

文章编号: 2095-2163(2020)09-0090-05

中图分类号: TP391

文献标志码: A

# 基于主成分分析对“禁瓶令”效果的研究

杜晋华

(中国地质大学(北京)信息工程学院, 北京 100083)

**摘要:** 目前世界各地对塑料垃圾的回收率较低,一些国家试图通过推行“禁瓶令”从源头上解决这一问题。基于国内外现有研究成果,本文归纳了可评估该项举措效果的多个指标,使用主成分分析对指标进行了降维,建立了举措效果评价模型。旨在通过从时间维度对采取“禁瓶令”前后评价结果变化的纵向比较,以及对不同区域开始“禁瓶令”后评价结果的横向比较,分析出“禁瓶令”的有效性以及适用范围,为有关国家和地区的相关政策出台提供参考。

**关键词:** 禁瓶令; 主成分分析; 评价模型

## Study on the effect of bottle prohibition order based on principal component analysis

DU Jinhua

(School of Information Engineering, China University of Geosciences (Beijing), Beijing 100083, China)

**[Abstract]** Recycling rates for plastic waste around the world are low, and some countries are trying to tackle the problem by enforcing bottling bans. Based on the existing research results at home and abroad, this paper summarizes a number of indicators that can be used to evaluate the effect of the initiative, which uses principal component analysis to reduce the dimension of the indicators, and establishes an evaluation model for the effect of the initiative. The aim is to analyze the effectiveness and scope of application of the bottle prohibition order through the longitudinal comparison of the evaluation results before and after the bottle prohibition order in the time dimension and the horizontal comparison of the evaluation results after the bottle prohibition order in different regions, so as to provide reference for relevant countries and regions to issue relevant policies.

**[Key words]** bottle prohibition order; principal component analysis; evaluation model

### 0 引言

生活水平的提高和经济的富裕使得人们更加追求高品质的生活和可持续性的发展,更加关注个人健康与生态平衡。瓶装水的存废因此成为一个备受争议的话题。部分地区出于对环境保护和能源节约的需要,采取“禁瓶令”的方式禁止瓶装水的生产,但受到该禁令限制的居民和企业并不看好“禁瓶令”<sup>[1]</sup>。如何衡量“禁瓶令”的有效性成为制约当地政府决策的一个重大问题。

本文通过建立有效的评价模型,针对同一区域给出采取“禁瓶令”前后的评价分数,针对不同区域给出采取“禁瓶令”后的评价分数,衡量“禁瓶令”的实际效果并判断其适用的场景。

### 1 模型假设和符号说明

#### 1.1 模型假设

(1) 所研究对象(如某城市)的平均购买力在研究阶段保持不变;

(2) 单瓶矿泉水的制作流程的全部环节在研究阶段内不发生明显变化。

#### 1.2 名词解释

(1) 人口流量:指在一定时间内人口数量的增量;

(2) 时间成本:指顾客为想得到所期望的商品或服务而必须耗费的时间换算而成的代价;

(3) 主成分载荷:各指标对主成分所做的贡献,其值越大,在主成分中所起的作用就越大,该评价指标也就越敏感。

#### 1.3 符号说明

符号说明见表1。

### 2 模型建立

#### 2.1 指标归纳

瓶装水的存废备受争议。一方面,其具有便携、健康、易运输等优点;而另一方面,瓶装水对水的私有化和商品化,破坏了人类的水生态系统,加剧了能源的消耗和环境的污染,损害了人类的利益。针对不同使用环境,不同使用人群,瓶装水的作用又有所不同。因此,“禁瓶令”的推行与否会有正负两方面影响。

