多元分析方法及实例应用

多元分析是多变量的统计分析方法，应用方面非常广阔，内容庞大，视角独特，实用性强，深受工程技术人员的青睐，在很多工程中应用广泛，且被不断地完善和改进。

1.5 因子分析

1.5.1 因子分析基本原理

因子分析：研究从变量群中提取共性因子的统计技术。通过研究众多变量之间存在的内部依赖关系，探求观测数据中的基本结构，并使用少数几个假想变量来表示其基本的数据结构。其中的几个假想变量可以反映出众多变量中的主要信息。原始变量是可观测的显在变量，而假想变量是不可观测的潜在变量，称之为因子。因子分析的主要目的是用来描述隐藏在一组测量到的变量中的一些更基本的，但又无法直接测量到的隐性变量 (latent variable, latent factor)。

因子分析可以看成主成分分析的推广，它也是多元统计分析中常用的一种降维方式，但因子分析所涉及的计算与主成分分析非常类似，差别也非常明显，主要有以下几方面：

1. 主成分分析将方差分为不同的正交成分，因子分析将方差划归为不同的起因因子。
2. 主成分分析是变量变换，而因子分析需要构造因子模型。
3. 主成分分析中原始变量的线性组合表示新的综合变量，即主成分。因子分析中的潜在的假想变量和随机影响变量的线性组合表示原始变量。

因子分析与回归分析不同因子分析中的因子较为抽象，而回归变量具有非常明确的现实意义。

1.数学模型

设 P 个变量 $X_i\left(i=1,2,3,\cdots,P\right)$ 可以表示为

$$X_{\mathrm{i}}=\mu_{\mathrm{i}}+\alpha_{\mathrm{i1}}F_1+...+\alpha_{\mathrm{ip}}F_m+\varepsilon_i,m\le p \tag{1.1}$$

或

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_P \end{bmatrix}=\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_P \end{bmatrix}+\begin{bmatrix} \alpha_{11} & \alpha_{12} & ... & \alpha_{1\mathrm{m}} \\ \alpha_{21} & \alpha_{22} & ... & \alpha_{2\mathrm{m}} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{p1} & \alpha_{p2} & \cdots & \alpha_{p\mathrm{m}} \end{bmatrix}\begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_{\mathrm{m}} \end{bmatrix}+\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_P \end{bmatrix} \tag{1.2}$$

或

$$X-\mu=\Lambda F+\varepsilon \tag{1.3}$$

其中，

$$X=\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_P \end{bmatrix},\mu=\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_P \end{bmatrix},\Lambda=\begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1\mathrm{m}} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2\mathrm{m}} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{P1} & \alpha_{P2} & \cdots & \alpha_{\mathrm{pm}} \end{bmatrix},F=\begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_p \end{bmatrix}，\ \varepsilon=\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{\mathrm{p}} \end{bmatrix} \tag{1.4}$$

称 $F_1,F_2,\cdots,F_P$ 为公共因子，是不可观测的变量，他们的系数称为载荷因子。ε_{i} 是特殊因子，是不能被前 m 个公共因子包含的部分。并且满足

$$\begin{aligned} &E\left(F\right)=0，\ E\left(\varepsilon\right)=0，\ \mathrm{Cov}\left(F\right)=I_{\mathrm{q}}, \\ &D\left(\varepsilon\right)=\mathrm{Cov}\left(\varepsilon\right)=\mathrm{diag}\left(\sigma_1^2，\ \sigma_2^2,\cdots，\ \sigma_{\mathrm{q}}^2\right)，\ \mathrm{Cov}\left(F,\varepsilon\right)=0 \end{aligned} \tag{1.5}$$

2.因子分析模型的性质

1) 原始变量 X 的协方差矩阵的分解。由 $X-\mu=\Lambda F+\varepsilon$ 得，

$$\mathrm{Cov}\left(X-\mu\right)=\Delta\mathrm{Cov}\left(F\right)\Lambda^T+\mathrm{Cov}\left(\varepsilon\right),$$

即

$$\mathrm{Cov}\left(X\right)=\Lambda\Lambda^T+\mathrm{diag}\left(\sigma_1^2,\sigma_2^2,\cdots,\sigma_{\mathrm{q}}^2\right) \tag{1.6}$$

