第 20 卷第 4 期
2001 年 7 月

无锡轻工大学学报
Journal of Wuxi University of Light Industry

Vol.20 No.4
Jul. 2001

文章编号:1009-038X(2001)04-0409-07

主成份投影法用于多目标决策与评价

吴有炜
(无锡轻工大学 计算科学与信息传播系,江苏 无锡 214036)

摘 要:提出了一种多目标决策与评价的新方法——主成份投影法。该方法概念清楚,涵义明确,本质上是一种加权方法,但与其它一些加权方法并非完全一致,且具有不同的经济涵义和数学涵义。

关键词:多目标决策;主成份分析;向量投影
中图分类号:O231　　**文献标识码**:A

A Principal Component Projection Method for Multiindicies Decision and Evaluation

WU You-wei
(Department of Calculation Science and Information Communication, Wuxi University of Light Industry, Wuxi 214036, China)

Abstract: A new method named principal component projection method for multiindices and evaluation was presented in the paper. The new method is a kind of additive weighting methods essentially. However, there is different mathematic and economic meaning between the new and the old.

Key words: multiindices decision; principal component analysis; vector projection

多目标决策或多目标综合评价是决策论中常见的问题,例如经济研究中经常会遇到多项经济指标评价经济效益的综合评价问题.作者首先对多项指标向量一次主观赋权后进行主成份分析,消除各项评价指标间相关性影响,再二次客观赋权,然后构造了理想决策向量,采用比较各决策向量在理想决策向量上的投影,得到了一种全新的决策方法,并成功用于多目标综合决策评价,取得比较满意的评价结果[1,2]。该主成份投影法概念清楚,借助于主成份分析软件,可操作性强,具有一定的推广和实用价值。

1 主成份投影法原理

设有 n 个决策方案(或被评价对象),由 m 个指标来描述,得到原始的决策数据矩阵:

$$
\begin{array}{cc}
 & \begin{array}{cccc} G_1 & G_2 & \cdots & G_m \end{array} \\
\boldsymbol{Z}=\begin{array}{c} A_1 \\ A_2 \\ \\ A_n \end{array} & \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ & \cdots\cdots\cdots & & \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}
\end{array}
$$

数据阵 Z 的第 i 行表示决策方案 A_i 的 m 个指标值($x_{i1}, x_{i2}, \cdots, x_{im}$)($i=1,2,\cdots,n$)

数据阵 Z 的第 j 列表示用评价指标 G_j 评价 n

收稿日期:2000-12-21; 修订日期:2001-04-22.
作者简介:吴有炜(1949-),男,江苏无锡人,理学学士,副教授.

个决策方案的评价值$(x_{1j},x_{2j},\cdots,x_{nj})^{\mathrm{T}}$，$(j=1,2,\cdots,m)$

在多目标决策或综合评价实践中，人们总希望选取的评价指标越多越好．但是过多的评价指标不仅会增加评价工作量，而且会因评价指标间的相关联系造成评价信息相互重叠和干扰，从而难以客观地反映被评价对象的相对地位．作者采用主成份分析有效地减少了评价指标的数目，消除了指标间的相关性影响．

一般而言，不同评价指标往往具有不同的量纲和量纲单位，为了消除由此带来的不可公度性，决策之前首先应将各评价指标作正向化和无量纲化处理．

通过对指标进行正向化处理，将各评价指标趋势取向．

若指标 G_j 是以大为好的正向指标，令

$$y_{ij}=x_{ij}-\min x_j \tag{1}$$

若指标 G_j 是以小为好的逆向指标，令

$$y_{ij}=\max x_j-x_{ij} \tag{2}$$

若指标 G_j 是以离标准值 x_{0j} 差的绝对值越小越好的适度指标，令

$$y_{ij}=\max_{1\leqslant l\leqslant n}|x_{lj}-x_{0j}|-|x_{ij}-x_{0j}| \tag{3}$$

若指标 G_j 是以落在稳定区间$[q_{1j},q_{2j}]$为好的区间指标，令

$$y_{ij}=\begin{cases}x_{ij}-\min\{x_{lj}\mid x_{lj}<q_{lj}\}，当\ x_{lj}<q_{lj}\ 时；\\ \max\{\max\{x_{lj}\mid x_{lj}>q_{2j}\}-q_{2j}，\\ \quad q_{1j}-\min\{x_{lj}\mid x_{lj}<q_{lj}\}\}，\\ \quad 当\ x_{lj}\in[q_{1j},q_{2j}]时；\\ \max\{x_{lj}\mid x_{lj}>q_{2j}\}-x_{ij}，\\ \quad 当\ x_{ij}>q_{2j}时．\end{cases} \tag{4}$$