作者简介:杜晋华(2000-),男,本科生,主要研究方向:机器学习、数值分析与计算。

收稿日期:2020-06-23

表 1 二级评价指标  
Tab. 1 Secondary evaluation index

一级指标	二级指标	二级指标解释
a1:生产指标	a11	原料额外费用支出比:生产一瓶矿泉水所消耗的原料较普通饮用水的费用比
	a12	水源地经济受损度:矿泉水水源供应地周边受水源开采造成的人均收入下降程度
	a13	生产安全系数:流水线生产一瓶矿泉水发生事故的可能性
	a14	供水层破坏度:供水层遭受人工取水影响的破坏程度
	a15	碳排放量:生产单个矿泉水瓶的碳排放量
a2:输送指标	a21	尾气平均排放量:运输单瓶矿泉水 1 公里排出的温室气体量
	a22	运输耗能:运输单瓶矿泉水 1 公里耗费的能源总费用
	a23	运费节省比:采用瓶装较其他封装方式平均节省的费用占其他封装方式平均耗费的比率
a3:加工指标	a31	平均加工费:生产单瓶矿泉水共需要的人工和器械费用
	a32	加工复杂度:加工完一瓶矿泉水的流程的复杂程度
a4:包装指标	a41	广告效益:从矿泉水广告上获得的收益
	a42	包装成本:包装一瓶矿泉水花费的包装费用以及材料耗费
a5:销售指标	a51	销售利润率:销售一瓶矿泉水获得的平均利润
	a52	同等饮料替代比:顾客在所有可选择饮料范围内选择矿泉水的可能性
a6:市场指标	a61	就业贡献率:矿泉水业全部从业人员占总从业人员的比重
	a62	税收贡献率:矿泉水业对市场缴纳税款占市场税款总额的比重
	a63	平均价格:单瓶矿泉水的平均销售价格
a7:使用指标	a71	直接污染占比:可视塑料瓶污染占总环境污染的比例
	a72	间接污染占比:潜在塑料瓶污染占总环境污染的比例
	a73	及时性:获得单瓶矿泉水的时间成本与获得日常用水的时间成本的比例
	a74	便利性:在大多数情况下,获得单瓶矿泉水的可能性较获得日常用水的可能性的比例
	a75	健康性:矿泉水相较自来水的纯净度、含矿物度等的平均程度
	a76	人均购买量:单人在单位时间内购买矿泉水的平均值
a8:回收指标	a81	降解复杂度:一个矿泉水瓶完全降解需要的时间年限
	a82	人工处理难度:人工降解一个矿泉水瓶需要投入的时间
	a83	回收难度:搜集单个矿泉水瓶需要的时间

注:除上述符号外,还涉及到:  $t$ : 以年为单位的某一确定的时间点。

正面影响:

- (1) 衍生出大批新的产业部门,带动瓶装水产业的人员就业;
- (2) 瓶装水广告效益拉动其他行业经济增长;
- (3) 瓶装水生产厂家从中获得收益,并为当地贡献税收;
- (4) 在自然灾害、社会险情中及时解决吃水困难;
- (5) 影响顾客的货架选择:在众多不健康饮品中选择瓶装水,利于健康;
- (6) 提高顾客用水满意度,促进人员流动,带动相关旅游业和人员外出地经济。

负面影响:

- (1) 在生产、输送、加工、包装过程中的能源损耗与浪费;
- (2) 生产瓶装水对水源地生态与经济的破坏;
- (3) 增加可饮用水二次加工的时间成本;
- (4) 在生产、输送、加工、包装过程中产生大量温室气体;
- (5) 空瓶回收与处理成本高昂;
- (6) 降解难度大,环境污染严重。

根据可比性和概括性的原则,针对上述影响可将衡量指标细化为 8 个一级指标及 26 个二级指标。

## 2.2 数据标准化

在即将建立的多指标评价模型中,由于各评价指标的性质不同,它们会具有不同的量纲和数量级。考虑到各指标间的水平相差很大,如果直接用原始指标值进行分析,就会突出数值较高的指标在综合分析中的作用,相对削弱数值水平较低指标的作用。为了保证结果的可靠性,需要对原始指标数据进行标准化处理。采用 min-max 标准化的归一化方法。

设每一个研究对象的上述 26 个指标构成行向量为式(1):

$$X = [x_1, x_2, \dots, x_{26}]. \quad (1)$$

对该向量中的值序列进行变换为式(2):

$$y_i = \frac{x_i - \min_{1 \leq j \leq 26} \{x_j\}}{\max_{1 \leq j \leq 26} \{x_j\} - \min_{1 \leq j \leq 26} \{x_j\}}. \quad (2)$$