$\sigma_1^2,\sigma_2^2,\cdots,\sigma_{\mathrm{m}}^2$ 的值越小，则公共因子分享的成分越多。

2） 载荷矩阵不是唯一的。设 T 为一个 $P \times P$ 的正交矩阵，令

$\tilde{\Lambda} = \Lambda T, \tilde{F} = T^T F$，则模型可以表示为

$$X = \mu + \tilde{\Lambda}\tilde{F} + \varepsilon \tag{1.7}$$

3. 因子载荷矩阵的几个统计性质

1) 因子载荷 α_{ij} 的统计意义

因子载荷 α_{ij} 是第 i 个变量和第 j 个公共因子的相关系数，反映了第 i 个变量和第 j 个公共因子的相关重要性。若其绝对值越大，则相关密切程度越高。

2) 变量共同度统计意义

变量 X_i 的共同都是因子载荷矩阵的第 i 行的元素的平方和，记为 $h^2 = \sum_{j=1}^{m} \alpha_{ij}^2$。

对(1.1)式进行左右两边求方差，得

$$Var\left(X_i\right) = a_{i1}^2 Var\left(F_1\right) + \cdots + a_{im}^2 Var\left(F_m\right) + Var\left(\varepsilon_i\right) \tag{1.8}$$

则

$$1 = \sum_{j=1}^{m} \alpha_{ij}^2 + \sigma_i^2$$

其中，特殊因子的方差 $\sigma_i^2\left(i = 1, 2, \cdots, p\right)$ 称为特殊方差。

3) 公共因子 F_j 方差贡献的统计意义

因子载荷矩阵中各列元素的平方和

$$S_j = \sum_{i=1}^{p} \alpha_{ij}^2 \tag{1.9}$$

称为 $F_j\left(j = 1, 2, 3, \cdots, m\right)$ 对所有的 X_i 的方差贡献和，用于衡量 F_j 的相对重要性。

1.5.2 因子载荷的矩阵的估计方法

因子分析的一个基本问题是如何估计因子载荷，即如何求解因子模型式(1.1)。

1. 主成分分析法

假设$\lambda_1 \ge \lambda_2 \ge \lambda_3 \ge \cdots \ge \lambda_p$为相关系数矩阵 R 特征值，$\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \cdots, \boldsymbol{\eta}_p$为相应的标准正交化特征向量。假设$m < p$，则因子载荷矩阵$\Lambda$为

$$\Lambda = \left[\sqrt{\lambda_1}\boldsymbol{\eta}_1, \sqrt{\lambda_2}\boldsymbol{\eta}_2, \cdots, \sqrt{\lambda_m}\boldsymbol{\eta}_m\right] \tag{1.10}$$

特殊因子的方差用$R - \Lambda\Lambda^T$的对角元来估计，即

$$\sigma_i^2 = 1 - \sum_{j=1}^{m} \alpha_{ij}^2 \tag{1.11}$$

2. 主因子法

主因子法是在对主成分分析法进行修改之后提出的一种新的方法，首先需要对变量及进行标准化变换，则

$$R = \Lambda\Lambda^T + D$$

其中，$D = \mathrm{diag}\left(\sigma_1^2, \sigma_2^2, \cdots, \sigma_{\mathrm{m}}^2\right)$ 。

则

$$R^* = \Lambda\Lambda^T = R = D \tag{1.12}$$

其中，R^*为约相关系数矩阵，且对角线上的元素是h_i^2。

估计特殊因子的方差方法如下：

1) 取$\hat{h}_i^2 = 1$，则主因子解与主成分解等价。

2) 取$\hat{h}_i^2 = \max_{j \ne i}\left|r_{ij}\right|$，取$X_i$与其余的$X_j$的简单系数的绝对值的最大者。

记

$$R^* = R - D = \begin{bmatrix} \hat{h}_1^2 & r_{12} & \cdots & r_{1p} \\ r_{21} & \hat{h}_2^2 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & \hat{h}_p^2 \end{bmatrix}$$

可以求得如下的因子载荷矩阵：

$$\mathbf{\Lambda} = \left[\sqrt{\lambda_1^*}\mathbf{u}_1^*, \sqrt{\lambda_2^*}\mathbf{u}_2^*, \cdots, \sqrt{\lambda_p^*}\mathbf{u}_p^* \right]$$

