第六章 因子分析

6.1 因子分析的引入

例子1： 影响学习成绩的因素分析 (Spearman, 1904)

33个学生6门功课成绩的相关系数矩阵

	古典语(x_1)	法语(x_2)	英语(x_3)	数学(x_4)	判别(x_5)	音乐(x_6)
古典语(x_1)	1.00	0.83	0.78	0.70	0.66	0.63
法语 (x_2)	0.83	1.00	0.67	0.67	0.65	0.57
英语 (x_3)	0.78	0.67	1.00	0.64	0.54	0.51
数学 (x_4)	0.70	0.67	0.64	1.00	0.54	0.51
判别 (x_5)	0.66	0.65	0.54	0.54	1.00	0.40
音乐 (x_6)	0.63	0.57	0.51	0.51	0.40	1.00

特征：同一行或列上两个相关系数的比值基本上相等.

$$\frac{0.67}{0.78}=0.859,\ \frac{0.67}{0.70}=0.957,\ \frac{0.65}{0.66}=0.985,\ \frac{0.57}{0.63}=0.905;$$

$$\frac{0.78}{0.83}=0.940,\ \frac{0.64}{0.67}=0.955,\ \frac{0.54}{0.65}=0.831,\ \frac{0.51}{0.57}=0.895.$$

Spearman推测：**对任意给定的第 j 列和第 k 列，若 $i\neq j,\ i\neq k$,**

$$\text{则}\ \frac{Cov(x_i,x_j)}{Cov(x_i,x_k)}\ \text{的值与}\ i\ \text{无关.}$$

Spearman进一步推测：**6门功课成绩所对应总体的协方差阵有如下结构**

$$\mathbf{a}\mathbf{a}'+diag(\sigma_1^2,\cdots,\sigma_6^2),$$

其中 $\mathbf{a}=(a_1,a_2,a_3,a_4,a_5,a_6)'$.

协差阵具体表示为

$$\begin{pmatrix} a_1^2+\sigma_1^2 & a_1a_2 & a_1a_3 & a_1a_4 & a_1a_5 & a_1a_6 \\ a_2a_1 & a_2^2+\sigma_2^2 & a_2a_3 & a_2a_4 & a_2a_5 & a_2a_6 \\ a_3a_1 & a_3a_2 & a_3^2+\sigma_3^2 & a_3a_4 & a_3a_5 & a_3a_6 \\ a_4a_1 & a_4a_2 & a_4a_3 & a_4^2+\sigma_4^2 & a_4a_5 & a_4a_6 \\ a_5a_1 & a_5a_2 & a_5a_3 & a_5a_4 & a_5^2+\sigma_5^2 & a_5a_6 \\ a_6a_1 & a_6a_2 & a_6a_3 & a_6a_4 & a_6a_5 & a_6^2+\sigma_6^2 \end{pmatrix}.$$

Spearman对6门功课成绩建立了如下模型：

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{pmatrix} f + \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \end{pmatrix},$$

其中, f 是一维的公共变量; $u_1, \cdots, u_6$ 是相互独立的随机变量, 方差分别是 $\sigma_1^2, \cdots, \sigma_6^2$; f 与 $(u_1, \cdots, u_6)$ 独立, $Var(f) = 1$.

记 $X = (x_1, \cdots, x_6)'$, $\mathbf{a} = (a_1, \cdots, a_6)'$, $U = (u_1, \cdots, u_6)'$,
则模型可写为

$$X = \mathbf{a}f + U,$$

或

$$x_i = a_i f + u_i, \quad 1 \leq i \leq 6.$$

Spearman模型的解释：

1) 每一门功课的成绩都由两部分组成.

前一部分中的 f 是对所有课程的成绩都有贡献的随机变量;

后一部分中 u_i 是仅对第 i 门课程的成绩有贡献的随机变量.

因此, 称 f 为**公共因子**, 称 $(u_1, \cdots, u_6)$ 为**特殊因子**.

2) a_i 称为第 i 门课程成绩的**因子载荷**, 理解为公共因子 f 对第 i 门课程成绩的作用程度.

3) 公共因子 f 可以理解为学生的阅读能力.

由于不同人的阅读能力有差异, 故它是一个随机变量.

阅读能力对不同课程的贡献率 a_i 不一样.

Spearman模型是最早的仅有一个公共因子的因子模型.

在多变量分析中, 某些变量间往往存在相关性.

是什么原因使变量间有关联呢? 是否存在不能直接观测到的、但影响可观测变量变化的公共因子?

因子分析法就是寻找这些公共因子的模型分析方法.
它的目的是构建若干意义较为明确的公共因子, 以它们为框架分解原变量, 以此考察原变量间的联系与区别.

Spearman模型就是在发现学生6门功课成绩的相关性后, 挖掘出影响学习成绩的公共因子, 并解释其对学习成绩影响的因子分析模型.