式(1)、(2)中 $\max x_j=\max\limits_{1\leqslant l\leqslant n}\{x_{lj}\}$，$\min x_j=\min\limits_{1\leqslant l\leqslant n}\{x_{lj}\}$，由于数据平移不影响指标自身的方差和指标间的相关性，因此指标正向化公式(1)～(4)在本质上没有影响指标自身的方差和指标间的相关情况[3]．

原始数据中包含的信息由两部分组成：一部分是每一指标体现在各决策方案变异程度上的变异信息，这由每一指标的方差大小反映；另一部分是各指标间相互影响程度上的相关信息，这由指标间的相关系数体现．

为了在无量纲化处理中保留以上两部分信息，采用均值(无量纲)化方法．

指标 G_j 的均值为：

$$\bar{y}_j=\frac{1}{n}\sum_{l=1}^{n}y_{lj}\quad(j=1,2,\cdots,m) \tag{5}$$

均值化处理后数据阵 $Z=(z_{ij})$的元素

$$z_{ij}=\frac{y_{ij}-\bar{y}_j}{\bar{y}_j}=\frac{y_{ij}}{\bar{y}_j}-1 \tag{6}$$

均值化后各指标的均值均为 1，数据的协方差矩阵 $V=\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}\triangleq(v_{ij})$，其中

$$v_{ij}=\sum_{l=1}^{n}z_{li}z_{lj}=\sum_{l=1}^{n}\frac{(y_{li}-\bar{y}_{li})(y_{lj}-\bar{y}_j)}{\bar{y}_i\bar{y}_j}=\frac{s_{ij}}{\bar{y}_i\bar{y}_j} \tag{7}$$

式中 $s_{ij}=\sum\limits_{l=1}^{n}(y_{li}-\bar{y}_{li})(y_{lj}-\bar{y}_j)$为未均值化前数据的协方差．

当 $i=j$ 时，$s_{ii}=\sum\limits_{l=1}^{n}(y_{li}-\bar{y}_i)^2$

因此

$$v_{ii}=\frac{s_{ii}}{\bar{y}_i^{\,2}}=\left(\frac{\sqrt{s_{ii}}}{\bar{y}_i}\right)^2 \tag{8}$$

即均值化后数据的协方差矩阵的对角元素是各指标变异系数$\sqrt{s_{ii}}/\bar{y}_i$ 的平方，它保留了各指标变异程度的信息．

未均值化前，原始指标的相互影响关系由相关系数 v_{ij}来反映，其计算公式是

$$r_{ij}=\frac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}}$$

而均值化后的相关系数 r'_{ij}应按如下公式计算：

$$r'_{ij}=\frac{v_{ij}}{\sqrt{v_{ii}}\sqrt{v_{jj}}}$$

将式(7)，(8)代入上式可知

$$r'_{ij}=\frac{s_{ij}}{\bar{y}_i\cdot\bar{y}_j}\Big/\left(\frac{\sqrt{s_{ii}}}{\bar{y}_i}\cdot\frac{\sqrt{s_{jj}}}{\bar{y}_j}\right)=\frac{s_{ij}}{\sqrt{s_{ii}}\cdot\sqrt{s_{jj}}}=r_{ij} \tag{9}$$

也就是说，均值化处理并不改变指标间的相关系数．由式(8)与式(9)可以看出，均值化后的协方差矩阵不仅消除了指标量纲和量纲单位的影响，而且能全面反映原始数据所包含的两部分信息[4]．

构造理想决策方案

$\boldsymbol{A}^*=(y_1^*,y_2^*,\cdots,y_m^*)$

若指标 G_j 是正向指标或逆向指标时，令

$$y_j^*=\max_{1\leqslant i\leqslant n}\{x_{ij}\}-\min_{1\leqslant i\leqslant n}\{x_{ij}\}$$

若指标 G_j 是离标准值 x_{0j} 差的绝对值越小越好的适度指标，令

$$y_j^*=\max_{1\leqslant i\leqslant n}|x_{ij}-x_{0j}|$$

若指标 G_j 是以落在稳定区间$[q_{1j},q_{2j}]$为好的区间指标，令

$$y_j^*=\max\{\max\{x_{lj}\mid x_{lj}>q_{2j}\}-q_{2j}，$$

$q_{1j}-\min\{x_{lj}\mid x_{lj}<q_{1j}\}\}$.

