

法律声明

□ 本课件包括：演示文稿，示例，代码，题库，视频和声音等，小象学院拥有完全知识产权的权利；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或机构不得盗版、复制、仿造其中的创意，我们将保留一切通过法律手段追究违反者的权利。

□ 课程详情请咨询

■ 微信公众号：小象

■ 新浪微博：ChinaHadoop



第六章 因子分析

主讲教师 周仕君

第六章 因子分析

- 第一节 因子分析模型
- 第二节 因子载荷矩阵求解
- 第三节 公因子重要性的分析

为了对因子分析的基本理论有一个完整的认识，我们先给出查尔斯·斯皮尔曼 1904 年用到的例子。斯皮尔曼在该例中研究了 33 名学生在古典语 (C)、法语 (F)、英语 (E)、数学 (M)、判别 (D) 和音乐 (Mu) 六门考试成绩之间的相关性，并得到如下相关矩阵：

	C	F	E	M	D	Mu
C	1.00	0.83	0.78	0.70	0.66	0.63
F	0.83	1.00	0.67	0.67	0.65	0.57
E	0.78	0.67	1.00	0.64	0.54	0.51
M	0.70	0.67	0.64	1.00	0.45	0.51
D	0.66	0.65	0.54	0.45	1.00	0.40
Mu	0.63	0.57	0.51	0.51	0.40	1.00

斯皮尔曼注意到上面相关矩阵中一个有趣的规律，即如果不考虑对角元素的话，任意两列的元素大致成比例，对 C 列和 E 列有

$$\frac{0.83}{0.67} \approx \frac{0.70}{0.64} \approx \frac{0.66}{0.54} \approx \frac{0.63}{0.51} \approx 1.2$$



于是斯皮尔曼指出每一科目的考试成绩都遵从以下形式：

$$X_i = a_i F + e_i \quad (6.1)$$

式中， X_i 为第 i 门科目标准化后的考试成绩，均值为 0，方差为 1； F 为公共因子，对各科考试成绩均有影响，也是均值为 0，方差为 1； e_i 为仅对第 i 门科目考试成绩有影响的特殊因子， F 与 e_i 相互独立。也就是说，每一门科目的考试成绩都可以看做是由一个公共因子（可以认为是一般智力）与一个特殊因子的和。在满足以上假定的条件下，就有

$$\text{cov}(X_i, X_j) = E(a_i F + e_i)(a_j F + e_j) = a_i a_j \text{var} F = a_i a_j$$

于是，有

$$\frac{\text{cov}(X_i, X_j)}{\text{cov}(X_i, X_k)} = \frac{a_j}{a_k} \quad (6.2)$$



- ❑ 因子分析(factor analysis)是研究观测变量变动的共同原因和特殊原因，从而达到简化变量结构的目的的多元统计分析方法。
- ❑ 因子分析的主要应用有两个方面，一寻求变量的基本结构，简化变量系统；二是用于分类，根据因子得分值，在因子轴所构成的空间中将变量或者样本进行分类（能够分析样品间差异的原因）。
- ❑ 常用的因子分析类型是R型因子分析和Q型因子分析。



第一节 因子分析模型

- 因子分析的数学模型
- 因子载荷阵的统计含义

一、因子分析的数学模型

(一) R型因子分析模型

□ R因子分析是将每一个变量都表示成公共因子的线性函数与特殊因子之和，即

$$X_i = a_{i1}F_1 + a_{i2}F_2 + \cdots + a_{im}F_m + \varepsilon_i, (i = 1, 2, \cdots, p) \quad (7.1)$$

□ (7.1) 式中的 $F_1, F_2, \cdots, F_m$ 称为公共因子, ε_i 称为 X_i 的特殊因子。该模型可用矩阵表示为：

$$X = AF + \varepsilon \quad (7.2)$$



用矩阵表示：

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix} + \underset{p \times m}{\mathbf{A}} \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix}$$

公共因子或潜在因子。该因子模型看似多元线性模型，但要注意公共因子是不可观测变量且只存在于某种理论意义之中。可理解为在高维空间中的互相垂直的m个坐标轴。



- 例如：在医学科学研究中，很多情况下要研究的变量是不能或不易直接测量得到的。
- 如，要研究家庭环境(x)、社会环境(y)和学校环境(z)对儿童智商(IQ)的发育影响问题，这四个变量都是不能或不易直接观测得到的。称这种不能或不易直接观测得到的变量为潜在变量(latent variable)或潜在因子(latent factor)，即公共因子。



- 虽然潜在变量不能直接测得，但是它一定与某些可测变量有着某种程度的关联。例如，儿童的智商这个潜在变量一定影响其在学校所取得的数学、语文、绘画、音乐、体育、自然常识等各科的成绩，因此，可以利用这些有关的指标去描述智商这个假设的理论变量。
- 我们更感兴趣于理论变量之间的关联性研究，而不是单一指标之间的关联性研究，因为前者包含了更多的综合信息。比如，研究家庭环境对孩子的智力的影响显然比仅研究孩子父亲的文化程度对孩子的数学成绩的影响要全面、深入，而且有意义得多。



