

城市工业企业经济效益的综合评判^{*}

冯旭东 陈方

(中国科技大学管理科学系, 安徽合肥 230027)

摘要 采用多元统计分析中的主成分分析方法改进型, 对城市工业企业的效益给予综合评判, 然后又对综合评判系数采用系统聚类分析方法对其作了分类, 用以找出各类城市的优势和差距, 便于针对性地制定提高经济效益的对策。

关键词 评判 经济效益 综合主成分法 聚类分析法

Synthetic Evaluations of Cities' Economic Results Using Improved Statistical Method

Feng Xudong Chen Fang

(Department of Management Science, USTC, Hefei Anhui 230027)

Abstract In this paper, synthetic evaluations of industrial enterprises' economic results of 35 cities is described by using improved principal components analysis method. Moreover, the use of hierarchical clustering analysis as post-processing is also described. The result shows that this improved method is better than traditional statistical method.

Keywords evaluation; economic results; synthetic principal components analysis method; hierarchical clustering analysis

目前, 我国评价工业经济效益用的是多个经济效益指标, 如资金利税率, 净产值率, 全员劳动生产率, 全部流动资金周转次数等等。这些指标虽然可以从不同角度对工业经济效益进行全面考核, 但由于有多个指标, 各指标的变化方向常常相互冲突, 因而给经济效益的综合评价的排序带来很大困难。本文采用综合主成分分析法对全国 35 个城市的全市全部独立核算工业企业的经济效益进行综合排序, 并用得出的各城市综合主成分量进行系统聚类分析。

1 综合主成分法

主成分法是把一组指标线性变换到一组为数不多且互不相关的综合指标上进行分析的方法。本文采用此法在保留原来所有信息的基础上把多个指标简化成一个综合指标, 再对样本进行排序。

1) 数据准备

确定经济效益体系中的代表性指标, 以评价全国 35 个城市的全市全部独立核算工业企业的经济效益为例。根据已有资料, 选取十个指标, 即产品销售率($x_1\%$), 资金利税率($x_2\%$), 净产值率($x_3\%$), 成本利润率($x_4\%$), 全员劳动生产率(x_5 元), 全部流动资金周转次数(x_6 次), 产值利税率($x_7\%$), 产值资金率($x_8\%$), 原值利税率($x_9\%$), 原值产值率($x_{10}\%$)。共抽取 35 个城市。即指标数 $N = 10$, 样本容量 $M = 35$, 原始数据见参考文献[2]。

^{*} 本文于 1997 年 3 月 27 日收到

2) 数据标准化

将一个城市的十个指标的原始数据写成矩阵形式 $X = (x_{ij})_{M \times N}$, 对其进行标准化处理得 $Y = (y_{ij})_{M \times N}$, 其中

$$y_{ij} = \frac{(x_{ij} - \bar{x}_j)}{\sqrt{s_{jj}}} \quad i = 1, 2, \dots, M; \quad j = 1, 2, \dots, N$$

$$\bar{x}_j = \frac{1}{M} \sum_{i=1}^M x_{ij}$$

$$s_{jj} = \frac{1}{(M-1)} \sum_{i=1}^M (x_{ij} - \bar{x}_j)^2$$

3) 求出相关矩阵 R 的极大似然无偏估计 $\hat{R} = (r_{ij})_{M \times N}$

y_i 与 y_j 相关系数 r_{ij} 的极大似然无偏估计为

$$\hat{r}_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}} = \frac{1}{(M-1)} \sum_{k=1}^M y_{ki}y_{kj}$$

相关矩阵 R 的极大似然无偏估计为

$$\hat{R} = (\hat{r}_{ij})_{N \times N} = \frac{1}{M-1} \sum_{k=1}^M (y_{ki}y_{kj})_{N \times N}$$

代入数据求得 \hat{R}

4) 求 R 阵的特征根与特征向量

由 R 的特征方程

$$|\hat{R} - \lambda I| = 0$$

求得 10 个非负特征根 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{10} \geq 0$ 为:

$$4.481 \quad 2.758 \quad 0.994 \quad 0.706 \quad 0.524 \quad 0.268 \quad 0.148 \quad 0.090 \quad 0.0190 \quad 0.012$$

而 10 个累积贡献率为:

$$0.448 \quad 0.724 \quad 0.823 \quad 0.894 \quad 0.946 \quad 0.973 \quad 0.988 \quad 0.997 \quad 0.999 \quad 1.000$$

以及 10 个特征向量, 为:

$$Z_1 = (0.098 \quad -0.413 \quad -0.085 \quad -0.334 \quad -0.383 \quad -0.382 \quad -0.304 \quad 0.360 \quad -0.383 \\ -0.201) \text{ 及 } Z_2, Z_3, \dots, Z_{10} \text{ (具体数值略)}.$$

5) 求综合主成分量

传统的主成分分析法综合评价经济效益时, 通常为计算上的方便取第一主成分分析。虽然第一主成分综合信息的能力强, 但它毕竟不能反映原有指标所有包含的信息。特别是当第一主成分的方差贡献率小于 0.75 时, 丢失的有用信息多, 以致评判结果受到变换导致得分数值的微小差异的影响, 难以精确判断。为了比较全面地反映问题, 我们概括原有所有的信息, 进行全面综合的评价。