3. 最大似然估计

1.5.3 因子旋转

建立因子分析数学模型可以找出公共因子以及对变量进行分组，还可以搞清楚每个公共因子的含义，便于进一步的分析。若每个公共因子的含义不清，则不利于进行实际背景的解释。而因子载荷矩阵不唯一，所以应该对因子载荷矩阵进行旋转。目的是为了简化载荷矩阵的结构，使载荷矩阵每列或行的元素平方值向 0 和 1 进行两级分化。正交旋转的主要方法如下。

1) 方差最大化。从简化因子载荷矩阵的每一列出发，使和每个因子有关的载荷的平方方差最大。当只有少数几个变量在某个因子上存在较高的载荷时，对因子进行解释最简单。方差最大的直观意义是希望通过旋转因子后，每个因子尽量分散开来，向两极靠近。
2) 四次方最大旋转。从简化载荷矩阵的每一行出发，通过旋转初始因子，使得每一个变量只在一个因子上存在较高的载荷，而在其他的因子上存在较低的载荷。若每一个变量只在一个因子上存在非零载荷，则因子解释是最为简单的。四次方最大法是使因子载荷矩阵中每一行的因子载荷平方的方差达到最大值。
3) 等量最大法。等量最大法是把四次方最大法和方差最大法结合起来，求其加权平均最大。
4) 对于两个因子的载荷矩阵来说

$$\Lambda = \left(\alpha_{ij}\right)_{p\times 2}, i=1,2,\cdots,p; j=1,2$$

取正交矩阵

$$T = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$$

为逆时针旋转。若作顺时针旋转，只需要将矩阵T次对角线上的两个元素进行对换即可。

若公共因子数$m>2$，可以考虑不同的两个因子的旋转，在m个因子中每次选择两个进行旋转，将有$\frac{m(m-2)}{2}$中选择方式，当完成$\frac{m(m-2)}{2}$次旋转就算是完成了一个循环，再重新开始第二个循环，一直循环到明确每一个因子的含义。

1.5.4 因子得分

1. 因子分析的基本概念

使用公共因子的线性组合来表示一组观测变量的有关问题，但是若需要使用这一部分因子做其他的研究，就需要对公共因子进行测度，即对公共因子进行赋值。

因子分析的数学模型为

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_P \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_P \end{bmatrix} + \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1\mathrm{m}} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2\mathrm{m}} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{p1} & \alpha_{p2} & \cdots & \alpha_{p\mathrm{m}} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_\mathrm{m} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_P \end{bmatrix} \tag{1.13}$$

因子得分函数

$$F_j = c_j + \beta_{j1}X_1 + \cdots + \beta_{jp}X_p, j=1,2,\cdots,m \tag{1.14}$$

若需要求每个因子的得分，就需要求得分函数的系数。但是(1.14)式中可见，$p>m$所以不能够得到精确的数值，只能通过估计来实现。

2. 巴特莱特因子得分（加权最小二乘法）

将 $X_i-\mu_i$ 视为因变量，将因子载荷矩阵

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1\mathrm{m}} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2\mathrm{m}} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{p1} & \alpha_{p2} & \cdots & \alpha_{p\mathrm{m}} \end{bmatrix}$$

视为自变量的观测。

则

$$\begin{cases} X_1-\mu_1=\alpha_{11}F_1+\alpha_{12}F_2+\cdots+\alpha_{1m}F_m+\varepsilon_1, \\ X_2-\mu_2=\alpha_{21}F_1+\alpha_{22}F_2+\cdots+\alpha_{2m}F_m+\varepsilon_2, \\ \vdots \\ X_p-\mu_p=\alpha_{p1}F_1+\alpha_{p2}F_2+\cdots+\alpha_{pm}F_m+\varepsilon_p. \end{cases} \tag{1.15}$$

使用加权最小二乘法求得分，使

$$\sum_{i=1}^{p}\left[\left(X_i-\mu_i\right)-\left(\alpha_{i1}\hat{F}_1+\alpha_{i2}\hat{F}_2+\cdots+\alpha_{im}\hat{F}_m\right)\right]^2/\sigma^2 \tag{1.16}$$