则得到新向量,式(3):

$$X' = [y_1, y_2, \dots, y_{26}]. \quad (3)$$

该向量中各元素均属于区间[0,1]且无量纲。

## 2.3 数据获取

以美国芝加哥市和康科德镇为例,通过搜集美国运输统计局、劳工统计局、人口统计局官方公布的数据以及美国零售业数据或者通过实地市场调查获得数据,直接确定多数指标取值。

部分指标需要通过二次计算才能得到,如式(4)和式(5)计算的水源地经济受损度和供水层破坏度。

$$\text{水源地经济受损度} = 1 - \frac{\text{地区采水后阶段 GDP}}{\text{地区采水前同阶段长度 GDP}}. \quad (4)$$

$$\text{供水层破坏度} = \frac{\text{供水层塌方体积}}{\text{采水地区供水层总体积}}. \quad (5)$$

为了使评价体系的适用性更强,能够反映不同地域范围、不同人口数量区域之间的差异,并通过对比同一区域在采取禁瓶令前后受到的影响程度,可以选取两个地域范围相差较大的地区(如旧金山市和康科德镇)采取禁瓶令前后的四组数据。

## 3 主成分分析

在待建立的评价模型中,过多的变量增加了分析问题的难度和复杂性,而诸变量之间是具有一定相关关系的,所以期望通过相关分析,利用较少的新变量代替原来较多的旧变量,同时尽可能多地保留原来变量所反映的信息。因此,本文选用具有降维思想的主成分分析法来构建模型。

设有  $n$  个区域样本,每个样本观测上述 23 个指标,将原始数据写成矩阵,式(6):

$$X = \begin{bmatrix} x_{1,a11} & x_{1,a12} & \dots & x_{1,a83} \\ x_{2,a11} & x_{2,a12} & \dots & x_{2,a83} \\ \dots & \dots & \dots & \dots \\ x_{n,a11} & x_{n,a12} & \dots & x_{n,a83} \end{bmatrix}. \quad (6)$$

(1) 将原始数据  $X$  标准化;

(2) 建立变量的相关系数阵,式(7)和式(8);

$$R = (r_{i,j})_{26 \times 26}, \quad (7)$$

$$r_{i,j} = \frac{\sum_{k=1}^n (x_{k,i} - \bar{x}_i)(x_{k,j} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{k,i} - \bar{x}_i)^2 \sum_{k=1}^n (x_{k,j} - \bar{x}_j)^2}}. \quad (8)$$

(3) 借助 MATLAB 等工具求  $R$  的特征根及相应的单位特征向量:

设求出的特征根为:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{26} \geq 0.$$

相应的特征向量为式:

$$a_1 = \begin{bmatrix} a_{1,1} \\ a_{2,1} \\ \dots \\ a_{26,1} \end{bmatrix}, a_2 = \begin{bmatrix} a_{1,2} \\ a_{2,2} \\ \dots \\ a_{26,2} \end{bmatrix}, \dots, a_{26} = \begin{bmatrix} a_{1,26} \\ a_{2,26} \\ \dots \\ a_{26,26} \end{bmatrix},$$

(4) 得到主成分,式(9);

$$F_i = a_{1,i}X_1 + a_{2,i}X_2 + \dots + a_{26,i}X_{26} (i = 1, 2, \dots, 26). \quad (9)$$

(5) 计算主成分贡献率及累计贡献率;

① 贡献率,式(10):

$$\frac{\lambda_i}{\sum_{k=1}^{26} \lambda_k} (i = 1, 2, \dots, 26). \quad (10)$$

② 累计贡献率,式(11):

$$\frac{\sum_{k=1}^i \lambda_k}{\sum_{k=1}^{26} \lambda_k} (i = 1, 2, \dots, 26). \quad (11)$$

(6) 对主成分进行取舍:取累计贡献率达 85%~95%的特征值  $\lambda_1, \lambda_2, \dots, \lambda_{26}$  所对应的第 1、第 2、...、第  $m$  ( $m \leq 26$ ) 个主成分;