均值化后理想决策方案

$$A^*=\left(\frac{y_1^*}{\bar{y}_1}-1,\frac{y_2^*}{\bar{y}_2}-1,\cdots,\frac{y_m^*}{\bar{y}_m}-1\right)\triangleq(z_1^*,z_2^*,\cdots,z_m^*)$$

易见　$z_{ij}\leqslant z_j^*$　$(i=1,2,\cdots,n;j=1,2,\cdots,m)$.　(10)

指标赋权是多目标决策和综合评价的重要步骤,通过对指标赋权确定各指标在综合评价中的主次地位.赋权有主观赋权法和客观赋权法两大类.客观赋权法通过提取指标变异信息量和指标间的相关信息量确定权重,突出了数据的数学含义,譬如本文后面提到的主成份分析法.而主观赋权法直接通过指标本身的经济含义来赋权,譬如专家主观赋权法.两者均具有各自不可取代的优越性.作者采用二次混合赋权,即先主观赋权,再采用主成份分析法客观赋权.

设通过对指标本身的含义评价已有指标的主观加权向量 $\vec{d}=(d_1,d_2,\cdots,d_m)$ $(d_j>0)$.若不需要采取主观赋权,可取 $\vec{d}=(1,1,\cdots,1)$,此时主观评价各原始指标地位平等,则本文的二次混合赋权转化为单纯的主成份分析法客观赋权.

在主观加权向量 $\vec{d}$ 的作用下,主观加权的决策数据阵 $C=(cij)=$

$$\begin{array}{c} \\ A_1\\ A_2\\ \\ A_n\end{array}\begin{array}{c}\begin{array}{cccc}G_1 & G_2 & \cdots & G_m\end{array}\\ \left[\begin{array}{cccc} d_1z_{11} & d_2z_{12} & \cdots & d_mz_{1m}\\ d_1z_{21} & d_2z_{22} & \cdots & d_mz_{2m}\\ & \cdots\cdots\cdots & & \\ d_1z_{n1} & d_2z_{n2} & \cdots & d_mz_{nm}\end{array}\right]\end{array}$$

对应的协方差阵

$\bar{U}=C^TC\triangleq(u_{ij})$

其中

$$u_{ij}=\sum_{l=1}^{n}d_id_jz_{li}z_{lj}\quad 1\leqslant i,j\leqslant m$$

计算 m 阶协方差阵 $\bar{U}$ 的 m 个特征值 $\lambda_1\geqslant\lambda_2\geqslant\lambda_3\geqslant\cdots\geqslant\lambda_m\geqslant0$ 及其相应的单位特征向量 $\vec{p}_1,\vec{p}_2,\cdots,\vec{p}_m$,得正交变换矩阵

$$P=[\vec{p}_1,\vec{p}_2\cdots\vec{p}_m]\tag{11}$$

对主观加权指标向量组 $C=(c_{lj})=[d_1\vec{z}_1,d_2\vec{z}_2,\cdots,d_m\vec{z}_m]$ 作正交旋转变换,得新的综合指标向量组为

$$\bar{W}=CP\triangleq[\vec{w}_1\vec{w}_2\cdots\vec{w}_m]\triangleq(w_{ij})_{m\times m}\tag{12}$$

第 j 个综合指标 $\vec{W}_j=\sum_{l=1}^{m}p_{lj}d_j\vec{z}_j$ 是原始指标向量组 $\{\vec{z}_1,\vec{z}_2,\cdots,\vec{z}_m\}$ 的线性组合.对应 $\vec{w}_j=\sum_{l=1}^{m}p_{lj}d_j\vec{z}_j$ 称为第 j 个主成份 $(j=1,2,\cdots,m)$.而第 i 个决策方案(或被评价对象) A_i 的 m 个综合指标,其中 $\vec{w}_{ij}=\sum_{l=1}^{m}d_lp_{lj}\vec{z}_{il}$ 评价值为 $(w_{i1},w_{i2},\cdots,w_{im})$.