其中的 $\hat{F}_1,\cdots,\hat{F}_m$ 表示相应个案的因子得分。

使用矩阵表达即

$$X-\mu=\Lambda F+\varepsilon$$

要是

$$\left(X-\mu-\Lambda F\right)^T D^{-1}\left(X-\mu-\Lambda F\right) \tag{1.17}$$

达到最小。

经计算可得，

$$\hat{F}=\left(\Lambda^T D^{-1}\Lambda\right)^{-1}\Lambda^T D^{-1}\left(X-\mu\right)$$

3. 回归方法

回归方法的基本思想为

假设

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_P \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_P \end{bmatrix} + \begin{bmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1\mathrm{m}} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2\mathrm{m}} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{p1} & \alpha_{p2} & \cdots & \alpha_{p\mathrm{m}} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_{\mathrm{m}} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_P \end{bmatrix}$$

因子得分函数为

$$\hat{F}_j = \beta_{j1} X_1 + \cdots + \beta_{jp} X_p,\ j = 1, 2, \cdots, m$$

而

$$\alpha_{ij} = \gamma_{X_i F_j} = E\left(X_i F_j\right) = E\left[X_i\left(\beta_{j1} X_1 + \cdots + \beta_{jp} X_p\right)\right]$$

$$= \beta_{j1}\gamma_{i1} + \cdots + \beta_{jp}\gamma_{ip} = \left[\gamma_{i1}, \cdots, \gamma_{ip}\right] \begin{bmatrix} \beta_{j1} \\ \beta_{j2} \\ \vdots \\ \beta_{jp} \end{bmatrix}$$

所以，存在

$$\begin{bmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1\mathrm{m}} \\ \gamma_{21} & \gamma_{22} & \dots & \gamma_{2\mathrm{m}} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p1} & \gamma_{p2} & \cdots & \gamma_{pm} \end{bmatrix} \begin{bmatrix} \beta_{j1} \\ \beta_{j2} \\ \vdots \\ \beta_{jm} \end{bmatrix} = \begin{bmatrix} \alpha_{1j} \\ \alpha_{2j} \\ \vdots \\ \alpha_{pj} \end{bmatrix},\ j = 1, 2, \cdots, m$$

其中，$\begin{bmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1\mathrm{m}} \\ \gamma_{21} & \gamma_{22} & \dots & \gamma_{2\mathrm{m}} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p1} & \gamma_{p2} & \cdots & \gamma_{pm} \end{bmatrix}, \begin{bmatrix} \beta_{j1} \\ \beta_{j2} \\ \vdots \\ \beta_{jm} \end{bmatrix}, \begin{bmatrix} \alpha_{1j} \\ \alpha_{2j} \\ \vdots \\ \alpha_{pj} \end{bmatrix}$ 分别为原始变量的相关系数矩阵、第 j 个因子得分函数的系数以及载荷矩阵的第 j 列。

使用矩阵表示为

$$\begin{bmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1m} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{p1} & \beta_{p2} & \cdots & \beta_{pm} \end{bmatrix} = R^{-1}\Lambda$$