(7) 利用主成分  $F_1, \dots, F_{26}$  做线性组合,并以每个主成分的贡献率作为权重构造综合评价函数,式(12):

$$y = a_1F_1 + \dots + a_mF_m. \quad (12)$$

## 4 模型求解

以美国的康科德镇和旧金山市为例,经过标准化处理后的数据见表 2。

表2 区域“禁瓶令”效果指标数据汇总实例

Tab. 2 An example of the effect index data summary of the regional bottle prohibition order

时间 区域	执行“禁瓶令”前		执行“禁瓶令”后	
	康科德镇	旧金山市	康科德镇	旧金山市
a11	0.000 1	0.000 1	0	0
a12	0.153 3	0.126 2	0.090 3	0.092 8
a13	0.000 1	0.000 1	0	0
a14	0.028 8	0.028 5	0	0
a15	0.004 2	0.004 2	0.006 8	0.005 1
a21	0.002 4	0.002 4	0.015	0.011 4
a22	0.009 8	0.009 7	0	0
a23	0.001 8	0.001 7	0	0
a31	0.000 1	0.000 1	0	0
a32	0.001	0.001	0	0
a41	0.000 2	0.000 2	0	0
a42	0.000 1	0.000 1	0	0
a51	0.000 9	0.000 7	0	0
a52	0.092	0.101	0	0
a61	0	0	0	0
a62	0.000 1	0.000 8	0	0
a63	0.001	0.001	0.757 6	0.576 6
a71	0.033 3	0.038 8	0.030 6	0.033 6
a72	0.084 2	0.135 9	0.645 1	0.799 2
a73	0	0.000 4	0.015	0.137
a74	0.013 7	0.102 3	0	0
a75	0.002 5	0.002 5	0.015	0.011 4
a76	0.000 7	0.008 1	0	0
a81	0.979 1	0.970 8	0	0
a82	0.002 4	0.002 4	0	0
a83	0.000 1	0.001 9	0	0

利用主成分分析对上述数据进行计算,可得到前三个主成分、主成分得分与特征根。设前三个主成分的各项系数为  $k_i = [k_1, k_2, \dots, k_{26}] (i = 1, 2, 3)$ , 则计算结果可表示为:

$$k_1 = [-0.480\ 7\ 0.636\ 9\ -0.480\ 7\ -0.271\ 9\ -0.444\ 5\ -0.450\ 0\ -0.410\ 1\ -0.468\ 6\ -0.480\ 7\ -0.474\ 1\ -0.480\ 0\ -0.480\ 7\ -0.475\ 6\ 0.224\ 6\ -0.481\ 4\ -0.478\ 1\ 0.226\ 1\ -0.183\ 6\ 0.092\ 0\ -0.396\ 4\ -0.055\ 2\ -0.449\ 3\ -0.449\ 1\ 0.648\ 9\ -0.463\ 9\ -0.474\ 1]$$

$$k_2 = [-0.401\ 3\ 0.127\ 0\ -0.401\ 3\ -0.431\ 2\ -0.361\ 7\ -0.306\ 3\ -0.411\ 4\ -0.403\ 1\ -0.401\ 3\ -0.402\ 3\ -0.401\ 4\ -0.401\ 3\ -0.402\ 1\ -0.501\ 7$$

$$-0.401\ 2\ -0.401\ 7\ 4.521\ 2\ -0.201\ 6\ 0.819\ 0\ 0.160\ 7\ -0.457\ 2\ -0.306\ 4\ -0.405\ 4\ -0.421\ 9\ -0.403\ 7\ -0.402\ 2]$$

$$k_3 = [0.013\ 3\ 0.021\ 9\ 0.013\ 3\ 0.014\ 7\ 0.019\ 7\ 0.026\ 6\ 0.013\ 8\ 0.013\ 4\ 0.013\ 3\ 0.013\ 3\ 0.013\ 3\ 0.013\ 3\ 0.013\ 3\ 0.013\ 4\ 0.013\ 6\ 0.013\ 3\ 0.013\ 0\ 0.677\ 0\ 0.001\ 0\ -0.582\ 2\ -0.439\ 1\ -0.024\ 3\ 0.026\ 6\ 0.010\ 1\ 0.062\ 1\ 0.013\ 4\ 0.012\ 5]$$