正交变换 $\bar{W}=CP$ 满足

(1)

$$\bar{W}^T\bar{W}\triangleq(\vec{w}_i^T\vec{w}_j)_{m\times m}=(CP)^T(CP)=P^T(C^TC)P=P^T\bar{U}P=\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m\end{bmatrix}\tag{13}$$

式(13)说明两点:新的综合指标向量 $\{\vec{w}_j\}$ 间彼此正交,避免了信息的重叠和相互干扰.其次由 $\|\vec{w}_j\|^2=\vec{w}_j^T\vec{w}_j=\lambda_j$,表明了特征值 λ_j 正是综合指标(主成份) w_j 所含的信息量,且各主成份所含信息量已经排序:$\|\vec{w}_1\|^2\geqslant\|\vec{w}_2\|^2\geqslant\cdots\geqslant\|\vec{w}_m\|^2$.

(2)对于第 i 个决策方案(被评价对象)在原始指标组和新的综合指标组下的两组评价值 $(z_{i1},z_{i2},\cdots,z_{im})$ 和 $(w_{i1},w_{i2},\cdots,w_{im})$ 成立:

$$\begin{aligned}&\|(w_{i1},w_{i2},\cdots,w_{im})\|^2=\\&\|(d_1z_{i1},d_2z_{i2},\cdots,d_mz_{im})P\|^2=\\&(d_1z_{i1},d_2z_{i2},\cdots,d_mz_{im})P^TP\\&(d_1z_{i1},d_2z_{i2},\cdots,d_mz_{im})^T=\\&(d_1z_{i1},d_2z_{i2},\cdots,d_mz_{im})\\&(d_1z_{i1},d_2z_{i2},\cdots,d_mz_{im})^T=\\&\|(d_1z_{i1},d_2z_{i2},\cdots,d_mz_{im})\|^2\end{aligned}$$

即决策方案(被评价对象) A_i 经一次主观赋权后的信息量在正交变换 $\bar{W}=CP$ 下保持不变,进而总信息量亦不变.

理想决策 A^* 经一次主观赋权后在正交变换 $\bar{W}=CP$ 下的评价值为

$$A^*=(d_1z_1^*,d_2z_2^*,\cdots,d_mz_m^*)P\triangleq(w_1^*,w_2^*,\cdots,w_m^*)$$

利用排序后的综合指标(主成份)所含信息量的大小,可进行二次客观赋权,并且可以进行降维处理.

各主成分的重要程度可以由它们所含的信息量的大小确定,第 j 个综合指标的客观赋权为

$$\beta_j=\frac{\lambda_j}{\sum_{l=1}^{m}\lambda_l}\quad(j=1,2,\cdots,m)$$

亦可以先进行降维处理再客观赋权,按累积信

息量的贡献率不低于某个阈值(譬如85%)的原则保留前 k 个主成份,即 $\frac{\lambda_1+\lambda_2+\cdots+\lambda_k}{\lambda_1+\lambda_2+\cdots+\lambda_m}\geqslant 85\%$ 来确定 k,此时客观赋权为:

$$\beta_j=\frac{\lambda_j}{\sum_{l=1}^{k}\lambda_l}\quad(j=1,2,\cdots,m)\tag{14}$$

经二次混合赋权后且保留前 k 个主成份的第 j 个决策方案(被评价对象)的评价值为

$$A_i=(\beta_1 w_{i1},\beta_2 w_{i2},\cdots,\beta_k w_{ik})=\left(\sum_{l=1}^{m}\beta_1 d_l p_{l1} z_{i1},\sum_{l=1}^{m}\beta_2 d_l p_{l2} z_{i1},\cdots,\sum_{l=1}^{m}\beta_k d_l p_{lk} z_{i1}\right)\quad(k\leqslant m)\tag{15}$$

由式(13)、(15)可见,二次混合赋权既重视了原始指标的经济含义,又经正交旋转消除了指标间的相关性,兼顾了综合指标数学含义上的信息量.