例子2：顾客感知因子模型

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ y_{10} \\ y_{11} \end{pmatrix} = \begin{pmatrix} b_{11} & 0 & 0 & 0 & 0 \\ b_{21} & 0 & 0 & 0 & 0 \\ b_{31} & 0 & 0 & 0 & 0 \\ 0 & b_{42} & 0 & 0 & 0 \\ 0 & b_{52} & 0 & 0 & 0 \\ 0 & 0 & b_{63} & 0 & 0 \\ 0 & 0 & b_{73} & 0 & 0 \\ 0 & 0 & b_{83} & 0 & 0 \\ 0 & 0 & 0 & b_{94} & 0 \\ 0 & 0 & 0 & 0 & b_{10,5} \\ 0 & 0 & 0 & 0 & b_{11,5} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \\ \delta_9 \\ \delta_{10} \\ \delta_{11} \end{pmatrix}.$$

$$Cov(y) = \mathbf{b}\mathbf{b}' + diag(\sigma_1^2, \cdots, \sigma_{11}^2).$$

$\eta_1, \eta_2, \eta_3, \eta_4, \eta_5$ **是潜变量**, 分别表示顾客对质量的感知、对价值的感知、顾客满意度、顾客抱怨和顾客忠诚度.

y_1, y_2, y_3 分别表示顾客对质量、质量可靠性和质量满足需求程度的评价, 它们与潜变量 η_1 (对质量感知)有关.

y_4, y_5 分别表示给定价格后顾客对质量的评价和给定质量后顾客对价格的评价, 它们与潜变量 η_2 (对价值的感知)有关.

y_6, y_7, y_8 分别表示顾客对企业的总评价、感知的质量与期望的比较, 以及感知的质量与理想中的质量的差距, 它们与潜变量 η_3 (顾客满意度)有关.

y_9 表示顾客正式或非正式的投诉行为对质量的评价和给定质量后顾客对价格的评价, 它与潜变量 η_4 (顾客抱怨)有关.

y_{10}, y_{11} 分别表示顾客重复购买的可能性, 顾客可承受的涨价幅度或促销(降价)对顾客购买的影响, 它们与潜变量 η_5 (顾客忠诚度)有关.

因子分析的目的是用有限个不可观测的隐变量来解释原始变量之间的相关关系．是一种把多个变量化为少数几个综合变量的多变量分析方法.

因子分析主要用于：

1) 减少分析变量个数;

2) 通过对变量间相关关系的探测, 将原始变量进行分类. 即可以将相关性高的变量分为一组, 并用共性因子代替该组变量.

记 $x = (x_1, \cdots, x_p)'$, $f = (f_1, \cdots, f_m)'$, $u = (u_1, \cdots, u_p)'$.

假定：

1) $f_1, \cdots, f_m$ 相互独立, 且 $Var(f_i) = 1$, $1 \le i \le m$;

2) $u_1, \cdots, u_p$, 相互独立, 且 $Var(u_i) = \sigma_i^2$, $1 \le i \le p$;

3) f 与 u 相互独立;

4) $Var(x_i) = 1$, $1 \le i \le p$.

记 $\mathbf{A} = (a_{ij})_{p \times m}$. 那么上述的模型可以一般化为

$$x_i = \sum_{j=1}^{m} a_{ij} f_j + u_i, \ 1 \le i \le p.$$

由于

$$1 = Var(x_i) = \sum_{j=1}^{m} a_{ij}^2 + \sigma_i^2,\ 1 \le i \le p,$$

记 $h_i^2 = \sum_{j=1}^{m} a_{ij}^2$, 则有 $1 = h_i^2 + \sigma_i^2,\ 1 \le i \le p.$

称 h_i^2 为**公共因子对** x_i **的贡献**, 它反映公共因子对 x_i 的影响.

h_i^2 越大则 σ_i^2 越小, 反之亦然. 说明变量 x_i 对公共因子和特殊因子的依赖此消彼长.

记 $g_j^2 = \sum_{i=1}^{p} a_{ij}^2$, 则称 g_j^2 **为公共因子** f_j **对** x **的贡献**, $1 \le j \le m$.

g_j^2 越大, 说明 f_j 对 x 的影响越大, 也就越重要.

对任意 $1 \le i \le p$, $1 \le j \le m$, 有

$$Cov(x_i,\, f_j) = \sum_{k=1}^{m} a_{ik} Cov(f_k, f_j) + Cov(u_i,\, f_j) = a_{ij} = \rho(x_i,\, f_j).$$

因此, 因子载荷矩阵 $\mathbf{A} = (a_{ij})$ 的统计意义为:

1) a_{ij} 是 x_i 与 f_j 的相关系数;

2) $\sum_{j=1}^{m} a_{ij}^2$ 是 x_i 对公共因子的依赖程度;

3) $\sum_{i=1}^{p} a_{ij}^2$ 是公共因子 f_j 对 x 的各个分量的影响总和.

