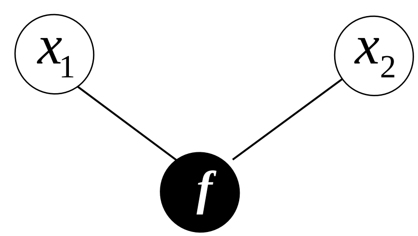
第十三讲 因子分析

2024.4.22



f: 潜变量/隐变量/因子

潜变量模型

潜变量

不可直接观测但可以用于解释可观测现象（manifest variable）的变量称为潜变量/隐变量/因子(latent variable, hidden variable).

物理学中的隐变量理论是由物理学家质疑量子力学完备性而提出的替代理论。一些物理学家例如爱因斯坦，认为量子力学并未完整地描述物理系统的状态，亦即质疑量子力学是不完备的。因此量子力学的背后应该隐藏了一个尚未发现的理论，可以完整解释物理系统所有可观测量的演化行为，而避免掉任何不确定性或随机性。

历史上爱因斯坦是隐变量理论的主要倡导者，出于对标准量子力学诠释的概率性解释的不满。他曾说：“我相信上帝不掷骰子。”

1935年，爱因斯坦与波多尔斯基、罗森共同提出的EPR佯谬（以姓氏字首为缩写）试图对哥本哈根诠释做出挑战，论文中指出隐变量应该加入量子力学中，俾使在量子纠缠现象中不会出现鬼魅般的超距作用。在提出后，这样的争辩仍停留在物理哲学的范畴，直到约翰·贝尔提出贝尔定理方得区分两者差异。透过实验证实：一定类型的局域隐变量理论与实验结果不相符，包括EPR佯谬中提出的诠释版本。

资产定价/投资收益的Fama-French 三因子模型假设投资组合的收益可以由市场风险、市值风险、账面a市值比风险等三个潜在（不可直接测量的）因素/因子来共同解释。

潜变量模型

潜变量模型通过少数独立的因子（潜变量）描述可观测随机变量的相关性。潜变量/隐变量统计模型大致可分如下类：

	Manifest variables	
Latent variables	Continuous	Categorical
Continuous	Factor analysis	Item response theory
Categorical	Latent profile analysis	Latent class analysis

例1 (Probit model - item response): 假设伯努利响应y通过连续的潜变量f与自变量$\mathbf{x}$有关，$f \sim N(a+\mathbf{b}^\top \mathbf{x}, 1)$。当$f>0$时，$y=1$。 则

$$p = P(y=1|\mathbf{x}) = P(f>c|\mathbf{x}) = \Phi(a+\mathbf{b}^\top \mathbf{x}),$$

Probit模型与logistic回归模型类似。

例2 (纵向数据的线性混合效应模型): 假设同一研究对象i有重复观测(y_{ij}, x_{ij}), $j=1,\ldots,n_i$, 假设线性模型

$$y_{ij} = a_i + bx_{ij} + \varepsilon_{ij}, \varepsilon_{ij} \text{ iid } \sim N(0,\sigma^2),$$

为了考虑进同一个研究对象的各次测量之间的相关性，假设截距

$$a_i \text{ iid } \sim N(a,\tau^2)\text{，} \quad a_i \perp\!\!\!\perp \varepsilon_{ij}$$

a_i称为随机效应，a_i的引入使得模型更合理和灵活

(1) a_i代表了研究对象i各次测量之间的相关性；

(2) a_i代表了不同的研究对象的异质性。

主成分变换的反变换：正交展开表示

正交基展开

随机向量$\mathbf{x}_{p\times 1}$的主成分变换$\mathbf{y} = V^\top \mathbf{x} \Rightarrow$ 正交展开$\mathbf{x} = V\mathbf{y}$：

$$\mathbf{x} = \mathbf{v}_1 y_1 + \cdots + \mathbf{v}_p y_p,\quad \mathrm{var}(y_1) = \lambda_1 \geq \mathrm{var}(y_2) = \lambda_2 \geq \ldots$$

- $\mathbf{x}$表示为正交基$\mathbf{v}_1, \ldots, \mathbf{v}_p$的线性组合，系数为主成分$y_1, \ldots, y_p$.
- $\mathbf{x}$由一组不相关的 r.v. $y_1, \ldots, y_p$生成

最佳逼近

PCA以上述正交展开的前m项逼近$\mathbf{x}$:

$$\mathbf{x} \approx \mathbf{v}_1 y_1 + \cdots + \mathbf{v}_m y_m$$

$$\mathbf{x} = \mathbf{v}_1 y_1 + \cdots + \mathbf{v}_m y_m + \boldsymbol{e}_m$$

由Eckart-Young-Mirsky定理，这是最佳逼近（参加后面SVD）。

Eckart-Young-Mirsky定理：在Σ的所有秩m近似中，

$$\hat{\Sigma} = V_m \Lambda_m V_m^\top = \lambda_1 \mathbf{v}_1 \mathbf{v}_1^\top + \ldots + \lambda_m \mathbf{v}_m \mathbf{v}_m^\top$$

使得误差平方和 $\min_{\mathrm{rank}(\hat{\Sigma})=m} \|\Sigma - \hat{\Sigma}\|_F^2$ 达到最小。换言之，

$\hat{\mathbf{x}} = \mathbf{v}_1 y_1 + \ldots + \mathbf{v}_m y_m$是$\mathbf{x}$的最佳逼近。

主成分是潜变量

- $\mathbf{x}$可以观测，主成分 $\mathbf{y}$不可观测，是从$\mathbf{x}$推断得到的。
- PCA以少数主成分（$\mathbf{y}$的一部分）解释$\mathbf{x}$。
- 少数主成分/潜变量可用于解释可观测变量$\mathbf{x}$的相关性。