将二次赋权的(降维)理想决策方案单位化:

$$(A^*)^0=\frac{1}{\sqrt{\sum_{l=1}^{k}(\beta_l w_l^*)^2}}(\beta_1 w_1^*,\beta_2 w_2^*,\cdots,\beta_k w_k^*)\quad(k\leqslant m)$$

求经过二次混合赋权后每一决策方案 A_i 在理想决策方案 A^* 上的投影:

$$D_i=A_i\cdot(A^*)^0=\frac{1}{\sqrt{\sum_{l=1}^{k}(\beta_l w_l^*)^2}}\sum_{l=1}^{k}\beta_l^2 w_{il} w_l^*\quad(i=1,2,\cdots,n)\tag{16}$$

其几何含意为决策方案 A_i 的模与该方案和理想方案 A^* 之间夹角的余弦的乘积,即方案 A_i 的综合评价值

$$D_i=\|A_i\|\cos(\widehat{A_i,A^*})\tag{17}$$

式(17)表示了方案 A_i 的理想化程度.

依据 D_i 的大小来确定各决策方案 A_i 的优劣顺序.

由式(10)知 $\|A_i\|\leqslant\|A^*\|\quad(i=1,2,\cdots,n)$

再由式(17)知,任何方案 A_i 的综合评价值 D_i 和理想方案 A^* 的综合评价值 D^* 相比,均有

$$D_i\leqslant D^*\quad(i=1,2,\cdots,n)$$

2 应用实例

作者以《中国工业经济统计年鉴》1993年提供的全国16个省、直辖市主要工业经济效益指标的统计资料[5]为基础数据,进行经济效益的比较和排序分析.很显然,此类问题是典型的多指标综合评价问题,已知被评价对象集为 $A=$ {北京,天津,上海,江苏,…,山西},共有16个被评价对象,指标集 $G=\{G_j\mid j=1,2,\cdots,f\}$,其中 G_1 为全员劳动生产率(元/人);G_2 为资金利税率(%);G_3 为百元销售收入实现利润(元);G_4 为百元工业产值占用流动资金(元);G_5 为产值利税率(%),共5个评价指标.根据实际意义,除 G_4 为逆向指标外,其余均为正向指标.具体工业经济效益指标值见表1.

表1 1992年全国部分省、直辖市主要工业经济效益指标

Tab.1 The industrial economic efficiency indexes of several provinces and major cities in 1992

省(直辖市)		全员劳动生产率 G_1/(元/人)	资金利税率 G_2/%	百元销售收入实现利润 G_3/元	百元工业产值占用流动资金 G_4/元	产值利税率 G_5/%
A_1	北京	47 177	16.61	8.89	31.05	15.77
A_2	天津	43 323	9.08	3.65	29.80	8.44
A_3	上海	59 023	13.84	6.06	26.55	12.87
A_4	江苏	46 821	10.59	3.51	22.46	7.41
A_5	浙江	41 646	13.24	4.64	24.33	9.33
A_6	安徽	26 446	10.16	2.38	26.80	9.85
A_7	福建	38 381	11.97	4.79	26.45	10.64
A_8	广东	57 808	10.29	4.54	23.00	9.23
A_9	辽宁	28 869	7.68	2.12	31.08	9.05
A_{10}	山东	38 812	8.92	3.38	25.68	8.73
A_{11}	湖北	30 721	10.87	4.15	30.36	11.44
A_{12}	湖南	24 848	10.77	2.42	30.71	11.37
A_{13}	河南	26 925	9.34	3.06	30.11	10.84
A_{14}	江西	23 269	8.25	2.58	32.57	8.62
A_{15}	河北	28 267	8.13	3.17	29.25	9.17
A_{16}	山西	21 583	7.41	4.66	35.35	11.27

本例一次主观赋权以权重向量为 $\vec{d}=(1,1,1,1,1)$,其最终结果实际上为单纯采用主成份分析法客观赋权的结果.