- 例：学生高考成绩，要考语文、数学、物理、化学、生物、历史、地理、外语等科目。从教育学上来看，学生各科成绩的好坏主要是各个学生的阅读能力、抽象能力、记忆能力和学习刻苦努力程度支配的。因此，学生各科成绩这些指标的变动就取决于上述四个共同因素和其它一些特殊因素的作用。
- 再例：对国民经济系统运行的观测指标有国内生产总值、国民收入、社会商品零售额、库存额、物价指数、货币发行量、失业率、固定资产投资率、城乡居民储蓄存款额、固定资产总额、财政收入、财政支出、进出口总额、国际收支余额等。这些指标的数值大小取决于该国的社会经济制度、银行制度、市场机制、人口数量、科学技术水平等因素。



用矩阵表示：

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix} + \underset{p \times m}{\mathbf{A}} \begin{bmatrix} F_1 \\ F_2 \\ \vdots \end{bmatrix}$$

这里：

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ a_{p1} & a_{p2} & \cdots & a_{pm} \end{bmatrix} = (A_1, A_2, \cdots, A_m)$$

 模型中的 a_{ij} 称为因子“载荷”，是第 i 个变量在第 j 个因子上的负荷，如果把变量 X_i 看成 m 维空间中的一个点，则 a_{ij} 表示它在坐标轴 F_j 上的投影，因此矩阵 A 称为因子载荷矩阵。



构造模型时满足：

(1) $m \leq p$;

(2) $\text{Cov}(F, \varepsilon) = 0$, 即公共因子与特殊因子是不相关的;

(3) $D_F = D(F) = \begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{bmatrix} = I_m$, 即各个公共因子不相关
且方差为1;



$$(4) \mathbf{D}_\varepsilon = D(\varepsilon) = \begin{bmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_p^2 \end{bmatrix}, \text{ 即各个特殊因子不相关,}$$

方差不要求相等。



（二）Q型因子分析（因子得分）

□ 类似地，Q型因子分析的数学模型可表示为：

$$X_i = a_{i1}F_1 + a_{i2}F_2 + \cdots + a_{im}F_m + \varepsilon_i, (i=1, 2, \cdots, n) \quad (7.3)$$

□ Q型因子分析与R型因子分析模型的差异体现在， $X_1, X_2, \cdots, X_n$ 表示的是 **n 个样品**。



主成份分析与因子分析的异同

相同点：

- 无论是R型或Q型因子分析，都用公共因子 F 代替 X ，一般要求 $m < p$ ， $m < n$ ，因此，因子分析与主成分分析一样，也是一种降低变量维数的方法。
- 因子分析的求解过程同主成分分析类似，也是从一个协方差阵出发的。



因子分析与主成份分析的区别

- 主成份分析的数学模型本质上是一种线性变换，是将原始坐标变换到变异程度大的方向上去，相当于从空间上转换观看数据的角度，突出数据变异的方向，归纳重要信息。
- 而因子分析从本质上看是从显在变量去“提炼”潜在因子的过程。



-
- 正因为因子分析是一个提炼潜在因子的过程，因子的个数 m 取多大是要通过一定规则确定的，并且因子的形式也不是唯一确定的。
 - 一般说来，作为“自变量”的因子 F_1 ， F_2 ，...， F_m 是不可直接观测的。



因子载荷阵不唯一的原因:

□ 第一，变量 X 的协差阵 Σ 的分解式为

$$\begin{aligned} D(X) &= D(AF + \varepsilon) = E[(AF + \varepsilon)(AF + \varepsilon)'] \\ &= AE(FF')A' + AE(F\varepsilon') + E(\varepsilon F')A' + E(\varepsilon\varepsilon') \\ &= AD(F)A' + D(\varepsilon) \end{aligned}$$

□ 由模型 (7.2) 式所满足的条件知

$$\Sigma = AA' + D_{\varepsilon} \quad (7.4)$$

□ 如果 X 为标准化的随机向量，则 Σ 就是相关矩阵

$$R = (\rho_{ij}), \quad \text{即 } R = AA' + D_{\varepsilon}$$



□ 第二，对于 $m \times m$ 的正交矩阵 T ，令 $A^* = AT, F^* = T'F$

模型可以表示为： $X = A^*F^* + \varepsilon$

□ 由于 $\text{Cov}(F^*, \varepsilon) = E(F^* \varepsilon') = T'E(F \varepsilon') = 0$

$$D(F^*) = T'D(F)T = T'T = I_{m \times m}$$

所以仍然满足模型的条件。

□ 同样 Σ 也可以分解为 $\Sigma = A^*A^{*'} + D_\varepsilon$

□ 因此，因子载荷矩阵 A 不是唯一的。



二、因子载荷阵的统计意义

1. 因子载荷 a_{ij} 的统计意义

□ 对于因子模型

$$X_i = a_{i1}F_1 + a_{i2}F_2 + \cdots + a_{ij}F_j + \cdots + a_{im}F_m + \varepsilon_i \\ (i = 1, 2, \cdots, p)$$

