第七章　主成分分析

§7.1　引　言

我们在作数据分析处理时，涉及的样品往往包含有多个测量指标(比如 p 个指标)，较多的指标会带来分析问题的复杂性。然而，这些指标彼此之间常常存在着一定程度的、有时甚至是相当高的相关性，这就使含在观测数据中的信息在一定程度上有所重迭。主成分分析就是一种通过降维技术把多个指标约化为少数几个综合指标的统计分析方法。这些综合指标能够反映原始指标的绝大部分信息，它们通常表示为原始 p 个指标的某种线性组合。为了使这些综合指标所含的信息互不重迭，应要求它们之间互不相关。

例如，考虑 $p=2$ 的情形，假设共有 n 个样品，每个样品都测量了两个指标(x_1,x_2)，它们大致分布在一个椭圆内，如图 7.1 所示。显然，在坐标系 x_1Ox_2 中，n 个点的坐标 x_1 和 x_2 呈现某种(线性)相关性。我们将该坐标系按逆时针方向旋转某个角度 θ 变成新坐标系 y_1Oy_2，这里 y_1 是椭圆的长轴方向，y_2 是短轴方向，如图 7.1 所示。旋转公式为

$$\begin{cases} y_1 = x_1\cos\theta + x_2\sin\theta \\ y_2 = -x_1\sin\theta + x_2\cos\theta \end{cases} \tag{7.1.1}$$

易见，n 个点在新坐标系下的坐标 y_1 和 y_2 几乎不相关。y_1 和 y_2 称为原始变量 x_1 和 x_2 的综合变量，n 个点在 y_1 轴上的方差达到最大，即在此方向上所含的有关 n 个样品间差异的信息是最多的。因此，若欲将二维空间的点投影到某个一维方向，则选择 y_1 轴方向能使信息的损失降低到最小。我们称 y_1 轴为第一主成分，而与 y_1

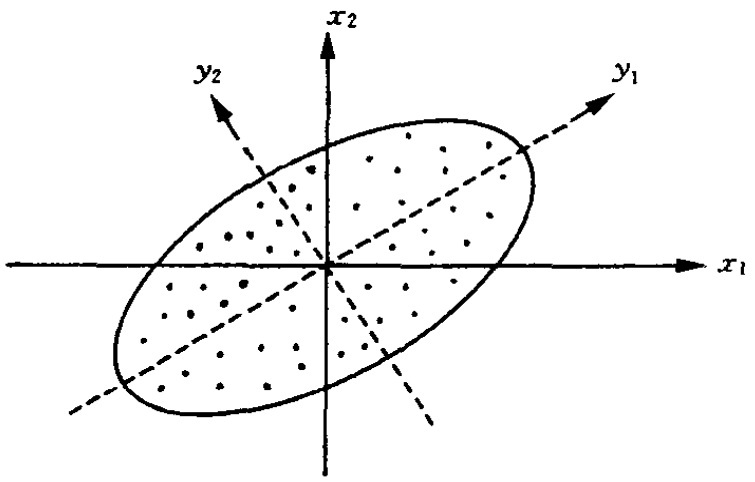


图 7.1

轴正交的 y_2 轴，有着较小的方差，称为第二主成分。图 7.1 中，第一主成分的效果与椭圆的形状有很大的关系，椭圆越是扁平，n 个点在 y_1 轴上的方差就相对越大，在 y_2 轴上的方差就相对越小，用第一主成分代替二维空间所造成的信息损失也就越小。考虑这样两种极端的情形：一种是椭圆的长轴与短轴的长度相等，即椭圆变成圆，第一主成分只含有二维空间点的约一半信息，若仅用这一个综合变量，则将损失约 50%的信息，这显然是不可取的。造成它的原因是，原始变量 x_1 和 x_2 的相关程度几乎为零，也就是说，x_1 和 x_2 所包含的信息几乎互不重迭，因此无法用一个一维的综合变量来代替它们。另一种是椭圆扁平到了极限，变成 y_1 轴上的一条线段，第一主成分包含有二维空间点的 100%信息，仅用这一个综合变量代替原始的二维变量不会有任何的信息损失，此时的主成分分析效果是非常理想的。其原因是，原始变量 x_1 和 x_2 可以相互确定，它们所含的信息是完全相同的，因此使用一个综合变量也就完全够了。

§7.2 总体的主成分

一、主成分的定义及导出

设 $x=(x_1,\cdots,x_p)'$ 为一个 p 维随机向量，并假定二阶矩存在，

记 $\mu=E(\boldsymbol{x})$，$\Sigma=V(\boldsymbol{x})$。考虑如下的线性变换

$$\begin{cases} y_1=a_{11}x_1+\cdots+a_{1p}x_p=\boldsymbol{a}_1'\boldsymbol{x} \\ \quad\vdots \\ y_p=a_{p1}x_1+\cdots+a_{pp}x_p=\boldsymbol{a}_p'\boldsymbol{x} \end{cases} \tag{7.2.1}$$

我们希望 y_1 是 $x_1,\cdots,x_p$ 的一切线性函数中方差最大的。因为 $V(\boldsymbol{a}_1'\boldsymbol{x})=\boldsymbol{a}_1'\Sigma\boldsymbol{a}_1$，对任意的常数 k，$V(k\boldsymbol{a}_1'\boldsymbol{x})=k^2V(\boldsymbol{a}_1'\boldsymbol{x})=k^2\boldsymbol{a}_1'\Sigma\boldsymbol{a}_1$，所以如不对 $\boldsymbol{a}_1$ 加以限制，就会使问题变得没有什么意义。于是常常限制

$$\boldsymbol{a}_i'\boldsymbol{a}_i=1,\quad i=1,\cdots,p \tag{7.2.2}$$

故我们希望在(7.2.2)式的条件下寻求向量 $\boldsymbol{a}_1$，使得 $V(y_1)=\boldsymbol{a}_1'\Sigma\boldsymbol{a}_1$ 达到最大，y_1 就称为第一主成分。

设 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_p\geqslant0$(因为 Σ 非负定)为 Σ 的特征值，$\boldsymbol{t}_1,\cdots,\boldsymbol{t}_p$ 为相应的单位特征向量，且相互正交。则由(1.6.6)式知，

$$\Sigma=T\begin{pmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_p\end{pmatrix}T'=\sum_{i=1}^{p}\lambda_i\boldsymbol{t}_i\boldsymbol{t}_i' \tag{7.2.3}$$

其中 $T=(\boldsymbol{t}_1,\cdots,\boldsymbol{t}_p)$ 为正交矩阵。

对 p 维单位向量 $\boldsymbol{a}$，有

$$\begin{aligned}\boldsymbol{a}'\Sigma\boldsymbol{a}&=\sum_{i=1}^{p}\lambda_i\boldsymbol{a}'\boldsymbol{t}_i\boldsymbol{t}_i'\boldsymbol{a}=\sum_{i=1}^{p}\lambda_i(\boldsymbol{a}'\boldsymbol{t}_i)^2\\ &\leqslant\lambda_1\sum_{i=1}^{p}(\boldsymbol{a}'\boldsymbol{t}_i)^2=\lambda_1\sum_{i=1}^{p}\boldsymbol{a}'\boldsymbol{t}_i\boldsymbol{t}_i'\boldsymbol{a}=\lambda_1\boldsymbol{a}'TT'\boldsymbol{a}\\ &=\lambda_1\boldsymbol{a}'\boldsymbol{a}=\lambda_1\end{aligned}$$

当取 $\boldsymbol{a}=\boldsymbol{t}_1$ 时，有

$$\boldsymbol{t}_1'\Sigma\boldsymbol{t}_1=\boldsymbol{t}_1'(\lambda_1\boldsymbol{t}_1)=\lambda_1 \tag{7.2.4}$$

所以，$y_1=\boldsymbol{t}_1'\boldsymbol{x}$ 就是所求的第一主成分，它的方差具有最大值 λ_1。如果第一主成分所含信息不够多，还不足以代表原始的 p 个变量，则需考虑使用 y_2，为了使 y_2 所含的信息与 y_1 不重迭，应要求

$$\operatorname{cov}(y_1,y_2)=0 \tag{7.2.5}$$

于是，我们在约束条件(7.2.2)式和(7.2.5)式下寻求向量 a_2，使得 $V(y_2)=a_2'\Sigma a_2$ 达到最大，所求的 y_2 称为第二主成分。类似地，我们可以再定义第三主成分、…、第 p 主成分。一般来说，x 的第 i 主成分 $y_i=a_i'x$ 是指：在约束条件(7.2.2)式和

$$\operatorname{cov}(y_k,y_i)=0,\quad k=1,\cdots,i-1 \tag{7.2.6}$$

下寻求 a_i，使得 $V(y_i)=a_i'\Sigma a_i$ 达到最大。

现在我们来求 p 维单位向量 a，使得 $y_2=a'x$ 为第二主成分。由(7.2.6)式知

$$\begin{aligned}\operatorname{cov}(y_1,y_2)&=\operatorname{cov}(t_1'x,a'x)=a'\Sigma t_1\\&=\lambda_1 a't_1=0\end{aligned}$$

于是

$$a't_1=0$$

从而

$$\begin{aligned}V(y_2)=a'\Sigma a&=\sum_{i=1}^{p}\lambda_i(a't_i)^2=\sum_{i=2}^{p}\lambda_i(a't_i)^2\\&\leqslant\lambda_2\sum_{i=2}^{p}(a't_i)^2=\lambda_2\sum_{i=1}^{p}(a't_i)^2\\&=\lambda_2 a'TT'a=\lambda_2 a'a=\lambda_2\end{aligned}$$

若取 $a=t_2$，则有

$$t_2'\Sigma t_2=t_2'(\lambda_2 t_2)=\lambda_2 \tag{7.2.7}$$

所以，$y_2=t_2'x$ 就是所求的第二主成分，具有方差 λ_2。一般地，我们可求得第 i 主成分为 $y_i=t_i'x$，它具有方差 λ_i，$i=1,\cdots,p$。

二、主成分的性质

1. 主成分的均值和协方差矩阵

记

$$y=\begin{pmatrix}y_1\\\vdots\\y_p\end{pmatrix},\quad \upsilon=Ey,\quad \Lambda=\begin{pmatrix}\lambda_1&&0\\&\ddots&\\0&&\lambda_p\end{pmatrix}$$

由于

$$\boldsymbol{y}=T'\boldsymbol{x} \tag{7.2.8}$$

故

$$\upsilon=E(T'\boldsymbol{x})=T'\mu \tag{7.2.9}$$

$$V(\boldsymbol{y})=T'V(\boldsymbol{x})T=T'\Sigma T=\Lambda \tag{7.2.10}$$

2. 主成分的总方差

由于

$$\mathrm{tr}(\Lambda)=\mathrm{tr}(T'\Sigma T)=\mathrm{tr}(\Sigma TT')=\mathrm{tr}(\Sigma)$$

所以

$$\sum_{i=1}^{p}\lambda_i=\sum_{i=1}^{p}\sigma_{ii} \tag{7.2.11}$$

或

$$\sum_{i=1}^{p}V(y_i)=\sum_{i=1}^{p}V(x_i) \tag{7.2.11}'$$

由此可以看出，主成分分析把 p 个原始变量 $x_1,\cdots,x_p$ 的总方差 $\mathrm{tr}(\Sigma)$分解成了 p 个不相关的变量 $y_1,\cdots,y_p$ 的方差之和 $\sum_{i=1}^{p}\lambda_i$。主成分分析的目的就是为了减少变量的个数，一般是不会使用所有 p 个主成分的，忽略一些带有较小方差的主成分将不会给总方差带来大的影响。我们称 $\lambda_k\Big/\sum_{i=1}^{p}\lambda_i$ 为主成分 y_k 的贡献率；第一主成分的贡献率最大，这表明 $y_1=t_i'\boldsymbol{x}$ 综合原始变量 $x_1,\cdots,x_p$ 的能力最强，而 $y_2,\cdots,y_p$ 的综合能力依次递减。若只取 $m(<p)$个主成分，则称 $\sum_{i=1}^{m}\lambda_i\Big/\sum_{i=1}^{p}\lambda_i$ 为主成分 $y_1,\cdots,y_m$ 的累计贡献率，累计贡献率表明 $y_1,\cdots,y_m$ 综合 $x_1,\cdots,x_p$ 的能力。通常取 m，使得累计贡献率达到一个较高的百分数(如 85%以上)。

3. 变量 x_i 与主成分 y_j 之间的相关系数

x_i 与 y_j 的相关系数

$$\rho(x_i,y_j)=\frac{\mathrm{cov}(x_i,y_j)}{\sqrt{V(x_i)}\sqrt{V(y_j)}}=\frac{\mathrm{cov}(x_i,y_j)}{\sqrt{\sigma_{ii}}\sqrt{\lambda_j}} \tag{7.2.12}$$

由(7.2.8)式知

$$\boldsymbol{x}=T\boldsymbol{y} \tag{7.2.13}$$

若记 $T=(t_{ij})$，则

$$x_i=t_{i1}y_1+\cdots+t_{ip}y_p \tag{7.2.13}'$$

所以

$$\operatorname{cov}(x_i,y_j)=\operatorname{cov}(t_{ij}y_j,y_j)=t_{ij}\lambda_j$$

代入(7.2.12)式，得

$$\rho(x_i,y_j)=\frac{\sqrt{\lambda_j}}{\sqrt{\sigma_{ii}}}\,t_{ij} \tag{7.2.14}$$

所有这些相关系数列于表 7.1 中。在实际应用中，通常我们只对 x_i $(i=1,\cdots,p)$与 $y_j(j=1,\cdots,m)$的相关系数感兴趣，因此往往只列出表 7.1 的前 m 列，即形成 $p\times m$ 表。

表 7.1　　变量 x_i 与主成分 y_j 之间的相关系数

主成分 / 原始变量	y_1	y_2	…	y_p
x_1	$\frac{\sqrt{\lambda_1}}{\sqrt{\sigma_{11}}}t_{11}$	$\frac{\sqrt{\lambda_2}}{\sqrt{\sigma_{11}}}t_{12}$	…	$\frac{\sqrt{\lambda_p}}{\sqrt{\sigma_{11}}}t_{1p}$
x_2	$\frac{\sqrt{\lambda_1}}{\sqrt{\sigma_{22}}}t_{21}$	$\frac{\sqrt{\lambda_2}}{\sqrt{\sigma_{22}}}t_{22}$	…	$\frac{\sqrt{\lambda_p}}{\sqrt{\sigma_{22}}}t_{2p}$
⋮	⋮	⋮		⋮
x_p	$\frac{\sqrt{\lambda_1}}{\sqrt{\sigma_{pp}}}t_{p1}$	$\frac{\sqrt{\lambda_2}}{\sqrt{\sigma_{pp}}}t_{p2}$	…	$\frac{\sqrt{\lambda_p}}{\sqrt{\sigma_{pp}}}t_{pp}$

4. m 个主成分对原始变量的贡献率

前面提到的累计贡献率这个概念度量了主成分 $y_1,\cdots,y_m$ 从原始变量 $x_1,\cdots,x_p$ 中提取信息的多少，那么，$y_1,\cdots,y_m$ 包含有 x_i $(i=1,\cdots,p)$的多少信息应该用什么指标来度量呢？这个指标就是

x_i 与 $y_1,\cdots,y_m$ 的复相关系数的平方、称为 m 个主成分 $y_1,\cdots,y_m$ 对原始变量 x_i 的贡献率，记为 $\rho^2_{i\cdot 1\cdots m}$。

由(3.3.15)式知

$$\rho^2_{i\cdot 1\cdots m}$$

$$=\frac{(\operatorname{cov}(x_i,y_1),\cdots,\operatorname{cov}(x_i,y_m))\begin{pmatrix}V(y_1) & & 0\\ & \ddots & \\ 0 & & V(y_m)\end{pmatrix}^{-1}\begin{pmatrix}\operatorname{cov}(x_i,y_1)\\ \vdots\\ \operatorname{cov}(x_i,y_m)\end{pmatrix}}{V(x_i)}$$

$$=(\rho(x_i,y_1),\cdots,\rho(x_i,y_m))\begin{pmatrix}\rho(x_i,y_1)\\ \vdots\\ \rho(x_i,y_m)\end{pmatrix}$$

$$=\sum_{j=1}^{m}\rho^2(x_i,y_j)=\sum_{j=1}^{m}\lambda_j t_{ij}^2/\sigma_{ii} \tag{7.2.15}$$

这些值列于表 7.2 中。

表 7.2　　主成分 $y_1,\cdots,y_m$ 对原始变量 x_i 的贡献率

$\rho^2_{1\cdot 1\cdots m}$	$\rho^2_{2\cdot 1\cdots m}$	$\cdots$	$\rho^2_{p\cdot 1\cdots m}$
$\sum_{j=1}^{m}\lambda_j t_{1j}^2/\sigma_{11}$	$\sum_{j=1}^{m}\lambda_j t_{2j}^2/\sigma_{22}$	$\cdots$	$\sum_{j=1}^{m}\lambda_j t_{pj}^2/\sigma_{pp}$