6.2 正交因子模型

设 p 维随机向量 x 可以表示为

$$x = \mu + \mathbf{A}f + u, \qquad (\mathbf{M})$$

其中, μ 是 p 维常数向量, $\mathbf{A}$ 是 $p \times m$ 阶常数矩阵, $f \overset{d}{\sim} N_m(0, I_m)$,

$m < p$, $u \overset{d}{\sim} N_p(0, D)$, $D = diag(\sigma_1^2, \cdots, \sigma_p^2)$, f 与 u 相互独立.

则称模型($\mathbf{M}$)为**正交因子模型**, 并称 f 为**公共因子**, u 为**特殊因子**,

$\mathbf{A}$为**因子载荷矩阵**.

也称 x 有**因子结构**, 其协差阵为

$$Cov(x) = \Sigma = \mathbf{A}\mathbf{A}' + D.$$

注意：因子载荷矩阵并不唯一, 因为对任意 m 阶正交矩阵 T,

有

$$\begin{aligned} x &= \mu + \mathbf{A}f + u \\ &= \mu + (\mathbf{A}T)(T'f) + u \\ &= \mu + (\mathbf{A}T)f^* + u, \end{aligned}$$

其中,

$$f^* = T'f \overset{d}{\sim} N_m(0, I_m),$$

$$\begin{aligned} Cov(x) &= \mathbf{A}\mathbf{A}' + D \\ &= (\mathbf{A}T)(\mathbf{A}T)' + D. \end{aligned}$$

因子载荷矩阵的表示：

在给定 x 的相关阵 R 和对角阵 D 的条件下, 如何求解 $\mathbf{A}$?

令 $R^* = R - D = \mathbf{A}\mathbf{A}'$, 则称 R^* 为约相关阵.

易知, R^* 的对角元素为 h_i^2, $1 \leq i \leq p$, 其它元素与 R 一样, 且非负定.

若记 $R^* = (r_{ij}^*)_{p\times p}$, 则 $r_{ij}^* = \sum_{k=1}^{m} a_{ik}a_{jk}$, $1 \leq i, j \leq p$.

目标： 求解 $\mathbf{A}$ 的各列, 使得“**贡献**” $g_1^2 \geq \cdots \geq g_m^2$.

要求： 使得 $g_1^2 = \sum_{i=1}^{p} a_{i1}^2$ 达到最大值的解.

记 $\lambda_1 \geq \cdots \geq \lambda_p \geq 0$ 为 R^* 的特征根, 其对应的正则正交特征向量分别为 $\alpha_1, \cdots, \alpha_p$.

则

$$\begin{aligned} R^* &= (\alpha_1, \cdots, \alpha_p) diag(\lambda_1, \cdots, \lambda_p)(\alpha_1, \cdots, \alpha_p)' \\ &= (\alpha_1, \cdots, \alpha_p) diag(\sqrt{\lambda_1}, \cdots, \sqrt{\lambda_p}) diag(\sqrt{\lambda_1}, \cdots, \sqrt{\lambda_p})(\alpha_1, \cdots, \alpha_p)'. \end{aligned}$$

那么有

$$\mathbf{A} = (\alpha_1, \cdots, \alpha_m) diag(\sqrt{\lambda_1}, \cdots, \sqrt{\lambda_m}),$$

其中 m 是 R^* 的秩.

6.3 因子载荷矩阵的估计

6.3.1 R主成份估计法

记 $\hat{R}$ 为样本相关阵. 对 $\hat{R}$ 做谱分解有

$$\hat{R} = \sum_{i=1}^{p} \hat{\lambda}_i \hat{\alpha}_i \hat{\alpha}_i',$$

其中, $\hat{\lambda}_1 > \cdots > \hat{\lambda}_p > 0$.

先取 $\hat{a}_1 = \sqrt{\hat{\lambda}_1}\hat{\alpha}_1$, 然后看 $\hat{R} - \hat{a}_1\hat{a}_1' = \sum_{i=2}^{p} \hat{\lambda}_i \hat{\alpha}_i \hat{\alpha}_i'$ 是否接近对对角阵. 如果接近对角阵, 表明剩下的都是特殊因子的影响, 只有一个公共因子. 否则一直进行下去.

实际操作中, 给定一个阈值 δ, 若

$$\frac{\sum_{i=1}^{m}\hat{\lambda}_i}{\sum_{i=1}^{p}\hat{\lambda}_i} \geq \delta,$$

则取

$$\hat{\mathbf{A}} = (\hat{\alpha}_1, \cdots, \hat{\alpha}_m) diag(\sqrt{\hat{\lambda}_1}, \cdots, \sqrt{\hat{\lambda}_m}).$$

6.3.2 极大似然估计法

假设 $x_1, \cdots, x_n$ 是来自正交因子模型($\mathbf{M}$)的独立样本.