□ 可以得到, X_i 与 F_j 的协方差为:

$$\begin{aligned} \text{Cov}(X_i, F_j) &= \text{Cov}\left(\sum_{k=1}^m a_{ik}F_k + \varepsilon_i, F_j\right) \\ &= \text{Cov}\left(\sum_{k=1}^m a_{ik}F_k, F_j\right) + \text{Cov}(\varepsilon_i, F_j) = a_{ij} \end{aligned}$$



-
- 如果对 X_i 作了标准化处理, X_i 的标准差为1, 且 F_j 的标准差为1, 有

$$r_{X_i, F_j} = \frac{\text{Cov}(X_i, F_j)}{\sqrt{D(X_i)}\sqrt{D(F_j)}} = \text{Cov}(X_i, F_j) = a_{ij} \quad (7.6)$$

- 那么, 对于标准化后的 X_i , a_{ij} 是 X_i 与 F_j 的相关系数, 表示 X_i 依赖 F_j 的分量(比重)。因此统计学上也称其为权。
- 由于历史的原因, 心理学家将它叫做载荷, 表示第 i 个变量在第 j 个公共因子上的负荷, 它反映了第 i 个变量在第 j 个公共因子上的相对重要性。



2. 变量共同度 的统计意义

□ 由因子模型，知

$$\begin{aligned} D(X_i) &= a_{i1}^2 D(F_1) + a_{i2}^2 D(F_2) + \cdots + a_{im}^2 D(F_m) + D(\varepsilon_i) \\ &= a_{i1}^2 + a_{i2}^2 + \cdots + a_{im}^2 + D(\varepsilon_i) \\ &= h_i^2 + \sigma_i^2 \end{aligned} \quad (7.8)$$

□ 变量 X_i 的共同度为： $h_i^2 = \sum_{j=1}^m a_{ij}^2 \quad i = 1, 2, \cdots, p$

📖 如果对 X_i 作了标准化处理，有

$$1 = h_i^2 + \sigma_i^2 \quad (7.9)$$



3、公因子 F_j 的方差贡献 g_j^2 的统计意义

□ 设因子载荷矩阵为 A ，称第 j 列元素的平方和，即：

$$g_j^2 = \sum_{i=1}^p a_{ij}^2 \quad j=1, 2, \dots, m$$

为公共因子 F_j 对 X 的贡献，即 g_j^2 表示同一公共因子 F_j 对各变量所提供的方差贡献之总和，反映该因子对全部变量的总影响。

□ 它是衡量每一个公共因子相对重要性的一个尺度。公共因子 F_j 的相对重要性可用 g_j^2/h 来衡量，它称为公共因子 F_j 对 X 的贡献率。



-
- 因子分析的目的就是要由原始随机向量的协差阵 Σ 或相关阵 R 求出因子分析模型的解，即求出载荷阵 A 和特性方差阵 D_ϵ ，并给公因子赋予有实际背景的解释。



第二节 因子载荷阵求解

- 因子载荷矩阵的求解
- 约相关阵的估计



一、因子载荷矩阵的求解-主因子法

- 实际应用中建立因子分析的具体模型，关键是根据样本数据估计载荷矩阵 A 。“主轴因子法”是一种常用的估计的方法。

$$\begin{aligned} D(X) &= D(AF + \varepsilon) = E[(AF + \varepsilon)(AF + \varepsilon)'] \\ &= AE(FF')A' + AE(F\varepsilon') + E(\varepsilon F')A' + E(\varepsilon\varepsilon') \\ &= AD(F)A' + D(\varepsilon) \end{aligned}$$

- 假定原始向量 $X = (X_1, X_2, \dots, X_p)'$ 已作了标准化变换。如果随机向量 X 满足因子模型 (7.2) 式，已知 X 的相关阵为 R ，由 (7.5) 式知

$$R = AA' + D_{\varepsilon}$$



□ 令 $\mathbf{R}^* = \mathbf{R} - \mathbf{D}_\varepsilon = \mathbf{A}\mathbf{A}'$

则称 $\mathbf{R}^*$ 为 $\mathbf{X}$ 的约相关阵。 $\mathbf{R}^*$ 中的主对角线的元素是 h_i^2 而不是1，非对角线的元素和 $\mathbf{R}$ 中的完全一样，并且 $\mathbf{R}^*$ 是一个非负定矩阵。

□ 这里记 $\mathbf{R}^* = (r_{ij}^*)_{p \times p}$ ，那么

$$r_{ij}^* = \sum_{k=1}^m a_{ik} a_{jk} = \begin{cases} \sigma_{ij} & i \neq j \\ \sigma_{ii} - \sigma_i^2 & i = j \end{cases} \quad (7.10)$$



反映了该公共因子对全部变量的总方差所作的贡献，即反映了该公共因子对全部变量的总影响程度。

□ A 的解是不唯一的，可以有許多。这种方法要求得到的解使第一公共因子 F_1 对 X 的贡献 $g_1^2 = \sum_{i=1}^p a_{i1}^2$ 达到最大，第二公共因子 F_2 对 X 的贡献 $g_2^2 = \sum_{i=1}^p a_{i2}^2$ 达到次之，……，第 m 个公共因子 F_m 对 X 的贡献最小。即相应的“贡献”依次 $g_1^2 \geq g_2^2 \geq \cdots \geq g_m^2$ 。