由(7.2.13)′式知，$y_1,\cdots,y_p$ 对 x_i 的贡献率 $\rho^2_{i\cdot 1\cdots p}=1$，所以

$$\sum_{j=1}^{p}\rho^2(x_i,y_j)=\sum_{j=1}^{p}\lambda_j t_{ij}^2/\sigma_{ii}=1 \tag{7.2.16}$$

例 7.2.1　设 $\boldsymbol{x}=(x_1,x_2,x_3)'$ 的协方差矩阵为

$$\Sigma=\begin{pmatrix}1 & -2 & 0\\ -2 & 5 & 0\\ 0 & 0 & 2\end{pmatrix}$$

其特征值为

$$\lambda_1=5.83,\quad \lambda_2=2.00,\quad \lambda_3=0.17$$

相应的特征向量为

$$t_1=\begin{pmatrix}0.383\\-0.924\\0.000\end{pmatrix},\quad t_2=\begin{pmatrix}0\\0\\1\end{pmatrix},\quad t_3=\begin{pmatrix}0.924\\0.383\\0.000\end{pmatrix}$$

若只取一个主成分，则贡献率为

$$5.83/(5.83+2.00+0.17)=0.72875=72.875\%$$

进一步计算主成分对每个变量的贡献率，并列于下表

i	$\rho(y_1,x_i)$	$\rho^2_{i\cdot 1}$	$\rho(y_2,x_i)$	$\rho^2_{i\cdot 12}$
1	0.925	0.855	0.000	0.855
2	−0.998	0.996	0.000	0.996
3	0.000	0.000	1.000	1.000

可见，y_1 对第三个变量的贡献率为零，这是因为 x_3 与 x_1 和 x_2 都不相关，在 y_1 中未包含一点有关 x_3 的信息，这时仅取一个主成分就显得不够了，故应再取 y_2，此时累计贡献率为

$$(5.83+2.00)/8=97.875\%$$

(y_1,y_2) 对每个变量 x_i 的贡献率分别为 $\rho^2_{1\cdot 12}=85.5\%$，$\rho^2_{2\cdot 12}=99.6\%$，$\rho^2_{3\cdot 12}=100\%$，都比较高。

三、载荷矩阵

(7.2.13)式可以表达为

$$\begin{pmatrix}x_1\\x_2\\\vdots\\x_p\end{pmatrix}=\begin{pmatrix}t_{11}&t_{12}&\cdots&t_{1p}\\t_{21}&t_{22}&\cdots&t_{2p}\\\vdots&\vdots&&\vdots\\t_{p1}&t_{p2}&\cdots&t_{pp}\end{pmatrix}\begin{pmatrix}y_1\\y_2\\\vdots\\y_p\end{pmatrix}$$

可见，x 的每一分量均可表示成主成分 $y_1,y_2,\cdots,y_p$ 的线性组合。如果我们选取前 m 个主成分，并记

$$\begin{pmatrix}\hat{x}_1\\\vdots\\\hat{x}_p\end{pmatrix}=\begin{pmatrix}t_{11}&\cdots&t_{1m}\\\vdots&&\vdots\\t_{p1}&\cdots&t_{pm}\end{pmatrix}\begin{pmatrix}y_1\\\vdots\\y_m\end{pmatrix}\qquad(7.2.17)$$

则有

$$\begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} = \begin{pmatrix} t_{11} & \cdots & t_{1m} \\ \vdots & & \vdots \\ t_{p1} & \cdots & t_{pm} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} + \begin{pmatrix} t_{1,m+1} & \cdots & t_{1p} \\ \vdots & & \vdots \\ t_{p,m+1} & \cdots & t_{pp} \end{pmatrix} \begin{pmatrix} y_{m+1} \\ \vdots \\ y_p \end{pmatrix}$$

$$\approx \begin{pmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_p \end{pmatrix} \tag{7.2.18}$$

即 $\boldsymbol{x}$ 的每一个分量 x_i 均可近似地表示为前 m 个主成分 $y_1, y_2, \cdots, y_m$ 的线性组合。由于 $y_1, y_2, \cdots, y_p$ 是不相关的，故从(7.2.13)′式与(7.2.17)式得

$$V(x_i) = \sum_{j=1}^{p} t_{ij}^2 V(y_j) = \sum_{j=1}^{p} \lambda_j t_{ij}^2 \tag{7.2.19}$$

和

$$V(\hat{x}_i) = \sum_{j=1}^{m} t_{ij}^2 V(y_j) = \sum_{j=1}^{m} \lambda_j t_{ij}^2 \tag{7.2.20}$$

比较(7.2.19)与(7.2.20)两式，可以看出用 $\hat{x}_i$ 去代替 x_i 时，$\hat{x}_i$ 一般能说明 x_i 方差的大部分，所占的比例为

$$\sum_{j=1}^{m} \lambda_j t_{ij}^2 \Big/ \sum_{j=1}^{p} \lambda_j t_{ij}^2 \tag{7.2.21}$$

可见，$\lambda_j (j=1, \cdots, m)$相对越大，上述比值一般就越大，说明用 m 个主成分 $y_1, y_2, \cdots, y_m$ 来综合反映原始变量 $x_1, x_2, \cdots, x_p$ 的效果也就越好。另一方面这个比值也取决于 t_{ij}，我们称 t_{ij} 为第 i 个原始变量 x_i 在第 j 个主成分 y_j 上的载荷，而称由矩阵 T 的前 m 列组成的 $p \times m$ 矩阵为主成分的载荷矩阵，记为 A，即

$$A = \begin{pmatrix} t_{11} & \cdots & t_{1m} \\ \vdots & & \vdots \\ t_{p1} & \cdots & t_{pm} \end{pmatrix} \tag{7.2.22}$$

我们来分析一下载荷矩阵 A 中元素 t_{ij} 代表的意义。A 的第 j 列反映了主成分 y_j 对原始变量 $\boldsymbol{x}$ 各分量的作用。如果 A 中出现了一列中只有一个非零元素，不妨设第 1 列为 $(1, 0, \cdots, 0)'$，这时

$$\begin{pmatrix}\hat{x}_1\\ \hat{x}_2\\ \vdots\\ \hat{x}_p\end{pmatrix}=\begin{pmatrix}1 & t_{12} & \cdots & t_{1m}\\ 0 & t_{22} & \cdots & t_{2m}\\ \vdots & \vdots & & \vdots\\ 0 & t_{p2} & \cdots & t_{pm}\end{pmatrix}\begin{pmatrix}y_1\\ y_2\\ \vdots\\ y_m\end{pmatrix}$$

即

$$\hat{x}_1=y_1+t_{12}y_2+\cdots+t_{1m}y_m$$

$$\hat{x}_i=t_{i2}y_2+\cdots+t_{im}y_m,\quad i=2,\cdots,p$$

则表明第一主成分只对原始变量 x_1 有作用，而对其它的原始变量 $x_2,\cdots,x_p$ 都不起作用；如果 A 中某一列的元素均不为零，则表明这一列相应的主成分对各原始变量 $x_1,\cdots,x_p$ 都起作用。因此，我们把前一种主成分称为特殊成分，而把后一种称为公共成分。由此可见，载荷矩阵的具体形式可供我们分析每一主成分对诸原始变量的贡献。所以，在主成分分析中，在求出主成分的同时，还应求出载荷矩阵。

实际应用中，一般先对 m 个主成分 $y_1,y_2,\cdots,y_m$ 的方差施行标准化，然后再求出主成分的载荷矩阵。即令

$$f=\begin{pmatrix}f_1\\ \vdots\\ f_m\end{pmatrix}=\begin{pmatrix}\frac{1}{\sqrt{\lambda_1}} & & 0\\ & \ddots & \\ 0 & & \frac{1}{\sqrt{\lambda_m}}\end{pmatrix}\begin{pmatrix}y_1\\ \vdots\\ y_m\end{pmatrix}$$

于是

$$V(f)=\begin{pmatrix}\frac{1}{\sqrt{\lambda_1}} & & 0\\ & \ddots & \\ 0 & & \frac{1}{\sqrt{\lambda_m}}\end{pmatrix}V\begin{pmatrix}y_1\\ \vdots\\ y_m\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{\lambda_1}} & & 0\\ & \ddots & \\ 0 & & \frac{1}{\sqrt{\lambda_m}}\end{pmatrix}=I_m$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_m} \end{pmatrix} \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}$$

所以，由(7.2.17)式得

$$\hat{\boldsymbol{x}} = \begin{pmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_p \end{pmatrix} = \begin{pmatrix} t_{11} & \cdots & t_{1m} \\ \vdots & & \vdots \\ t_{p1} & \cdots & t_{pm} \end{pmatrix} \begin{pmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_m} \end{pmatrix} \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}$$

$$= Bf \tag{7.2.23}$$

其中 $B=A\mathrm{diag}(\sqrt{\lambda_1},\cdots,\sqrt{\lambda_m})$，这是用标准化的主成分近似表示原始变量 $x_1,x_2,\cdots,x_p$ 的公式。此时 $\hat{x}_i$ 的方差可表示为

$$V(\hat{x}_i)=\sum_{j=1}^{m} b_{ij}^2, \quad i=1,2,\cdots,p \tag{7.2.24}$$

这里 $B=(b_{ij})$，且有 $b_{ij}=t_{ij}\sqrt{\lambda_j}$，称 B 为标准化主成分 $\boldsymbol{f}$ 的载荷矩阵。

如果对标准化的主成分施行一个正交变换，即令

$$\boldsymbol{h}=\begin{pmatrix} h_1 \\ \vdots \\ h_m \end{pmatrix}=\Gamma'\begin{pmatrix} \boldsymbol{f}_1 \\ \vdots \\ f_m \end{pmatrix}=\Gamma'\boldsymbol{f} \tag{7.2.25}$$

其中 Γ 为一正交矩阵，则有 $V(\boldsymbol{h})=\Gamma'V(\boldsymbol{f})\Gamma=\Gamma'\Gamma=I_m$，故 h 仍然是标准化的主成分。又 $\boldsymbol{f}=\Gamma h$，因此 $\hat{\boldsymbol{x}}$ 也可用 $\boldsymbol{h}$ 来表示，即有

$$\hat{\boldsymbol{x}}=B\boldsymbol{f}=B\Gamma\boldsymbol{h}=C\boldsymbol{h} \tag{7.2.26}$$

其中 $C=B\Gamma=(c_{ij})$ 为标准化主成分 h 的载荷矩阵。与(7.2.24)式同样的道理，有

$$V(\hat{x}_i)=\sum_{j=1}^{m} c_{ij}^2, \quad i=1,2,\cdots,p \tag{7.2.27}$$

故

$$\sum_{j=1}^{m} b_{ij}^2 = \sum_{j=1}^{m} c_{ij}^2, \quad i=1,2,\cdots,p \tag{7.2.28}$$

这表明标准化的主成分经过正交变换之后，$\hat{x}_i$ 的方差及其表达形式都是不变的。这种不变的性质在很大程度上允许我们寻求这样标准化主成分的正交变换，使得变换后的载荷矩阵具有更鲜明的实际意义。

四、从相关矩阵出发求主成分

我们前面讨论的主成分是从协方差矩阵 Σ 出发求得的，其结果受原始 p 个变量单位的影响。不同的变量往往有不同的单位，对同一变量使用不同的单位会产生不同的主成分，主成分会过于照顾方差(σ_{ii})大的变量 x_i，而对方差小的变量却照顾得不够。为使主成分分析能够均等地对待每一个原始变量，消除由于单位的不同而可能带来的一些不合理的影响，常常将各原始变量作标准化处理，即令

$$x_i^* = \frac{x_i - E(x_i)}{\sqrt{V(x_i)}}, \quad i=1,\cdots,p \tag{7.2.29}$$

显然，$\boldsymbol{x}^* = (x_1^*,\cdots,x_p^*)'$ 的协方差矩阵就是 $\boldsymbol{x}$ 的相关矩阵 R。

从 R 出发求得主成分的方法与从 Σ 出发是完全类似的，并且主成分的一些性质具有更简洁的数学形式。首先对 R 进行谱分解，即存在正交矩阵 $T^* = (t_1^*,\cdots,t_p^*) = (t_{ij}^*)$，使得

$$R = T^* \Lambda^* T^{*\prime} = \sum_{i=1}^{p} \lambda_i^* t_i^* t_i^{*\prime} \tag{7.2.30}$$

这里 $\Lambda^* = \mathrm{diag}(\lambda_1^*,\cdots,\lambda_p^*)$，$\lambda_1^* \geqslant \cdots \geqslant \lambda_p^* \geqslant 0$ 为 R 的 p 个特征值。由此得到 p 个主成分 $y_1^* = t_1^{*\prime}\boldsymbol{x}^*,\cdots,y_p^* = t_p^{*\prime}\boldsymbol{x}^*$。记 $\boldsymbol{y}^* = (y_1^*,\cdots,y_p^*)'$，于是

$$\boldsymbol{y}^* = T^{*\prime}\boldsymbol{x}^* \tag{7.2.31}$$

上述主成分具有的性质可概括如下：

(1) $E(\boldsymbol{y}^*) = \boldsymbol{0}$，$V(\boldsymbol{y}^*) = \Lambda^*$。

(2) $\sum_{i=1}^{p}\lambda_i^*=p$。

(3) 变量 x_i^* 与主成分 y_j^* 之间的相关系数

$$\rho(x_i^*,y_j^*)=\sqrt{\lambda_j^*}\,t_{ij}^*$$

(4) 主成分 $y_1^*,\cdots,y_m^*$ 对变量 x_i^* 的贡献率

$$\rho_{i\cdot 1\cdots m}^2=\sum_{j=1}^{m}\rho^2(x_i^*,y_j^*)=\sum_{j=1}^{m}\lambda_j^*\,t_{ij}^{*2}$$

(5) $\sum_{j=1}^{p}\rho^2(x_i^*,y_j^*)=\sum_{j=1}^{p}\lambda_j^*\,t_{ij}^{*2}=1$。

§7.3 样本的主成分

从上一节的讨论可知，我们可以从协方差矩阵 Σ 或相关矩阵 R 出发求得主成分。但在实际问题中，Σ 或 R 一般都是未知的，需要通过样本来进行估计。设数据矩阵为

$$X=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)'=\begin{pmatrix}x_{11}&\cdots&x_{1p}\\ \vdots& &\vdots\\ x_{n1}&\cdots&x_{np}\end{pmatrix}$$

则样本离差矩阵、样本协方差矩阵和样本相关矩阵分别为

$$A=(a_{ij})=\sum_{i=1}^{n}(\boldsymbol{x}_i-\bar{\boldsymbol{x}})(\boldsymbol{x}_i-\bar{\boldsymbol{x}})'$$

$$S=\frac{1}{n-1}A=(s_{ij})$$

$$\hat{R}=(r_{ij}),\qquad r_{ij}=\frac{a_{ij}}{\sqrt{a_{ii}}\sqrt{a_{jj}}}=\frac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}}$$

其中 $\bar{\boldsymbol{x}}=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i=(\frac{1}{n}\sum_{i=1}^{n}x_{i1},\cdots,\frac{1}{n}\sum_{i=1}^{n}x_{ip})'=(\bar{x}_1,\cdots,\bar{x}_p)'$ 为样本均值。可以用 S 代替 Σ，用 $\hat{R}$ 代替 R，然后从 S 或 $\hat{R}$ 出发按上一节的方法求得主成分。若从 S 出发，则主成分分析将使得方差

大的那些变量与具有大特征值的主成分有较密切的联系，而方差小的另一些变量同具有小特征值的主成分有较强的联系。因此，在实际应用中，一般我们是从 $\hat{R}$ 出发来求得主成分的，除非原始变量所测量的单位是可比较的，或者这些变量已用某些方法标准化了。以下我们只讨论从 $\hat{R}$ 出发的情形，为简化符号，仍将 $\hat{R}$ 记为 R。

首先我们对原始数据作标准化处理，即令

$$x_{ij}^{*}=\frac{x_{ij}-\bar{x}_j}{\sqrt{s_{jj}}},\quad i=1,\cdots,n,\quad j=1,\cdots,p$$

以后仍将 x_{ij}^{*} 记为 x_{ij}，因此数据矩阵中存放的是标准化了的数据。

设相关矩阵 R 的 p 个特征值为 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_p\geqslant 0$，$t_1,t_2,\cdots,t_p$ 为相应的正交单位特征向量，则第 j 个主成分为

$$y_j=t_j'x,\quad j=1,\cdots,p \tag{7.3.1}$$

原始变量 x_i 与主成分 y_j 的样本相关系数列于表 7.3 中。主成分 $y_1,\cdots,y_m(m<p)$ 对 x_i 的贡献率列于表 7.4 中。