PCA拓展

以概率模型表达PCA，我们有一般的因子模型（概率PCA模型）：

对于任何$\mathbf{x}_{p\times 1}$，因子模型假设存在m个潜变量/因子$\mathbf{f}$，使得

$$\mathbf{x} = L\mathbf{f} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \perp\!\!\!\perp \mathbf{f}$$

随机函数或随机过程存在与PCA类似的Karhunen-Loève展开：

任一随机函数（随机过程）可表示为相互正交的常函数的线性组合，组合系数互不相关。

Karhunen-Loève定理. 假设 $x_t = x(t)$ 是平方可积的随机过程， $t \in R$ ，

$$E(x_t) = 0, \text{协方差函数} K(s,t) = cov(x_s, x_t),$$

则存在相互正交模长为1的常函数$v_k(t), k = 1,2, \ldots$, 使得 x_t有正交展开

$$x_t = \sum_{k=1}^{\infty} y_k v_k(t)$$

其中$v_k(t)$满足

$$\int K(s,t) v_k(s) ds = \lambda_k v_k(t), k = 1,2, \ldots$$

即λ_k，$v_k(t)$是线性变换 $Tf = \int K(s,\cdot) f(s) ds$ 特征根和特征函数。另外，$y_k = \int x_t v_k(t) dt$, 满足 $E(y_k) = 0, var(y_k) = \lambda_k, cov(y_k, y_j) = 0$，这些不相关的$y_k$称为主成分。

$Mercer$定理(谱分解): 对称正定的二元连续函数$K(s,t)$存在谱分解

$$K(s,t) = \sum_{k=1}^{\infty} \lambda_k v_k(s) v_k(t)$$

因子分析

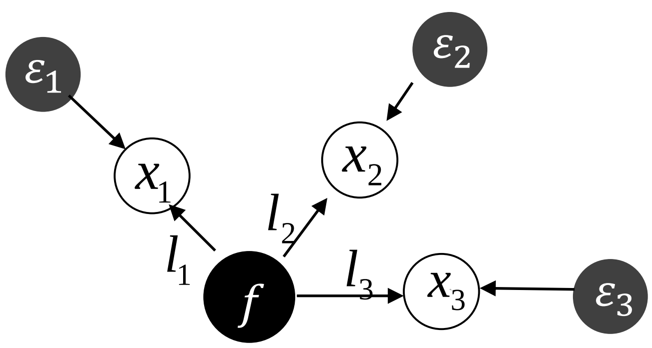
因子分析：以潜变量解释多个变量的相关性

例2 (Spearman,1904).若干儿童三门课Classics (x_1), French (x_2), English (x_3)
成绩的相关系数矩阵：

$$R=\begin{pmatrix}1 & 0.83 & 0.78\\ & 1 & 0.67\\ & & 1\end{pmatrix}$$

Spearman认为3门成绩高相关是因为它们都与一个不可观测的潜变量（称为公共因子，比如语言能力）有关，另外每门课还有其特殊性（特殊因子）。考试的目的是通过三门成绩推断能力。

一个因子(能力f)影响三个变量(课程x_i): $x_i = l_i f + \varepsilon_i$



$$\begin{cases}x_1 = l_1 f + \varepsilon_1\\ x_2 = l_2 f + \varepsilon_2, \quad \varepsilon_i \sim (0,\psi_i)\\ x_3 = l_3 f + \varepsilon_3\end{cases}$$

f :公共因子，ε_i :特殊因子

ε_i与 f 独立，ε's独立

l_i :因子载荷 / 系数

因为$l_i f = (2l_i)(f/2)$，为了模型可识别（参数可估计），
还需要一些假定/约束，通常假设：$E(f)=0,\ \mathrm{var}(f)=1$

估计参数：因子模型两边同时求方差或协方差：

$$\rho_{ij} = \mathrm{cov}(x_i, x_j) = \mathrm{cov}(l_i f + \varepsilon_i, l_j f + \varepsilon_j) = l_i l_j,\quad i \neq j,$$

$$1 = \mathrm{var}(x_i) = l_i^2 + \psi_i, \text{等等},$$

得到参数的方程

$$\begin{cases} l_i^2 + \psi_i = 1,\quad i = 1,2,3 \\ l_1 l_2 = 0.83,\quad l_1 l_3 = 0.78,\quad l_2 l_3 = 0.67 \end{cases} \qquad (*)$$

$$\Rightarrow l_1 = 0.98,\quad l_2 = 0.84,\quad l_3 = 0.80;\quad \psi_1 = 0.03,\ \psi_2 = 0.29,\ \psi_3 = 0.37$$

拉丁classics的载荷最大（0.98），与潜在能力最相关。

因子模型的一般形式：$\mathbf{x} = L\mathbf{f} + \boldsymbol{\varepsilon}$，描述了$\mathbf{x}$的方差结构，即(*)式：

$$\mathrm{var}(\mathbf{x}) = \Sigma = \mathrm{var}(L\mathbf{f}) + \mathrm{var}(\boldsymbol{\varepsilon}) + \Psi,$$

通常没有(唯一)解,可使用极大似然法或主成分方法求近似解。

例1是Spearman 研究的下述相关系数矩阵的一部分。

Spearman's correlation matrix for six measures of school performance. The bottom row shows the g loadings

	Classics	French	English	Math	Pitch	Music
Classics	–					
French	.83	–				
English	.78	.67	–			
Math	.70	.67	.64	–		
Pitch	.66	.65	.54	.45	–	
Music	.63	.57	.51	.51	.40	–
g-loading	.958	.882	.803	.750	.673	.646