考虑所有的主成分 Z_1, Z_2, \dots, Z_{10} , 把它们各自的贡献率 CR_i , 看作它们包含所有信息的多少即权数。这样就可得到一个综合主成分。

$$\begin{aligned} Z &= \sum_i CR_i z_i \\ &= \sum_i CR_i \sum_j L_{ij} Y_j \\ &= \sum_j \left(\sum_i CR_i L_{ij} \right) Y_j \\ &= \sum_j L^* Y_j \\ L^* &= \sum_i CR_i L_{ij} \end{aligned}$$

代入数据, 算出综合特征矢量

$$L^* = (l_j^*)_N$$

$$= (0.117, -0.096, 0.093, -0.106, -0.166, -0.140, -0.033, 0.242, -0.259, -0.282)$$

将样品在各个指标变量上的标准化取值 y_{ij} 乘以相应主成分载荷, 得主成分值

$$z_i = \sum_j L^* Y_j$$

按这种方法的具体步骤可计算得到综合评价系数, 并且此值对各城市进行排序, 其中评价系数大, 名次在后, 得如表 1 结果:

表 1

城市名	石家庄	唐山	太原	呼和浩特	包头	沈阳	大连
评价系数	3.96	4.23	4.63	3.50	4.30	4.56	3.98
名次	21	26	30	15	27	29	22
城市名	鞍山	抚顺	长春	吉林	哈尔滨	佳木斯	南京
评价系数	3.78	3.99	3.65	4.55	5.48	5.67	2.90
名次	19	23	17	28	33	34	8
城市名	无锡	苏州	杭州	合肥	福州	南昌	济南
评价系数	1.75	1.24	2.00	3.14	2.94	2.66	3.23
名次	2	1	3	11	9	5	12
城市名	淄博	郑州	武汉	长沙	南宁	成都	重庆
评价系数	2.87	3.38	3.77	3.09	2.67	3.53	4.20
名次	7	14	18	10	6	16	25
城市名	贵阳	昆明	西安	兰州	西宁	银川	乌鲁木齐
评价系数	3.29	2.14	5.03	3.88	5.82	4.63	4.10
名次	13	4	32	20	35	31	24

6) 排序结果分析

由表 1 中结果可以看出, 苏州的独立核算工业企业的综合经济效益最好, 其相应的综合评价系数最小, 为 1.2, 无锡次之, 杭州第三, 而西宁的独立核算工业企业的综合经济效益最差, 其相应的综合评价系数最大, 为 5.8, 其次是佳木斯。综合评价系数最大与最小之间相差 4.6, 较大而不是十分悬殊, 这表明我国各城市的独立核算工业企业的综合经济效益分布不平衡但没有到极点。

从相邻位次城市的综合评价系数的差距看, 最大差距为昆明与南昌之间, 也不超过 0.52 (如昆明和南昌之间)。因此我国的城市发展虽呈区域性的特点, 但其城市间的综合经济效益排序不是一成不变的, 而是具有可变性和竞争性。一个城市只要稍微提高对综合经济效益的重视程度, 着力解决一些现存问题, 其效益位次便可迅速上升。

2 聚类分析法

聚类分析就是要用某种准则将多维空间上的靠近的点集归为一类而不同类间的距离尽可能大。本文采用此法对上述综合得分进行最优分割, 把样本按综合经济效益水平高低分成四类。再通过求各类中各指标的平均值来找出各类城市的优势和差距, 以便各城市制定提高经济效益的针对性对策。

1) 聚类原理

选用聚类分析中的 Q 型系统聚类法。

采用欧氏距离:

$$d_{ij} = |z_i - z_j|$$

表示样品 i 和样品 j 之间的距离。

并采用系统聚类法中常使用的由 Sokal 和 Michener 提出的类平均距离公式:

$$D_{pq}^2 = \frac{1}{n_p n_q} \sum_{i \in G_p, j \in G_q} d_{ij}^2$$

其中 n_p 与 n_q 为类 G_p 与 G_q 中元素的个数, d_{ij} 为 G_p 中元素 z_i 与 G_q 中元素 z_j 间的欧氏距离。

2) 数据准备

把上述所得的综合评判系数进行消除量纲的标准化处理, 记为 z_1, z_2, \dots, z_{35} 。

3) 聚类结果

对 z 按上述原理进行聚类, 可得聚类谱系图(略)。对该图进行分析, 我们可将样本分成四类, 各类所包含的样品如下:

I. 苏州 无锡 杭州 昆明

II. 南昌 南宁 淄博 南京 福州 长沙 合肥 济南 贵阳 郑州 呼和浩特 成都 长春 武汉 鞍山

III. 兰州 石家庄 大连 抚顺 乌鲁木齐 重庆 唐山 包头 吉林 沈阳 太原 银川 西安

IV. 哈尔滨 佳木斯 西宁

4) 分类结果分析

上述分类中, 第一类城市的综合经济效益水平高, 其综合评价系数小于 2.2; 第二类城市的综合经济效益水平较高, 其综合评价系数介于 2.6 和 3.8, 第三类城市的综合经济效益水平一般, 其综合评价系数介于 3.8 和 5.1, 第四类城市的综合经济效益水平低下, 其综合评价系数大于 5.4。算出这四类城市相应于各指标的平均值如表 2:

表 2

城市类	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
第一类	95.3	25.1	25.3	4.6	13377	2.04	11.5	69.3	25.7	267.0
第二类	95.6	21.9	27.7	5.1	9859	1.65	12.6	99.0	20.3	166.7
第三类	97.1	16.8	27.4	4.0	8737	1.53	10.5	119.8	11.8	115.2
第四类	98.8	6.6	26.3	-0.37	7062	0.94	6.7	158.4	7.4	114.8

对上述分析可以得出以下结果:

1) 从第一类到第四类, 递增的效益指标有 x_1 和 x_8 。

其中 x_1 为产品销售率(%), 它对综合经济效益理应是正的贡献, 这说明综合经济效益好的城市反而没有做好销售工作, 但从它们的差距来看, 只相差几个百分点, 因此它对整个排序没有多大影响。

而 x_8 为产值资金率(%), 它对综合经济效益是负的贡献, 即综合经济效益好的其值应小, 而上面数据恰恰反映了这一点。

2) 第一类城市对各指标的平均值与第二类比较, 可以看出两者各有千秋, 但综合来看, 第一类优于第二类。

第一类城市对第二、五、六、八、九、十指标的平均值都明显优于第二类城市相应的平均值, 而前者对第一、三、四、七指标的平均值都稍逊于后者, 因此从总体来看, 前者综合经济效益较高于后者。

3) 第一类城市对第三个指标值的平均值是所有四类中最差的, 这说明它的综合经济效益水平虽是最高的, 但它的平均净产值率却不行。对其所含的四个城市在第三个指标即净产值率(%)上进行分析, 发现除昆明排在三十五个城市中第一外, 其它三个排在三十名后, 如苏州排在最后一, 无锡排在倒数第二, 杭州排在第三十名。

4) 总体来看, 除对第一个指标完全异常及第三、第四和第七对第一类城市异常外, 其它各指标都呈现出从第一类城市向第四类城市递减的态势。

从上表可以看出,第一类城市的资金和产值利税率比第四类城市分别高出 18.5 和 4.8 个百分点,成本利润率高出 4.9 个百分点,全员劳动生产率多 4315 元,全部流动资金周转次数多 1.1 次,百元产值流动资金占用率少 89.1 个百分点,百元固定资产原值实现的利税和产值分别高出 18.3 和 152.2 个百分点。

这说明我们用综合主成分来评判是有效的,而且用系统聚类法分割出的四类城市间的工业经济效益差异是明显的,我们可以按上述几点分析,各城市针对性地制定相应政策,发挥原有优势,弥补不足,以提高综合经济效益。

3 对应用方法的思考

我们用了多元统计方法的综合主成分分析和系统聚类分析对全国 35 个城市的全市全部独立核算工业企业的经济效益进行综合排序和分类。

本文采用综合评判方法,其最大的优点是考虑了所有成分包含的信息,因而比传统采用累积贡献率 0.85 的主成分法更全面、更准确。当然其信息之间的汇总采用加权求和的经验法缺乏理论基础。

从本文中算出的主成分可以看出,其最终结果的经济意义并不明确,这一方面是源于主成分方法自身的理论局限,如主成分只是对初始指标的线性组合,同时主成分分析对样本量的要求比较苛刻,尤其是指标多时,要求样本量远远大于指标个数才能得到有意义的结果;另一方面,经济效益的模糊性也有一定的影响。

尽管如此,我们用系统聚类法对综合主成分量进行最优分割再求各类平均值并进行分析后,可以认为,作为多元统计分析方法的综合主成分法,仍不失为综合评价经济效益的有效方法。

参考文献

- 1 M. 肯德尔. 多元分析. 科学出版社, 1983: 16~ 58
- 2 呼和浩特市、包头市、银川市统计局. 92 年全国部分重点城市经济社会资料简编, 1993: 41~ 42
- 3 孟生旺. 用主成分分析法进行多指标综合评价应注意的问题. 统计研究, 1992(4)
- 4 彭家生, 陈述云, 黄茹. 经济效益评价的灰色多元方法. 统计与决策, 1992(2)

(上接第 43 页)

参考文献

- 1 王光远. 工程软设计理论. 北京: 科学出版社, 1992
- 2 张淑华等. 串联系统的模糊可靠性. 系统工程理论与实践, 1994, 14(9)
- 3 张淑华. 考虑结构失效相关性工程系统的全局优化. 哈尔滨建筑大学博士学位论文, 1993
- 4 王光远, 张淑华. 串联工程系统的全局优化. 哈尔滨建筑大学学报, 1991(4)
- 5 王光远, 张淑华. 考虑结构失效相关时串联工程系统造价可靠度间的关系. 哈尔滨建筑大学学报, 1992(4)
- 6 张淑华, 王光远. 工程并联子系统的优化. 哈尔滨建筑大学学报, 1993(6)