表 7.3　　原始变量 x_i 与主成分 y_j 的样本相关系数

主成分 / 原始变量	y_1	y_2	$\cdots$	y_p
x_1	$\sqrt{\lambda_1}\ t_{11}$	$\sqrt{\lambda_2}\ t_{12}$	$\cdots$	$\sqrt{\lambda_p}\ t_{1p}$
x_2	$\sqrt{\lambda_1}\ t_{21}$	$\sqrt{\lambda_2}\ t_{22}$	$\cdots$	$\sqrt{\lambda_p}\ t_{2p}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$
x_p	$\sqrt{\lambda_1}\ t_{p1}$	$\sqrt{\lambda_2}\ t_{p2}$	$\cdots$	$\sqrt{\lambda_p}\ t_{pp}$

表 7.4　　主成分 $y_1,\cdots,y_m$ 对原始标准化变量 x_i 的贡献率

$r^2_{1\cdot 1\cdots m}$	$r^2_{2\cdot 1\cdots m}$	$\cdots$	$r^2_{p\cdot 1\cdots m}$
$\sum_{j=1}^{m}\lambda_j t_{1j}^2$	$\sum_{j=1}^{m}\lambda_j t_{2j}^2$	$\cdots$	$\sum_{j=1}^{m}\lambda_j t_{pj}^2$

若将第 i 个观测值 $\boldsymbol{x}_i$ 代入(7.3.1)式，得

$$y_{ij}=\boldsymbol{t}_j'\boldsymbol{x}_i,\quad i=1,\cdots,n,\quad j=1,\cdots,p \tag{7.3.2}$$

则 $\boldsymbol{y}_{(i)}=(y_{i1},\cdots,y_{ip})'$，$i=1,\cdots,n$ 称为主成分得分。由于数据矩阵中的数据是被标准化了的，所以

$$\begin{aligned}\bar{\boldsymbol{y}}&=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{y}_{(i)}=\left(\frac{1}{n}\sum_{i=1}^{n}y_{i1},\cdots,\frac{1}{n}\sum_{i=1}^{n}y_{ip}\right)'\\&=\left(\frac{1}{n}\boldsymbol{t}_1'\sum_{i=1}^{n}\boldsymbol{x}_i,\cdots,\frac{1}{n}\boldsymbol{t}_p'\sum_{i=1}^{n}\boldsymbol{x}_i\right)'=\boldsymbol{0}\end{aligned} \tag{7.3.3}$$

我们在确定主成分的个数 m 时，既要使 $y_1,\cdots,y_m$ 的累计贡献率 $\sum_{j=1}^{m}\lambda_j\Big/p$ 达到一定的百分比，比如至少 85%，也应考虑到尽量使 m 是一个较小的自然数，以利于作实际分析。我们常常需要在两者之间作出某种折衷。

例 7.3.1　表 7.5 中给出的是美国 50 个州每 100 000 个人中七种犯罪的比率数据。这七种犯罪是：murder(杀人罪)，rape(强奸罪)，robbery(抢劫罪)，assault(斗殴罪)，burglary(夜盗罪)，larceny(偷盗罪)，auto(汽车犯罪)。很难直接从这七个变量出发来评价各州的治安和犯罪情况，而使用主成分分析却可以把这些变量概括为两个或三个综合变量(即主成分)，以便帮助我们较简便地分析这些数据。

表 7.5　　美国 50 个州七种犯罪的比率数据

state	murder	rape	robbery	assault	burglary	larceny	auto
Alabama	14.2	25.2	96.8	278.3	1135.5	1881.9	280.7
Alaska	10.8	51.6	96.8	284.0	1331.7	3369.8	753.3

续表

state	murder	rape	robbery	assault	burglary	larceny	auto
Arizona	9.5	34.2	138.2	312.3	2346.1	4467.4	439.5
Arkansas	8.8	27.6	83.2	203.4	972.6	1862.1	183.4
California	11.5	49.4	287.0	358.0	2139.4	3499.8	663.5
Colorado	6.3	42.0	170.7	292.9	1935.2	3903.2	477.1
Connecticut	4.2	16.8	129.5	131.8	1346.0	2620.7	593.2
Delaware	6.0	24.9	157.0	194.2	1682.6	3678.4	467.0
Florida	10.2	39.6	187.9	449.1	1859.9	3840.5	351.4
Georgia	11.7	31.1	140.5	256.5	1351.1	2170.2	297.9
Hawaii	7.2	25.5	128.0	64.1	1911.5	3920.4	489.4
Idaho	5.5	19.4	39.6	172.5	1050.8	2599.6	237.6
Illinois	9.9	21.8	211.3	209.0	1085.0	2828.5	528.6
Indiana	7.4	26.5	123.2	153.5	1086.2	2498.7	377.4
Iowa	2.3	10.6	41.2	89.8	812.5	2685.1	219.9
Kansas	6.6	22.0	100.7	180.5	1270.4	2739.3	244.3
Kentucky	10.1	19.1	81.1	123.3	872.2	1662.1	245.4
Louisiana	15.5	30.9	142.9	335.5	1165.5	2469.9	337.7
Maine	2.4	13.5	38.7	170.0	1253.1	2350.7	246.9
Maryland	8.0	34.8	292.1	358.9	1400.0	3177.7	428.5
Massachusetts	3.1	20.8	169.1	231.6	1532.2	2311.3	1140.1
Michigan	9.3	38.9	261.9	274.6	1522.7	3159.0	545.5
Minnesota	2.7	19.5	85.9	85.8	1134.7	2559.3	343.1
Mississippi	14.3	19.6	65.7	189.1	915.6	1239.9	144.4
Missouri	9.6	28.3	189.0	233.5	1318.3	2424.2	378.4
Montana	5.4	16.7	39.2	156.8	804.9	2773.2	309.2
Nebraska	3.9	18.1	64.7	112.7	760.0	2316.1	249.1
Nevada	15.8	49.1	323.1	355.0	2453.1	4212.6	559.2
New Hampshire	3.2	10.7	23.2	76.0	1041.7	2343.9	293.4
New Jersey	5.6	21.0	180.4	185.1	1435.8	2774.5	511.5
New Mexico	8.8	39.1	109.6	343.4	1418.7	3008.6	259.5
New York	10.7	29.4	472.6	319.1	1728.0	2782.0	745.8
North Carolina	10.6	17.0	61.3	318.3	1154.1	2037.8	192.1

续表

state	murder	rape	robbery	assault	burglary	larceny	auto
ohio	7.8	27.3	190.5	181.1	1216.0	2696.8	400.4
North Dakota	0.9	9.0	13.3	43.8	446.1	1843.0	144.7
Oklahoma	8.6	29.2	73.8	205.0	1288.2	2228.1	326.8
Oregon	4.9	39.9	124.1	286.9	1636.4	3506.1	388.9
Pennsylvania	5.6	19.0	130.3	128.0	877.5	1624.1	333.2
Rhode Island	3.6	10.5	86.5	201.0	1489.5	2844.1	791.4
South Carolina	11.9	33.0	105.9	485.3	1613.6	2342.4	245.1
South Dakota	2.0	13.5	17.9	155.7	570.5	1704.4	147.5
Tennessee	10.1	29.7	145.8	203.9	1259.7	1776.5	314.0
Texas	13.3	33.8	152.4	208.2	1603.1	2988.7	397.6
Utah	3.5	20.3	68.8	147.3	1171.6	3004.6	334.5
Vermont	1.4	15.9	30.8	101.2	1348.2	2201.0	265.2
Virginia	9.0	23.3	92.1	165.7	986.2	2521.2	226.7
Washington	4.3	39.6	106.2	224.8	1605.6	3386.9	360.3
West Virginia	6.0	13.2	42.2	90.9	597.4	1341.7	163.3
Wisconsin	2.8	12.9	52.2	63.7	846.9	2614.2	220.7
Wyoming	5.4	21.9	39.7	173.9	811.6	2772.2	282.0

算得的相关矩阵列于表 7.6，它的前三个特征值和特征向量列于表 7.7。

表 7.6　　七个变量间的相关矩阵

1.0000	0.6012	0.4837	0.6486	0.3858	0.1019	0.0688
0.6012	1.0000	0.5919	0.7403	0.7121	0.6140	0.3489
0.4837	0.5919	1.0000	0.5571	0.6372	0.4467	0.5907
0.6486	0.7403	0.5571	1.0000	0.6229	0.4044	0.2758
0.3858	0.7121	0.6372	0.6229	1.0000	0.7921	0.5580
0.1019	0.6140	0.4467	0.4044	0.7921	1.0000	0.4442
0.0688	0.3489	0.5907	0.2758	0.5580	0.4442	1.0000

表 7.7　　表 7.6 的前三个特征值和特征向量

特征向量	t_1	t_2	t_3
murder	0.300	−.629	0.178
rape	0.431	−.169	−.244
robbery	0.396	0.042	0.495
assault	0.396	−.343	−.069
burglary	0.440	0.203	−.209
larceny	0.357	0.402	−.539
auto	0.295	0.502	0.568
特征值	4.115	1.239	0.726
贡献率	0.588	0.177	0.104
累计贡献率	0.588	0.765	0.869

我们从相关矩阵出发进行主成分分析。从表 7.7 可以看出，前两个主成分的累计贡献率已达 76.5%，前三个主成分的累计贡献率达 86.9%，因此可以考虑只取前面两个或三个主成分，它们能够很好地概括这组数据。

由于第一主成分对所有变量都有近似相等的载荷，因此可认为是对所有犯罪率的度量。第二主成分在变量 auto 和 larceny 上有高的正载荷，而在变量 murder 和 assault 上有高的负载荷；在 burglary 上存在小的正载荷，而在 RAPE 上存在小的负载荷。可以认为这个主成分是用于度量暴力犯罪在犯罪性质上占的比重。第三主成分很难给出明显的解释。

§7.4　SAS 程序及输出

在例 7.3.1 中编制 SAS 程序如下：

```
data crime;
    input state $ 1—15 murder rape robbery
        assault burglary larceny auto;
    cards;
```

Alabama	14.2	25.2	96.8	278.3	1135.5	1881.9	280.7
Alaska	10.8	51.6	96.8	284.0	1331.7	3369.8	753.3
Arizona	9.5	34.2	138.2	312.3	2346.1	4467.4	439.5
Arkansas	8.8	27.6	83.2	203.4	972.6	1862.1	183.4
California	11.5	49.4	287.0	358.0	2139.4	3499.8	663.5
Colorado	6.3	42.0	170.7	292.9	1935.2	3903.2	477.1
Connecticut	4.2	16.8	129.5	131.8	1346.0	2620.7	593.2
Delaware	6.0	24.9	157.0	194.2	1682.6	3678.4	467.0
Florida	10.2	39.6	187.9	449.1	1859.9	3840.5	351.4
Georgia	11.7	31.1	140.5	256.5	1351.1	2170.2	297.9
Hawaii	7.2	25.5	128.0	64.1	1911.5	3920.4	489.4
Idaho	5.5	19.4	39.6	172.5	1050.8	2599.6	237.6
Illinois	9.9	21.8	211.3	209.0	1085.0	2828.5	528.6
Indiana	7.4	26.5	123.2	153.5	1086.2	2498.7	377.4
Iowa	2.3	10.6	41.2	89.8	812.5	2685.1	219.9
Kansas	6.6	22.0	100.7	180.5	1270.4	2739.3	244.3
Kentucky	10.1	19.1	81.1	123.3	872.2	1662.1	245.4
Louisiana	15.5	30.9	142.9	335.5	1165.5	2469.9	337.7
Maine	2.4	13.5	38.7	170.0	1253.1	2350.7	246.9
Maryland	8.0	34.8	292.1	358.9	1400.0	3177.7	428.5
Massachusetts	3.1	20.8	169.1	231.6	1532.2	2311.3	1140.1
Michigan	9.3	38.9	261.9	274.6	1522.7	3159.0	545.5
Minnesota	2.7	19.5	85.9	85.8	1134.7	2559.3	343.1
Mississippi	14.3	19.6	65.7	189.1	915.6	1239.9	144.4
Missouri	9.6	28.3	189.0	233.5	1318.3	2424.2	378.4
Montana	5.4	16.7	39.2	156.8	804.9	2773.2	309.2
Nebraska	3.9	18.1	64.7	112.7	760.0	2316.1	249.1
Nevada	15.8	49.1	323.1	355.0	2453.1	4212.6	559.2

New Hampshire	3.2	10.7	23.2	76.0	1041.7	2343.9	293.4
New Jersey	5.6	21.0	180.4	185.1	1435.8	2774.5	511.5
New Mexico	8.8	39.1	109.6	343.4	1418.7	3008.6	259.5
New York	10.7	29.4	472.6	319.1	1728.0	2782.0	745.8
North Carolina	10.6	17.0	61.3	318.3	1154.1	2037.8	192.1
North Dakota	0.9	9.0	13.3	43.8	446.1	1843.0	144.7
Ohio	7.8	27.3	190.5	181.1	1216.0	2696.8	400.4
Oklahoma	8.6	29.2	73.8	205.0	1288.2	2228.1	326.8
Oregon	4.9	39.9	124.1	286.9	1636.4	3506.1	388.9
Pennsylvania	5.6	19.0	130.3	128.0	877.5	1624.1	333.2
Rhode Island	3.6	10.5	86.5	201.0	1489.5	2844.1	791.4
South Carolina	11.9	33.0	105.9	485.3	1613.6	2342.4	245.1
South Dakota	2.0	13.5	17.9	155.7	570.5	1704.4	147.5
Tennessee	10.1	29.7	145.8	203.9	1259.7	1776.5	314.0
Texas	13.3	33.8	152.4	208.2	1603.1	2988.7	397.6
Utah	3.5	20.3	68.8	147.3	1171.6	3004.6	334.5
Vermont	1.4	15.9	30.8	101.2	1348.2	2201.0	265.2
Virginia	9.0	23.3	92.1	165.7	986.2	2521.2	226.7
Washington	4.3	39.6	106.2	224.8	1605.6	3386.9	360.3
West Virginia	6.0	13.2	42.2	90.9	597.4	1341.7	163.3
Wisconsin	2.8	12.9	52.2	63.7	846.9	2614.2	220.7
Wyoming	5.4	21.9	39.7	173.9	811.6	2772.2	282.0

```
;
proc princomp out=crimcomp;
run;

proc sort;
     by prin1;
run;
```

```
proc print;
     id state;
     var prin1 prin2 murder rape robbery
          assault burglary larceny auto;
run;
proc sort;
     by prin2;
run;
proc print;
     id state;
     var prin1 prin2 murder rape robbery
          assault burglary larceny auto;
run;

proc plot;
     plot prin2 * prin1=state;
run;
```

屏幕输出：

输出 7.1　　对美国 50 个州七种犯罪比率的主成分分析

Principal Component Analysis

50 Observations
7 Variables

(1)Simple Statistics

	MURDER	RAPE	ROBBERY	ASSAULT
Mean	7.444000000	25.73400000	124.0920000	211.3000000
StD	3.866768941	10.75962995	68.3485672	100.2530492

	BURGLARY	LARCENY	AUTO
Mean	1291.904000	2671.288000	377.5260000
StD	432.455711	725.908707	193.3944175

(2)Correlation Matrix

	MURDER	RAPE	ROBBERY	ASSAULT	BURGLARY	LARCENY	AUTO
MURDER	1.0000	0.6012	0.4837	0.6486	0.3858	0.1019	0.0688
RAPE	0.6012	1.0000	0.5919	0.7403	0.7121	0.6140	0.3469
ROBBERY	0.4837	0.5919	1.0000	0.5571	0.6372	0.4467	0.5907
ASSAULT	0.6486	0.7403	0.5571	1.0000	0.6229	0.4044	0.2758
BURGLARY	0.3858	0.7121	0.6372	0.6229	1.0000	0.7921	0.5580
LARCENY	0.1019	0.6140	0.4467	0.4044	0.7921	1.0000	0.4442
AUTO	0.0688	0.3489	0.5907	0.2758	0.5560	0.4442	1.0000

Eigenvalues of the Correlation Matrix

	(3) Eigenvalue	(4) Difference	(5) Proportion	(6) Cumulative
PRIN1	4.11498	2.87624	0.587851	0.58765
PRIN2	1.23672	0.51291	0.176960	0.76481
PRIN3	0.72562	0.40936	0.103688	0.66850
PRIN4	0.31643	0.05846	0.045205	0.91370
PRIN5	0.25797	0.03593	0.036853	0.95056
PRIN6	0.22204	0.09796	0.031720	0.98228
PRIN7	0.12406	.	0.017722	1.00000