Pitch： pitch discrimination
g-loading: 载荷，各门课与潜在因子的相关系数

因子分析的起源

- 因子分析(FA: factor analysis)模型是PCA/SVD的一种概率模型表达，有人称之为概率PCA模型。
- 因子分析起源于20世纪初Pearson和Spearman的心理学、教育学研究, 也与Fisher的方差成分模型有关

Charles Edward Spearman, FRS (1863 – 1945)， 英国实验心理学家, *g*-factor， Spearman's rho是他最有名的贡献。

Spearman在研究智力/认知测试(intelligence test/cognitive test)的过程中发现各种不同的测试（比如语言、数学、绘画、音乐等）与两种因素(因子)有关，一个是影响各种测试成绩的普通因子（*g*-factor，General intelligence factor），另外就是影响具体科目的特殊因子 （*s*-factor，specific factor），由此发展形成了单因子分析方法 。IQ （Intelligence quotient）分数（均值100，std 15）可以认为是g-factor 的体现。

His statistical work was not appreciated by his University College colleague K. Pearson and there was a long feud between them.（wiki）

正交因子模型

正交因子模型

假设随机向量$\mathbf{x}_{p\times1}$, $E(\mathbf{x})=\boldsymbol{\mu}, \mathrm{var}(\mathbf{x})=\Sigma$, $m\le p$, 正交因子模型假设如下线性模型：

$$\mathbf{x}_{p\times1}=\boldsymbol{\mu}_{p\times1}+L_{p\times m}\mathbf{f}_{m\times1}+\boldsymbol{\varepsilon}_{p\times1}$$

Common factors

- 公共因子$\mathbf{f}=(f_1,\dots,f_m)^{\mathsf{T}}$: $\mathbf{f}$ 各分量不相关(故称为正交因子)

$$E(\mathbf{f})=\mathbf{0},\quad \mathrm{var}(\mathbf{f})=I_m$$

loadings

- 载荷矩阵L: $p\times m$常数矩阵;

Specific factors

uniqueness

- 特殊因子$\boldsymbol{\varepsilon}=(\varepsilon_1,\dots,\varepsilon_p)^{\mathsf{T}}$: 假设

$$E(\boldsymbol{\varepsilon})=\mathbf{0},\quad \mathrm{var}(\boldsymbol{\varepsilon})=\Psi=diag(\psi_1,\dots,\psi_p)$$

- $\boldsymbol{\varepsilon}$与$\mathbf{f}$独立。

注. 每个样本都假设满足因子模型：$\underset{p\times1}{\mathbf{x}_i}=\underset{p\times m}{L}\,\underset{m\times1}{\mathbf{f}_i}+\underset{p\times1}{\boldsymbol{\varepsilon}_i}, i=1,\dots,n.$

则数据矩阵X的因子模型表达为：$X=FL^{\mathsf{T}}+E$

其中$X=(\mathbf{x}_1,\dots,\mathbf{x}_n)^{\mathsf{T}}, F=(\mathbf{f}_1,\dots,\mathbf{f}_n)^{\mathsf{T}}, E=(\boldsymbol{\varepsilon}_1,\dots,\boldsymbol{\varepsilon}_n)^{\mathsf{T}}.$

方差结构

命题1.假设随机向量$\mathbf{x}_{p\times1}$，$E(\mathbf{x})=\boldsymbol{\mu}, \mathrm{var}(\mathbf{x})=\Sigma$, 满足正交因子模型

$$\mathbf{x}_{p\times1}=\boldsymbol{\mu}_{p\times1}+L_{p\times m}\mathbf{f}_{m\times1}+\boldsymbol{\varepsilon}_{p\times1},\ m<p,$$

则$\Sigma=LL^{\mathrm{T}}+\Psi, L=\mathrm{cov}(\mathbf{x},\mathbf{f})$.

证明：

（1）$\mathrm{var}(\mathbf{x})=\mathrm{var}(\boldsymbol{\mu}+L\mathbf{f}+\boldsymbol{\varepsilon})=\mathrm{var}(L\mathbf{f})+\mathrm{var}(\boldsymbol{\varepsilon})=L\,\mathrm{var}(\mathbf{f})L^{\mathrm{T}}+\Psi=LL^{\mathrm{T}}+\Psi$

（2）$\mathrm{cov}(\mathbf{x},\mathbf{f})=\mathrm{cov}(\boldsymbol{\mu}+L\mathbf{f}+\boldsymbol{\varepsilon},\mathbf{f})=\mathrm{cov}(L\mathbf{f},\mathbf{f})+\mathrm{cov}(\boldsymbol{\varepsilon},\mathbf{f})=L\,\mathrm{cov}(\mathbf{f},\mathbf{f})=L$

该命题说明，模型$\mathbf{x}=\boldsymbol{\mu}+L\mathbf{f}+\boldsymbol{\varepsilon}$蕴含了

(1) Σ具有简单结构：$\Sigma=LL^{\mathrm{T}}+\Psi$，其中$L$低秩，$\Psi$对角，

(2) 载荷L为原数据$\mathbf{x}$与潜在因子$\mathbf{f}$的协方差，

(3) 当$\mathbf{x}$是标准化的，$L=(l_{ij})$是相关系数矩阵，l_{ij}为变量i与因子j的相关系数。

方差结构

$$\mathbf{x}=\boldsymbol{\mu}+L\mathbf{f}+\boldsymbol{\varepsilon} \quad \Rightarrow \quad x_i=\mu_i+L_i^{\mathsf{T}}\mathbf{f}+\varepsilon_i \qquad L=(L_1,\ldots,L_p)^{\mathsf{T}}=(l_{ij})$$

$$\Downarrow \qquad\qquad\qquad \Downarrow$$

$$\Sigma=LL^{\mathsf{T}}+\Psi \quad \Rightarrow \quad \sigma_{ii}=L_i^{\mathsf{T}}L_i+\Psi_i$$

$$\underbrace{\sigma_{ii}}_{\mathrm{var}(x_i)}=\underbrace{L_i^{\mathsf{T}}L_i}_{\mathrm{var}(L_i\mathbf{f})}+\underbrace{\Psi_i}_{\mathrm{var}(\varepsilon_i)}=\underbrace{l_{i1}^2+\ldots+l_{im}^2}_{h_i^2\text{,共性方差}}+\underbrace{\psi_i}_{\text{特有方差/个性方差}}$$

