

基于聚类分析和因子分析对我国 31 个主要城市空气质量的研究

刘 琴, 葛梅梅

(滁州学院数学与金融学院, 安徽 滁州 239000)

摘 要:随着社会经济的发展,空气质量的优劣越来越受到人们的关注。根据可监测的空气质量指标利用聚类分析将我国主要的 31 个省、直辖市、自治区进行空气质量优劣分类,同时利用因子分析将影响空气质量的指标综合为少数指标,并对因子进行命名,综合得出影响空气质量的主要因素。

关键词:空气质量;聚类分析;因子分析

中图分类号:X823 **文献标志码:**A **文章编号:**1673-1891(2021)01-0062-04

Study on Air Qualities of 31 Major Cities in China Based on Cluster Analysis and Factor Analysis

LIU Qin, GE Meimei

(School of Mathematics and Finance, Chuzhou University, Chuzhou, Anhui 239000, China)

Abstract: With social and economic development, people increasingly pay attention to air quality. In this paper we classify the air quality of 31 provincial capitals and municipalities according to the detectable air quality indexes through cluster analysis. At the same time, we use factor analysis to integrate the indexes of the factors affecting air quality into minority key indexes, and also designate the factors. With these we can comprehensively work out the major factors affecting the air quality.

Keywords: air quality; cluster analysis; factor analysis

0 引言

随着工业发展、人口密集、生活污染物排放等现象的出现,空气质量越来越差,严重影响着人们的生活。只有找到影响空气质量的主要因素,才能使人们的生活质量得以改善。近年来,魏臻、林芳^[1]利用聚类分析和主成分分析对城市空气质量进行了评价;胡锐玲^[2]在相关性分析的前提下,利用因子分析对各省市环境综合水平进行了研究;孙晓玲、朱家明^[3]等人利用模糊综合评价和因子分析相结合的方法,建立空气质量指标体系,对北京、天津、石家庄等 31 个城市的空气质量进行分析评价。本文根据中国 31 个主要城市 2019 年空气质量指标相关数据,对各个指标数据标准化之后先进行聚类分析,将全国 31 个主要城市根据环境现状以及数据处理结果进行空气质量优劣分类;再利用因子分析,从影响空气质量的 6 个主要指标中提取 2 个因

子,使影响空气质量的因素具有命名解释性,用 2 个因子来主要概括影响空气质量的好坏。最后根据聚类分析以及因子分析的结果提出对城市环境保护方面的政策性建议。

1 数据来源

此次全国 31 个主要城市的空气质量指标数据来源于中国统计年鉴环保重点城市空气质量情况,然后对数据进行了标准化处理。

2 研究方法

2.1 系统聚类法

系统聚类法又叫“分层聚类法”,其基本思想是先将各个样品各看成一类,然后规定类与类之间的距离,选择距离最小的一类合并成新的的一类,再计算新类与其他类之间的距离,再将距离最近的两类合并,如此继续下去,直到所有的样品合并成一类

收稿日期:2020-05-04

基金项目:大学生创新训练项目(2020CXXL143);安徽省省级教学一般项目(2018jyxm0587)。

作者简介:刘琴(1999—),女,安徽六安人,本科,研究方向:经济统计学。

表 2 因子解释原有影响因素总方差的情况

成分	初始特征值			提取平方和载入			旋转平方和载入		
	合计	方差/%	累积/%	合计	方差/%	累积/%	合计	方差/%	累积/%
1	4.081	68.019	68.019	4.081	68.019	68.019	2.891	48.179	48.179
2	1.012	16.861	84.880	1.012	16.861	84.880	2.202	36.701	84.880
3	0.442	7.366	92.246						
4	0.241	4.020	96.266						
5	0.154	2.567	98.833						
6	0.070	1.167	100.000						

两个因子。用极大方差法对因子载荷矩阵进行旋转,得到表 2 结果。

表 3 旋转后的因子载荷矩阵

	成分	
	1	2
Zscore(SO ₂)	0.071	0.902
Zscore(NO ₂)	0.847	0.327
Zscore(PM ₁₀)	0.663	0.691
Zscore(PM _{2.5})	0.859	0.399
Zscore(CO)	0.382	0.802
Zscore(O ₃)	0.919	0.052

从旋转后的因子载荷矩阵可以看出,第一个因子在 O₃、PM_{2.5}、NO₂这几个指标上的载荷量大,第二个因子在 SO₂、CO 上的载荷量大。根据因子的实际意义,第一个因子可以命名为“呼吸道影响因子”;第二个因子可命名为“刺激性影响因子”。

从因子载荷矩阵表中,可以得出因子分析的模型式(1)(2)。

$$F_1 = 0.071 Z_{SO_2} + 0.847 Z_{NO_2} + 0.663 Z_{PM_{10}} + 0.859 Z_{PM_{2.5}} + 0.382 Z_{CO} + 0.919 Z_{O_3} \quad (1)$$

$$F_2 = 0.902 Z_{SO_2} + 0.327 Z_{NO_2} + 0.691 Z_{PM_{10}} + 0.399 Z_{PM_{2.5}} + 0.802 Z_{CO} + 0.052 Z_{O_3} \quad (2)$$