。



□ 现在求 $A_1 = (a_{11}, a_{21}, \dots, a_{p1})'$ 向量，在条件

$r_{ij}^* = \sum_{k=1}^m a_{ik} a_{jk}$, $i, j = 1, 2, \dots, p$ 下，使得 $g_1^2 = \sum_{i=1}^p a_{i1}^2$ 达到最大值。

□ 这是一个条件极值问题，在此构造目标函数为

$$\varphi(a_{11}, a_{21}, \dots, a_{p1}) = \frac{1}{2} g_1^2 - \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p \lambda_{ij} \left(\sum_{k=1}^m a_{ik} a_{jk} - r_{ij}^* \right)$$

其中 R^* 是拉氏系数，由于 λ_{ij} 是对称阵，所以 $\lambda_{ij} = \lambda_{ji}$



$$\square \text{ 于是有: } \begin{cases} \frac{\partial \varphi}{\partial a_{i1}} = a_{i1} - \sum_{j=1}^p \lambda_{ij} a_{j1} = 0, & i = 1, 2, \dots, p \quad \textcircled{1} \\ \frac{\partial \varphi}{\partial a_{it}} = -\sum_{j=1}^p \lambda_{ij} a_{jt} = 0, & t = 2, \dots, k \quad \textcircled{2} \end{cases}$$

$\square$ 两式合并，得到

$$\sum_{j=1}^p \lambda_{ij} a_{jt} - \delta_{1t} a_{i1} = 0, \quad i = 1, 2, \dots, p; t = 1, 2, \dots, m \quad (7.12)$$

$$\square \text{ 其中: } \delta_{1t} = \begin{cases} 1 & t = 1 \\ 0 & t \neq 1 \end{cases}$$



$$\sum_{j=1}^p \lambda_{ij} a_{jt} - \delta_{1t} a_{i1} = 0$$

□ 用 a_{i1} 乘 (7.12) 式, 并对 i 求和, 得

$$\sum_{j=1}^p (\sum_{i=1}^p \lambda_{ij} a_{i1}) a_{jt} - \delta_{1t} \sum_{i=1}^p a_{i1}^2 = 0, t = 1, 2, \dots, m$$

□ 应该注意到,

$\lambda_{ij} = \lambda_{ji}$, 由①式得:

$$a_{i1} - \sum_{j=1}^p \lambda_{ij} a_{j1} = 0$$

$$g_1^2 = \sum_{i=1}^p a_{i1}^2, \sum_{j=1}^p \lambda_{ji} a_{j1} = a_{i1} \Rightarrow \sum_{i=1}^p \lambda_{ij} a_{i1} = a_{j1}$$

□ 即有

$$\sum_{j=1}^p a_{j1} a_{jt} - \delta_{1t} g_1^2 = 0, t = 1, 2, \dots, m \quad (7.13)$$



$$\sum_{j=1}^p a_{j1} a_{jt} - \delta_{1t} g_1^2 = 0$$

□ 用 a_{it} 乘 (7.13) 式, 并对 t 求和, 得

$$\sum_{j=1}^p a_{j1} \left(\sum_{t=1}^m a_{jt} a_{it} \right) - \sum_{t=1}^m \delta_{1t} a_{it} g_1^2 = 0, i = 1, 2, \dots, p$$

□ 由于 $r_{ij}^* = \sum_{t=1}^m a_{it} a_{jt}$, $\because \delta_{1t} = \begin{cases} 1 & t = 1 \\ 0 & t \neq 1 \end{cases}$ 那么

$$\sum_{j=1}^p r_{ij}^* a_{j1} = a_{i1} g_1^2, i = 1, 2, \dots, p$$

□ 用向量表示为 $(r_{i1}^*, r_{i2}^*, \dots, r_{ip}^*) \begin{pmatrix} a_{11} \\ \vdots \\ a_{p1} \end{pmatrix} = a_{i1} g_1^2, i = 1, 2, \dots, p$



□ 则有 $(\mathbf{R}^* - \mathbf{I}g_1^2)\mathbf{A}_1 = 0$

因此, g_1^2 是约相关阵 $\mathbf{R}^*$ 的最大特征根, $\mathbf{A}_1$ 是相应于 g_1^2 的特征向量

□ 如果记约相关阵 $\mathbf{R}^*$ 的最大特征根为 λ_1^* , 相应的单位特征向量为 $\mathbf{t}_1^*$ 。考虑约束条件 $g_1^2 = \sum_{i=1}^p a_{i1}^2 = \mathbf{A}_1' \mathbf{A}_1 = \lambda_1^*$

且 $\mathbf{t}_1^{*'} \mathbf{t}_1^* = 1$, 则 $\mathbf{A}_1$ 应取为 $\mathbf{A}_1 = \sqrt{\lambda_1^*} \mathbf{t}_1^*$ 。

□ 显然, $\mathbf{A}_1$ 仍然是相应于 λ_1^* 的一个特征向量, 且满足

$$\mathbf{A}_1' \mathbf{A}_1 = \lambda_1^* \mathbf{t}_1^{*'} \mathbf{t}_1^* = \lambda_1^* = g_1^2$$