(7)Eigenvectors

	PRIN1	PRIN2	PRIN3	PRIN4	PRIN5	PRIN6	PRIN7
MURDER	0.30027	-.62917	0.17824	-.23211	0.53812	0.25911	0.26759
RAPE	0.43175	-.16943	-.24419	0.06221	0.16847	-.77327	-.29648
ROBBERY	0.39687	0.04224	0.49586	-.55796	-.51997	-.11438	-.00390
ASSAULT	0.39665	-.34352	-.06951	0.62980	-.50665	0.17236	0.19174
BURGLARY	0.44015	0.20334	-.20989	-.05755	0.10103	0.53598	-.64811
LARCENY	0.35736	0.40231	-.53923	-.23469	0.03009	0.03940	0.60169
AUTO	0.29517	0.50242	0.56838	0.41923	0.36975	-.05729	0.14704

输出 7.2　　按第一主成分排序的美国 50 个州

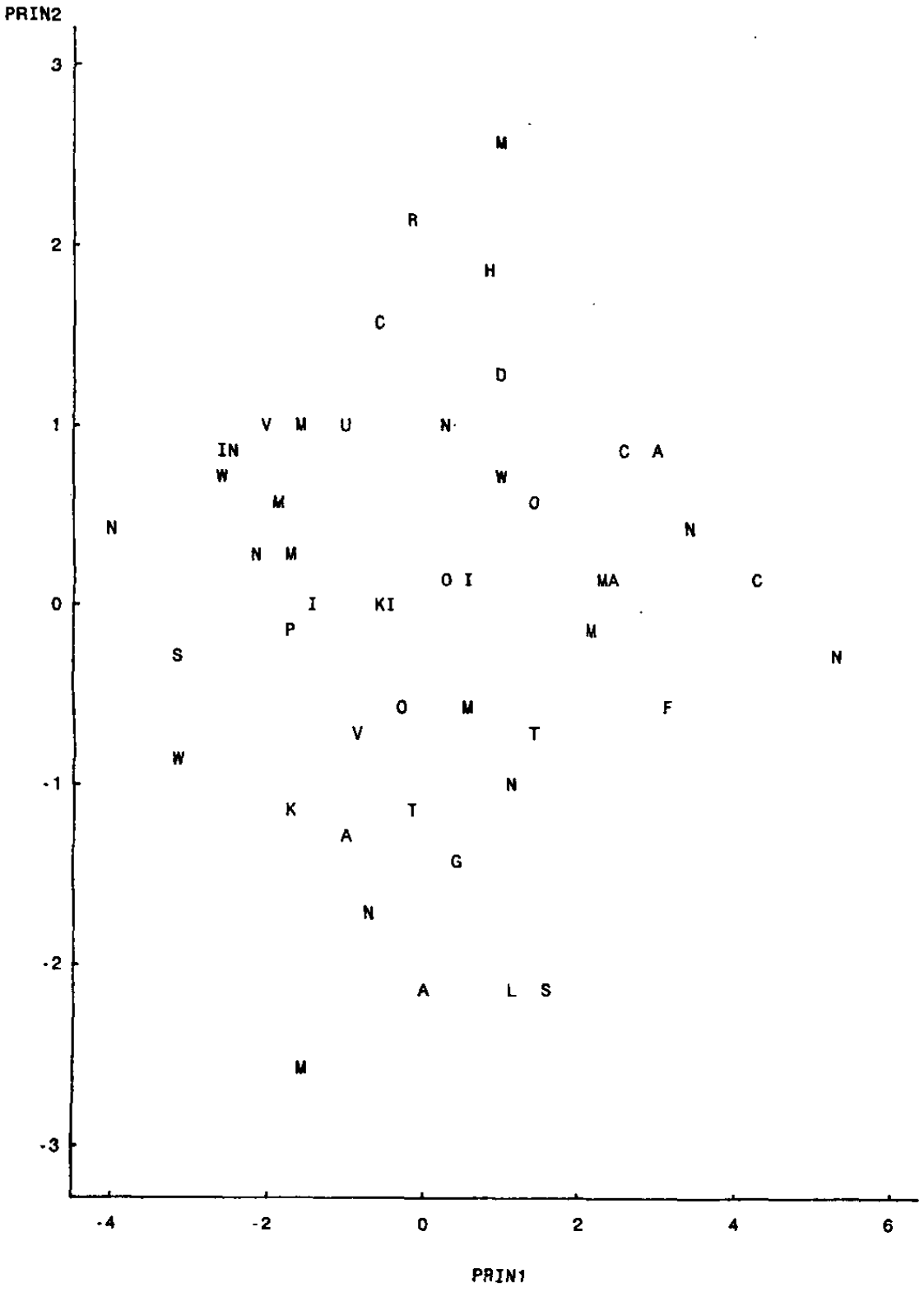
STATE	PRIN1	PRIN2	MURDER	RAPE	RDBBERY	ASSAULT	BURGLARY	LARCENY	AUTO
North Dakota	-3.9640	0.3876	0.9	9.0	13.3	43.8	446.1	1843.0	144.7
South Dakota	-3.1720	-0.2544	2.0	13.5	17.9	155.7	570.5	1704.4	147.5
West Virginia	-3.1477	-0.8142	6.0	13.2	42.2	90.9	597.4	1341.7	163.3
Iowa	-2.5815	0.8247	2.3	10.6	41.2	89.8	812.5	2685.1	219.9
Wisconsin	-2.5029	0.7808	2.8	12.9	52.2	63.7	846.9	2614.2	220.7
New Hampshire	-2.4656	0.8250	3.2	10.7	23.2	76.0	1041.7	2343.9	293.4
Nebraska	-2.1507	0.2257	3.9	18.1	64.7	112.7	760.0	2316.1	249.1
Vermont	-2.0643	0.9449	1.4	15.9	30.6	101.2	1348.2	2201.0	265.2
Maine	-1.8263	0.5787	2.4	13.5	38.7	170.0	1253.1	2350.7	246.9
Kentucky	-1.7269	-1.1466	10.1	19.1	81.1	123.3	872.2	1662.1	245.4
Pennsylvania	-1.7200	-0.1959	5.6	19.0	130.3	128.0	877.5	1624.1	333.2
Montana	-1.6680	0.2709	5.4	16.7	39.2	156.8	804.9	2773.2	309.2
Minnesota	-1.5543	1.0564	2.7	19.5	85.9	85.8	1134.7	2559.3	343.1
Mississippi	-1.5073	-2.5467	14.3	19.8	65.7	189.1	915.6	1239.9	144.4
Idaho	-1.4324	-0.0080	5.5	19.4	39.6	172.5	1050.8	2599.6	237.8
Wyoming	-1.4246	0.0626	5.4	21.9	39.7	173.9	811.6	2772.2	282.0
Arkansas	-1.0544	-1.3454	8.8	27.6	83.2	203.4	972.6	1862.1	183.4
Utah	-1.0499	0.9365	3.5	20.3	68.8	147.3	1171.8	3004.6	334.5
Virginia	-0.9162	-0.6926	9.0	23.3	92.1	165.7	986.2	2521.2	226.7
North Carolina	-0.6992	-1.6702	10.6	17.0	61.3	318.3	1154.1	2037.8	192.1
Kansas	-0.6340	-0.0280	6.6	22.0	100.7	180.5	1270.4	2739.3	244.3
Connecticut	-0.5413	1.5012	4.2	16.8	129.5	131.8	1346.0	2620.7	593.2
Indiana	-0.4999	0.0000	7.4	26.5	123.2	153.5	1086.2	2498.7	377.4
Oklahoma	-0.3213	-0.6242	8.6	29.2	73.8	205.0	1288.2	2228.1	326.8
Rhode Island	-0.2015	2.1465	3.6	10.5	86.5	201.0	1489.5	2844.1	791.4
Tennessee	-0.1366	-1.1349	10.1	29.7	145.8	203.9	1259.7	1776.5	314.0
Alabama	-0.0498	-2.0961	14.2	25.2	96.8	278.3	1135.5	1881.9	280.7
New Jersey	0.2178	0.9642	5.6	21.0	180.4	185.1	1435.8	2774.5	511.5
Ohio	0.2395	0.0905	7.8	27.3	190.5	181.1	1216.0	2696.8	400.4
Georgia	0.4904	-1.3807	11.7	31.1	140.5	256.5	1351.1	2170.2	297.9
Illinois	0.5129	0.0942	9.9	21.8	211.3	209.0	1085.0	2828.5	528.8
Missouri	0.5563	-0.5585	9.6	28.3	189.0	233.5	1318.3	2424.2	378.4
Hawaii	0.8231	1.8239	7.2	25.5	128.0	64.1	1911.5	3920.4	489.4
Washington	0.9305	0.7377	4.3	39.6	106.2	224.8	1605.6	3386.9	360.3
Delaware	0.9645	1.2967	6.0	24.9	157.0	194.2	1682.6	3678.4	467.0
Massachusetts	0.9784	2.6310	3.1	20.8	169.1	231.6	1532.2	2311.3	1140.1
Louisiana	1.1202	-2.0832	15.5	30.9	142.9	335.5	1165.5	2469.9	337.7
New Mexico	1.2141	-0.9507	8.8	39.1	109.6	343.4	1418.7	3008.6	259.5
Texas	1.3969	-0.6813	13.3	33.8	152.4	208.2	1603.1	2988.7	397.6
Oregon	1.4490	0.5860	4.9	39.9	124.1	296.9	1636.4	3506.1	388.9
South Carolina	1.6033	-2.1621	11.9	33.0	105.9	485.3	1613.6	2342.4	245.1
Maryland	2.1828	-0.1947	8.0	34.8	292.1	358.9	1400.0	3177.7	428.5
Michigan	2.2733	0.1548	9.3	38.9	261.9	274.6	1522.7	3159.0	545.5
Alaska	2.4215	0.1665	10.8	51.6	96.8	284.0	1331.7	3369.8	753.3
Colorado	2.5092	0.9166	6.3	42.0	170.7	292.9	1935.2	3903.2	477.1
Arizona	3.0141	0.8449	9.5	34.2	138.2	312.3	2346.1	4467.4	439.5
Florida	3.1117	-0.6039	10.2	39.6	187.9	449.1	1859.9	3840.5	351.4
New York	3.4524	0.4328	10.7	29.4	472.6	319.1	1728.0	2782.0	745.8
California	4.2838	0.1431	11.5	49.4	287.0	358.0	2139.4	3499.8	663.5
Nevada	5.2669	-0.2526	15.8	49.1	323.1	355.0	2453.1	4212.6	559.2

输出 7.3　　按第二主成分排序的美国 50 个州

STATE	PRIN1	PRIN2	MURDER	RAPE	ROBBERY	ASSAULT	BURGLARY	LARCENY	AUTO
Mississippi	-1.5073	-2.5467	14.3	19.6	65.7	189.1	915.6	1239.9	144.4
South Carolina	1.6033	-2.1621	11.9	33.0	105.9	485.3	1613.6	2342.4	245.1
Alabama	-0.0498	-2.0961	14.2	25.2	96.8	278.3	1135.5	1881.9	280.7
Louisiana	1.1202	-2.0832	15.5	30.9	142.9	335.5	1165.5	2469.9	337.7
North Carolina	-0.6992	-1.6702	10.6	17.0	61.3	318.3	1154.1	2037.8	192.1
Georgia	0.4904	-1.3807	11.7	31.1	140.5	256.5	1351.1	2170.2	297.9
Arkansas	-1.0544	-1.3454	8.8	27.6	83.2	203.4	972.6	1862.1	183.4
Kentucky	-1.7269	-1.1466	10.1	19.1	81.1	123.3	872.2	1662.1	245.4
Tennessee	-0.1366	-1.1349	10.1	29.7	145.8	203.9	1259.7	1776.5	314.0
New Mexico	1.2141	-0.9507	8.8	39.1	109.6	343.4	1418.7	3008.6	259.5
West Virginia	-3.1477	-0.8142	6.0	13.2	42.2	90.9	597.4	1341.7	163.3
Virginia	-0.9162	-0.6926	9.0	23.3	92.1	165.7	966.2	2521.2	226.7
Texas	1.3969	-0.6813	13.3	33.8	152.4	208.2	1603.1	2988.7	397.6
Oklahoma	-0.3213	-0.6242	6.6	29.2	73.8	205.0	1288.2	2228.1	326.8
Florida	3.1117	-0.6039	10.2	39.6	187.9	449.1	1859.9	3840.5	351.4
Missouri	0.5563	-0.5585	9.6	28.3	189.0	233.5	1318.3	2424.2	378.4
South Dakota	-3.1720	-0.2544	2.0	13.5	17.9	155.7	570.5	1704.4	147.5
Nevada	5.2669	-0.2526	15.8	49.1	323.1	355.0	2453.1	4212.6	559.2
Pennsylvania	-1.7200	-0.1959	5.6	19.0	130.3	128.0	877.5	1624.1	333.2
Maryland	2.1828	-0.1947	8.0	34.8	292.1	358.9	1400.0	3177.7	428.5
Kansas	-0.6340	-0.0280	6.6	22.0	100.7	180.5	1270.4	2739.3	244.3
Idaho	-1.4324	-0.0080	5.5	19.4	39.6	172.5	1050.8	2599.6	237.6
Indiana	-0.4999	0.0000	7.4	26.5	123.2	153.5	1086.2	2498.7	377.4
Wyoming	-1.4246	0.0626	5.4	21.9	39.7	173.9	811.6	2772.2	282.0
Ohio	0.2395	0.0905	7.8	27.3	190.5	181.1	1216.0	2696.8	400.4
Illinois	0.5129	0.0942	9.9	21.8	211.3	209.0	1085.0	2826.5	528.8
California	4.2838	0.1431	11.5	49.4	287.0	358.0	2139.4	3499.8	663.5
Michigan	2.2733	0.1548	9.3	38.9	261.9	274.6	1522.7	3159.0	545.5
Alaska	2.4215	0.1665	10.8	51.6	96.8	284.0	1331.7	3369.8	753.3
Nebraska	-2.1507	0.2257	3.9	18.1	64.7	112.7	760.0	2316.1	249.1
Montana	-1.6680	0.2709	5.4	16.7	39.2	156.8	804.9	2773.2	309.2
North Dakota	-3.9640	0.3876	0.9	9.0	13.3	43.8	446.1	1843.0	144.7
New York	3.4524	0.4328	10.7	29.4	472.6	319.1	1728.0	2782.0	745.8
Maine	-1.8263	0.5787	2.4	13.5	38.7	170.0	1253.1	2350.7	246.9
Oregon	1.4490	0.5860	4.9	39.9	124.1	286.9	1636.4	3508.1	388.9
Washington	0.9305	0.7377	4.3	39.6	106.2	224.8	1605.6	3386.9	360.3
Wisconsin	-2.5029	0.7908	2.8	12.9	52.2	63.7	846.9	2614.2	220.7
Iowa	-2.5815	0.8247	2.3	10.8	41.2	89.8	812.5	2685.1	219.9
New Hampshire	-2.4656	0.8250	3.2	10.7	23.2	76.0	1041.7	2343.9	293.4
Arizona	3.0141	0.8449	9.5	34.2	138.2	312.3	2346.1	4467.4	439.5
Colorado	2.5092	0.9166	6.3	42.0	170.7	292.9	1935.2	3903.2	477.1
Utah	-1.0499	0.9365	3.5	20.3	68.8	147.3	1171.6	3004.6	334.5
Vermont	-2.0643	0.9449	1.4	15.9	30.8	101.2	1348.2	2201.0	265.2
New Jersey	0.2178	0.9642	5.6	21.0	180.4	185.1	1435.8	2774.5	511.5
Minnesota	-1.5543	1.0564	2.7	19.5	85.9	85.8	1134.7	2559.3	343.1
Delaware	0.9645	1.2967	6.0	24.9	157.0	194.2	1682.6	3678.4	467.0
Connecticut	-0.5413	1.5012	4.2	16.8	129.5	131.8	1346.0	2620.7	593.2
Hawaii	0.8231	1.8239	7.2	25.5	128.0	64.1	1911.5	3920.4	489.4
Rhode Island	-0.2015	2.1465	3.6	10.5	86.5	201.0	1489.5	2844.1	791.4
Massachusetts	0.9784	2.6310	3.1	20.8	169.1	231.6	1532.2	2311.3	1140.1

输出 7.4　　按第一和第二主成分输出的散点图

Plot of PRIN2*PRIN1. Symbol is value of STATE.



NOTE: 1 obs hidden.