利用公式(1)、(2),经正向化后的决策数据阵为:

$$\overline{\boldsymbol{Y}}=\begin{array}{cc} & \begin{array}{ccccc} G_1 & G_2 & G_3 & G_4 & G_5 \end{array} \\ \begin{array}{c} A_1\\ A_2\\ A_3\\ A_4\\ A_5\\ A_6\\ A_7\\ A_8\\ A_9\\ A_{10}\\ A_{11}\\ A_{12}\\ A_{13}\\ A_{14}\\ A_{15}\\ A_{16} \end{array} & \left[\begin{array}{ccccc} 21594 & 9.2 & 6.7 & 4.3 & 8.36 \\ 21740 & 1.67 & 1.53 & 5.55 & 1.03 \\ 37450 & 6.43 & 3.94 & 8.8 & 5.46 \\ 25238 & 3.18 & 1.39 & 12.89 & 0 \\ 20063 & 5.83 & 2.52 & 11.02 & 1.92 \\ 4863 & 2.75 & 0.26 & 8.55 & 2.44 \\ 16798 & 4.56 & 2.67 & 8.9 & 3.23 \\ 36225 & 2.88 & 2.42 & 12.35 & 1.82 \\ 7286 & 0.27 & 0 & 4.27 & 1.64 \\ 17229 & 1.51 & 1.26 & 9.67 & 1.32 \\ 9138 & 3.46 & 2.03 & 4.99 & 4.03 \\ 3265 & 3.36 & 0.3 & 4.64 & 3.96 \\ 5342 & 1.93 & 0.94 & 5.24 & 3.43 \\ 1686 & 0.84 & 0.46 & 2.78 & 1.21 \\ 6684 & 0.72 & 1.05 & 6.1 & 1.76 \\ 0 & 0 & 2.54 & 0 & 3.86 \end{array}\right] \end{array}$$

利用公式(5)、(6),均值化后的决策数据阵为:

$$\boldsymbol{Z}=\begin{array}{cc} & \begin{array}{ccccc} G_1 & G_2 & G_3 & G_4 & G_5 \end{array} \\ \begin{array}{c} A_1\\ A_2\\ A_3\\ A_4\\ A_5\\ A_6\\ A_7\\ A_8\\ A_9\\ A_{10}\\ A_{11}\\ A_{12}\\ A_{13}\\ A_{14}\\ A_{15}\\ A_{16} \end{array} & \left[\begin{array}{ccccc} 0.716232 & 2.02943 & 2.60106 & -0.37483 & 1.94172 \\ 0.457793 & -0.450092 & -0.18617 & -0.193095 & -0.637563 \\ 1.51165 & 1.11731 & 1.09574 & 0.279418 & 0.92127 \\ 0.69236 & 0.0471308 & -0.260638 & 0.874056 & -1 \\ 0.345345 & 0.91974 & 0.340426 & 0.60218 & -0.324389 \\ -0.673906 & -0.0944624 & -0.861702 & 0.24307 & -0.14141 \\ 0.126407 & 0.501546 & 0.420213 & 0.293956 & 0.136576 \\ 1.42911 & -0.0516552 & 0.287234 & 0.795546 & -0.359577 \\ -0.51143 & -0.911093 & -1 & -0.379192 & -0.422915 \\ 0.155308 & -0.502778 & -0.329787 & 0.405905 & -0.535517 \\ -0.387242 & 0.139331 & 0.0797872 & -0.274512 & 0.41808 \\ -0.781062 & 0.106402 & -0.840426 & -0.325398 & 0.393449 \\ -0.641787 & -0.36447 & -0.5 & -0.238165 & 0.206952 \\ -0.886944 & -0.723399 & -0.755319 & -0.59582 & -0.574224 \\ -0.551797 & -0.762914 & -0.441489 & -0.113131 & -0.380689 \\ -1 & -1 & 0.351064 & -1 & 0.358261 \end{array}\right] \end{array}$$

由公式(7),相应的协方差阵 $\boldsymbol{V}=\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}$ 为:

$$\boldsymbol{V}=\begin{bmatrix} 9.6675 & 5.8873 & 6.3846 & 4.2501 & 1.1780 \\ 5.8873 & 10.0326 & 8.8779 & 1.9895 & 6.0386 \\ 6.3846 & 8.8779 & 12.1464 & 0.2518 & 7.3312 \\ 4.2501 & 1.9895 & 0.2518 & 4.0751 & -2.0191 \\ 1.1780 & 6.0386 & 7.3312 & -2.0191 & 7.7397 \end{bmatrix}$$

$\boldsymbol{V}$ 的特征值由大到小依次为:$\lambda_1=28.9659$,$\lambda_2=10.7395$,$\lambda_3=2.59531$,$\lambda_4=0.943181$,$\lambda_5=0.417399$.由对应的单位特征向量构成的正交阵,即公式(11)中的正交阵.