这样就得到A阵中的第一列 $\mathbf{A}_1$



□ 为了求得载荷矩阵 A 中其余 $m-1$ 列，应该注意
到约相关阵 R^* 的谱分解式为

$$R^* = AA' = (A_1, \dots, A_m) \begin{pmatrix} A'_1 \\ \vdots \\ A'_m \end{pmatrix} = \sum_{t=1}^m A_t A'_t$$

$$R^* = T^* \begin{bmatrix} \lambda_1^* & & 0 \\ & \ddots & \\ 0 & & \lambda_p^* \end{bmatrix} T^{*'} = \left(\sqrt{\lambda_1^*} t_1^*, \dots, \sqrt{\lambda_p^*} t_p^* \right) \begin{bmatrix} \sqrt{\lambda_1^*} t_1^{*'} \\ \vdots \\ \sqrt{\lambda_p^*} t_p^{*'} \end{bmatrix} = AA'$$

$$A'_1 A_1 = \lambda_1^* t_1^{*'} t_1^* = \lambda_1^* = g_1^2$$



□ 约相关阵 R^* 还可以分解为（谱分解）：

$$R^* = \sum_{i=1}^p \lambda_i^* \mathbf{t}_i^* \mathbf{t}_i^{*'} = A_1 A_1' + \sum_{i=2}^p \lambda_i^* \mathbf{t}_i^* \mathbf{t}_i^{*'}$$

□ 因此，求出 A_1 后，将 R^* 减去 $A_1 A_1'$ ，就得

$$R^* - A_1 A_1' = \sum_{t=2}^m A_t A_t'$$



□ 对于矩阵 $R^* - A_1 A_1'$ 重复上面的讨论必然有：

$$\text{即有} \left[g_2^2 I - (R^* - A_1 A_1') \right] A_2 = 0$$

$$(g_2^2 I - R^*) A_2 + A_1 A_1' A_2 = 0$$



□ 因此，可以看出，需取

$g_2^2 = \lambda_2^*, A_2 = \sqrt{\lambda_2^*} t_2^*$ ，即 g_2^2 是约相关阵 R^* 的第二

大特征根 λ_2^* ， A_2 是相应于 λ_2^* 且满足

$g_2^2 = \sum_{i=1}^p a_{i2}^2 = A_2' A_2 = \lambda_2^*$ 的特征向量。依此类推
，可以求得 $g_t^2 = \lambda_t^* \quad A_t = \sqrt{\lambda_t^*} t_t^* \quad t=1, 2, \dots, m$

□ 其中， λ_t^* 约相关阵 R^* 的第 t 大特征根， t_t^* 为相应的单位特征向量。



□ 这样就求得载荷矩阵为：

$$\mathbf{A} = (\sqrt{\lambda_1^*} \mathbf{t}_1^* \quad \sqrt{\lambda_2^*} \mathbf{t}_2^*, \quad \dots, \quad \sqrt{\lambda_m^*} \mathbf{t}_m^*)$$

$$= (\mathbf{t}_1^* \quad \mathbf{t}_2^*, \quad \dots, \quad \mathbf{t}_m^*) \begin{bmatrix} \sqrt{\lambda_1^*} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_m^*} \end{bmatrix}$$

□ 这样在模型上就解决了从约相关阵 $\mathbf{R}^*$ 出发求载荷矩阵 $\mathbf{A}$ 。



二、约相关阵的估计

- 上面的分析是以首先得到约相关阵 R^* 为基础的，在实际应用中，相关矩阵 R 和个性方差矩阵 D_ε 一般是未知的。 R 可以用样本相关阵来估计，而对 D_ε 的估计则较困难。由 (7.9) 式知

$$\sigma_i^2 = 1 - h_i^2, \quad i = 1, 2, \dots, p$$

- 所以，估计个性方差 σ_i^2 等价于估计共性方差 h_i^2 。



个性方差 σ_i^2 (或共性方差 h_i^2) 的常用初始估计方法

(1) $\hat{h}_i^2$ 取为原始变量 X_i 与其它所有原始变量

$X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_p$ 的复相关系数的平方, 则

$$\hat{\sigma}_i^2 = 1 - \hat{h}_i^2。 \quad (\text{即变量 } X_i \text{ 与其余 } p-1 \text{ 个变量线性回}$$

归的判定系数)

(2) 取 $\hat{h}_i^2 = \max_{i \neq j} |r_{ij}|$, 其中 r_{ij} 为 R 中的元素, 则

$$\hat{\sigma}_i^2 = 1 - \hat{h}_i^2。 \quad (\text{即取 } X_i \text{ 与其余变量的简单相关系}$$

数的最大值)。



(3) 设 r_{ik}, r_{il} 为 $\mathbf{R}$ 的第 i 行上主对角线以外的两个最大值, 取 $\hat{h}_i^2 = \frac{r_{ik}r_{il}}{r_{kl}}$, 则 $\hat{\sigma}_i^2 = 1 - \hat{h}_i^2$ 。