共性方差
特殊方差

载荷矩阵L的第i行 L_i的平方和

$$h_i^2=L_i^{\mathsf{T}}L_i=l_{i1}^2+\ldots+l_{im}^2$$

称为共性方差(communality), h_i^2代表了所有公共因子能解释的第i个变量的方差。$\psi_i=\sigma_{ii}-h_i^2$称为特殊方差/个性方差(specific variance), 代表了方差$\sigma_{ii}=\mathrm{var}(x_i)$中公共因子所不能解释的部分。

注：因子j对变量i的方差贡献率：l_{ij}^2/σ_{ii}(一般不计算这个贡献率）

所有公共因子的方差贡献率

- h_i^2/σ_{ii} :所有公共因子所能解释 $\mathrm{var}(x_i)$的比例.
- 总方差的方差分解： $tr(\Sigma)=tr(L^{\mathrm{T}}L)+tr(\Psi)=h_1^2+\ldots+h_p^2+tr(\Psi)$,

所有公共因子的方差贡献率(因子所能解释的总方差的比例)：

$$\frac{\mathrm{tr}(L^{\mathrm{T}}L)}{\mathrm{tr}(\Sigma)}=\frac{h_1^2+\ldots.+h_p^2}{\sigma_{11}+\ldots+\sigma_{pp}}.$$

另一方面，我们还可以计算每一个公共因子的方差贡献率。

记L的各列：$L=(L_{(1)},\ldots,L_{(m)})$,总方差也可以表述为

$$\mathrm{tr}(\Sigma)=\|L_{(1)}\|^2+\ldots+\|L_{(m)}\|^2+\mathrm{tr}(\Psi)\triangleq SS_1+\ldots+SS_m+\mathrm{tr}(\Psi)$$

其中$SS_j=\|L_{(j)}\|^2=l_{1j}^2+\ldots+l_{pj}^2=L$第$j$列的平方和.

单个因子的方差贡献率

第j个因子对总方差的贡献率：$\dfrac{SS_j}{\mathrm{tr}(\Sigma)}=\dfrac{l_{1j}^2+\ldots+l_{pj}^2}{\mathrm{tr}(\Sigma)}$， $j=1,\ldots,m.$

天才是1%的灵感加99%的努力 –爱迪生

方差贡献率(汇总）

$$
L=\begin{array}{cc}
 & \text{因子} \\
\text{变量} & \begin{pmatrix}
l_{11} & \cdots & l_{1j} & \cdots & l_{1m} \\
\vdots & & \vdots & & \vdots \\
l_{i1} & \cdots & l_{ij} & \cdots & l_{im} \\
\vdots & & \vdots & & \vdots \\
l_{p1} & \cdots & l_{pj} & \cdots & l_{pm}
\end{pmatrix}
\end{array}
$$

$h_i^2=l_{i1}^2+\ldots+l_{im}^2$ (第i个变量的共性方差)

$=$所有因子能解释的变量i的方差.

$SS_j=l_{1j}^2+\ldots+l_{pj}^2$ (第j个因子载荷平方和)

$=$第j个因子所能解释的总方差

- ~~因子j对变量i的方差贡献率：l_{ij}^2/σ_{ii}~~
- 第i个变量中所有公共因子所能解释的方差的比例：h_i^2/σ_{ii}
- 单个公共因子j的对总方差贡献比例：$SS_j/\mathrm{tr}(\Sigma)$
- 所有公共因子总的方差贡献率：$\dfrac{SS_1+\ldots+SS_m}{\mathrm{tr}(\Sigma)}=\dfrac{h_1^2+\ldots.+h_p^2}{\mathrm{tr}(\Sigma)}$

因子模型的参数估计方法1：主因子法

PCA记号：假设Σ的特征根$\lambda_1 \geq \ldots \geq \lambda_p \geq 0$，对应的正交单位特征向量

$\mathbf{v}_1, \ldots, \mathbf{v}_p$， 记$\Lambda = diag(\lambda_1, \ldots, \lambda_p), V = (\mathbf{v}_1, \ldots, \mathbf{v}_p) = (v_{ij})$， $V^\top V = VV^\top = I_p$

$$\Sigma = V\Lambda V^\top = \mathbf{v}_1\mathbf{v}_1^\top\lambda_1 + \ldots + \mathbf{v}_p\mathbf{v}_p^\top\lambda_p$$

主成分/主成分变换： $\mathbf{y} = V^\top\mathbf{x}$， $\mathrm{var}(\mathbf{y}) = V^\top\Sigma V = \Lambda$， $y_1, \ldots, y_p$不相关。

记$\mathbf{y}_m = (y_1, \ldots, y_m)^\top$， $V_m = (\mathbf{v}_1, \ldots, \mathbf{v}_m)$， $\Lambda_m = diag(\lambda_1, \ldots, \lambda_m)$.