输出 7.1 的说明如下：

(1) 简单描述统计量。它包括每个变量的均值和标准差。

(2) 相关矩阵。

(3) 相关矩阵的特征值。

(4) 相邻两特征值之差。

(5) 每个主成分的贡献率。

(6) 前几个主成分的累计贡献率。

(7) 特征向量。

输出 7.2 是把 50 个州按第一主成分得分从小到大的顺序重新排序后的输出结果。输出 7.3 是把 50 个州按第二主成分得分从小到大的顺序重新排序后的输出。输出 7.4 输出的是第一和第二主成分的散点图，它对各州的综合犯罪比率和暴力犯罪的比重有较直观地描述。图中各散点都是用相应州的字头输出的，对重复的字符，可对照输出 7.2 和输出 7.3 加以辨别。从散点图中可以看出，Nevada 和 California 在最右边，综合犯罪比率是最高的，但暴力犯罪在犯罪性质上的比重(第二主成分)处于平均水平。North Dakota、South Dakota 和 West Virginia 在最左边，综合犯罪比率是最低的。来自东南部的州大多在散点图的底部，这表明暴力犯罪的比率相对一般犯罪的比率较高。Massachusetts 在最上部，表明暴力犯罪的比率较一般犯罪的比率为低。

小　结

1. 主成分分析是通过降维技术用少数几个综合变量来代替原始多个变量的一种统计分析方法。这些综合变量集中了原始变量的大部分信息。

2. 第一主成分所包含的信息量最大，第二主成分其次，其它主成分依次递减，各主成分之间互不相关，这就保证了各主成分所含的信息互不重复。

3. 取多少个主成分，既要考虑到前几个主成分的累计贡献率达到一定比例，也要考虑到应选取尽可能少的主成分以较好地达到降维的目的。

4. 当各变量的单位不相同时，应从相关矩阵出发进行主成分分析。

5. 计算出主成分之后，应对要使用的前若干个主成分作出符合实际背景和意义的解释。

习 题

7.1 在例 7.3.1 中，试从协方差矩阵出发进行主成分分析，并与例 7.3.1 中的方法进行比较。两种方法哪一种较为合理？

7.2 在选择的美国 64 个城市中，一月份和七月份的日平均温度列于下表，试从协方差矩阵出发进行主成分分析，并画出两个原始变量及两个主成分的散点图。

城市	一月份	七月份	城市	一月份	七月份
Mobile	51.2	81.6	Baltimore	33.4	76.6
Phoenix	51.2	91.2	Boston	29.2	73.3
Little Rock	39.5	81.4	Detroit	25.5	73.3
Sacramento	45.1	75.2	Sault Ste Marie	14.2	63.8
Denver	29.9	73.0	Duluth	8.5	65.6
Hartford	24.8	72.7	Minneapolis	12.2	71.9
Wilmington	32.0	75.8	Jackson	47.1	81.7
Washington DC	35.6	78.7	Kansas City	27.8	78.8
Jacksonville	54.6	81.0	St Louis	31.3	78.6
Miami	67.2	82.3	Great Falls	20.5	69.3
Atlanta	42.4	78.0	Omaha	22.6	77.2
Boise	29.0	74.5	Reno	31.9	69.3
Chicago	22.9	71.9	Concord	20.6	69.7
Peoria	23.8	75.1	Atlantic City	32.7	75.1
Indianapolis	27.9	75.0	Albuquerque	35.2	78.7
Des Moines	19.4	75.1	Albany	21.5	72.0
Wichita	31.3	80.7	Buffalo	23.7	70.1

续表

城市	一月份	七月份	城市	一月份	七月份
Louisville	33.3	76.9	New York	32.2	76.6
New Orleans	52.9	81.9	Charlotte	42.1	78.5
Portland, Maine	21.5	68.0	Raleigh	40.5	77.5
Bismarck	8.2	70.8	Nashville	38.3	79.6
Cincinnati	31.1	75.6	Dallas	44.8	84.8
Cleveland	26.9	71.4	El Paso	43.6	82.3
Columbus	28.4	73.6	Houston	52.1	83.3
Oklahoma City	36.8	81.5	Salt Lake City	28.0	76.7
Portland, Oreg	38.1	67.1	Burlington	16.8	69.8
Philadelphia	32.3	76.8	Norfolk	40.5	78.3
Pittsburgh	28.1	71.9	Richmond	37.5	77.9
Providence	28.4	72.1	Spokane	25.4	69.7
Columbia	45.4	81.2	Charleston, WV	34.5	75.0
Sioux Falls	14.2	73.3	Milwaukee	19.4	69.9
Memphis	40.5	79.6	Cheyenne	26.6	69.1

第八章　因子分析

§8.1　引　言

因子分析是主成分分析的推广，它也是一种把多个变量化为少数几个综合变量的多元分析方法，其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。

例 8.1.1　Linden 对二次大战以来奥林匹克十项全能比赛的得分作了分析研究，他收集了 160 组数据，这十个全能项目依次为：100 米跑、跳远、铅球、跳高、400 米跑、110 米跨栏、铁饼、撑杆跳高、标枪、1500 米跑。但总的来说基本上可归结于他们的短跑速度、爆发性臂力、爆发性腿力和耐力这四个方面，每一方面都称为一个因子。用 $x_1,\cdots,x_{10}$ 分别表示十个项目的得分，它们可以表示为含有上述四个因子的线性模型：

$$x_i=\mu_i+a_{i1}f_1+a_{i2}f_2+a_{i3}f_3+a_{i4}f_4+\varepsilon_i$$

$$i=1,2,\cdots,10$$

其中 f_1,f_2,f_3,f_4 表示四个因子，称为公因子，a_{ij} 称为第 i 个变量在第 j 个因子上的载荷，μ_i 是总平均，ε_i 是第 i 项得分不能被四个公因子解释的部分，称之为特殊因子。这个模型形式上与线性回归模型几乎一样，但它们有着本质的区别：回归模型中自变量是可以被观测得到的，而上述因子模型中的 f_1,f_2,f_3,f_4 是不可观测的隐变量，这使得该模型理解起来较为困难；再者，两个模型的参数意义也很不相同。

例 8.1.2　为了评价高中学生将来进大学时的学习能力，抽了 200 名高中生进行问卷调查，共 50 个问题。所有这些问题可简

单地归结为阅读理解、数学水平和艺术修养这三个方面。这也是一个因子分析模型，每一方面就是一个因子。

例 8.1.3　公司老板对 48 名申请工作的人进行面试，并给出申请人在 15 个方面所得的分数，这 15 个方面是：(1)申请信的形式；(2)外貌；(3)专业能力；(4)讨人喜欢的能力；(5)自信心；(6)洞察力；(7)诚实；(8)推销能力；(9)经验；(10)驾驶汽车本领；(11)抱负；(12)理解能力；(13)潜力；(14)对工作要求强烈程度；(15)适应性。这些问题可以归结为如下的几个方面：申请者外露的能力，讨人喜欢的程度，申请者的经验，专业能力。每一方面都是因子模型中的一个因子。

§8.2　因子模型

一、数学模型

设 p 维可观测的随机向量 $\boldsymbol{x}=(x_1,x_2,\cdots,x_p)'$ 的均值为 $\mu=(\mu_1,\mu_2,\cdots,\mu_p)'$，协方差矩阵为 $\Sigma=(\sigma_{ij})$，因子分析的一般模型为

$$\begin{cases} x_1=\mu_1+a_{11}f_1+a_{12}f_2+\cdots+a_{1m}f_m+\varepsilon_1 \\ x_2=\mu_2+a_{21}f_1+a_{22}f_2+\cdots+a_{2m}f_m+\varepsilon_2 \\ \quad\vdots \\ x_p=\mu_p+a_{p1}f_1+a_{p2}f_2+\cdots+a_{pm}f_m+\varepsilon_p \end{cases} \tag{8.2.1}$$

其中 $f_1,f_2,\cdots,f_m$ 为公因子，$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_p$ 为特殊因子，它们都是不可观测的随机变量。公因子 $f_1,f_2,\cdots,f_m$ 出现在每一个原始变量 $x_i(i=1,2,\cdots,p)$ 的表达式中，可理解为原始变量共同具有的公共因素；每个公因子 $f_j(j=1,2,\cdots,m)$ 至少对两个原始变量有作用，否则它将归入特殊因子。每个特殊因子 $\varepsilon_i(i=1,2,\cdots,p)$ 仅仅出现在与之相应的第 i 个原始变量 x_i 的表达式中，它只对这个原始变量有作用。(8.2.1)式可用矩阵表示为

$$\boldsymbol{x}=\mu+A\boldsymbol{f}+\boldsymbol{\varepsilon} \tag{8.2.2}$$

式中 $\boldsymbol{f}=(f_1,f_2,\cdots,f_m)'(m\leqslant p)$ 为公因子向量，$\boldsymbol{\varepsilon}=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_p)'$

为特殊因子向量，$A=(a_{ij}):p\times m$ 称为因子载荷矩阵，并假设 A 的秩为 m。通常假定

$$\begin{cases} E(\boldsymbol{f})=\boldsymbol{0} \\ E(\boldsymbol{\varepsilon})=\boldsymbol{0} \\ V(\boldsymbol{f})=I \\ V(\boldsymbol{\varepsilon})=D=\mathrm{diag}(\sigma_1^2,\sigma_2^2,\cdots,\sigma_p^2) \\ \mathrm{cov}(\boldsymbol{f},\boldsymbol{\varepsilon})=E(\boldsymbol{f}\boldsymbol{\varepsilon}')=0 \end{cases} \tag{8.2.3}$$

由上述假定可以看出，公因子彼此不相关且具有单位方差，特殊因子彼此不相关且和公因子也不相关。

因子分析与主成分分析是多元分析中两种重要的降维方法，但两者有很大的不同。主成分分析不能作为一个模型来描述，它只能作为一般的变量变换，主成分是可观测的原始变量的线性组合；而因子分析需要构造一个因子模型，公因子一般不能表示为原始变量的线性组合。

二、因子模型的性质

1. $\boldsymbol{x}$ 的协方差矩阵 Σ 的分解

由(8.2.2)式知

$$\begin{aligned} V(\boldsymbol{x})&=V(A\boldsymbol{f}+\boldsymbol{\varepsilon})=E[(A\boldsymbol{f}+\boldsymbol{\varepsilon})(A\boldsymbol{f}+\boldsymbol{\varepsilon})'] \\ &=AE(\boldsymbol{f}\boldsymbol{f}')A'+AE(\boldsymbol{f}\boldsymbol{\varepsilon}')+E(\boldsymbol{\varepsilon}\boldsymbol{f}')A'+E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') \\ &=AV(\boldsymbol{f})A'+V(\boldsymbol{\varepsilon}) \end{aligned}$$

再由(8.2.3)式可得

$$\Sigma=AA'+D \tag{8.2.4}$$

这就是 Σ 的一个分解。如果 $\boldsymbol{x}$ 为标准化了的随机向量，则 Σ 就是相关矩阵 $R=(\rho_{ij})$，即有

$$R=AA'+D \tag{8.2.5}$$

2. 模型不受单位的影响

将 $\boldsymbol{x}$ 的单位作变化，就是作一变换 $\boldsymbol{x}^*=\Delta\boldsymbol{x}$，这里 $\Delta=\mathrm{diag}(\delta_1,\delta_2,\cdots,\delta_p)$，$\delta_i>0$，$i=1,2,\cdots,p$，于是

$$\boldsymbol{x}^*=\Delta\mu+\Delta A\boldsymbol{f}+\Delta\boldsymbol{\varepsilon}$$

令 $\mu^*=\Delta\mu$，$A^*=\Delta A$，$f^*=f$，$\varepsilon^*=\Delta\varepsilon$，则有

$$x^*=\mu^*+A^*f^*+\varepsilon^*$$

这个模型能满足完全类似于(8.2.3)式的假定，即

$$\begin{cases}E(f^*)=\mathbf{0}\\E(\varepsilon^*)=\mathbf{0}\\V(f^*)=I\\V(\varepsilon^*)=D^*\\\mathrm{cov}(f^*,\varepsilon^*)=E(f^*\varepsilon^{*\prime})=0\end{cases}$$

其中 $D^*=\mathrm{diag}(\sigma_1^{*2},\sigma_2^{*2},\cdots,\sigma_p^{*2})$，$\sigma_i^{*2}=\delta_i^2\sigma_i^2$，$i=1,2,\cdots,p$。

3. 因子载荷是不唯一的

设 T 为任一 $m\times m$ 正交矩阵，令 $A^*=AT$，$f^*=T'f$，则模型(8.2.2)式能表示为

$$x=\mu+A^*f^*+\varepsilon \tag{8.2.6}$$

因为

$$E(f^*)=T'E(f)=\mathbf{0}$$

$$V(f^*)=T'V(f)T=T'T=I$$

$$\mathrm{cov}(f^*,\varepsilon)=E(f^*\varepsilon')=T'E(f\varepsilon')=0$$

所以仍满足条件(8.2.3)式。从(8.2.4)式可以看出，Σ 也可分解为

$$\Sigma=A^*A^{*\prime}+D \tag{8.2.7}$$

因此，因子载荷矩阵 A 不是唯一的，在实际应用中常常利用这一点，通过因子的变换，使得新的因子有更好的实际意义。

三、因子载荷矩阵的统计意义

1. A 的元素 a_{ij}——原始变量 x_i 与公因子 f_j 之间的协方差函数

(8.2.1)式可以表示为

$$x_i=\mu_i+a_{i1}f_1+a_{i2}f_2+\cdots+a_{im}f_m+\varepsilon_i,\quad i=1,2,\cdots,p \tag{8.2.8}$$

故

$$\operatorname{cov}(x_i,f_j)=\sum_{\alpha=1}^{m}a_{i\alpha}\operatorname{cov}(f_\alpha,f_j)+\operatorname{cov}(\varepsilon_i,f_j)=a_{ij}$$

(8.2.9)

即 a_{ij}是 x_i 与 f_j 之间的协方差函数。若 $\boldsymbol{x}$ 为标准化了的随机向量，即 $V(\boldsymbol{x})=I$，则 x_i 与 f_j 的相关系数

$$\rho(x_i,f_j)=\frac{\operatorname{cov}(x_i,f_j)}{\sqrt{V(x_i)V(f_j)}}=\operatorname{cov}(x_i,f_j)=a_{ij} \quad (8.2.10)$$

此时 a_{ij}表示 x_i 与 f_j 之间的相关系数。

2. A 的行元素平方和 $h_i^2=\sum_{j=1}^{m}a_{ij}^2$——原始变量 x_i 对公因子依赖的程度

对(8.2.8)式两边取方差

$$\begin{aligned}V(x_i)&=a_{i1}^2V(f_1)+a_{i2}^2V(f_2)+\cdots+a_{im}^2V(f_m)+V(\varepsilon_i)\\&=a_{i1}^2+a_{i2}^2+\cdots+a_{im}^2+\sigma_i^2,\quad i=1,2,\cdots,p\end{aligned} \quad (8.2.11)$$

令

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

于是

$$\sigma_{ii}=h_i^2+\sigma_i^2,\quad i=1,2,\cdots,p \quad (8.2.12)$$

h_i^2 反映了公因子对 x_i 的影响，可以看成是公因子对 x_i 的方差贡献，称为共性方差；而 σ_i^2 是特殊因子 ε_i 对 x_i 的方差贡献，称为个性方差。当 $\boldsymbol{x}$ 为标准化了的随机向量时，$\sigma_{ii}=1$，此时有

$$h_i^2+\sigma_i^2=1,\quad i=1,2,\cdots,p \quad (8.2.13)$$

3. A 的列元素平方和 $g_j^2=\sum_{i=1}^{p}a_{ij}^2$——公因子 f_j 对 $\boldsymbol{x}$ 的贡献

由(8.2.11)式得

$$\sum_{i=1}^{p}V(x_i)=\sum_{i=1}^{p}a_{i1}^2V(f_1)+\sum_{i=1}^{p}a_{i2}^2V(f_2)+\cdots$$

$$+\sum_{i=1}^{p}a_{im}^2V(f_m)+\sum_{i=1}^{p}V(\varepsilon_i)$$

$$=g_1^2V(f_1)+g_2^2V(f_2)+\cdots+g_m^2V(f_m)+\sum_{i=1}^{p}\sigma_i^2$$

$$=g_1^2+g_2^2+\cdots+g_m^2+\sum_{i=1}^{p}\sigma_i^2 \qquad (8.2.14)$$

其中

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

从(8.2.14)式可见，A 的第 j 列元素的平方和 g_j^2 是 $V(f_j)$ 的系数，g_j^2 的值越大，反映了 f_j 对 x 的影响越大，g_j^2 是衡量公因子 f_j 重要性的一个尺度，可视为公因子 f_j 对 x 的贡献。

§8.3 参数估计

设 $x_1,x_2,\cdots,x_n$ 是一组 p 维样本，则 μ 和 Σ 可分别估计为

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i \quad 和 \quad S=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})'$$