因子得分的估计为

$$\hat{F} = \left(\hat{F}_{ij}\right)_{n\times m} = X_0 R^{-1}\Lambda$$

其中，$\hat{F}_{ij}$ 表示第 i 个样本点对第 j 个因子 $\hat{F}$ 得分的估计值，X_0 为 $n\times m$ 的原

始数据矩阵。

1.5.5 因子分析的步骤与主成分分析的对比

1) 因子分析的步骤：

1. 选择分析的步骤；
2. 计算所选原始变量的相关系数矩阵；
3. 提出公共因子；
4. 因子旋转；
5. 得出结论。

2) 主成分分析法与因子分析法的数学模型的异同点

1. 相同点

指标的标准化、相关系数矩阵、相关系数特征值、相关系数特征向量、累计贡献率确定主成分、因子个数、综合主成分的分析评价、综合因子的分析评价。

2. 不同点

不同之处见下表 1.6。

表 1.6 主成分分析与因子分析法的不同点

主成分分析数学模型	因子分析的一种数学模型
$F_i = a_{i1}x_1 + \cdots + a_{pi}x_i = a_i^T x, i = 1, \cdots, m$	$x_j = b_{j1}F_1 + b_{j2}F_2 + \cdots + b_{jm}F_m + \varepsilon_j$ $j = 1, 2, 3, \cdots, p$
$A = \left(a_{ij}\right)_{p\times m} = \left[a_1, a_2, \cdots, a_m\right], Ra_i = \lambda a_i$ R 为相关系数矩阵，λ, a_i 表示相应的特征值和单位特征值，$\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p \ge 0$	因子载荷矩阵 $B = \left(b_{ij}\right)_{p\times m} = \hat{B}C$ $, \hat{B} = \left[\sqrt{\lambda_1}a_1, \cdots, \sqrt{\lambda_m}a_m\right]$ 为初等因子载荷矩阵，C 为正交旋转矩阵
$A^I A = I$， A 为正交矩阵	$B^I B \ne I$， B 为正交矩阵

用 A 的第 i 列绝对值大的对应变量对 F_i 进行命名	将 B 的第 j 列绝对值大的对应变量归为 F_j 一类并由此对其命名
$\lambda_1,\lambda_2,\cdots,\lambda_m$ 互不相同，a_{ij} 唯一	相关系数 $r_{X_iF_j}=b_{ij}$ 不唯一
协方差 $Cov(F_i,F_j)=\lambda_i\delta_{ij},\delta_{ij}=\begin{cases}1,i\neq j,\\0,i=j\end{cases}$	协方差 $Cov(F_i,F_j)=\lambda_i\delta_{ij},\delta_{ij}=\begin{cases}1,i\neq j,\\0,i=j\end{cases}$
λ_i（特征值）表示主成分 F_i 的特征值	$S_i=\sum_{k=1}^{p}b_{ki}^2$ 为 F_i 对 x 的贡献未必等于 λ_i
主成分函数 $[F_1,F_2,\cdots,F_m]^T=A^Tx$	因子得分函数 $[F_1,F_2,\cdots,F_m]=xR^{-1}B$
主成分 F_i 中 x 的系数平方和 $\sum_{k=1}^{p}a_{ki}^2=1$，无特殊因子	$\sum_{k=1}^{p}b_{ki}^2+\sigma_j^2=h_j^2+\sigma_j^2=1$ 其中 h_j^2 表示共和度，σ_j^2 表示特殊方差
综合主成分函数 $F=\sum_{\lambda=1}^{m}(\lambda_i/p)F_i$，其中 $p=\sum_{\lambda=1}^{m}\lambda_i$	综合因子得分函数 $F=\sum_{\lambda=1}^{m}(S_i/p)F_i$，其中 $p=\sum_{\lambda=1}^{m}S_i$

1.6 人体生化检验指标因子分析

1.6.1 问题重述

表 1.7 资料使 25 名健康人的 7 项生化检验结果，7 项生化检验指标依次为 $x_1,x_2,\cdots,x_7$，请对该资料进行因子分析。

表 1.7 检验数据

x1	x2	x3	x4	x5	x6	x7
3.76	3.66	0.54	5.28	9.77	13.74	4.78

8.59	4.99	1.34	10.02	7.5	10.16	2.13
6.22	6.14	4.52	9.84	2.17	2.73	1.09
7.57	7.28	7.07	12.66	1.79	2.1	0.82
9.03	7.08	2.59	11.76	4.54	6.22	1.28
5.51	3.98	1.3	6.92	5.33	7.3	2.4
3.27	0.62	0.44	3.36	7.63	8.84	8.39
8.74	7	3.31	11.68	3.53	4.76	1.12
9.64	9.49	1.03	13.57	13.13	18.52	2.35
9.73	1.33	1	9.87	9.87	11.06	3.7
8.59	2.98	1.17	9.17	7.85	9.91	2.62
7.12	5.49	3.68	9.72	2.64	3.43	1.19
4.69	3.01	2.17	5.98	2.76	3.55	2.01
5.51	1.34	1.27	5.81	4.57	5.38	3.43
1.66	1.61	1.57	2.8	1.78	2.09	3.72
5.9	5.76	1.55	8.84	5.4	7.5	1.97
9.84	9.27	1.51	13.6	9.02	12.67	1.75
8.39	4.92	2.54	10.05	3.96	5.24	1.43
4.94	4.38	1.03	6.68	6.49	9.06	2.81
7.23	2.3	1.77	7.79	4.39	5.37	2.27
9.46	7.31	1.04	12	11.58	16.18	2.42
9.55	5.35	4.25	11.74	2.77	3.51	1.05
4.94	4.52	4.5	8.07	1.79	2.1	1.29
8.21	3.08	2.42	9.1	3.75	4.66	1.72
9.41	6.44	5.11	12.5	2.45	3.1	0.91