最小二乘

正交因子模型$\mathbf{x}_{p\times1} = L_{p\times m}\mathbf{f}_{m\times1} + \boldsymbol{\varepsilon}_{p\times1}$下（不妨假设$E(\mathbf{x}) = 0$），$\mathbf{x}$的协方差矩阵

$$\Sigma = LL^\top + \Psi \quad , \quad \Psi = \mathrm{var}(\boldsymbol{\varepsilon}) = \mathrm{diag}(\Psi_1, \ldots, \Psi_p)$$

我们希望$LL^\top$与Σ差距尽量小，为此极小化误差平方和

$$\| \Sigma - LL^\top \|_F^2 = tr(\Sigma - LL^\top)(\Sigma - LL^\top)^\top$$

由Eckart - Young - Mirsky Theorem (P4)，当

$$LL^\top = V_m\Lambda_m V_m^\top = \left(V_m\Lambda_m^{1/2}\right)\left(V_m\Lambda_m^{1/2}\right)^\top$$

时误差平方和最小，得主因子解(principal factor) $\hat{L} = V_m\Lambda_m^{1/2}$.

拟合线性模型$\mathbf{x}=\hat{L}\mathbf{f}+\boldsymbol{\varepsilon}$,得$\mathbf{f}$的LS预测：$\hat{\mathbf{f}}=(\hat{L}^{\mathsf{T}}\hat{L})^{-1}\hat{L}^{\mathsf{T}}\mathbf{x}=\Lambda_m^{-1/2}\mathbf{y}_m$.

因此主成分正交展开：

$$\mathbf{x}=Vy=\mathbf{v}_1y_1+\ldots+\mathbf{v}_py_p=\mathbf{v}_1y_1+\ldots+\mathbf{v}_my_m+(\mathbf{v}_{m+1}y_{m+1}+\ldots+\mathbf{v}_py_p)$$

$$\overset{\Delta}{=}V_m\mathbf{y}_m+\mathbf{e}_m=(V_m\Lambda_m^{1/2})(\Lambda_m^{-1/2}\mathbf{y}_m)+\mathbf{e}_m$$

$$=\hat{L}\hat{\mathbf{f}}+\mathbf{e}_m$$

的前m项就是因子模型的解,其中$\hat{L}=V_m\Lambda_m^{1/2}$，$\hat{\mathbf{f}}=\Lambda_m^{-1/2}\mathbf{y}_m$。

因子模型的参数估计2：极大似然方法

正态因子模型

假设$\mathbf{x}$的实现/样本$\mathbf{x}_1,\ldots,\mathbf{x}_n$满足因子模型：

$$\mathbf{x}_i=\boldsymbol{\mu}+L\mathbf{f}_i+\boldsymbol{\varepsilon}_i,\quad \mathbf{f}_i\text{和}\boldsymbol{\varepsilon}_i\text{独立}$$

假设$\mathbf{f}_i,\boldsymbol{\varepsilon}_i$都服从正态分布即$\mathbf{f}_i\sim N_m(\mathbf{0},I_m),\boldsymbol{\varepsilon}_i\sim N_p(\mathbf{0},\Psi)$.

正态因子模型假设下，$\mathbf{x}_i\sim N_p(\boldsymbol{\mu},\Sigma),\ \Sigma=LL^{\mathrm{T}}+\Psi,i=1,\ldots,n.$

似然函数

$$L(\boldsymbol{\mu},L,\Psi)=\prod_{i=1}^{n}\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(\mathbf{x}_i-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}_i-\boldsymbol{\mu})\right),$$

需要估计的参数：$\boldsymbol{\mu},L,\Psi$,我们已知$\boldsymbol{\mu}$的估计$\hat{\boldsymbol{\mu}}=\bar{\mathbf{x}}$,

而L,Ψ可通过极大化 $l(\Omega)=\log|\Omega|-tr\left(\Omega S\right)$得到，其中

$\Omega=\Sigma^{-1}=(LL^{\mathrm{T}}+\Psi)^{-1}$，$S=\sum(\mathbf{x}_i-\boldsymbol{\mu})(\mathbf{x}_i-\boldsymbol{\mu})^{\mathrm{T}}/n.$

解不唯一：如果L，Ψ满足模型方程

$$\mathbf{x}=\boldsymbol{\mu}+L\mathbf{f}+\boldsymbol{\varepsilon}$$

则对任一正交矩阵$O,\mathbf{L}^*=LO,\mathbf{f}^*=O^{\mathrm{T}}\mathbf{f}$也满足方程。

极大似然法

求解L,Ψ的极大似然估计等价于优化问题：

$$\min_{L,\Psi}\left(\log|LL^{\mathrm{T}}+\Psi|+tr(S(LL^{\mathrm{T}}+\Psi)^{-1})\right)$$

因为解不唯一，增加约束：$L^{\mathrm{T}}\Psi^{-1}L=$对角矩阵

如果参数的极大似然估计为$\tilde{L}$，$\tilde{\Psi}$，拟合的协方差阵$\tilde{L}\tilde{L}^{\mathrm{T}}+\tilde{\Psi}$接近于$S$，那么认为因子模型拟合效果良好。更严格地，下述似然比检验统计量($Wilks$统计量)度量了拟合优度，其p值越大,说明拟合得越好.