(4) 取 $\hat{h}_i^2 = 1$, 则 $\hat{\sigma}_i^2 = 0$ 。这样得到的 $\hat{\mathbf{A}}$, 实际上是针对 $\mathbf{R}$ 主成分分解 (主成份法)。

□ 这些方法都要求同时给出公共因子数 m 的值。在实践中, 当 m 不能被事先给定时, 可按约相关阵 $\mathbf{R}^*$ 的特征根的大小次序抽取, 当 $\sum_{j=1}^m \lambda_j$ 与 $\sum_{i=1}^p h_i^2 = \text{tr}(\mathbf{R}^*)$

接近时终止。



□ 在进行因子分析时，因子的个数 m 应该取为多少呢？一般可以采用确定主成份个数的原则，也就是寻找一个使得 $\sum_{j=1}^m \lambda_j^* / \sum_{j=1}^p \lambda_j^*$ 达到较大百分比（如至少85%）的自然数 m 。取 $\hat{R}^*$ 的前 m 个正特征值 $\lambda_1^* \geq \lambda_2^* \geq \dots \geq \lambda_m^* > 0$ 及相应的正交单位特征向量 $t_1^*, t_2^*, \dots, t_m^*$ ，可以分解式 $\hat{R}^* = \hat{A}\hat{A}'$ 。



□ 其中 $\hat{\mathbf{A}} = (\sqrt{\hat{\lambda}_1^*} \hat{\mathbf{t}}_1^* \quad \sqrt{\hat{\lambda}_2^*} \hat{\mathbf{t}}_2^*, \dots, \sqrt{\hat{\lambda}_m^*} \hat{\mathbf{t}}_m^*) = (\hat{a}_{ij})_{p \times m}$, 这样可以得到 σ_i^2 的最终估计为

$$\hat{\sigma}_i^2 = 1 - \hat{h}_i^2 = 1 - \sum_{j=1}^m \hat{a}_{ij}^2 \quad i = 1, 2, \dots, p$$

□ 在此, 我们称 $\hat{\mathbf{A}}$ 和 $\hat{\mathbf{D}} = \text{diag}(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \dots, \hat{\sigma}_p^2)$ 为因子模型的主因子解。



例 2. 对于某地固定资产投资率 x_1 , 通货膨胀率 x_2 , 失业率 x_3 的相关矩阵:

$$R = \begin{pmatrix} 1 & \frac{1}{5} & -\frac{1}{5} \\ \frac{1}{5} & 1 & -\frac{2}{5} \\ -\frac{1}{5} & -\frac{2}{5} & 1 \end{pmatrix}$$

若用 $\hat{h}_i^2 = \max_{j \neq i} |r_{ij}|$ 来估计 h_i^2 , 则有 $\hat{h}_1^2 = \frac{1}{5}$, $\hat{h}_2^2 = \frac{2}{5}$, $\hat{h}_3^2 = \frac{2}{5}$ 。约相关矩阵为:

$$R^* = \begin{pmatrix} \frac{1}{5} & \frac{1}{5} & -\frac{1}{5} \\ \frac{1}{5} & \frac{2}{5} & -\frac{2}{5} \\ -\frac{1}{5} & -\frac{2}{5} & \frac{2}{5} \end{pmatrix} = \frac{1}{5} \begin{pmatrix} 1 & 1 & -1 \\ 1 & 2 & -2 \\ -1 & -2 & 2 \end{pmatrix}$$



R^* 的三个特征根为：

$$\lambda_1 = \frac{1}{5}\eta_1 = \frac{1}{5} \times \frac{5 + \sqrt{17}}{2} = 0.9123$$

$$\lambda_2 = \frac{1}{5}\eta_2 = \frac{1}{5} \times \frac{5 - \sqrt{17}}{2} = 0.0877$$

$$\lambda_3 = \frac{1}{5}\eta_3 = 0$$

对应着非零特征根的两个单位特征向量为：

$$\gamma_1 = (0.369, 0.657, -0.657)'$$

$$\gamma_2 = (0.929, -0.261, 0.261)'$$



由此得因子载荷阵为：

$$\begin{aligned}
 A &= (\gamma_1 \quad \gamma_2) \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{bmatrix} \\
 &= \begin{bmatrix} 0.369 & 0.929 \\ 0.657 & -0.261 \\ -0.657 & 0.261 \end{bmatrix} \begin{bmatrix} \sqrt{0.9123} & 0 \\ 0 & \sqrt{0.0877} \end{bmatrix} \\
 &= \begin{bmatrix} 0.352 & 0.275 \\ 0.628 & -0.077 \\ -0.628 & 0.077 \end{bmatrix}
 \end{aligned}$$

皇
=



三、主成份法

- 由推导可知，用主成份法得到的各个公共因子的载荷与相应的主成份的系数仅相差一个常数 $\sqrt{\lambda_j}$ 倍即因子载荷阵为：

$$A = (a^{(1)}\sqrt{\lambda_1}, a^{(2)}\sqrt{\lambda_2}, \dots, a^{(p)}\sqrt{\lambda_p})$$