为了建立因子模型，首先要估计因子载荷矩阵 $A=(a_{ij}):p\times m$ 和个性方差矩阵 $D=\mathrm{diag}(\sigma_1^2,\sigma_2^2,\cdots,\sigma_p^2)$。常用的参数估计方法有如下三种：主成分法，主因子法和极大似然法。

一、主成分法

设样本协方差矩阵 S 的特征值依次为 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_p\geqslant0$，相应的正交单位特征向量为 $t_1,t_2,\cdots,t_p$。选取相对较小的主成分个数 m，并使得累计贡献率 $\sum_{i=1}^{m}\lambda_i\Big/\sum_{i=1}^{p}\lambda_i$ 达到一个较高的百分比，则 S 可作如下的近似分解

$$\begin{aligned}S&=\lambda_1t_1t_1'+\cdots+\lambda_mt_mt_m'+\lambda_{m+1}t_{m+1}t_{m+1}'+\cdots+\lambda_pt_pt_p'\\&\approx\lambda_1t_1t_1'+\cdots+\lambda_mt_mt_m'+\hat{D}\\&=\hat{A}\hat{A}'+\hat{D}\end{aligned} \qquad (8.3.1)$$

其中 $\hat{A}=(\sqrt{\lambda_1}t_1,\cdots,\sqrt{\lambda_m}t_m)=(\hat{a}_{ij})$ 为 $p\times m$ 矩阵，$\hat{D}=\mathrm{diag}(\hat{\sigma}_1^2,\cdots,\hat{\sigma}_p^2)$，$\hat{\sigma}_i^2=s_{ii}-\sum_{j=1}^{m}\hat{a}_{ij}^2$，$i=1,2,\cdots,p$。

这里的 $\hat{A}$ 和 $\hat{D}$ 就是因子模型的一个解。因子载荷矩阵 $\hat{A}$ 的第 j 列与 S 的第 j 个主成分的系数仅相差一个倍数 $\sqrt{\lambda_j}$ $(j=1,2,\cdots,m)$，因此这个解就称为主成分解。

若 p 个原始变量的单位不同，则我们首先对原始变量作标准化变换，此时的样本协方差矩阵即为原始变量的样本相关矩阵 $\hat{R}$，用 $\hat{R}$ 代替(8.3.1)式中的 S，可类似地求得主成分解。

例 8.3.1　下表给出的数据是在洛杉矶十二个标准大都市居民统计地区中进行人口调查获得的。它有五个社会经济变量，它们分别是人口总数(x_1)、居民的教育程度或中等教育的年数(x_2)、佣人总数(x_3)、各种服务行业的人数(x_4)和中等的房价(x_5)。

表 8.1　　五个社会经济变量

编号	x_1	x_2	x_3	x_4	x_5
1	5700	12.8	2500	270	25000
2	1000	10.9	600	10	10000
3	3400	8.8	1000	10	9000
4	3800	13.6	1700	140	25000
5	4000	12.8	1600	140	25000
6	8200	8.3	2600	60	12000
7	1200	11.4	400	10	16000
8	9100	11.5	3300	60	14000
9	9900	12.5	3400	180	18000
10	9600	13.7	3600	390	25000
11	9600	9.6	3300	80	12000
12	9400	11.4	4000	100	13000

采用主成分法进行参数估计。经计算，相关矩阵为

$$\hat{R}=\begin{pmatrix}1.0000 & 0.0098 & 0.9725 & 0.4389 & 0.0224\\ 0.0098 & 1.0000 & 0.1543 & 0.6914 & 0.8631\\ 0.9725 & 0.1543 & 1.0000 & 0.5147 & 0.1219\\ 0.4389 & 0.6914 & 0.5147 & 1.0000 & 0.7777\\ 0.0224 & 0.8631 & 0.1219 & 0.7777 & 1.0000\end{pmatrix}$$

$\hat{R}$ 的特征值为

$$\lambda_1=2.8733,\quad \lambda_2=1.7967,\quad \lambda_3=0.2148$$

$$\lambda_4=0.0999,\qquad \lambda_5=0.0153$$

前两个特征值的累计贡献率达 93.4%，故可取 $m=2$，相应的结果列于表 8.2。

表 8.2　　当 $m=2$ 时的主成分解

变量	$\hat{h}_i^2$	$\hat{a}_{i1}$	$\hat{a}_{i2}$
x_1	0.9878	0.5810	0.8064
x_2	0.8851	0.7670	−0.5448
x_3	0.9793	0.6724	0.7261
x_4	0.8802	0.9324	−0.1043
x_5	0.9375	0.7912	−0.5582

从上表可见：(1) 五个变量在第一个公因子上都具有大的正载荷，尤其是 x_4 的载荷特别大。在第二个公因子上变量 x_1 和 x_3 都有较大的正载荷，x_2 和 x_5 都有较大的负载荷，x_1、x_3 与 x_2、x_5 形成了鲜明的对照，而在 x_4 上的载荷非常小。(2) 两个公因子对所有变量的共性方差估计都很大，在 0.8802 到 0.9878 之间。

二、主因子法

主因子法是因子分析中一种最简单、最有效的方法，它已得到了最普遍的应用。我们这里假定原始向量 $\boldsymbol{x}$ 已作了标准化变换。如果随机向量 $\boldsymbol{x}$ 满足因子模型(8.2.2)式，则有

$$R=AA'+D$$

其中 R 为 $\boldsymbol{x}$ 的相关矩阵，令

$$R^* = R - D = AA' \tag{8.3.2}$$

则称 R^* 为 x 的约相关矩阵。易见，R^* 中的对角元素是 h_i^2，而不是1，非对角元素和 R 中是完全一样的，并且 R^* 是一个非负定矩阵。我们首先在相关矩阵 R 及个性方差矩阵 $D=\mathrm{diag}(\sigma_1^2,\sigma_2^2,\cdots,\sigma_p^2)$ 已知的条件下，求出因子载荷矩阵 A。

由上一节因子模型的性质3知，A 的解是不唯一的，可以有许多。主因子法就是要求得到的解能使第一个公因子 f_1 对 x 的贡献 $g_1^2=\sum_{i=1}^{p} a_{i1}^2$ 达到最大，第二个公因子 f_2 对 x 的贡献 $g_2^2=\sum_{i=1}^{p} a_{i2}^2$ 次之，…，第 m 个公因子 f_m 对 x 的贡献最小。

由于 $\mathrm{rank}(R^*)=\mathrm{rank}(AA')=\mathrm{rank}(A)=m$，所以 R^* 有 m 个正特征值，依次记为 $\lambda_1^*\geqslant\lambda_2^*\geqslant\cdots\geqslant\lambda_m^*>0$，相应的正交单位特征向量为 $t_1^*,t_2^*,\cdots,t_m^*$，故 R^* 的谱分解为

$$\begin{aligned} R^* &= \lambda_1^* t_1^* t_1^{*\prime} + \lambda_2^* t_2^* t_2^{*\prime} + \cdots + \lambda_m^* t_m^* t_m^{*\prime} \\ &= (\sqrt{\lambda_1^*}\,t_1^*,\cdots,\sqrt{\lambda_m^*}\,t_m^*)\begin{pmatrix} \sqrt{\lambda_1^*}\,t_1^{*\prime} \\ \vdots \\ \sqrt{\lambda_m^*}\,t_m^{*\prime} \end{pmatrix} = AA' \end{aligned} \tag{8.3.3}$$

其中

$$A=(\sqrt{\lambda_1^*}\,t_1^*,\cdots,\sqrt{\lambda_m^*}\,t_m^*) \tag{8.3.4}$$

它就是我们所要求的主因子解。A 中第 j 列元素的平方和为 $(\sqrt{\lambda_j^*}\,t_j^*)'(\sqrt{\lambda_j^*}\,t_j^*)=\lambda_j^*$，即

$$\lambda_j^* = g_j^2 = \sum_{i=1}^{p} a_{ij}^2 \tag{8.3.5}$$

在实际应用中，相关矩阵 R 和个性方差矩阵 D 一般都是未知的，它们可通过一组样本 $x_1,x_2,\cdots,x_n$ 来进行估计。为了符号上的方便，我们将 R（或 R^*）的估计值仍记为 R（或 R^*）。估计个性方差 σ_i^2 等价于估计共性方差 h_i^2，这是因为由(8.2.13)式知

$$\sigma_i^2=1-h_i^2,\quad i=1,2,\cdots,p \tag{8.3.6}$$

σ_i^2(或 h_i^2)的较好估计一般很难直接得到,通常是先给出它的一个初始估计 $\hat{\sigma}_i^2$(或 $\hat{h}_i^2$),待载荷矩阵 A 估计好之后再作出 σ_i^2(或 h_i^2)的最终估计。

个性方差 σ_i^2(或共性方差 h_i^2)的常用初始估计方法有如下几种:

(1) $\hat{h}_i^2$ 取为原始变量 x_i 与其它所有原始变量 $x_1,\cdots,x_{i-1},x_{i+1},\cdots,x_p$ 的复相关系数的平方,则 $\hat{\sigma}_i^2=1-\hat{h}_i^2$。

(2) 取 $\hat{\sigma}_i^2=1/r^{ii}$,其中 R^{ii}是 R^{-1}的对角元素。

(3) 取 $\hat{h}_i^2=\max\limits_{j\neq i}|r_{ij}|$,则 $\hat{\sigma}_i^2=1-\hat{h}_i^2$。

(4) 取 $\hat{h}_i^2=1$,则 $\hat{\sigma}_i^2=0$,得到的 $\hat{A}$ 是一个主成分解。

因子的个数 m 应选取为多少呢?一般可采用主成分分析中确定主成分个数的原则,即寻求一个较小的自然数 m,使得 $\sum\limits_{j=1}^{m}\lambda_j^*\Big/\sum\limits_{j=1}^{p}\lambda_j^*$ 达到一个较高的百分比(如至少达到 85%)。需要指出的是,R^*的部分特征值可能是负的,这是因为 R^*是通过用样本估计 R 和 D 而得到的,可能已不再是非负定矩阵了。

最后,取 R^*的前 m 个正特征值 $\lambda_1^*\geqslant\lambda_2^*\geqslant\cdots\geqslant\lambda_m^*>0$ 及其相应的正交单位特征向量 $t_1^*,t_2^*,\cdots,t_m^*$,可以得到近似分解式

$$R^*\approx\hat{A}\hat{A}'$$

其中 $\hat{A}=(\sqrt{\lambda_1^*}\,t_1^*,\cdots,\sqrt{\lambda_m^*}\,t_m^*)=(\hat{a}_{ij})$,$\sigma_i^2$ 的最终估计为

$$\hat{\sigma}_i^2=1-\hat{h}_i^2=1-\sum_{j=1}^{m}\hat{a}_{ij}^2,\quad i=1,2,\cdots,p \tag{8.3.7}$$

我们称这样求得的 $\hat{A}$ 和 $\hat{D}=\mathrm{diag}(\hat{\sigma}_1^2,\hat{\sigma}_2^2,\cdots,\hat{\sigma}_p^2)$为因子模型的主因子解。

如果我们希望求得近似程度更好的解,则可以采用迭代主因子法,即利用(8.3.7)式中的 $\hat{\sigma}_i^2$ 再作为个性方差的初始估计,重复上述步骤,直至解稳定为止。

例 8.3.2　在例 8.3.1 中采用主因子法。

选用原始变量 x_i 与其它四个原始变量的复相关系数平方作为 $\hat{h}_i^2$ 的初始估计值。计算得

$$\hat{h}_1^2=0.9686,\quad \hat{h}_2^2=0.8223,\quad \hat{h}_3^2=0.9692$$

$$\hat{h}_4^2=0.7857,\quad \hat{h}_5^2=0.8470$$

于是约相关矩阵为

$$R^*=\begin{pmatrix}0.9686 & 0.0098 & 0.9725 & 0.4389 & 0.0224\\ 0.0098 & 0.8223 & 0.1543 & 0.6914 & 0.8631\\ 0.9725 & 0.1543 & 0.9692 & 0.5147 & 0.1219\\ 0.4389 & 0.6914 & 0.5147 & 0.7857 & 0.7777\\ 0.0224 & 0.8631 & 0.1219 & 0.7777 & 0.8470\end{pmatrix}$$

R^* 的特征值为

$$\lambda_1^*=2.7343,\quad \lambda_2^*=1.7161,\quad \lambda_3^*=0.0396$$

$$\lambda_4^*=-0.0245,\quad \lambda_5^*=-0.0726$$

故取 $m=2$，相应的计算结果列于表 8.3。

表 8.3 给出的结果与表 8.2 是类似的。在第一个公因子上，x_4 有最大的正载荷，而 x_1 有最小的正载荷。在第二个公因子上，x_1 和 x_3 有大的正载荷，而 x_2 和 x_5 有大的负载荷，x_4 有小的负载荷。所有的共性方差估计都很接近于初始的共性方差。

表 8.3　　当 $m=2$ 时的主因子解

变量	最终的 $\hat{h}_i^2$	$\hat{a}_{i1}$	$\hat{a}_{i2}$
x_1	0.9781	0.6253	0.7662
x_2	0.8176	0.7137	−0.5552
x_3	0.9720	0.7145	0.6794
x_4	0.7977	0.8790	−0.1585
x_5	0.8850	0.7422	−0.5781

三、极大似然法

设公因子 $f\sim N_m(\mathbf{0},I)$，特殊因子 $\varepsilon\sim N_p(\mathbf{0},D)$，且相互独立，则原始向量 $x\sim N_p(\mu,\Sigma)$。由样本 $x_1,x_2,\cdots,x_n$ 计算得到的似然函

数是 μ 和 Σ 的函数 $L(\mu,\Sigma)$。由于 $\Sigma=AA'+D$，故似然函数可更确切地表示为 $L(\mu,A,D)$。记 (μ,A,D) 的极大似然估计为 $(\hat{\mu},\hat{A},\hat{D})$，即有

$$L(\hat{\mu},\hat{A},\hat{D})=\max\ L(\mu,A,D)$$

可以证明，$\hat{\mu}=\bar{x}$，而 $\hat{A}$ 和 $\hat{D}$ 满足以下方程组

$$\begin{cases}\hat{\Sigma}\hat{D}^{-1}\hat{A}=\hat{A}(I_m+\hat{A}'\hat{D}^{-1}\hat{A})\\ \hat{D}=\mathrm{diag}(\hat{\Sigma}-\hat{A}\hat{A}')\end{cases}\tag{8.3.8}$$

其中 $\hat{\Sigma}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})'$，由于 A 的解是不唯一的，为了得到唯一解，可附加计算上方便的唯一性条件：

$$A'D^{-1}A\text{ 是对角矩阵}\tag{8.3.9}$$

(8.3.8)式中的 $\hat{A}$ 和 $\hat{D}$ 一般可用迭代方法解得。

极大似然法在正态性假定能较好地被满足或者在大样本的情况下，能给出比主因子法更好的估计，并且有令人满意的渐近性质。极大似然法的计算量大约是主因子法的 100 倍，这是由于极大似然估计需要用迭代方法计算并且要试着提取不同个数的因子。实际应用中，在使用极大似然法之前，一般先使用主因子法进行分析，以便给出因子个数的初步估计。

例 8.3.3　在例 8.3.1 中采用极大似然估计。

从例 8.3.2 的主因子法中得知，应取 $m=2$。用极大似然法求出载荷矩阵，列于表 8.4。$\hat{h}_i^2$ 的初始估计值与例 8.3.2 相同。

表 8.4　　当 $m=2$ 时的极大似然估计

变量	最终的 $\hat{h}_i^2$	$\hat{a}_{i1}$	$\hat{a}_{i2}$
x_1	1.0000	1.0000	0.0000
x_2	0.8101	0.0098	0.9000
x_3	0.9596	0.9725	0.1180
x_4	0.8156	0.4389	0.7893
x_5	0.9219	0.0224	0.9599

§8.4 因子旋转

因子模型的参数估计完成之后，还必须对模型中的公因子进行合理的解释。进行这种解释通常需要一定的专业知识和经验，要对每个公因子给出具有实际意义的一种名称，它可用来反映在预测每个可观测的原始变量时这个公因子的重要性，也就是相应于这个因子的载荷。因子的解释带有一定的主观性，我们常常通过旋转公因子的方法来减少这种主观性。

公因子是否易于解释，很大程度上取决于因子载荷矩阵 A 的元素结构。假设 A 是从相关矩阵 R 出发求得的，则 $\sum_{j=1}^{m} a_{ij}^2 = h_i^2 \leqslant 1$，故有 $|a_{ij}| \leqslant 1$，即 A 的所有元素均在 -1 和 1 之间。如果载荷矩阵 A 的所有元素都接近 0 或 ± 1，则模型的公因子就容易解释。这时可将原始变量 $x_1, x_2, \cdots, x_p$ 分成 m 个部分，第一部分对应第一个公因子 f_1，…，第 m 部分对应第 m 个公因子 f_m。反之，如果载荷矩阵 A 的多数元素居中，不大不小，则对模型的公因子将难以作出解释，此时必须进行因子旋转，使得旋转之后的载荷矩阵在每一列上元素的绝对值尽量地拉开大小距离，也就是尽可能地使其中的一些元素接近于 0，另一些元素接近于 ± 1。