多元正态的似然比检验一般（近似地）具有形式

$$-2\log(\max_{H_0}L/\max_{H_1}L)=n\log\left(|\hat{\Sigma}_0|/|\hat{\Sigma}_1|\right)\overset{H_0}{\sim}\chi^2_{\nu_1-\nu_0}$$

其中下标0,1分别表示在原假设下和对立假设下的方差估计或参数个数。

拟合优度

拟合优度：$W=n\log\left(|\tilde{L}\tilde{L}^{\mathrm{T}}+\tilde{\Psi}|/|S|\right)$,

$$\text{p-value}=\mathrm{P}(\chi^2_{df}>W),\ \ df=\frac{1}{2}[(p-m)^2-p-m].$$

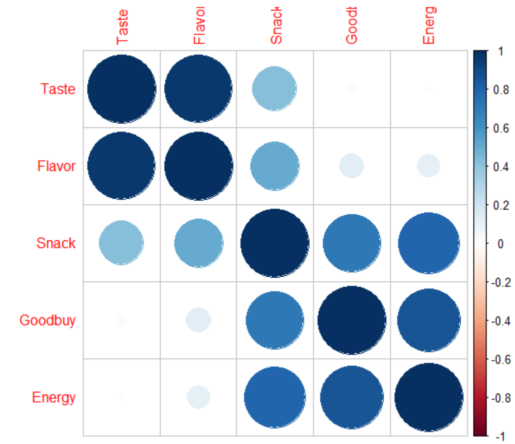
其中 $v_1=p(p+1)/2$,$v_0=pm+p-m(m-1)/2$,

其中L参数个数pm, ψ参数个数p, 约束"$L^{\mathrm{T}}\Psi^{-1}L$为对角矩阵"的个数$m(m-1)/2$.

例2（课本例9.3）. 一项消费偏好研究随机抽取若干消费者对一个新食品的 5项指标进行评价打分 (1-7)，指标为

Taste(口味,酸甜苦咸，不含辣！)， Good buy (价格合理), Flavor(辣，触感等本质特征), Suitable for snack(适合小吃)，Provides lots of energy(热量)

$$
相关系数矩阵 R = \begin{array}{l} \\ \text{Taste} \\ \text{Goodbuy} \\ \text{Flavor} \\ \text{Snack} \\ \text{Energy} \end{array}
\begin{array}{c} \begin{array}{ccccc} \text{Taste} & \text{Goodbuy} & \text{Flavor} & \text{Snack} & \text{Energy} \end{array} \\
\begin{pmatrix} 1 & 0.02 & 0.96 & 0.42 & 0.01 \\ 0.02 & 1 & 0.13 & 0.71 & 0.85 \\ 0.96 & 0.13 & 1 & 0.50 & 0.11 \\ 0.42 & 0.71 & 0.50 & 1 & 0.79 \\ 0.01 & 0.85 & 0.11 & 0.79 & 1 \end{pmatrix} \end{array}
$$



Taste, Flavor 的相关系数较大(0.96)，构成一个组；
Goodbuy, Snack, Energy三者相关系数较大(0.71,0.85,0.79)

1.极大似然法（两因子模型正态因子模型），结果如下：

Variable	Loading F1	Loading F2	Communality h_i^2	Uniqueness $\psi_i = 1 - h_i^2$
Taste	0	0.985	0.972	0.028
Goodbuy	0.873	0	0.763	0.237
Flavor	0.131	0.971	0.960	0.040
Snack	0.817	0.405	0.832	0.168
Energy	0.973	0	0.948	0.052
SS/variance	2.396	2.078		
proportion	0.479	0.416		
Cumu prop	0.479	0.895		

因子含义：

F2：与Taste, Flavor 的相关性较大，“美味”因子

F1：与Goodbuy, Snack, Energy相关性较大，“实用性”因子

两因子可解释总方差的89.5%，可解释Taste方差的97.2%，对Goodbuy解释率最低，76.3%。

R函数: factanal （极大似然法）

> factanal(covmat=R， factors=2，n.obs=100)

输出结果如下：

Uniquenesses:

Taste	Goodbuy	Flavor	Snack	Energy
0.028	0.237	0.040	0.168	0.052

Loadings:

	Factor1	Factor2
Taste	0	0.985
Goodbuy	0.873	0
Flavor	0.131	0.971
Snack	0.817	0.405
Energy	0.973	0

共性方差$h_1^2=0^2+0.985^2$,

个性方差$\psi_1=\sigma_{11}-h_1^2=1-0.985^2=0.028$

	Factor1	Factor2
SS loadings	2.396	2.078
Proportion Var	0.479	0.416
Cumulative Var	0.479	0.895

$F1$的$SS_1=0^2+0.873^2+...+0.973^2=2.396$

总方差$=tr(R)=5$, F1的方差贡献率$=2.396/5=0.479$.

Test of the hypothesis that 2 factors are sufficient. The chi square statistic is 2.22 on 1 degree of freedom. The p-value is 0.136

2.主因子法（不假设正态），结果如下：

Variable	Loading F1	Loading F2	Communality h_i^2	Uniqueness $\psi_i = 1 - h_i^2$
Taste	0.560	0.816	0.980	0.021
Goodbuy	0.777	-0.524	0.879	0.121
Flavor	0.645	0.748	0.976	0.024
Snack	0.939	-0.105	0.893	0.107
Energy	0.798	-0.543	0.932	0.068
SS/variance	2.853	1.806		
proportion	0.571	0.361		
Cumu prop	0.571	0.932		

第一因子解释总方差的57.1%，两因子解释93.2%。

共性方差$\tilde{h}_i^2$分别为：0.98,0.88,...,0.93 两因子模型($m=2$)解释了各个分量方差的88%以上，最高达98%

R软件：主因子方法使用主成分函数： princomp

```
> princomp(covmat=R)
```

```
Importance of components:
                    Comp.1  Comp.2  Comp.3  Comp.4  Comp.5
Standard deviation   1.69    1.34    0.45    0.32    0.18
```

$\sqrt{\lambda_1}=1.69, \lambda_1=1.69^2=2.85$

Loadings:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
[1,]	0.33	-0.61		0.14	0.70
[2,]	0.46	0.39	0.74	-0.28	
[3,]	0.38	-0.56	0.17	0.12	-0.71
[4,]	0.56		-0.60	-0.57	
[5,]	0.47	0.40	-0.22	0.75	

空白处为0

注意：主因子解的loading是$\mathbf{v}_i\sqrt{\lambda_i}$，但主成分分析给出的loading是特征向量$\mathbf{v}_i$而主因子不是$\mathbf{v}_i\sqrt{\lambda_i}$。

将上述loading中的第一列乘上$\sqrt{\lambda_1}=1.69$，即得到因子分析中的第一列loading,

第2列乘上$\sqrt{\lambda_2}=1.34$，即得到因子分析中的第2列loading.