- 在主成份法下，给定公共因子数目 $q < p$ ，此时要求公共因子的方差

$$h_i^2 = \frac{\sum_{j=1}^p \sum_{i=1}^m (a_i^{(j)^2} \lambda_j)}{\sum_{i=1}^p \sigma_{ii}} = \frac{\sum_{j=1}^m \lambda_j}{\sum_{i=1}^p \sigma_{ii}} \geq 85\%$$

- 主成份法的缺点是：有时误差比较大。



例 1. 假设某地固定资产投资率 x_1 , 通货膨胀率 x_2 , 失业率 x_3 的相关矩阵为:

$$R = \begin{bmatrix} 1 & \frac{1}{5} & -\frac{1}{5} \\ \frac{1}{5} & 1 & -\frac{2}{5} \\ -\frac{1}{5} & -\frac{2}{5} & 1 \end{bmatrix}$$

试用主成份解法求因子分析模型。

相关阵的特征多项式为:

$$|\lambda I - R| = \begin{vmatrix} \lambda - 1 & -\frac{1}{5} & \frac{1}{5} \\ -\frac{1}{5} & \lambda - 1 & \frac{2}{5} \\ \frac{1}{5} & \frac{2}{5} & \lambda - 1 \end{vmatrix}$$



各特征根分别为：

$$\lambda_1 = \frac{6 + \sqrt{3}}{5} = 1.55$$

$$\lambda_2 = \frac{6 - \sqrt{3}}{5} = 0.85$$

$$\lambda_3 = \frac{3}{5} = 0.6$$

分别用每个特征根求出相应的单位化特征向量为：

$$\alpha^{(1)} = (0.457, 0.629, -0.629)'$$

$$\alpha^{(2)} = (0.883, -0.331, 0.331)'$$

$$\alpha^{(3)} = (0, 0.707, 0.707)'$$

三个主成份分别为：

$$y_1 = 0.457x_1 + 0.629x_2 - 0.629x_3$$

$$y_2 = 0.883x_1 - 0.331x_2 + 0.331x_3$$

$$y_3 = 0.707x_2 + 0.707x_3$$



因子载荷矩阵为：

$$\tilde{A} = AD(\sqrt{\lambda_j}) = \begin{bmatrix} 0.457 & 0.883 & 0 \\ 0.629 & -0.331 & 0.707 \\ -0.629 & 0.331 & 0.707 \end{bmatrix}$$
$$\begin{bmatrix} \sqrt{1.55} & 0 & 0 \\ 0 & \sqrt{0.85} & 0 \\ 0 & 0 & \sqrt{0.60} \end{bmatrix} = \begin{bmatrix} 0.569 & 0.814 & 0 \\ 0.783 & -0.305 & 0.548 \\ 0.783 & 0.305 & 0.548 \end{bmatrix}$$

因子分析模型为：

$$x_1 = 0.569f_1 + 0.814f_2$$

$$x_2 = -0.783f_1 + 0.305f_2 + 0.548f_3$$

可取前两个因子 f_1 和 f_2 为公共因子，第一个因子 f_1 对 X 的贡献为 $\lambda_1 = 1.55$ ，第二个因子对 X 的贡献为 $\lambda_2 = 0.85$ 。共性方差 $h_1^2 = 1$, $h_2^2 = 0.706$, $h_3^2 = 0.706$ 。第一个因子 f_1 为物价就业因子，第二个因子 f_2 为投资因子。



第三节 公因子重要性的分析

- 因子旋转
- 因子得分



一、因子旋转

- 因子旋转方法有正交旋转和斜交旋转两类，其中，正交旋转是对公共因子作正交旋转，就是对载荷矩阵 A 作一正交变换，右乘正交矩阵 Γ (即将因子轴旋转)，使得 $A\Gamma$ 有更鲜明的实际意义。
- 即使旋转后的因子载荷阵结构简化，便于对公共因子进行解释。所谓结构简化就是使每个变量仅在一个公共因子上有较大的载荷，而在其余公共因子上的载荷比较小，至多是中等大小



- 旋转以后的公共因子向量为 $F^* = \Gamma'F$ ，它的各个分量 $F_1^*, F_2^*, \dots, F_m^*$ 也是互不相关的公共因子。
- 根据正交矩阵 Γ 的不同选取方式，将构造出不同的正交旋转的方法。常用的方法是最大方差旋转法（varimax）。



正交变换后各变量的共性方差不变

令

$$A^* = A\Gamma = (a_{ij}^*)_{p \times m}, \quad d_{ij} = a_{ij}^* / h_i, \quad \bar{d}_j = \frac{1}{p} \sum_{i=1}^p d_{ij}^2$$

□ 由于各变量 X_i 在某一个因子 F_j 上的载荷 a_{ij}^* 的平方是该因子对该变量的共性方差的贡献，而各个变量的共性方差一般又是互不相同的，若某个变量 X_i 的共性方差较大，则分配在各个因子上的载荷就大些；反之就小些。因此，为了消除各个变量的共性方差大小不同的影响，计算某一因子上的载荷的方差时，要先将各个载荷的平方除以共性方差，即类似与将其标准化，然后计算标准化后的载荷的方差。