$rank(\mathbf{A}) = m$. 则 $(\mu, \mathbf{A}, D)$ 的似然函数为

$$L(\mu, \mathbf{A}, D) = |\mathbf{A}\mathbf{A}' + D|^{-n/2} \exp\left\{-\frac{n}{2} tr[(\mathbf{A}\mathbf{A}' + D)^{-1}(S + (\bar{x} - \mu)(\bar{x} - \mu)')]\right\},$$

其中, $\bar{x}$ 样本均值, S 是样本协差阵.

对 μ 求极大后得 $(\mathbf{A},\ D)$ 的对数似然函数为

$$l(\mathbf{A}, D) = -\frac{n}{2} \ln |\mathbf{A}\mathbf{A}' + D| - \frac{n}{2} tr\left[(\mathbf{A}\mathbf{A}' + D)^{-1} S)\right].$$

极大似然估计 $(\hat{\mathbf{A}}, \hat{D})$ 满足如下的方程组

$$\begin{cases} \frac{dl(\hat{\mathbf{A}}, \hat{D})}{d\mathbf{A}} = 0; \\ \\ \frac{dl(\hat{\mathbf{A}}, \hat{D})}{dD} = 0. \end{cases}$$

化简后可知极大似然估计满足下面的方程组

$$\begin{cases} \mathbf{A} = S(\mathbf{A}\mathbf{A}' + D)^{-1}\mathbf{A}; \\ \\ diag(S) = diag(\mathbf{A}\mathbf{A}' + D). \end{cases}$$

再化简：

$$\begin{aligned} & \mathbf{A}(I_m + \mathbf{A}'D^{-1}\mathbf{A}) = DD^{-1}\mathbf{A} + \mathbf{A}\mathbf{A}'D^{-1}\mathbf{A} = (D + \mathbf{A}\mathbf{A}')D^{-1}\mathbf{A} \\ \Rightarrow \quad & (D + \mathbf{A}\mathbf{A}')^{-1}\mathbf{A} = D^{-1}\mathbf{A}(I_m + \mathbf{A}'D^{-1}\mathbf{A})^{-1} \end{aligned}$$

再化简：

$$\mathbf{A}(I_m+\mathbf{A}'D^{-1}\mathbf{A}) = DD^{-1}\mathbf{A}+\mathbf{A}\mathbf{A}'D^{-1}\mathbf{A} = (D+\mathbf{A}\mathbf{A}')D^{-1}\mathbf{A}$$

$$\Rightarrow \quad (D+\mathbf{A}\mathbf{A}')^{-1}\mathbf{A} = D^{-1}\mathbf{A}(I_m+\mathbf{A}'D^{-1}\mathbf{A})^{-1},$$

因此有

$$\mathbf{A}=S(\mathbf{A}\mathbf{A}'+D)^{-1}\mathbf{A} \quad \Rightarrow \quad \mathbf{A}=SD^{-1}\mathbf{A}(I_m+\mathbf{A}'D^{-1}\mathbf{A})^{-1}$$

$$\Rightarrow \quad SD^{-1}\mathbf{A}=\mathbf{A}(I_m+\mathbf{A}'D^{-1}\mathbf{A}).$$

似然方程化为

$$\begin{cases} SD^{-1}\mathbf{A}=\mathbf{A}(I_m+\mathbf{A}'D^{-1}\mathbf{A}); & (1) \\ diag(S)=diag(\mathbf{A}\mathbf{A}'+D). & (2) \end{cases}$$

方程(1)还可表示为

$$(D^{-1/2}SD^{-1/2})(D^{-1/2}\mathbf{A}) = (D^{-1/2}\mathbf{A})(I_m+\mathbf{A}'D^{-1}\mathbf{A}).$$

注意到: 因子负荷矩阵 A 不唯一. 这是因为

$$\Sigma = \mathbf{A}\mathbf{A}' + D = (\mathbf{A}T)(\mathbf{A}T)' + D,$$

对任意正交矩阵 T 成立. 由于正交矩阵 T 有

$$m^2 - m - m(m-1)/2 = m(m-1)/2$$

个变量(自由度).

因此, 为使 $\mathbf{A}$ 的解唯一, 需对 $\mathbf{A}$ 加上 $\dfrac{m(m-1)}{2}$ 个约束.