1.6.2 符号规定与基本假设

1. 符号规定

1. $x_1, x_2, x_3, \cdots, x_7$ 分别表示 7 个评价指标；

2. a_{ij} 表示第 $i\left(i=1,2,3,4,\cdots,25\right)$ 第 j 个 $\left(j=1,2,3,4,5,6,7\right)$ 指标变量的取值；

3. μ_j 表示第 j 个指标变量的样本均值；；

4. s_j 表示第 j 个指标变量的样本标准差；

2. 基本假设

1. 假设原始数据不存在偏差；
2. 假设原始数据经过完全统计；
3. 假设原始数据变量之间不存在联系。

1.6.3 模型的建立与分析

1. 首先，对原始数据进行数据标准化。将各项指标值 a_{ij} 转换成 $\tilde{a}_{ij}$，即

$$\tilde{a}_{ij}=\frac{a_{ij}-\mu_j}{s_j},=1,2,\cdots,25;\ j=1,2,\cdots,7 \tag{1.18}$$

其中，$\mu_j=\frac{1}{25}\sum_{i=1}^{25}a_{ij}, s_j=\sqrt{\frac{1}{24}\sum_{i=1}^{25}\left(a_{ij}-\mu_j\right)^2}$ 。

标准化指标变量为

$$\tilde{x}_j=\frac{x_j-\mu_j}{s_j},\ j=1,2,3,\cdots,7 \tag{1.19}$$

2. 接着，计算相关系数矩阵 R。

相关系数矩阵为 $R=\left(r_{ij}\right)_{7\times 7}$

$$r_{ij}=\frac{\sum_{k=1}^{25}\tilde{a}_{ki}\cdot\tilde{a}_{kj}}{25-1},i,\ j=1,2,3,\cdots,7$$

其中，$r_{ii}=1, r_{ij}=r_{ji}$，r_{ji} 指的是第 i 个指标与第 j 个指标的相关系数。

3. 然后，计算初等载荷矩阵。计算相关系数矩阵 R 特征值

$\lambda_1\ge\lambda_2\ge\lambda_3\ge\lambda_4\ge\lambda_5\ge\lambda_6\ge\lambda_7\ge 0$，及相应的标准化特征向量 $u_1,u_2,\cdots,u_7$

其中，$\mathbf{u}_j=\left(u_{1j},u_{2j},u_{3j},\cdots,u_{7j}\right)^T$，则初等载荷矩阵为

$$\Lambda=\left[\sqrt{\lambda_1}u_1,\sqrt{\lambda_2}u_2,\cdots,\sqrt{\lambda_7}u_7\right]$$

可得，特征根与各因子的贡献如表 1.8 所示

表 1.8 特征根与特征因子的贡献

特征值	34.496	18.983	2.506	0.7988	3413	0.0379	0.0042
贡献率	60.316	33.192	4.4248	1.3968	0.5968	0.0663	0.0074
累计贡献率	60.316	93.508	97.933	99.29	99.926	99.993	100

4. 最后，选择 m 个主因子。由上表可知，选择前三个主因子。对提取的因子载荷矩阵进行旋转，进而构造因子模型

$$\begin{cases}\hat{x}_1=\alpha_{11}\hat{F}_1+\alpha_{12}\hat{F}_2+\alpha_{13}\hat{F}_3,\\ \vdots\\ \hat{x}_7=\alpha_{71}\hat{F}_1+\alpha_{72}\hat{F}_2+\alpha_{73}\hat{F}_3,\end{cases}$$

5. 对上述构造模型进行求解。

1.6.4 模型的求解

将上述模型进行 MATLAB 软件编写程序：

```
clc,clear
dd=load('data803.txt');
sy=zscore(dd);
r=corrcoef(sy);
[vec1,val,con]=pcacov(r)
f1=repmat(sign(sum(vec1)),size(vec1,1),1);
vec2=vec1.*f1;
f2=repmat(sqrt(val)',size(vec2,1),1);
a=vec2.*f2
num=input('请选择主因子个数:');
am=a(:,[1:num]);
[b,t]=rotatefactors(am,'method','varimax')
bt=[b,a(:,[num+1:end])];
degree=sum(b.^2,2)
contr=sum(bt.^2)
rate=contr(1:num)/sum(contr)
coef=inv(r)*b
```