令 $\mathbf{A}^* = \mathbf{A}\mathbf{\Gamma} = (a_{ij}^*)_{p \times m}$, $d_{ij} = a_{ij}^* / h_i$ $\bar{d}_j = \frac{1}{p} \sum_{i=1}^p d_{ij}^2$
则 $\mathbf{A}^*$ 的第 j 列元素平方的相对方差可定义为:

$$V_j = \frac{1}{p} \sum_{i=1}^p (d_{ij}^2 - \bar{d}_j)^2$$

□ 所谓最大方差旋转法就是选择正交矩阵 $\mathbf{\Gamma}$, 使得
矩阵 $\mathbf{A}^*$ 所有 m 个列元素平方的相对方差之和

$$V = V_1 + V_2 + \cdots + V_m \quad (7.16)$$

达到最大。



- 对公因子 F_l 和 F_k 进行旋转，就是对 A 的第 l 和 k 两列进行正交变换，使这两列元素平方的相对方差之和达到最大，而其余各列不变，其正交变换矩阵为

$$A \Gamma_{lk} = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & \cos \theta & & -\sin \theta & \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \\ & & \sin \theta & & \cos \theta & \\ & & & & & 1 \\ & & & & & & \ddots \\ & & & & & & & 1 \end{pmatrix} = A^* \begin{matrix} a_{il}^* \\ a_{ik}^* \\ a_{ij}^* \\ \vdots \\ a_{kj}^* \\ \vdots \\ a_{kl}^* \end{matrix} \begin{matrix} = a_{il} \cos \theta + a_{ik} \sin \theta \\ = -a_{il} \sin \theta + a_{ik} \cos \theta \\ = a_{ij} \quad (j \neq k, l) \\ \\ \\ \\ \\ \end{matrix} \quad C_m^2$$

其中 θ 是因子轴 F_k 和 F_l 的旋转角度



- 每次取两个全部配对旋转，共 C_m^2 次，算一个循环，此时因子载荷矩阵为 $A_{(1)}^* = A\Gamma_{12}^{(1)}\Gamma_{13}^{(1)}\dots\Gamma_{(m-1)m}^{(1)}$
- 然后在第一次旋转的基础上再进行第二轮循环、第三轮循环，这样依次继续下去，每次循环得到的因子载荷阵的各因子载荷总方差将会逐步增大，即有单调上升序列。因为载荷阵的元素的绝对值均不大于1，因此该方差序列有上界，即此序列一定会收敛到某一个极限。在实践中，经过若干次循环后，当载荷阵的方差改变不大时，就可停止选择。



二、因子得分

- 在因子分析模型 $X = AF + \varepsilon$ 中，如果不考虑特殊因子的影响，当 $m=p$ 且 A 可逆时，可以从每个样品的指标取值 X 计算出其在因子 F 上的相应取值： $F = A^{-1}X$ ，即该样品在因子 F 上的“得分”情况，简称为该样品的因子得分。
- 但是因子分析模在实际应用中要求 $m < p$ ，因此，不能精确计算出因子的得分情况，只能对因子得分进行估计。用来求因子得分值的常用方法有汤姆生法和巴特莱特法两种。



汤姆生法求因子得分

- 该方法假设公共因子可对 p 个原始变量作回归，
即

$$\hat{F}_j = b_{j0} + b_{j1}X_1 + \cdots + b_{jp}X_p, \quad j = 1, \cdots, m \quad (7.18)$$

- 如果 F_j, X_i 都标准化了，回归的常数项为零，
即 $b_{j0} = 0$



□ 由因子载荷的统计意义 (7.6) 式知道, 对于任意的 $i=1, \dots, p, j=1, \dots, m$ 都有

$$\begin{aligned} a_{ij} = r_{X_i, F_j} &= E(X_i F_j) = E[X_i (b_{j1} X_1 + \dots + b_{jp} X_p)] \\ &= b_{j1} E(X_i X_1) + \dots + b_{jp} E(X_i X_{jp}) = b_{j1} r_{i1} + \dots + b_{jp} r_{ip} \end{aligned}$$

□ 记 $B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1p} \\ b_{21} & b_{22} & \dots & b_{2p} \\ \dots & \dots & \dots & \dots \\ b_{m1} & b_{m2} & \dots & b_{mp} \end{bmatrix}$, 则上式可写成矩阵形式为

$$A = RB' \text{ 或 } B = A'R^{-1}$$



□ 于是

$$\hat{\mathbf{F}} = \begin{bmatrix} \hat{F}_1 \\ \vdots \\ \hat{F}_m \end{bmatrix} = \begin{bmatrix} b'_1 \mathbf{X} \\ \vdots \\ b'_m \mathbf{X} \end{bmatrix} = \mathbf{B}\mathbf{X} = \mathbf{A}'\mathbf{R}^{-1}\mathbf{X}$$

□ 即得因子得分的估算公式

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

其中 $\mathbf{R}$ 是 $\mathbf{X}$ 的相关系数矩阵。



疑问

□ 问题答疑：<http://www.xxwenda.com/>

■ 可邀请老师或者其他人回答问题

联系我们

小象学院：互联网新技术在线教育领航者

- 微信公众号：小象
- 新浪微博：ChinaHadoop