因子旋转方法有正交旋转和斜交旋转两类，本书中我们只讨论正交旋转。对公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 T，使 AT 能有更鲜明的实际意义。旋转后的公因子向量为 $f^* = T'f$，它的各分量 $f_1^*, f_2^*, \cdots, f_m^*$ 也是互不相关的公因子。正交矩阵 T 的不同选取法构成了正交旋转的各种不同方法，在这些方法中使用最普遍的是最大方差旋转法(Varimax)，本节仅介绍这一种正交旋转法。

令

$$A^* = AT = (a_{ij}^*), \quad d_{ij} = a_{ij}^* / h_i$$

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

则 A^* 的第 j 列元素平方的相对方差可定义为

$$V_j=\frac{1}{p}\sum_{i=1}^{p}(d_{ij}^2-\bar{d}_j)^2 \tag{8.4.1}$$

用 a_{ij}^* 除以 h_i 是为了消除各个原始变量 x_i 对公因子依赖程度不同的影响，选择除数 h_i 是因为 A^* 的第 i 行平方和

$$h_i^{*2}=\sum_{j=1}^{m}a_{ij}^{*2}=(a_{i1}^*,\cdots,a_{im}^*)\begin{pmatrix}a_{i1}^*\\ \vdots\\ a_{im}^*\end{pmatrix}$$

$$=(a_{i1},\cdots,a_{im})TT'\begin{pmatrix}a_{i1}\\ \vdots\\ a_{im}\end{pmatrix}=\sum_{j=1}^{m}a_{ij}^2=h_i^2$$

取 d_{ij}^2 是为了消除 d_{ij} 符号不同的影响。所谓最大方差旋转法就是选择正交矩阵 T，使得矩阵 A^* 所有 m 个列元素平方的相对方差之和

$$V=V_1+V_2+\cdots+V_m \tag{8.4.2}$$

达到最大。

当 $m=2$ 时，设已求出的因子载荷矩阵为

$$A=\begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\\ \vdots & \vdots\\ a_{p1} & a_{p2}\end{pmatrix}$$

现选取正交变换矩阵 T 进行因子旋转，T 可以表示为

$$T=\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}$$

这里 θ 是坐标平面上因子轴按顺时针方向旋转的角度，只要求出 θ，也就求出了 T。

$$A^*=AT=\begin{pmatrix} a_{11}\cos\theta+a_{12}\sin\theta & -a_{11}\sin\theta+a_{12}\cos\theta \\ a_{21}\cos\theta+a_{22}\sin\theta & -a_{21}\sin\theta+a_{22}\cos\theta \\ \vdots & \vdots \\ a_{p1}\cos\theta+a_{p2}\sin\theta & -a_{p1}\sin\theta+a_{p2}\cos\theta \end{pmatrix}$$

$$=\begin{pmatrix} a_{11}^* & a_{12}^* \\ a_{21}^* & a_{22}^* \\ \vdots & \vdots \\ a_{p1}^* & a_{p2}^* \end{pmatrix}.$$

$$d_{ij}=a_{ij}^*/h_i,\quad i=1,2,\cdots,p,\quad j=1,2$$

$$\bar{d}_j=\frac{1}{p}\sum_{i=1}^{p}d_{ij}^2,\quad j=1,2$$

再由(8.4.1)式和(8.4.2)式即可求得 A^* 各列元素平方的相对方差之和 V。显然,V 是旋转角度 θ 的函数,按照最大方差旋转法的原则,应求出 θ,使 V 达到最大。由微积分中求极值的方法,将 V 对 θ 求导,并令其为零,可以推得 θ 满足

$$\mathrm{tg}4\theta=\frac{D-2AB/p}{C-(A^2-B^2)/p} \tag{8.4.3}$$

其中

$$A=\sum_{i=1}^{p}u_i,\quad B=\sum_{i=1}^{p}v_i$$

$$C=\sum_{i=1}^{p}(u_i^2-v_i^2),\quad D=2\sum_{i=1}^{p}u_iv_i$$

而

$$u_i=\left(\frac{a_{i1}}{h_i}\right)^2-\left(\frac{a_{i2}}{h_i}\right)^2,\quad v_i=2\frac{a_{i1}a_{i2}}{h_i^2}$$

当 $m>2$ 时,我们可以逐次对每两个公因子进行上述的旋转。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

$$T_{lk}=\begin{matrix}\begin{pmatrix}1 & & & & & & & & & \\ & \ddots & & & & & & & & \\ & & 1 & & & & & & & \\ & & & \cos\theta & & & & -\sin\theta & & \\ & & & & 1 & & & & & \\ & & & & & \ddots & & & & \\ & & & & & & 1 & & & \\ & & & \sin\theta & & & & \cos\theta & & \\ & & & & & & & & 1 & \\ & & & & & & & & & \ddots \\ & & & & & & & & & & 1\end{pmatrix} & \begin{matrix}\\ \\ \\ l \\ \\ \\ \\ k \\ \\ \\ \\ \end{matrix} \\ \begin{matrix} & & & l & & & & k & & & \end{matrix} & \end{matrix}$$

其中 θ 是因子轴 f_l 和 f_k 的旋转角度，矩阵中其余位置上的元素全为 0。m 个公因子的两两配对旋转共需进行 $\binom{m}{2}=\frac{1}{2}m(m-1)$ 次，称其为完成了第一轮旋转，并记第一轮旋转后的因子载荷矩阵为 $A^{(1)}$。然后再重新开始，进行第二轮的 $\binom{m}{2}$ 次配对旋转，新的因子载荷矩阵记为 $A^{(2)}$。如此继续旋转下去，记第 s 轮旋转后的因子载荷矩阵为 $A^{(s)}$，得到的一系列因子载荷矩阵为

$$A^{(1)},A^{(2)},\cdots,A^{(s)},\cdots$$

记 $V^{(s)}$ 为 $A^{(s)}$ 各列元素平方的相对方差之和，则必然有

$$V^{(1)}\leqslant V^{(2)}\leqslant\cdots\leqslant V^{(s)}\leqslant\cdots$$

这是一个有界的单调上升数列，因此一定会收敛到某一极限。在实际应用中，当 $V^{(s)}$ 的值变化不大时，即可停止旋转。

例 8.4.1　对例 8.3.2 的主因子法($m=2$)用最大方差旋转法求得的因子载荷矩阵列于表 8.5。

表 8.5　　最大方差旋转后的因子载荷矩阵

变量	$\hat{h}_i^2$	$\hat{a}_{i1}$	$\hat{a}_{i2}$
x_1	0.9781	0.0226	0.9887
x_2	0.8176	0.9042	0.0006
x_3	0.9720	0.1463	0.9750
x_4	0.7977	0.7909	0.4151
x_5	0.8850	0.9407	−0.0001

从上表可以看出，在第一个公因子上，x_2，x_4 和 x_5 有大的正载荷，而 x_1 和 x_3 的载荷很小，这个因子可解释为福利条件因子。在第二个公因子上，x_1 和 x_3 有大的正载荷，x_4 有较小的正载荷，而 x_2 和 x_5 只有很小的载荷，这个因子可解释为人口因子。

§8.5　因子得分

一、因子得分的概念

我们再回过来看一下因子模型(8.2.2)式，即

$$\boldsymbol{x}=\boldsymbol{\mu}+A\boldsymbol{f}+\boldsymbol{\varepsilon}$$

设 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 为一组样本。在前面的讨论中，我们根据这一组样本估计出了公因子个数 m、因子载荷矩阵 A 和个性方差矩阵 D，并试图对公因子 $f_1,f_2,\cdots,f_m$ 进行合理的解释，即给出具有实际意义的名称。如果对这些公因子难以作出解释，则可以通过因子旋转的方法使得旋转后的公因子有着更鲜明的实际意义。实际上，还有一个问题是令我们非常感兴趣的，就是给出每一个体 $\boldsymbol{x}_i$ 对 m 个公因子的得分。必须指出的是，因子得分的计算并不是通常意义下的参数估计，而是对不可观测的随机变量 $f_1,f_2,\cdots,f_m$ 作出估计。因子模型(8.2.1)式意味着这些公因子一般不是可观测原始变量 $x_1,x_2,\cdots,x_p$ 的线性组合，而是非线性组合。这些公因子的得分是无法直接计算得到的，但它们可用各种不同的方法来进行估计。为了数学上处理的方便，人们一般还是将公因子得分的估计值表达

为原始变量 $x_1,x_2,\cdots,x_p$ 的线性函数。以下我们介绍两种常用的因子得分估计方法。

二、巴特莱特(Bartlett)因子得分

因子模型(8.2.1)式可以写为

$$\begin{cases} x_1-\mu_1=a_{11}f_1+a_{12}f_2+\cdots+a_{1m}f_m+\varepsilon_1 \\ x_2-\mu_2=a_{21}f_1+a_{22}f_2+\cdots+a_{2m}f_m+\varepsilon_2 \\ \quad\vdots \\ x_p-\mu_p=a_{p1}f_1+a_{p2}f_2+\cdots+a_{pm}f_m+\varepsilon_p \end{cases} \tag{8.5.1}$$

其中 $V(\varepsilon_i)=\sigma_i^2$，$i=1,2,\cdots,p$。我们可以采用与求解线性回归模型相同的方法来求得因子得分 $f_1,f_2,\cdots,f_m$。由于 p 个个性方差不全相等，因此应采用加权的最小二乘估计法，也就是寻求 $f_1,f_2,\cdots,f_m$ 的一组取值 $\hat{f}_1,\hat{f}_2,\cdots,\hat{f}_m$ 使得加权的“残差”平方和

$$\sum_{i=1}^{p}[(x_i-\mu_i)-(a_{i1}\hat{f}_1+a_{i2}\hat{f}_2+\cdots+a_{im}\hat{f}_m)]^2/\sigma_i^2 \tag{8.5.2}$$

达到最小，这样求得的解 $\hat{f}_1,\hat{f}_2,\cdots,\hat{f}_m$ 就称为巴特莱特因子得分。

(8.5.1)式用矩阵来表示就是

$$\boldsymbol{x}-\boldsymbol{\mu}=A\boldsymbol{f}+\boldsymbol{\varepsilon} \tag{8.5.3}$$

(8.5.2)式可用矩阵表示为

$$(\boldsymbol{x}-\boldsymbol{\mu}-A\hat{\boldsymbol{f}})'D^{-1}(\boldsymbol{x}-\boldsymbol{\mu}-A\hat{\boldsymbol{f}}) \tag{8.5.4}$$

其中 $\hat{\boldsymbol{f}}=(\hat{f}_1,\hat{f}_2,\cdots,\hat{f}_m)'$。用微分学求极值的方法可以解得巴特莱特因子得分为

$$\hat{\boldsymbol{f}}=(A'D^{-1}A)^{-1}A'D^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \tag{8.5.5}$$

在实际应用中，用估计值 $\bar{\boldsymbol{x}}$、$\hat{A}$ 和 $\hat{D}$ 分别代替上述公式中的 μ、A 和 D，并将每个样品的数据 $\boldsymbol{x}_i$ 代入，便可得到相应的因子得分 $\hat{f}$。

若将 f 和 ε 不相关的假定加强为相互独立，则在 f 值已知的条件下，由(8.5.5)式和(8.5.3)式可得因子得分 $\hat{\boldsymbol{f}}$ 的条件数学期望

$$\begin{aligned} E(\hat{\boldsymbol{f}}\mid\boldsymbol{f})&=(A'D^{-1}A)^{-1}A'D^{-1}E(A\boldsymbol{f}+\boldsymbol{\varepsilon}\mid\boldsymbol{f}) \\ &=(A'D^{-1}A)^{-1}A'D^{-1}A\boldsymbol{f} \end{aligned}$$

$$=f \tag{8.5.6}$$

因此，从条件意义上来说巴特莱特因子得分 $\hat{f}$ 是无偏的。我们再来计算反映 $\hat{f}$ 估计精度的平均预报误差 $E[(\hat{f}-f)(\hat{f}-f)']$，由(8.5.5)式和(8.5.3)式得

$$\begin{aligned}\hat{f}-f &=(A'D^{-1}A)^{-1}A'D^{-1}(Af+\varepsilon)-f\\ &=(A'D^{-1}A)^{-1}A'D^{-1}\varepsilon\end{aligned}$$

故

$$\begin{aligned}&E[(\hat{f}-f)(\hat{f}-f)']\\ &=(A'D^{-1}A)^{-1}A'D^{-1}E(\varepsilon\varepsilon')D^{-1}A(A'D^{-1}A)^{-1}\\ &=(A'D^{-1}A)^{-1}A'D^{-1}DD^{-1}A(A'D^{-1}A)^{-1}\\ &=(A'D^{-1}A)^{-1}\end{aligned} \tag{8.5.7}$$

三、汤姆森(Thompson)因子得分

在因子模型(8.2.2)式中，假设$\begin{pmatrix} f \\ x \end{pmatrix}$服从$(m+p)$元正态分布，由条件(8.2.3)式得

$$E\begin{pmatrix} f \\ x \end{pmatrix}=\begin{pmatrix} E(f) \\ E(x) \end{pmatrix}=\begin{pmatrix} 0 \\ \mu \end{pmatrix} \tag{8.5.8}$$

$$\begin{aligned}V\begin{pmatrix} f \\ x \end{pmatrix}&=E\left[\begin{pmatrix} f \\ x-\mu \end{pmatrix}(f',(x-\mu)')\right]\\ &=\begin{pmatrix} E(ff') & E[f(x-\mu)'] \\ E[(x-\mu)f'] & E[(x-\mu)(x-\mu)'] \end{pmatrix}\\ &=\begin{pmatrix} I & E[f(Af+\varepsilon)'] \\ E[(Af+\varepsilon)f'] & \Sigma \end{pmatrix}\\ &=\begin{pmatrix} I & A' \\ A & \Sigma \end{pmatrix}\end{aligned} \tag{8.5.9}$$

由(3.2.6)式知，在 x 给定的条件下，f 的条件数学期望

$$E(f|x)=A'\Sigma^{-1}(x-\mu) \tag{8.5.10}$$

再由(8.2.4)式知，$\Sigma=AA'+D$，因此(8.5.10)式也可表示为

$$E(f|x)=A'(AA'+D)^{-1}(x-\mu) \tag{8.5.11}$$

或者

$$\tilde{f}=(I+A'D^{-1}A)^{-1}A'D^{-1}(x-\mu) \qquad (8.5.12)$$

上面两式相等，这是因为(可以直接验证)

$$A'(AA'+D)^{-1}=(I+A'D^{-1}A)^{-1}A'D^{-1}$$

称 $\tilde{f}$ 为汤姆森因子得分。在实际应用中，用 $\hat{\mu}=\bar{x}$、$\hat{A}$ 和 $\hat{D}$ 代替(8.5.12)式中的 μ、A 和 D 得因子得分。

由(8.5.12)式和(8.5.3)式得

$$\begin{aligned}E(\tilde{f}|f)&=(I+A'D^{-1}A)^{-1}A'D^{-1}E(Af+\varepsilon|f)\\&=(I+A'D^{-1}A)^{-1}A'D^{-1}Af\\&=(I+A'D^{-1}A)^{-1}(A'D^{-1}A+I-I)f\\&=f-(I+A'D^{-1}A)^{-1}f\end{aligned} \qquad (8.5.13)$$

所以，汤姆森因子得分是有偏的。

$$\begin{aligned}\tilde{f}-f&=(I+A'D^{-1}A)^{-1}A'D^{-1}(Af+\varepsilon)-f\\&=(I+A'D^{-1}A)^{-1}A'D^{-1}\varepsilon-(I+A'D^{-1}A)^{-1}f\end{aligned}$$

故 $\tilde{f}$ 的平均预报误差

$$\begin{aligned}&E[(\tilde{f}-f)(\tilde{f}-f)']\\&=(I+A'D^{-1}A)^{-1}A'D^{-1}E(\varepsilon\varepsilon')D^{-1}A(I+A'D^{-1}A)^{-1}\\&\quad+(I+A'D^{-1}A)^{-1}E(ff')(I+A'D^{-1}A)^{-1}\\&=(I+A'D^{-1}A)^{-1}(A'D^{-1}A+I)(I+A'D^{-1}A)^{-1}\\&=(I+A'D^{-1}A)^{-1}\end{aligned} \qquad (8.5.14)$$

比较(8.5.14)与(8.5.7)两式，由于 $(A'D^{-1}A)^{-1}-(I+A'D^{-1}A)^{-1}$ 是正定矩阵(见习题 8.2)，因此汤姆森因子得分比巴特莱特因子得分有更小的平均预报误差。