所以使用两因子模型即可几乎完美复原来的相关系数矩阵.

$$\tilde{\mathbf{L}}\tilde{\mathbf{L}}' + \tilde{\mathbf{\Psi}} = \begin{bmatrix} .56 & .82 \\ .78 & -.53 \\ .65 & .75 \\ .94 & -.10 \\ .80 & -.54 \end{bmatrix} \begin{bmatrix} .56 & .78 & .65 & .94 & .80 \\ .82 & -.53 & .75 & -.10 & -.54 \end{bmatrix}$$

$$+ \begin{bmatrix} .02 & 0 & 0 & 0 & 0 \\ 0 & .12 & 0 & 0 & 0 \\ 0 & 0 & .02 & 0 & 0 \\ 0 & 0 & 0 & .11 & 0 \\ 0 & 0 & 0 & 0 & .07 \end{bmatrix} = \begin{bmatrix} 1.00 & .01 & .97 & .44 & .00 \\ & 1.00 & .11 & .79 & .91 \\ & & 1.00 & .53 & .11 \\ & & & 1.00 & .81 \\ & & & & 1.00 \end{bmatrix} \approx \Sigma = \begin{bmatrix} 1.00 & .02 & .96 & .42 & .01 \\ .02 & 1.00 & .13 & .71 & .85 \\ .96 & .13 & 1.00 & .50 & .11 \\ .42 & .71 & .50 & 1.00 & .79 \\ .01 & .85 & .11 & .79 & 1.00 \end{bmatrix}$$

因子旋转

例1（课本例9.8）．$n=220$个男生6门课考试成绩的相关系数矩阵如下：

$$\mathbf{R}=\begin{array}{c} \begin{array}{cccccc} \text{Gaelic} & \text{English} & \text{History} & \text{Arithmetic} & \text{Algebra} & \text{Geometry} \end{array} \\ \begin{bmatrix} 1.0 & .439 & .410 & .288 & .329 & .248 \\ & 1.0 & .351 & .354 & .320 & .329 \\ & & 1.0 & .164 & .190 & .181 \\ & & & 1.0 & .595 & .470 \\ & & & & 1.0 & .464 \\ & & & & & 1.0 \end{bmatrix} \end{array}$$

Gaelic：爱尔兰凯尔特人的盖尔语

$m=2$时，极大似然估计得到loading的估计如下

Variable	Estimated factor loadings F_1	F_2	Communalities $\hat{h}_i^2$
1. Gaelic	.553	.429	.490
2. English	.568	.288	.406
3. History	.392	.450	.356
4. Arithmetic	.740	−.273	.623
5. Algebra	.724	−.211	.569
6. Geometry	.595	−.132	.372

F_1的loading都>0, F_1与各门成绩都正相关，与算数、代数相关性最大。F_1是一般智力因子。
F_2与文科类正相关，与数学类负相关，F_2是数学因子（或文科因子）

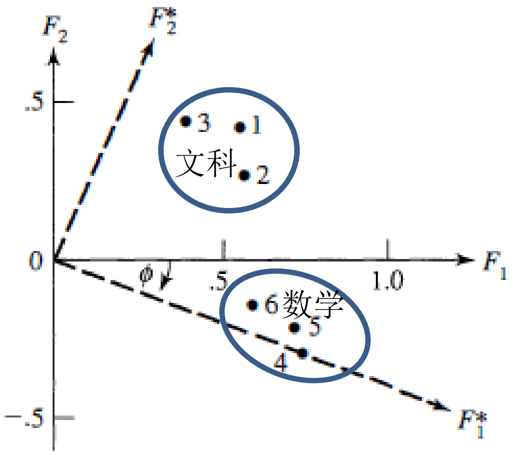
右图(F_1 - F_2 坐标系)是前两列载荷的散点图

$$(l_{i1}, l_{i2}), i = 1, \ldots, 5.$$

明显地，变量1,2,3是一簇(文科),

变量4,5,6是一簇(数学),

F_2能很好地区分这两类, F_2代表文理差别 (“文理差别”不好解释，适当旋转后$F2$更容易解释)。



旋转坐标系（O为m阶正交阵）：$L\mathbf{F} = LOO^T\mathbf{F} = L^*\mathbf{F}^*$

$$L_{p\times m} \to L^*{}_{p\times m} = LO, \quad \mathbf{F}_{m\times 1} \to \mathbf{F}^*{}_{m\times 1} = O^{\mathrm{T}}\mathbf{F},$$

