

103 主成分分析赋权法 (principal component analysis)

在计算评价体系的发展水平时, 权重方法的确定对评价决策结果产生直接影响。若方法过于主观, 会使评价结果产生误差, 尤其当评价体系中指标数目较多时, 评价结果可能与实际情况出现巨大偏差。主成分分析法基本上消除了主观性, 因此在实践中得到广泛使用。

(1) 基本原理

主成分分析是把多项指标转化为少数几个综合指标的一种统计分析方法。基本思想是从众多的观测变量中综合出携带原始数据信息最多且相互独立的几个因素解释原有数据变量, 目的是降维, 简化数据结构, 给研究和分析问题带来方便。

(2) 具体步骤

① 建立指标体系的原始数据矩阵。

每个对象均观察 P 项指标, 记为 X_1, X_2, \dots, X_p , 建立数据矩阵 X 。

② 评价指标的标准化处理。

为了避免变量的量纲不同所带来的影响, 对原始数据 (指标) $X_j = (X_{1j}, X_{2j}, \dots, X_{nj})$ 进行标准化处理, 使得每个变量的平均值为 0、方差为 1,

即 $Z_{ij} = \frac{x_{ij} - \bar{x}_{0j}}{\sqrt{s_{jj}}}$ (其中 $\bar{x}_{0j} = \frac{\sum_{i=1}^n x_{ij}}{n}$ 、 $s_{jj} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_{0j})^2}{n-1}$ 、 $j=1, 2, \dots, p$), 得到

$Z_j = (Z_{1j}, Z_{2j}, \dots, Z_{nj})$ 。

③ 由标准化数据计算相关矩阵 R 。

计算已作标准化处理的评价指标的相关系数矩阵 R , 元素 r_{ij} 表示原始变量 X_i 和 X_j 的相关系数。

$$r_{ij} = \frac{s_{jk}}{\sqrt{s_{jj}} \sqrt{s_{kk}}} = \frac{\frac{1}{n-1} \sum_{i=1}^p (x_{ij} - \bar{x}_{0j})(x_{ik} - \bar{x}_{0k})}{\sqrt{s_{jj}} \sqrt{s_{kk}}} = \frac{1}{n-1} \sum_{i=1}^n Z_{ij} Z_{ik}$$

$$\left(\text{其中 } s_{kk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ik} - \bar{x}_{0k})^2, \bar{x}_{0k} = \frac{1}{n} \sum_{i=1}^n x_{ik}, Z_{ik} = \frac{x_{ik} - \bar{x}_{0k}}{\sqrt{s_{kk}}} \right)$$

$$\therefore R = (\gamma_{ij})_{p \times p} = \frac{1}{n-1} \left(\sum_{i=1}^n Z_{ij} Z_{ik} \right)_{p \times p} = \frac{1}{n-1} (Z)^T Z$$

④计算 R 的特征值并对特征值进行排序。

求解特征方程 $|R - \lambda E| = 0$ ，计算相关矩阵 R 的特征值 λ 。

⑤确定主成分个数 m ，阈值一般取 85%。

$$sum = \sum_{j=1}^m \lambda_j / \sum_{i=1}^p \lambda_i \geq \text{阈值} 85\%$$

⑥计算 m 个相应的单位特征向量。

$$a_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{p1} \end{bmatrix} \quad a_2 = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{p2} \end{bmatrix} \quad \dots \quad a_m = \begin{bmatrix} a_{1m} \\ a_{2m} \\ \vdots \\ a_{pm} \end{bmatrix}$$

⑦写出主成分。

$$Z_k = a_{1k} X_1 + a_{2k} X_2 + \dots + a_{pk} X_p, k = 1, 2, \dots, m$$

主成分 Z_k 的应用，除了降维外，还可以通过 Z_i 与 Z_j 的对应取值变化，了解综合指标变量之间的关系和变化趋势；通过 Z_i 对 X_1, X_2, \dots, X_p 的贡献率

$v_i = \sum_{k=1}^m r^2(Z_k, X_i)$ ，然后找出 v_i 最大的指标变量 X_i ，视 X_i 为 Z_i 影响较大的指标，

其中 v_i 定义如下。

$$r^2(Z_k, X_i) = \frac{\sum_{m=1}^p z_{km} x_{jm}}{\sqrt{\sum_{m=1}^p z_{km}^2} \sqrt{\sum_{m=1}^p x_{jm}^2}}$$

(3) 赋权方法评判

主成分分析方法能够在损失少量信息的基础上减少变量，减少了运算的工作量，但是这种方法确定的新变量的意义通常不易确定。主成分分析赋权一方面体现了指标区分评价对象能力的大小，另一方面删除了一些重复信息，使得综合评价指标相互独立，主要用于评价指标数目较多并且指标之间有复杂关系的评价系统中。