对应的主成分为式(13):

$$z_i = \sum_{j=1}^{26} k_{i,j} x_j (i = 1, 2, 3). \quad (13)$$

第一主成分贡献率为49.87%;第二主成分贡献率为49.01%;第三主成分贡献率为0.99%,前两个主成分累计贡献率达98.88%,故可选取它们为新的因子。

第一新因子  $Z_1$  主要代表变量为  $a12$ (水源地经济受损度)、 $a81$ (降解复杂度),其权重系数分别为0.636 9、0.648 9,反映了这两个变量与生态环境水平密切相关;第二新因子  $Z_2$  主要代表变量为  $a52$ (同等饮料替代比)、 $a63$ (平均价格),其权重系数分别为0.501 7、0.521 2,反映了这两个变量与瓶装水主要用途密切相关。

最后的评价函数为式(14):

$$y = 0.498\ 7z_1 + 0.490\ 1z_2. \quad (14)$$

### 5 模型改进

上述模型在常见地域内计算出的评价分数能够在较长时间内与实际情况相符合。但在一些包含特殊场所(如机场)的区域,其适用性就有待质疑。

具体原因是:机场和静态地域不同,虽然在机场内瓶装水“禁瓶令”的影响因素没有发生较大改变,但机场十分突出的特点就是人员的流动性大。如美国旧金山常住人口约为88.5万,而旧金山机场年接待游客数达1900万。上述模型中对这一因素没有体现,所以可以从人口方面进行改进。

人口数量满足以下公式(15):

$$\text{地区人口} = \text{常住人口} + \text{流动人口}. \quad (15)$$

考虑到常住人口和流动人口购买力的不同, $a76$ 不应该是购买总量与地区人口的比值的直接结果。

设人均相对购买比为式(16):

$$\omega = \text{流动人口人均购买量} / \text{常住人口人均购买量}. \quad (16)$$

则购买总量应为式(17):

$$M = (1 + \omega) * \text{常住人口购买量}. \quad (17)$$

更新后的  $a76$  为式(18):



$$a_{76} = \frac{M}{\text{地区人口数}} \quad (18)$$

其中,  $\omega$  通常情况下借助零售业有关部门的反馈数据求得。但其时间成本远大于人口统计局人口数量统计工作, 所以无法及时获得数据。

为解决这一问题, 考虑引入人口相对数量比, 式(19):

$$\beta = \text{人口流量} / \text{常住人口数} \quad (19)$$

需要获得  $\beta$  与  $\omega$  的关系。受不同地域的影响, 人均相对购买比与人口相对数量比跟当地的经济水平和繁荣程度呈正相关, 而这两个指标与时间有明显的数量关系。为提供本模型除评价外的预测能力, 考虑利用拟合的思路, 借助往年的数据, 分析  $\beta$  与  $\omega$  所满足的函数关系。

利用最小二乘拟合的方法可求出  $\omega/\beta$  与时间  $t$  的非线性最小二乘拟合的函数, 设为式(20):

$$\omega/\beta = f(t) \quad (20)$$

## 6 误差分析与敏感性分析

本文建立的评价模型采用了主成分分析的方法。由于主成分分析的核心思想是降维, 虽然有效降低了评价模型的复杂度, 但势必造成了一定程度的数据信息丢失。在大多数情况下用大于 85% 贡献度的主成分作为新的指标, 误差可以在允许范围内, 对最终的评价结果不会造成足够大的影响, 但在特殊场合下, 如以某人口数量锐减的乡镇作为研究对象, 该模型可能失效。

评价模型的主成分选择是基于主成分载荷的大

小是否不小于 85% 确定的。主成分载荷越大, 该指标就越敏感, 就可以作为敏感性分析的指标。

设在理想情况下, 求出  $y$  的目标值为  $y_0$ 。为尽可能简化分析复杂度, 选择单因素敏感性分析, 取主载荷最大的成分为变量, 不改变其他指标的值, 观察该变量变化对整体结果的影响程度。

