 0.83 0.166 1.1
maze     0.13 0.43 0.20 0.798 1.2
reading  0.93 0.20 0.91 0.089 1.1
vocab    0.80 0.23 0.69 0.313 1.2

                       PA1  PA2
SS loadings           1.83 1.75
Proportion Var        0.30 0.29
Cumulative Var        0.30 0.60
Proportion Explained  0.51 0.49
Cumulative Proportion 0.51 1.00
```

- 结果显示因子变得更好解释了。阅读和词汇在第一因子上载荷较大，画图、积木图案和迷宫在第二因子上载荷较大，非语言的普通智力测量在两个因子上载荷较为平均，这表明存在一个语言智力因子和一个非语言智力因子。使用正交旋转将人为地强制两个因子不相关。如果想允许两个因子相关该怎么办呢？此时可以使用斜交转轴法，比如promax

用斜交旋转提取因子


```
# Listing 14.8 - Factor extraction with oblique rotation
> fa.promax <- fa(correlations, nfactors=2, rotate="promax", fm="pa")
> fa.promax
```

```
             PA1   PA2   h2    u2 com
general     0.37  0.48 0.57 0.432 1.9
picture    -0.03  0.63 0.38 0.623 1.0
blocks     -0.10  0.97 0.83 0.166 1.0
maze        0.00  0.45 0.20 0.798 1.0
reading     1.00 -0.09 0.91 0.089 1.0
vocab       0.84 -0.01 0.69 0.313 1.0

                       PA1  PA2
SS loadings           1.83 1.75
Proportion Var        0.30 0.29
Cumulative Var        0.30 0.60
Proportion Explained  0.51 0.49
Cumulative Proportion 0.51 1.00

 With factor correlations of
     PA1  PA2
PA1 1.00 0.55
PA2 0.55 1.00
```

- 对正交旋转，因子分析的重点在于因子结构矩阵（变量与因子的相关系数），而对于斜交旋转，因子分析会考虑三个矩阵：因子结构矩阵、因子模式矩阵和因子关联矩阵。
- 因子模式矩阵即标准化的回归系数矩阵。它列出了因子预测变量的权重。因子关联矩阵即因子相关系数矩阵。在代码清单14-8中，PA1和PA2栏中的值组成了因子模式矩阵。它们是标准化的回归系数，而不是相关系数。注意，矩阵的列仍用来对因子进行命名（虽然此处存在一些争论）。你同样可以得到一个语言因子和一个非语言因子。因子关联矩阵显示两个因子的相关系数为0.57，相关性很大。如果因子间的关联性很低，你可能需要重新使用正交旋转来简化问题。

变量与因子间的相关系数计算


```
# calculate factor loading matrix
> fsm <- function(oblique) {
  if (class(oblique)[2]=="fa" & is.null(oblique$Phi)) {
    warning("Object doesn't look like oblique EFA")
  } else {
    P <- unclass(oblique$loading)
    F <- P %*% oblique$Phi
    colnames(F) <- c("PA1", "PA2")
    return(F)
  }
}
> fsm(fa.promax)
```

正交或者斜交结果的图形绘制


```
# plot factor solution
> factor.plot(fa.promax, labels=rownames(fa.promax$loadings))
> fa.diagram(fa.promax, simple=FALSE)
```

因子分析应用案例

例 17-3　某医院为了合理地评价该院各月的医疗工作质量，搜集了 3 年有关门诊人次、出院人数、病床利用率、病床周转次数、平均住院天数、治愈好转率、病死率、诊断符合率、抢救成功率等 9 个指标数据，见表 17-5。试采用因子分析法，探讨其综合评价指标体系。

表 17-5　某医院 3 年的医疗工作质量有关指标实测值

年月 X_0	门诊人次 X_1	出院人数 X_2	病床利用率 X_3	病床周转次数 X_4	平均住院天数 X_5	治愈好转率 X_6（%）	病死率 X_7（%）	诊断符合率 X_8（%）	抢救成功率 X_9（%）
91.01	4.34	389	99.06	1.23	25.46	93.15	3.56	97.51	61.66
91.02	3.45	271	88.28	0.85	23.55	94.31	2.44	97.94	73.33
91.03	4.38	385	103.97	1.21	26.54	92.53	4.02	98.48	76.79
91.04	4.18	377	99.48	1.19	26.89	93.86	2.92	99.41	63.16
91.05	4.32	378	102.01	1.19	27.63	93.18	1.99	99.71	80.00
91.06	4.13	349	97.55	1.10	27.34	90.63	4.38	99.03	63.16
91.07	4.57	361	91.66	1.14	24.89	90.60	2.73	99.69	73.53
91.08	4.31	209	62.18	0.52	31.74	91.67	3.65	99.48	61.11
91.09	4.06	425	83.27	0.93	26.56	93.81	3.09	99.48	70.73
91.10	4.43	458	92.39	0.95	24.26	91.12	4.21	99.76	79.07
91.11	4.13	496	95.43	1.03	28.75	93.43	3.50	99.10	80.49

……

因子分析应用案例 代码


```
> library(psych)
> example17_3  <- read.table ("example17_3.csv", header=TRUE, sep=",")
> example17_3
> fa.parallel(example17_3, fa="fa", n.iter=100, main="Screen plots with parallel
analysis")
> fa  <- fa(example17_3, nfactors=4, rotate="none", fm="ml", score=TRUE)
> fa
> fa$weights
> fa$scores
> factor.plot(fa, labels=rownames(fa$loadings))
> fa.diagram(fa, simple=FALSE)
> fa2  <- fa(example17_3, nfactors=4, rotate="varimax", fm="ml", score=TRUE)
> fa2
> fa2$weights
> fa2$scores
> factor.plot(fa2, labels=rownames(fa$loadings))
> fa.diagram(fa2, simple=FALSE)
```

主要参考文献

【1】Robert I. Kabacoff 著, 《R语言实战 》(第2版), 人民邮电出版社, 2016

【2】Peter Dalgaard 著, 《R语言统计入门》》(第2版), 人民邮电出版社, 2014

【3】薛毅 陈立萍 著,《R语言实用教程》, 清华大学出版社, 2014

【4】张铁军 陈兴栋 刘振球 著,《R语言与医学统计图形》, 人民卫生出版社, 2018

【5】汪海波 萝莉 汪海玲 著, 《R语言统计分析与应用》， 人民邮电出版社，2018

Thanks!

感谢您的观看！