$$\boldsymbol{P}=\begin{bmatrix} 0.4129 & 0.6187 & -0.3720 & 0.4999 & -0.2419 \\ 0.5472 & 0.0065 & 0.6956 & -0.1549 & -0.4390 \\ 0.6134 & -0.1709 & -0.4877 & -0.5665 & 0.1891 \\ 0.0895 & 0.5515 & 0.3591 & -0.1331 & 0.7357 \\ 0.3819 & -0.5328 & 0.1047 & 0.6225 & 0.4145 \end{bmatrix} \tag{18}$$

对决策方案数据阵 Z 作正交变换 $\overline{W}=ZP$ 得公式(12)中的在新的综合指标 $\{w_j\}$ 下的决策数据阵

$$\boldsymbol{W}=\begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \\ A_6 \\ A_7 \\ A_8 \\ A_9 \\ A_{10} \\ A_{11} \\ A_{12} \\ A_{13} \\ A_{14} \\ A_{15} \\ A_{16} \end{array}\begin{array}{c} \begin{array}{ccccc} W_1 & W_2 & W_3 & W_4 & W_5 \end{array} \\ \begin{bmatrix} 3.6097 & -1.2295 & -0.0547 & -0.1712 & -0.0433 \\ -0.4322 & 0.5453 & -0.5287 & 0.0328 & -0.3547 \\ 2.2845 & 0.4184 & -0.1228 & 0.4981 & -0.0616 \\ -0.1520 & 1.4880 & 0.1115 & -0.2524 & -0.0089 \\ 0.7847 & 0.6664 & 0.5275 & -0.4448 & -0.1144 \\ -0.8907 & -0.0609 & 0.6777 & 0.0455 & 0.1617 \\ 0.6628 & 0.0990 & 0.2168 & -0.2067 & 0.1016 \\ 0.6718 & 1.4650 & -0.4596 & 0.2299 & 0.1675 \\ -1.5185 & -0.1352 & -0.1362 & 0.2392 & -0.1197 \\ -0.5815 & 0.6583 & -0.1569 & -0.0450 & 0.1974 \\ 0.1004 & -0.6264 & 0.1472 & 0.0364 & 0.0189 \\ -0.6586 & -0.7280 & 0.6988 & 0.3574 & -0.0930 \\ -0.7134 & -0.5556 & 0.1652 & 0.1794 & 0.1313 \\ -1.4980 & -0.4470 & -0.0790 & -0.1816 & -0.2870 \\ -1.0716 & -0.1304 & -0.1906 & -0.1295 & 0.1439 \\ -0.6974 & -1.4275 & -0.8164 & -0.1877 & 0.1602 \end{bmatrix} \end{array} \tag{19}$$

均值化后的理想决策方案为

$A^*=(Z_1^*,Z_2^*,\cdots,Z_m^*)=(1.5117,2.0294,2.6011,0.8741,1.9417)$

经正交变换得理想决策方案 $A^*=Z^*P$ 为

$A^*=(W_1^*,W_2^*,W_3^*,W_4^*,W_5^*)=(4.1499,-0.0487,0.0979,0.0602,0.6831)$

本例在公式(14)中,只要取 $k=2$,即保留第 1 主成份 W_1、第 2 主成份 W_2,信息的累计提取率已达

$$\frac{\lambda_1+\lambda_2}{\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5}=90.94\%$$

利用式(12)和(18),可得入选主成份 W 和均值化后的原始指标 Z 的关系为:

$W_1=0.4129z_1+0.5472z_2+0.6134z_3+0.0895z_4+0.3819z_5$

$W_2=0.6187z_1+0.0065z_2-0.1709z_3+0.5515z_4-0.5328z_5$

利用公式(14)得二次客观赋权入选主成份相应的权重为

$$\beta_1=\frac{\lambda_1}{\lambda_1+\lambda_2}=0.7295\quad \beta_2=\frac{\lambda_2}{\lambda_1+\lambda_2}=0.2705$$

最终由公式(16)得综合评价值 $D^{(1)}$ 的计算式:

$$D^{(1)}=A_i\cdot(A^*)^{\circ}=\frac{1}{\sqrt{(\beta_1W_1^*)^2+(\beta_2W_2^*)^2}}(\beta_1^2w_{i1}\,w_1^*+\beta_2^2w_{i2}\,w_2^*)=0.7295w_{i1}-0.00177w_{i2}\quad(i=1,2,\cdots,16) \tag{20}$$

将式(19)中前 2 列的数据代入式(20),得 16 个省、直辖市的综合评价值 $D^{(1)}$.