对 $\mathbf{A}$ 的约束为

$$\mathbf{A}'D^{-1}\mathbf{A} = \Gamma = diag(\gamma_1, \cdots, \gamma_m) > 0.$$

如果加上约束条件: $\mathbf{A}'D^{-1}\mathbf{A} = \Gamma = diag(\gamma_1, \cdots, \gamma_m) > 0$,
则由方程(1)推知

$$(D^{-1/2}SD^{-1/2})(D^{-1/2}\mathbf{A}) = (D^{-1/2}\mathbf{A})(I_m + \mathbf{A}'D^{-1}\mathbf{A})$$

$$(D^{-1/2}SD^{-1/2})(D^{-1/2}\mathbf{A})\Gamma^{-1/2} = (D^{-1/2}\mathbf{A})(I_m + \Gamma)\Gamma^{-1/2}$$

$$(D^{-1/2}SD^{-1/2})(D^{-1/2}\mathbf{A})\Gamma^{-1/2} = (D^{-1/2}\mathbf{A})\Gamma^{-1/2}(I_m + \Gamma)$$

$$(D^{-1/2}SD^{-1/2})(D^{-1/2}\mathbf{A}\Gamma^{-1/2}) = (D^{-1/2}\mathbf{A}\Gamma^{-1/2})(I_m + \Gamma),$$

而且有

$$(D^{-1/2}\mathbf{A}\Gamma^{-1/2})'(D^{-1/2}\mathbf{A}\Gamma^{-1/2}) = I_m.$$

即下面的式子成立:

$$\begin{cases} (D^{-1/2}SD^{-1/2})(D^{-1/2}\mathbf{A}\Gamma^{-1/2}) = (D^{-1/2}\mathbf{A}\Gamma^{-1/2})(I_m + \Gamma), \\ (D^{-1/2}\mathbf{A}\Gamma^{-1/2})'(D^{-1/2}\mathbf{A}\Gamma^{-1/2}) = I_m. \end{cases}$$

表明, $1+\gamma_1,\cdots,1+\gamma_m$ 是 $D^{-1/2}SD^{-1/2}$ 的 m 个正特征根,
$D^{-1/2}\mathbf{A}\Gamma^{-1/2}$ 是它们对应的正则正交特征向量矩阵.

又由于

$$\begin{aligned} D^{-1/2}\Sigma D^{-1/2} &= D^{-1/2}(\mathbf{A}\mathbf{A}'+D)D^{-1/2} \\ &= (\mathbf{A}'D^{-1/2})'(\mathbf{A}'D^{-1/2})+I_p, \end{aligned}$$

而 $(\mathbf{A}'D^{-1/2})'(\mathbf{A}'D^{-1/2})$ 与 $(\mathbf{A}'D^{-1/2})(\mathbf{A}'D^{-1/2})'=\mathbf{A}'D^{-1}\mathbf{A}=\Gamma$
有相同的非零特征根, 因此 $D^{-1/2}\Sigma D^{-1/2}$ 的前 m 个特征根为
$(1+\gamma_1)\geq\cdots\geq(1+\gamma_m)>0$, 其它全是 1.

因此知, 在约束条件 $\mathbf{A}'D^{-1}\mathbf{A}=\Gamma$ 下, $D^{-1/2}SD^{-1/2}$ 和 $D^{-1/2}\Sigma D^{-1/2}$
有相同的前 m 个正特征根.

在给定 D 的条件下, 可计算出 $D^{-1/2}SD^{-1/2}$ 的 m 个非零特征根 $(1+\gamma_1) \geq \cdots \geq (1+\gamma_m) > 0$, 以及它们对应的正则正交特征向量矩阵 $\mathbf{E}$. 令 $\Gamma = diag(\gamma_1, \cdots, \gamma_m)$.

则方程(1)的解为

$$D^{-1/2}\mathbf{A}\Gamma^{-1/2} = \mathbf{E} \quad \Longrightarrow \quad \mathbf{A} = D^{1/2}\mathbf{E}\Gamma^{1/2}.$$

又由 $D^{-1/2}SD^{-1/2} = \mathbf{E}(I_m + \Gamma)\mathbf{E}'$ 知

$$\begin{aligned} S &= D^{1/2}\mathbf{E}(I_m + \Gamma)\mathbf{E}'D^{1/2} \\ &= D + D^{1/2}\mathbf{E}\Gamma\mathbf{E}'D^{1/2} \\ &= D + (D^{1/2}\mathbf{E}\Gamma^{1/2})(D^{1/2}\mathbf{E}\Gamma^{1/2})' \\ &= D + \mathbf{A}\mathbf{A}', \end{aligned}$$

即方程(2)满足. 因此$\mathbf{A}$是方程的解, 且在约束条件下是唯一解.

极大似然估计的迭代算法

(1) 给出 D 的一个初始估计. 可由主成份法等确定 D 的一个粗估计, 或简单地由各分量的方差估计作为 D 初始估计.