衡量影响程度的变量及其计算为公式(21):

$$\kappa = 100\% * |y(a'_1) - y(a_1)| / |y(a_1)| \quad (21)$$

若  $\kappa \geq 60\%$ , 则该指标影响程度过大, 应该作为政府“禁瓶令”中重点考虑因素; 反之, 该指标影响程度较低, 政府“禁瓶令”决策的重心可从此处偏离。

## 7 结束语

为了解“禁瓶令”是否会给施行地更多带来有益影响, 本文通过建立举措效果评价模型为不同地域施行“禁瓶令”前后进行打分。相关人员可根据分值变化与差距对是否在某地区施行“禁瓶令”进行决策。通过本文的实例样本, 可以初步得到以下结论: 大型城市和流动性较强的地域, 如美国旧金山及其机场采取“禁瓶令”是有效的; 而其他地域采用“禁瓶令”, 短期内效果不明显。

## 参考文献

- [1] Rossi, Christina. Survey: More Cities Cut Bottled water Spending [EB/OL]. (2010) [2019-12-18]. <http://www.stopcorporateabuse.org/survey-more-cities-cut-bottled-water-spending>.
- [2] 宁井铭, 张正竹, 宛晓春. 基于气相色谱技术的普洱熟茶 7572 挥发性物质指纹图谱建立及应用[J]. 食品与发酵工业, 2016, 42(10): 176-183.
- [3] 周丽静, 侯彩云, 乔艳慧, 等. 不同贮藏年份普洱茶功能特性研究[J]. 食品科学, 2010, 31(13): 19-22.
- [4] XIE Guoxiang, WANG Yungang, YE Mao. Characterization of Pu-erh Tea Using Chemical and Metabolic Profiling Approaches[J]. Journal of Agricultural and Food Chemistry, 2009, 57(8): 3406-3054.
- [5] 湛滢, 欧行畅, 张杨波, 等. 不同储藏年份普洱生茶感官品质的分析[J]. 食品科技, 2018, 43(4): 48-52.
- [6] 任奇锋, 王俊. 自动进样与恒温控制型电子舌检测系统[J]. 农业机械学报, 2016, 47(4): 186-192.
- [7] 刘双印, 徐龙琴, 李振波, 等. 基于 PCA-MCAFA-LSSVM 的养殖水质 pH 值预测模型[J]. 农业机械学报, 2014, 45(5): 239-246.
- [8] 史庆瑞, 国婷婷, 殷廷家, 等. 基于电子舌检测的橙汁贮藏品质研究[J]. 食品与机械, 2017, 33(11): 137-142, 203.
- [9] 邵婷婷, 白宗文, 周美丽. 基于离散小波变换的信号分解与重构[J]. 计算机技术与发展, 2014, 24(11): 159-161.
- [10] Jian, Ling, et al. Budget online learning algorithm for least squares SVM[J]. IEEE transactions on neural networks and learning systems 2016, 28(9): 2076-2087.

(上接第 89 页)

的普洱茶进行定性分析。运用 DWT 对电子舌输出信号进行预处理, 利用 DWT 输出特征数据作为输入, 建立 LSSVM 分类模型, 利用 PSO 算法对 LSSVM 的  $(c, \sigma)$  参数进行寻优操作。实验表明电子舌结合 PSO-LSSVM 模型可以有效地分辨普洱茶的不同存储时间。该研究可为茶叶类产品的质量检测提供新的研究思路。

## 参考文献

- [1] 宁井铭, 张正竹, 宛晓春. 基于气相色谱技术的普洱熟茶 7572 挥发性物质指纹图谱建立及应用[J]. 食品与发酵工业, 2016, 42(10): 176-183.
- [2] 周丽静, 侯彩云, 乔艳慧, 等. 不同贮藏年份普洱茶功能特性研究[J]. 食品科学, 2010, 31(13): 19-22.
- [3] XIE Guoxiang, WANG Yungang, YE Mao. Characterization of Pu-erh Tea Using Chemical and Metabolic Profiling Approaches[J].