类似,可保留前 3 个主成份 W_1、W_2、W_3,其累计信息提取率高达

$$\frac{\lambda_1+\lambda_2}{\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5}=96.88\%$$

相应综合评价值 $D^{(2)}$ 的计算式:

$$D_i^{(2)}=\frac{1}{\sqrt{\sum_{l=1}^{3}(\beta_l W_l{}^*)^2}}\sum_{n=1}^{3}\beta_k{}^2 w_{il}w_k{}^*=0.6848w_{i1}-0.001005w_{i2}+0.000129w_{i3}\quad(i=1,2,\cdots,16)\qquad(21)$$

将式(19)中的前 3 列数据代入式(21),可得 16 个省、直辖市按前 3 个主成份得到的综合评价值 $D^{(2)}$. 比较 $D_i^{(1)}$,$D_i^{(2)}$,($i=1,2,\cdots,16$)各自的排序,发现两者的序完全一致,理由很简单,第 3 主成份 W_3 的作用相对于第 1、2 主成份已很次要,这由其占总信息量的比例可以看出. 本例可以肯定地推断,即使保留全部主成份得到的综合评价值也不能改变仅保留 2 个主成份产生的排序. 16 个省、直辖市的综合评价值 $D^{(1)}$、$D^{(2)}$ 及排序号见表 2.

由表 2 中的评价结果可以看出,北京作为我国的政治、经济和文化中心,上海作为我国的最大工业城市,其综合工业经济效益水平实力不凡,分别居于第 1 位、第 2 位. 而排名第 3 位到第 9 位的基本上是沿海率先实现改革开放的省、市,体现了经济发达、技术先进、管理水平高、经济效益较好的区域优势. 江西作为革命老区经济基础薄弱、技术落后;辽宁作为重工业基地省份,设备陈旧老化、资金匮乏、技改措施跟不上等众多主客观因素影响,这两个省份分列最末一、二位. 以上结论是单就 1992 年度的资料分析而言,在局部排序上带有一定的随机性,但总的来讲,本文的评价结论与人们的直观判断和习惯认识基本一致,有一定的可信度和决策参考价值. 特别在具体操作中,若能由有关经济专家对各项经济指标确定合适的一次客观赋权,则本文的二次混合赋权的分析结论定能更贴近实际情况.

表 2　1992 年全国部分省、直辖市主要工业经济效益综合评价值及排序

Tab.2　The complex values of the industrial economic efficiency indexes of several provinces and major cities and their ranks in 1992

省(直辖市)	前两个主成份综合评价值 $D^{(1)}$	前三个主成份综合评价值 $D^{(2)}$	排序号
北京	2.708	2.542	1
天津	−0.316	−0.297	8
上海	1.666	1.564	2
江苏	−0.114	−0.106	7
浙江	0.571	0.537	3
安徽	−0.650	−0.610	13
福建	0.483	0.454	5
广东	0.488	0.458	4
辽宁	−1.108	−1.040	16
山东	−0.425	−0.399	9
湖北	0.074	0.070	6
湖南	−0.479	−0.450	10
河南	−0.519	−0.488	12
江西	−1.092	−1.025	15
河北	−0.782	−0.734	14
山西	−0.506	−0.476	11

参考文献:

[1] 孟生旺. 多指标综合评价中权数的选择[J]. 统计研究,1993,(2):10～15.
[2] 陈述云. 多指标综合评价方法新探[J]. 统计与预测,1993(3):23～26.
[3] 邓勃. 分析测试数据的统计处理方法[M]. 北京:清华大学出版社,1995.
[4] 邱东. 多指标综合评价方法的系统分析[M]. 北京:中国统计出版社,1991.
[5] 国家统计局,工业交通统计局. 中国工业经济统计年鉴[M]. 北京:中国统计出版社,1993.

(责任编辑:李春丽)