运行结果为：

```
vec1 =
    0.4051    0.2921   -0.6699    0.0988   -0.2219   -0.4378    0.2261
    0.4322    0.2222    0.6960   -0.0366   -0.4326   -0.1813    0.2409
    0.3847   -0.3565    0.1506    0.6287    0.5142   -0.1902    0.0792
    0.4942    0.2320   -0.1120    0.2039   -0.1143    0.6478   -0.4638
   -0.1271    0.5752    0.0292    0.1116    0.3894    0.3588    0.5995
   -0.0961    0.5801    0.1737   -0.0071    0.3560   -0.4361   -0.5542
   -0.4811    0.1309    0.0242    0.7345   -0.4558   -0.0259   -0.0538
val =
    3.3952
    2.8063
    0.4365
    0.2762
    0.0812
    0.0042
    0.0004
con =
   48.5026
   40.0903
    6.2355
    3.9463
    1.1599
    0.0595
    0.0058
a =
    0.7465    0.4893   -0.4426    0.0519   -0.0632    0.0282    0.0046
    0.7964    0.3722    0.4598   -0.0192   -0.1233    0.0117    0.0049
    0.7089   -0.5973    0.0995    0.3305    0.1465    0.0123    0.0016
    0.9105    0.3887   -0.0740    0.1072   -0.0326   -0.0418   -0.0094
   -0.2342    0.9635    0.0193    0.0586    0.1110   -0.0231    0.0121
   -0.1771    0.9717    0.1148   -0.0038    0.1014    0.0281   -0.0112
   -0.8864    0.2192    0.0160    0.3860   -0.1299    0.0017   -0.0011
请输入主因子的个数：3
b =
    0.9642    0.1354    0.2110
    0.3858    0.0643    0.9117
    0.2303   -0.8169    0.3858
    0.8102    0.0024    0.5737
    0.1466    0.9785    0.0673
    0.1283    0.9700    0.1770
   -0.5536    0.5444   -0.4809
t =
    0.6956   -0.3853    0.6064
    0.3340    0.9207    0.2019
```

```
    -0.6361    0.0621    0.7691
degree =
     0.9925
     0.9843
     0.8692
     0.9856
     0.9836
     0.9888
     0.8341
contr =
     2.1324    2.8846    1.6209    0.2762    0.0812    0.0042    0.0004
rate =
     0.3046    0.4121    0.2316
coef =
     0.8561    0.0129   -0.6114
    -0.4626    0.0971    0.9792
    -0.0709   -0.2622    0.2590
     0.3406    0.0137    0.0602
     0.0386    0.3454    0.0615
    -0.0879    0.3552    0.2405
    -0.1788    0.1748   -0.1144
```

1.6.5 结果分析

经过上述 MATLAB 程序运算，可得因子载荷等估计如下表 1.9 所示。

不难发现可得三个因子，第一个因子是 x_1，第二个因子是 x_5 因子，第三个因子是 x_2 因子。

表 1.9 因子载荷等估计表

变量	旋转因子载荷估计			旋转后得分函数			共同度
	$\hat{F}_1$	$\hat{F}_2$	$\hat{F}_3$	因子 1	因子 2	因子 3	
1	0.9642	0.1354	0.2110	0.8561	0.0129	-0.6114	0.9925
2	0.3858	0.0643	0.9117	-0.4626	0.0971	0.9792	0.9843
3	0.2303	-0.8169	0.3858	-0.0709	-0.2622	0.2590	0.8692
4	0.8102	0.0024	0.5737	0.3406	0.0137	0.0602	0.9856
5	0.1466	0.9785	0.0673	0.0386	0.3454	0.0615	0.9836
6	0.1283	0.9700	0.1770	-0.0879	0.3552	0.2405	0.9888
7	-0.5536	0.5444	-0.4809	-0.1788	0.1748	-0.1144	0.8341
可解释方差	0.3046	0.4121	0.2316				