记 $D_0 = diag(\sigma_{1,0}, \cdots, \sigma_{p,0})$;

(2) 计算 $D_0^{-1/2} S D_0^{-1/2}$ 的前 m 个特征根 $\lambda_{1,1} > \cdots > \lambda_{1,m}$, 以及它们所对应的正则正交特征向量 $e_{1,1}, \cdots, e_{1,m}$.
记 $\Gamma_1 = diag(\lambda_{1,1} - 1, \cdots, \lambda_{1,m} - 1)$, $\mathbf{E}_1 = (e_{1,1}, \cdots, e_{1,m})$.
则因子载荷矩阵的估计为 $\mathbf{A}_1 = D_0^{1/2} \mathbf{E}_1 \Gamma_1^{1/2}$.
这里要求 $\lambda_{1,1} > \cdots > \lambda_{1,m} > 1$, 否则减小 m, 或停止;

(3) 由方程(2)给出 D 的迭代估计 $D_1 = diag(S) - diag(\mathbf{A}_1 \mathbf{A}_1')$.
这里要求 $D_1 > 0$, 否则停止;

(4) 返回(1), 直至迭代停止.

6.4 因子旋转

由于因子载荷矩阵不唯一, 因此可以寻找一个正交矩阵 T 使得 $\mathbf{A}T$ 对应的公共因子具有更好的实际意义而便于解释.

方差最大的正交旋转 (Varimax 旋转)

先考虑两个因子的正交旋转. 设因子载荷矩阵和正交矩阵为

$$\mathbf{A} = (a_1, a_2), \quad T = \begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix}, \quad \mathbf{B} = \mathbf{A}T = (b_1, b_2).$$

目标: 旋转后, 因子的“贡献”越分散越好.

结果: x 可分为两部分, 一部分主要与第一因子有关,

另一部分主要与第二因子有关.

定义 b_1 和 b_2 的相对方差为

$$V_i(\varphi) = \frac{1}{p}\sum_{j=1}^{p}\left(\frac{b_{ji}^2}{h_j^2}\right)^2 - \left(\frac{1}{p}\sum_{j=1}^{p}\frac{b_{ji}^2}{h_j^2}\right)^2, \quad i = 1,\ 2.$$

要求使得总方差最大, 即求

$$\hat{\varphi} = argmax_{\varphi}(V_1(\varphi) + V_2(\varphi)).$$

记

$$\begin{aligned}
\mu_j &= \left(\frac{a_{j1}}{h_j}\right)^2 - \left(\frac{a_{j2}}{h_j}\right)^2, \quad \nu_j = 2\left(\frac{a_{j1}}{h_j}\right)\left(\frac{a_{j2}}{h_j^2}\right), \quad 1 \le j \le p; \\
A &= \sum_{j=1}^{p}\mu_j, \qquad\qquad B = \sum_{j=1}^{p}\nu_j; \\
C &= \sum_{j=1}^{p}(\mu_j^2 - \nu_j^2), \quad D = \sum_{j=1}^{p}2\mu_j\nu_j.
\end{aligned}$$

由此得解为

$$\tan(4\hat{\varphi}) = \frac{D - 2AB/p}{C - (A^2 - B^2)/p},$$

进而得正交矩阵

$$T = \begin{pmatrix} \cos(\hat{\varphi}) & -\sin(\hat{\varphi}) \\ \sin(\hat{\varphi}) & \cos(\hat{\varphi}) \end{pmatrix}.$$

更多公共因子的情形

当公共因子个数 $m > 2$ 时, 可以逐次对每两个公共因子对进行 Varimax旋转. 这样做 $C_m^2 = m(m-1)/2$ 次后, 所有公共因子对都旋转一次, 记为旋转一轮. 如果旋转一轮后未达到目标, 再进行新一轮旋转.

记第 j 次旋转后因子载荷矩阵的相对方差总和为 $G^{(j)}$,

$$G^{(j)} = \sum_{i=1}^{m} V_i^{(j)}.$$

由Varimax的性质知, $G^{(1)} \le G^{(2)} \le \cdots \le G^{(n)} \le G^{(n+1)} \le \cdots$. 而因子载荷矩阵的元素绝对值均不大于1, 因此上述单调序列有上界, 故收敛.

实际中, 若认为因子载荷矩阵的相对方差总和已经变化很小时, 则停止旋转.

6.5 正交因子模型协方差结构的检验

设 $x_1, \cdots, x_n$ 是来自总体 $N_p(\mu, \Sigma)$ 的样本, 其中, $n > p$, $\Sigma > 0$. 有关正交因子模型($\mathbf{M}$)的检验问题为

$\mathbf{H}_0: \quad \Sigma = \mathbf{A}\mathbf{A}' + D,$

其中, $\mathbf{A}$ 是秩为 m 的 $p \times m$ 矩阵, $D = diag(\sigma_1^2, \cdots, \sigma_p^2) > 0.$