旋转后的坐标系F_1*-F_2*(虚线)中：

- 点4,5,6基本在F_1*轴上，所以F_1*可理解为数学因子;
- 点1,2,3靠近F_2*轴，所以F_2*可理解为文科因子。

因子旋转原则：旋转后变量的载荷坐标最好在坐标轴附近，即最好某些载荷为0，其它同号。

旋转之后的载荷

Variable	Estimated rotated factor loadings F_1^*	F_2^*	Communalities $\hat{h}_i^{*2} = \hat{h}_i^2$
1. Gaelic	.369	$\boxed{.594}$	.490
2. English	.433	$\boxed{.467}$	.406
3. History	.211	$\boxed{.558}$	.356
4. Arithmetic	$\boxed{.789}$	.001	.623
5. Algebra	$\boxed{.752}$	.054	.568
6. Geometry	$\boxed{.604}$	.083	.372

F_1^* :数学因子　　F_2^*：文科因子

注：上述旋转使得载荷的每一列尽量趋于两极化，最好变量的载荷坐标在坐标轴附近，即部分变量的载荷接近0，其它相对较大且最好同号。

varimax方法：旋转极大化每列载荷的绝对值或平方值的方差。

```
R:  factanal(x, rotation = ″varimax”, ...)
```

因子旋转

假设现有载荷和因子为$L,\mathbf{f}$,对任何正交矩阵$O_{m\times m}$

旋转载荷矩阵和因子向量：$\tilde{L}_{p\times m}=L_{p\times m}O_{m\times m}$， $\tilde{\mathbf{f}}=O^{\mathrm{T}}\mathbf{f}$

则$L\mathbf{f}=\tilde{L}\tilde{\mathbf{f}}$,模型中除了载荷和因子，其它都不受影响。

Varimax 准则

假设正交O旋转之后的载荷矩阵$\tilde{L}=LO=(\tilde{l}_{ij}\ ,1\le i\le p,1\le j\le m)$,

令$l_{ij}^{*}=\tilde{l}_{ij}\ /h_i$(标准化），令第$j$列载荷的平方$(l_{1j}^{*})^2,\ldots(l_{pj}^{*})^2$的样本方差

$$v_j=\frac{1}{p}\sum_{i=1}^{p}\left(l_{ij}^{*2}-\mu\right)^2,\ \mu=\sum_{i=1}^{p}l_{ij}^{*2}/p$$

令$V_{\mathbf{T}}=v_1+\ldots+v_m$. varimax准则寻找正交变换$\mathbf{T}$使得$V_{\mathbf{T}}$最大。

注：斜交旋转（oblique rotation）不要求正交矩阵旋转，即新的坐标轴不必正交。R中提供了promax斜交旋转函数。

因子得分

教育、心理测量学中的一个重要问题是通过考试测量潜在的能力，潜在因子是随机变量，其预测值称为因子得分(score)。

主因子法

第i个个体的因子j的估计(标准化的主成分):

$$\hat{f}_{ij} = y_{ij} / \sqrt{\lambda_j} = u_{ij}, U\text{的}(i, j)\text{元}$$

极大似然法

假设正态因子模型, 个体i的因子得分通常由如下方法预测:

- Regression: $\mathbf{f}_i$的加权最小二乘估计为($L, \boldsymbol{\mu}, \Psi$代入极大似然估计):

$$\hat{\mathbf{f}}_i = (L^{\mathrm{T}} \Psi^{-1} L)^{-1} L^{\mathrm{T}} \Psi^{-1} (\mathbf{x}_i - \boldsymbol{\mu}),$$

- Bartlett: $\mathbf{f}_i$的条件期望预测

$$\hat{\mathbf{f}}_i = E(\mathbf{f}_i \mid \mathbf{x}_i) = (I_m + L^{\mathrm{T}} \Psi^{-1} L)^{-1} L^{\mathrm{T}} \Psi^{-1} (\mathbf{x}_i - \boldsymbol{\mu})\text{。}$$

注：异方差模型$\mathbf{x}_i = \boldsymbol{\mu} + L\mathbf{f}_i + \boldsymbol{\varepsilon}_i$，$\mathrm{var}(\boldsymbol{\varepsilon}_i) = \Psi$中，将$\mathbf{f}_i$看作固定参数，应用广义/加权LS方法即得第一种预测。Bartlett预测是$E(\mathbf{f}_i \mid \mathbf{x}_i)$, 后验估计。

```
R:  factanal(x, scores = c(“regression”,” Bartlett”))
```

求解Barlett预测：

假设因子模型：$\mathbf{x}=\boldsymbol{\mu}+L\mathbf{f}+\boldsymbol{\varepsilon}, \mathrm{var}(\boldsymbol{\varepsilon})=\Psi$，$\mathbf{f},\boldsymbol{\varepsilon}$独立。

假设正态模型：$\mathbf{f}\sim N_m(0,I_m), \boldsymbol{\varepsilon}\sim N_p(0,\Psi)$,不妨设$\boldsymbol{\mu}=0.$

$$\text{则}\begin{pmatrix}\mathbf{f}\\ \mathbf{x}\end{pmatrix}\sim N_{m+p}\left(0,\begin{pmatrix}I_m & L^\top\\ L & LL^\top+\Psi\end{pmatrix}\right),$$

$$\mathbf{f}\mid\mathbf{x}\sim N_m\left(L^\top(LL^\top+\Psi)^{-1}\mathbf{x}, I_m-L^\top(LL^\top+\Psi)^{-1}L\right),$$

$$E(\mathbf{f}\mid\mathbf{x})=L^\top(LL^\top+\Psi)^{-1}\mathbf{x}=(I_m+L^\top\Psi^{-1}L)^{-1}L^\top\Psi^{-1}\mathbf{x}$$

Sherman - Morrison - Woodbury :

$$(LL^\top+\Psi)^{-1}=\Psi^{-1}-\Psi^{-1}L(I_m+L^\top\Psi^{-1}L)^{-1}L^\top\Psi^{-1}$$