例 8.5.1　在例 8.4.1 中，

$$\hat{A}=\begin{bmatrix}0.0226 & 0.9887\\0.9042 & 0.0006\\0.1463 & 0.9750\\0.7909 & 0.4151\\0.9407 & -0.0001\end{bmatrix}$$

$\hat{D}=$

$$\begin{pmatrix} 1-0.9781 & & & & 0 \\ & 1-0.8176 & & & \\ & & 1-0.9720 & & \\ & & & 1-0.7977 & \\ 0 & & & & 1-0.8850 \end{pmatrix}$$

$$=\mathrm{diag}(0.0219,\ 0.1824,\ 0.0280,\ 0.2023,\ 0.1150)$$

因此，巴特莱特因子得分为

$$\hat{\boldsymbol{f}}=(\hat{A}'\hat{D}^{-1}\hat{A})^{-1}\hat{A}'\hat{D}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}})$$

$$=\begin{pmatrix} 0.7340 & 0.0846 & 0.3104 & -0.0548 & -0.0766 \\ -0.4986 & -0.0573 & 0.5593 & 0.2912 & 0.4923 \end{pmatrix}\begin{pmatrix} x_1-\bar{x}_1 \\ x_2-\bar{x}_2 \\ x_3-\bar{x}_3 \\ x_4-\bar{x}_4 \\ x_5-\bar{x}_5 \end{pmatrix}$$

汤姆森因子得分为

$$\tilde{\boldsymbol{f}}=(I+\hat{A}'\hat{D}^{-1}\hat{A})^{-1}\hat{A}'\hat{D}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}})$$

$$=\begin{pmatrix} 0.7164 & 0.0826 & 0.3112 & -0.0508 & -0.0700 \\ -0.4626 & -0.0532 & 0.5311 & 0.2742 & 0.4637 \end{pmatrix}\begin{pmatrix} x_1-\bar{x}_1 \\ x_2-\bar{x}_2 \\ x_3-\bar{x}_3 \\ x_4-\bar{x}_4 \\ x_5-\bar{x}_5 \end{pmatrix}$$

洛杉矶十二个标准大都市居民统计地区的因子得分列于表8.6。

表 8.6　　十二个居民统计地区的因子得分

编号	巴特莱特因子得分		汤姆森因子得分	
	f_1	f_2	$\mathcal{F}_1$	$\mathcal{F}_2$
1	−966.59	4344.77	−903.72	4089.62
2	−3843.21	−1833.90	−3798.89	−1771.95
3	−1881.06	−3299.03	−1885.21	−3133.43
4	−2602.28	4806.83	−2507.18	4508.04
5	−2486.59	4651.22	−2395.09	4362.44
6	1906.16	−3306.18	1838.81	−3099.43
7	−4217.95	908.06	−4137.88	811.52
8	2631.12	−2379.09	2561.66	−2216.76
9	2936.52	−718.12	2879.85	−646.07
10	2230.89	3050.25	2226.54	2902.43
11	3150.03	−3607.00	3058.71	−3369.91
12	3142.96	−2617.78	3062.40	−2436.49

§8.6　SAS 程序及输出

对例 8.3.1 等编制 SAS 程序如下：

```
data socecon;
    input x1—x5;
    cards;
5700    12.8    2500    270    25000
1000    10.9    .600     10    10000
3400     8.8    1000     10     9000
3800    13.6    1700    140    25000
4000    12.8    1600    140    25000
8200     8.3    2600     60    12000
1200    11.4     400     10    16000
```

```
9100      11.5      3300       60      14000
9900      12.5      3400      180      18000
9600      13.7      3600      390      25000
9600       9.6      3300       80      12000
9400      11.4      4000      100      13000
;
proc factor data=socecon simple corr;
run;
proc factor data=socecon priors=smc preplot
        rotate=varimax reorder plot;
run;
proc factor data=socecon method=ml heywood n=2;
run;
```

屏幕输出：

输出 8.1　　均值、标准差及相关矩阵

(1) Means and Standard Deviations from 12 observations

	X1	X2	X3	X4	X5
Mean	6241.6667	11.441667	2333.3333	120.83333	17000
Std Dev	3439.9943	1.7865448	1241.2115	114.92751	6367.5313

(2) Correlations

	X1	X2	X3	X4	X5
X1	1.00000	0.00975	0.97245	0.43887	0.02241
X2	0.00975	1.00000	0.15428	0.69141	0.86307
X3	0.97245	0.15428	1.00000	0.51472	0.12193
X4	0.43887	0.89141	0.51472	1.00000	0.77765
X5	0.02241	0.86307	0.12193	0.77765	1.00000

输出 8.2　　主成分法的输出结果

Initial Factor Method: Principal Components

(3) Prior Communality Estimates: ONE

(4) Eigenvalues of the Correlation Matrix: ① Total = 5 ② Average = 1

	1	2	3	4	5
③ Eigenvalue	2.8733	1.7987	0.2148	0.0999	0.0153
④ Difference	1.0787	1.5818	0.1149	0.0847	
⑤ Proportion	0.5747	0.3593	0.0430	0.0200	0.0031
⑤ Cumulative	0.5747	0.9340	0.9770	0.9969	1.0000

(5) 2 factors will be retained by the MINEIGEN criterion.

(8) Factor Pattern

	FACTOR1	FACTOR2
X1	0.58096	0.80642
X2	0.76704	-0.54478
X3	0.67243	0.72605
X4	0.93239	-0.10431
X5	0.79116	-0.55618

(7) Variance explained by each factor

FACTOR1	FACTOR2
2.873314	1.796660

(8) Final Communality Estimates: Total = 4.889974

X1	X2	X3	X4	X5
0.987826	0.885106	0.979306	0.880236	0.937500

输出 8.3　　主因子法的输出结果

Initial Factor Method: Principal Factors

Prior Communality Estimates: SMC

X1	X2	X3	X4	X5
0.968592	0.S222S5	0.969181	0.785724	0.847019

Eigenvalues of the Reduced Corralation Matrix:

Total = 4.39280116 Average = 0.87856023

	1	2	3	4	5
Eigenvalue	2.7343	1.7161	0.0396	-0.0245	-0.0726
Difference	1.0182	1.6765	0.0841	0.0481	
Proportion	0.6225	0.3907	0.0090	-0.0056	-0.0165
Cumulative	0.6225	1.0131	1.0221	1.0165	1.0000

2 factors will be retainad by the PROPORTION criterion.

Factor Pattern

	FACTOR1	FACTOR2
X4	0.87899	-0.15847
X5	0.74215	-0.57606
X3	0.71447	0.67936
X2	0.71370	-0.55515
X1	0.62533	0.76621

Variance axplainad by each factor

FACTOR1	FACTOR2
2.734301	1.716069

Final Communality Eatimatea: Total = 4.450370

X1	X2	X3	X4	X5
0.976113	0.817564	0.971999	0.797743	0.884950

输出 8.4　　没有旋转的因子模型图

```
Initial Factor Method: Principal Factors

                (9) Plot of Factor Pattern for FACTOR1 and FACTOR2

                                        FACTOR1
                                           1

                                      D   .9

                                          .8

                                    E
                                     B    .7                    C
                                                                 A
                                          .6

                                          .5

                                          .4

                                          .3

                                          .2

                                                                        F
                                          .1                            A
                                                                        C
-1 -.9-.8-.7-.6-.5-.4-.3-.2-.1  0 .1 .2 .3 .4 .5 .6 .7 .6 .9 1.0       T
                                                                        O
                                         -.1                            R
                                                                        2
                                         -.2

                                         -.3

                                         -.4

                                         -.5

                                         -.6

                                         -.7

                                         -.8

                                         -.9

                                          -1

X1      =A    X2      =B    X3      =C    X4      =D    X5      =E
```

输出 8.5　　方差最大旋转法的结果

Rotation Method: Varimax

(10) Orthogonal Transformation Matrix

	1	2
1	0.78895	0.61446
2	-0.61446	0.78895

(11) Rotated Factor Pattern

	FACTOR1	FACTOR2
X5	0.94072	-0.00004
X2	0.90419	0.00055
X4	0.79085	0.41509
X1	0.02255	0.98874
X3	0.14625	0.97499

(12) Variance explained by each factor

FACTOR1	FACTOR2
2.349857	2.100513

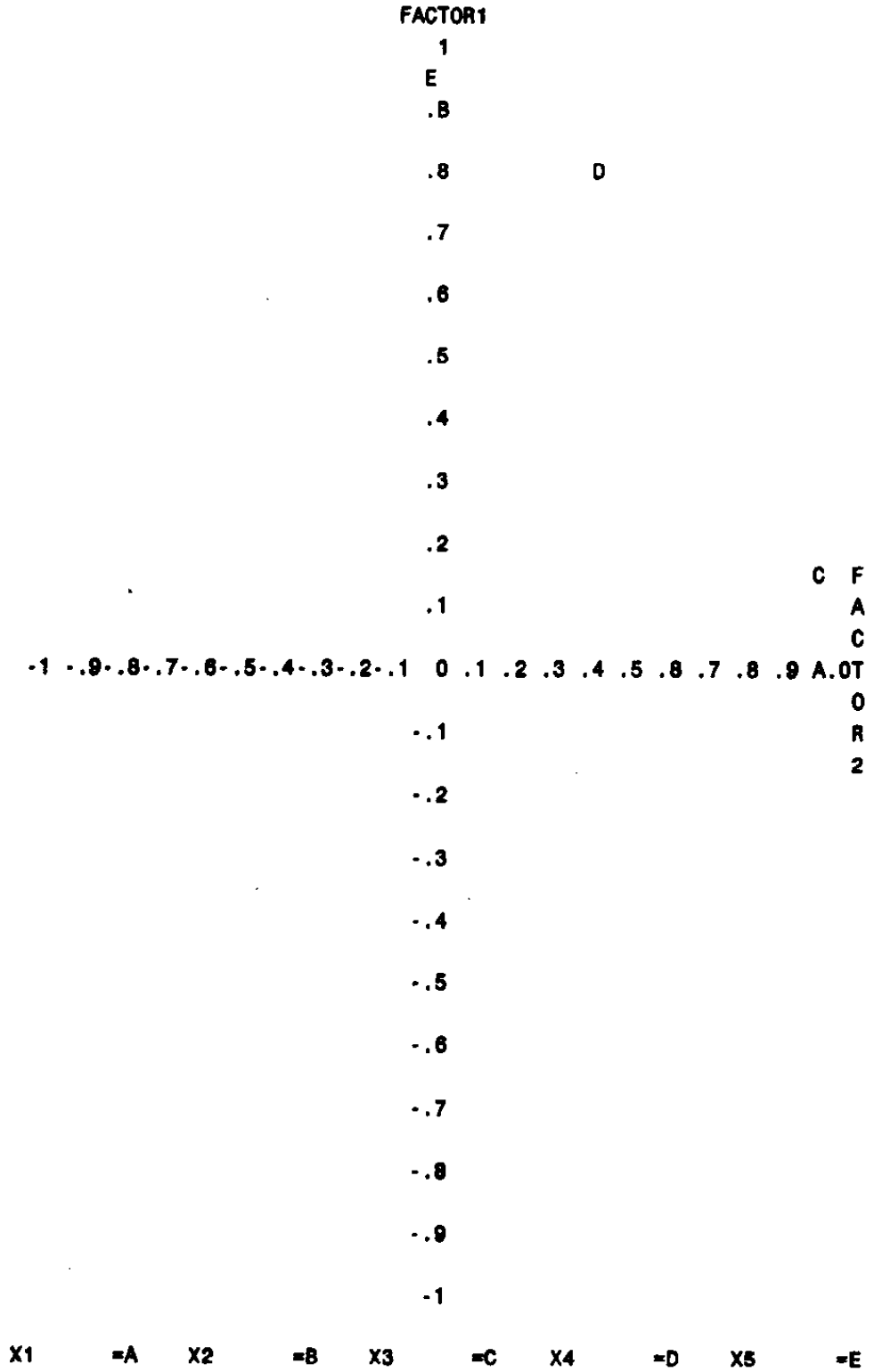
Final Communality Estimates: Total = 4.450370

X1	X2	X3	X4	X5
0.978113	0.817564	0.971999	0.797743	0.884950

输出 8.6　　旋转后的因子模型图

Rotation Method: Varimax

(13) Plot of Factor Pattern for FACTOR1 and FACTOR2



输出 8.7　　极大似然估计法的结果

Initial Factor Method: Maximum Likelihood

Prior Communality Estimates: SMC

X1	X2	X3	X4	X5
0.968592	0.822285	0.969181	0.785724	0.847019

Initial Factor Method: Maximum Likelihood

Factor Pattern

	FACTOR1	FACTOR2
X1	1.00000	0.00000
X2	0.00975	0.90003
X3	0.97245	0.11797
X4	0.43887	0.78930
X5	0.02241	0.95989

Variance axplained by each factor

	FACTOR1	FACTOR2
Unweighted	2.138881	2.368353

Final Communality Estimates and Variable Weights

Total Communality: Unweighted = 4.507214

Communality

X1	X2	X3	X4	X5
1.000000	0.610145	0.959571	0.815603	0.921894

输出说明如下：

(1) 每个变量的均值和标准差。

(2) 相关矩阵。

(3) 初始共性方差估计值。

(4) 相关矩阵的特征值。包括：

① 特征值总和；
② 平均特征值；
③ 特征值；
④ 相邻两特征值之差；
⑤ 被解释的方差比例；
⑥ 方差的累计比例。
(5) 确定的因子数目。
(6) 因子模型。
(7) 每个因子解释的方差。
(8) 最终共性方差估计。
(9) 因子模型图。
(10) 正交变换矩阵。
(11) 旋转后的因子模型。
(12) 旋转后每个因子说明的方差。
(13) 旋转后的因子模型图。

小 结

1. 因子分析是主成分分析的推广，它也是一种降维技术，其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。

2. 因子模型在形式上与线性回归模型很相似，但两者有着本质的区别：回归模型中的自变量是可观测到的，而因子模型中的各公因子是不可观测的隐变量。而且，两个模型的参数意义很不相同。

3. 因子载荷矩阵不是唯一的，利用这一点通过因子的旋转，可以使得旋转后的因子有更鲜明的实际意义。

4. 因子载荷矩阵的元素及一些元素组合有很明确的统计意义。

5. 因子模型中常用的参数估计方法有：主成分法，主因子法和极大似然法。

6. 在实际应用中，常从相关矩阵 R 出发进行因子模型分析。

7. 常用的因子得分估计方法有：巴特莱特因子得分和汤姆森因子得分两种方法。

习 题

8.1 比较因子分析和主成分分析的关系，说明它们的相似和不同之处。

8.2 在(8.5.14)和(8.5.7)两式中，试证 $(A'D^{-1}A)^{-1}-(I+A'D^{-1}A)^{-1}$ 是正定矩阵。

8.3 公司老板与48名申请工作的人进行面谈，然后就申请人十五个方面进行了打分，十五个变量为：申请信的形式(x_1)、外貌(x_2)、专业能力(x_3)、讨人喜欢的能力(x_4)、自信心(x_5)、洞察力(x_6)、诚实(x_7)、推销本领(x_8)、经验(x_9)、驾驶汽车本领(x_{10})、志向(x_{11})、领会能力(x_{12})、潜在能力(x_{13})、对工作要求强烈程度(x_{14})和是否合适该工作(x_{15})。根据数据算得的样本相关矩阵列于下表，试作因子分析。

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.00	0.24	0.04	0.31	0.09	0.23	−0.11	0.27	0.55	0.35	0.28	0.34	0.37	0.47	0.59
2		1.00	0.12	0.38	0.43	0.37	0.35	0.48	0.14	0.34	0.55	0.51	0.51	0.28	0.38
3			1.00	0.00	0.00	0.08	−0.03	0.05	0.27	0.09	0.04	0.20	0.29	−0.32	0.14
4				1.00	0.30	0.48	0.65	0.35	0.14	0.39	0.35	0.50	0.61	0.69	0.33
5					1.00	0.81	0.41	0.82	0.02	0.70	0.84	0.72	0.67	0.48	0.25
6						1.00	0.36	0.83	0.15	0.70	0.76	0.88	0.78	0.53	0.42
7							1.00	0.23	−0.16	0.28	0.21	0.39	0.42	0.45	0.00
8								1.00	0.23	0.81	0.86	0.77	0.73	0.55	0.55
9									1.00	0.34	0.20	0.30	0.35	0.21	0.69
10										1.00	0.78	0.71	0.79	0.61	0.62
11											1.00	0.78	0.77	0.55	0.43
12												1.00	0.88	0.55	0.53
13													1.00	0.54	0.57
14														1.00	0.40
15															1.00