记 $(\mathbf{A}, D)$ 的极大似然估计为 $(\hat{\mathbf{A}}, \hat{D})$, 则有

$$L(\hat{\mathbf{A}}, \hat{D}) = |\hat{\mathbf{A}}\hat{\mathbf{A}}' + \hat{D}|^{-n/2} \exp\left\{-\frac{n}{2}tr\left[(\hat{\mathbf{A}}\hat{\mathbf{A}}' + \hat{D})^{-1}S)\right]\right\}.$$

由于极大似然估计满足似然方程, 故有

$$
\begin{cases}
\hat{\mathbf{A}} = S(\hat{\mathbf{A}}\hat{\mathbf{A}}' + \hat{D})^{-1}\hat{\mathbf{A}}; & (3) \\
diag(S) = diag(\hat{\mathbf{A}}\hat{\mathbf{A}}' + \hat{D}). & (4)
\end{cases}
$$

又对任意方阵 Q 和正定的对角阵 P, 有

$$tr[diag(QP)] = tr[QP] = tr[(diag(Q))P],$$

因此由 $\hat{D}$ 是对角阵与方程(4)知

$$
\begin{aligned}
tr[(S - \hat{\mathbf{A}}\hat{\mathbf{A}}')\hat{D}^{-1}] &= tr[diag((S - \hat{\mathbf{A}}\hat{\mathbf{A}}')\hat{D}^{-1})] \\
&= tr[(diag(S - \hat{\mathbf{A}}\hat{\mathbf{A}}'))\hat{D}^{-1}] \\
&= tr[(diag(\hat{D}))\hat{D}^{-1}] \quad (\text{由方程}(4)) \\
&= tr[diag(\hat{D}\hat{D}^{-1})] \\
&= p.
\end{aligned}
$$

进而再利用方程(3)可推知

$$
\begin{aligned}
tr[S(\hat{\mathbf{A}}\hat{\mathbf{A}}' + \hat{D})^{-1}] &= tr[S(\hat{\mathbf{A}}\hat{\mathbf{A}}' + \hat{D})^{-1}(\hat{\mathbf{A}}\hat{\mathbf{A}}' + \hat{D} - \hat{\mathbf{A}}\hat{\mathbf{A}}')\hat{D}^{-1}] \\
&= tr[S\hat{D}^{-1} - (S(\hat{\mathbf{A}}\hat{\mathbf{A}}' + \hat{D})^{-1}\hat{\mathbf{A}})\hat{\mathbf{A}}'\hat{D}^{-1}] \\
&= tr[S\hat{D}^{-1} - (\hat{\mathbf{A}})\hat{\mathbf{A}}'\hat{D}^{-1}] \quad \text{方程(3)} \\
&= tr[(S - \hat{\mathbf{A}}\hat{\mathbf{A}}')\hat{D}^{-1}] \\
&= p.
\end{aligned}
$$

因此有

$$
L(\hat{\mathbf{A}}, \hat{D}) = |\hat{\mathbf{A}}\hat{\mathbf{A}}' + \hat{D}|^{-n/2} \exp\left\{-\frac{np}{2}\right\}.
$$

正交因子模型检验的似然比为

$$
\begin{aligned}
\lambda &= \frac{\sup\limits_{\mu,\,\Sigma=\mathbf{A}\mathbf{A}'+D} |\Sigma|^{-n/2} \exp\left\{-\frac{n}{2} tr[\Sigma^{-1}(S+(\bar{x}-\mu)(\bar{x}-\mu)')]\right\}}{\sup\limits_{\mu,\,\Sigma} |\Sigma|^{-n/2} \exp\left\{-\frac{n}{2} tr[\Sigma^{-1}(S+(\bar{x}-\mu)(\bar{x}-\mu)')]\right\}} \\
&= \left(\frac{|S|}{|\hat{\mathbf{A}}\hat{\mathbf{A}}'+\hat{D}|}\right)^{n/2}.
\end{aligned}
$$

计算参数空间的自由度：

(1) $dim(H) = p + \dfrac{p(p+1)}{2}$;

(2) 由于 $\mathbf{A}'D^{-1}\mathbf{A}$ 为对角阵, 即 $\mathbf{A}$ 有 $\dfrac{m(m-1)}{2}$ 个约束. 因此

$$dim(H_0) = mp - \frac{m(m-1)}{2} + 2p;$$

(3)

$$\begin{aligned} f &= dim(H) - dim(H_0) \\ &= \left[p + \frac{p(p+1)}{2}\right] - \left[mp + 2p - \frac{m(m-1)}{2}\right] \\ &= \frac{(p-m)^2 - (p+m)}{2}. \end{aligned}$$

由Wilks定理知

$$-2\ln\lambda = -n\left(\ln S - \ln|\hat{\mathbf{A}}\hat{\mathbf{A}}' + \hat{D}|\right) \xrightarrow{d} \chi^2(f).$$

检验方案:

当 $-2\ln\lambda > \chi^2_{1-\alpha}(f)$ 时拒绝零假设, 其犯第一类错误的概率近似为 α.

正交因子模型的识别性问题

当自由度 $f \leq 0$ 时, 正交因子模型存在可识别性问题.

1) $f < 0$, 分解 $\Sigma = \mathbf{A}\mathbf{A}' + D$ 不唯一;

2) $f = 0$, 分解 $\Sigma = \mathbf{A}\mathbf{A}' + D$ 或不存在, 或存在唯一.

6.6 斜交因子模型 (斜交旋转)

设 p 维随机向量 x 可以表示为

$$x = \mu + \mathbf{A} f + u, \qquad (\mathbf{M}^*)$$

其中, μ 是 p 维常数向量, $\mathbf{A}$ 是 $p \times m$ 阶常数矩阵, $f \overset{d}{\sim} N_m(0, R)$, $m < p$, $R > 0$ 为相关阵, $u \overset{d}{\sim} N_p(0, D)$, $D = diag(\sigma_1^2, \cdots, \sigma_p^2)$, f 与 u 相互独立. 则称模型($\mathbf{M}^*$) 为**斜交因子模型**, 称 f 为**公共因子**, u 为**特殊因子**, $\mathbf{A}$ 为**因子载荷矩阵**.

事实上, 存在满秩阵 T, 使得 $R = TT'$. 若令 $\mathbf{B} = \mathbf{A}T$, $g = T^{-1} f$, 则

$$\begin{aligned} x &= \mu + \mathbf{A} f + u \\ &= \mu + (\mathbf{A}T)(T^{-1} f) + u \\ &= \mu + \mathbf{B} g + u, \end{aligned}$$

易知,$Cov(g) = T^{-1}Cov(f)(T')^{-1} = T^{-1}R(T')^{-1} = I_m$, 则

$$x = \mu + \mathbf{B}g + u$$

是正交因子模型, g是公共因子, $\mathbf{B}$ 是正交因子载荷矩阵.

由于 T 非正交矩阵, 我们称公共因子 $f = Tg$ 为正交公共因子 g 的**斜交旋转**.

6.7 因子得分

因子模型是将变量表示为公共因子的线性组合.

由于公共因子能反映原始变量的相关关系, 用公共因子代表原始变量时, 有时更有利于描述研究对象的特征.

因而往往需要反过来将公共因子表示成变量的线性组合.
即

$$f = \mathbf{B}x, \tag{$*$}$$

其中, $B = (b_{ij})_{m\times p}$, f 为 m 维公共因子, x 为 p 维变量.

则称 $(*)$ 式为**因子得分函数**, $\mathbf{B}$ 为**因子得分矩阵**.

由因子得分函数知

$$f_j = \sum_{i=1}^{p} b_{ji}x_i,\ 1 \le j \le m.$$

因子得分的计算

计算因子得分矩阵 $\mathbf{B}$

假定变量 x 已作标准化处理, 即 $E(x_i)=0$, $Var(x_i)=1$, $1 \le i \le p$.
令 $R = Cov(x)$, R 也是 x 的相关阵. 记 $R = (r_{ij})_{p\times p}$.
假定因子载荷矩阵 $\mathbf{A}$ 和 相关阵 R 已知.

对任意 $1 \le i \le p$, $1 \le j \le m$, 有

$$\begin{aligned} a_{ij} &= E(x_i f_j) \\ &= \sum_{k=1}^{p} b_{jk} E(x_i x_k) \\ &= \sum_{k=1}^{p} r_{ik} b_{jk}, \end{aligned}$$

因此对 $1 \le j \le m$, 有

$$\begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{pmatrix} \begin{pmatrix} b_{j1} \\ b_{j2} \\ \vdots \\ b_{jp} \end{pmatrix} = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{pj} \end{pmatrix}.$$

因子得分的估计：回归法

假设已有因子载荷矩阵的估计 $\hat{\mathbf{A}} = (\hat{a}_{ij})$ 和相关阵的估计 $\hat{R}$.
则因子得分矩阵的估计为

$$\begin{pmatrix} \hat{b}_{j1} \\ \hat{b}_{j2} \\ \vdots \\ \hat{b}_{jp} \end{pmatrix} = \hat{R}^{-1} \begin{pmatrix} \hat{a}_{1j} \\ \hat{a}_{2j} \\ \vdots \\ \hat{a}_{pj} \end{pmatrix}, \ 1 \le j \le m,$$

即

$$\hat{\mathbf{B}} = (\hat{b}_{ij})_{m \times p} = \hat{\mathbf{A}}' \hat{R}^{-1}.$$

则因子得分估计为

$$\hat{f} = \hat{\mathbf{B}} x.$$

在估计出公共因子的得分后, 可以利用因子得分进行进一步的分析.

(1) 根据得分值大小, 可以判断公共因子的重要度;

(2) 可以根据因子得分对样本进行聚类.