Z_{SO₂}、Z_{NO₂}、Z_{PM₁₀}、Z_{PM_{2.5}}、Z_{CO}、Z_{O₃}是在原始数据标准化后的数据,各个变量前的系数是经过最大方差旋转后各因子的载荷量。最后建立综合评价模型,由因子方差贡献率计算得到各省市综合得分结果,计算公式为式(3)。

$$F = \frac{0.48179 \times F_1 + 0.36701 \times F_2}{0.18179 + 0.36701} \quad (3)$$

计算结果如表 4 所示。从表 4 中可以看出综合得分排在前五的分别是石家庄、太原、西安、郑州、济南,排名越靠前,得分越高,说明这些城市的空气质量越差,而排在后五名的分别是昆明、贵阳、福州、拉萨、海口,排名越靠后的城市说明这些城市的空气质量较好。

表 4 各个城市空气质量评价汇总

城市	因子得分		综合得分	综合排名
	F ₁	F ₂		
石家庄	1.767 08	1.394 50	1.605 981	1
太原	0.771 10	2.207 28	1.392 085	2
西安	1.318 56	0.628 51	1.020 191	3
郑州	1.522 36	0.193 67	0.947 852	4
济南	0.919 51	0.835 03	0.882 982	5
兰州	0.162 27	1.327 21	0.665 975	6
乌鲁木齐	0.363 92	0.888 08	0.590 560	7
天津	1.021 17	-0.513 30	0.357 685	8
西宁	-0.424 86	1.298 85	0.320 450	9
成都	0.858 93	-0.579 72	0.236 877	10
北京	1.006 68	-0.845 08	0.206 003	11
银川	-0.808 47	1.353 45	0.126 316	12
武汉	0.691 92	-0.738 85	0.073 274	13
呼和浩特	-0.776 08	1.152 42	0.057 778	14
合肥	0.684 57	-0.878 19	0.008 853	15
南京	0.423 63	-0.567 02	-0.004 710	16
沈阳	-0.497 98	0.478 77	-0.075 650	17
杭州	0.247 18	-0.732 86	-0.176 580	18
长沙	0.38784	-1.038 93	-0.229 080	19
重庆	0.183 91	-0.849 21	-0.262 800	20
哈尔滨	-0.611 69	0.054 12	-0.323 800	21
广州	0.092 69	-1.209 02	-0.470 150	22
南昌	-0.845 53	-0.057 44	-0.504 770	23
长春	-1.066 86	0.117 99	-0.554 550	24
上海	-0.107 69	-1.184 86	-0.573 440	25
南宁	-0.715 95	-0.423 51	-0.589 500	26
昆明	-0.997 38	-0.300 05	-0.695 860	27
贵阳	-1.082 69	-0.229 70	-0.713 870	28
福州	-1.033 75	-0.817 38	-0.940 190	29
拉萨	-1.579 84	-0.166 59	-0.968 770	30
海口	-1.874 55	-0.798 20	-1.409 150	31

4 结果与分析

根据我国城市分布情况,综合聚类分析的分类结果以及因子分析的各个城市空气质量评价汇总情况,得出结果与分析。

第一类:石家庄、太原、西安、郑州、济南、北京、天津。这些城市的因子得分最高,从地理位置上看,这些城市大多位于京津冀地区以及周边城市。近些年河北地区的工业发展迅速,地势西北高、东南低。京津冀地带工业污染较轻但人口密集,生活中的污染和汽车尾气问题严重,尤其是经常发生的雾霾现象。要想治理京津冀地区的空气污染情况,相关部门应及时找到引起污染的各种因素,根据污染源采取针对性解决措施。

第二类:呼和浩特、沈阳、兰州、西宁、银川、乌鲁木齐。这些城市的因子得分较高,空气质量较差。这些城市几乎分布在东北和西北地区,我国的沙漠主要集中在西北地区,近几年新疆和内蒙古草原的过度开垦,使得地面裸露程度变高,造成沙尘天气。而我国地势西高东低导致西北地区的沙尘天气顺势而下,影响到我国广大东北地区。所以要想提高这些城市的空气质量,需要制定政策,合理开发草原和森林,在西北地区建造防护林,减少对其他城市的影响。

第三类:上海、南京、杭州、合肥、武汉、长沙、广州、重庆、成都。从因子分析结果看,这些城市的因

子得分属于中等水平,综合得分在0.8以下,不是太高,这些城市大多属于长三角地区、珠三角地区。长三角对于全国经济发展具有较大的带动作用,长三角、珠三角第二产业比重不仅高于全国平均水平,也是世界各国中的最高水平,所以工业的发展不可避免地导致了工业废气废水等污染物较大排放量,造成了较差空气质量。因此要改善长三角地区的空气质量,要坚定不移地调整产业结构,淘汰落后产能。

第四类:长春、哈尔滨、福州、南昌、南宁、海口、贵阳、昆明、拉萨。从因子得分看,这些城市的因子得分最低,综合排名在20名以后,综合得分在0以下,空气质量最好。这些城市主要位于东南、西南地区的沿海城市,属于亚热带季风气候,常年温和多雨,对空气中的粉尘等污染物起到净化作用^[4],所以空气质量较好。

5 结语

本文对我国31个主要城市的空气质量进行了聚类分析以及因子分析,根据空气质量优劣情况,将31个城市分成了4类,根据影响空气质量状况的指标,将6个指标简化为2个概括性指标。根据分类情况,对于空气质量差的城市,应从多方面寻找原因,积极采取相应的措施,如减少二氧化硫等有毒气体的排出,控制工业排放量,减少露天焚烧,增加除烟除尘设备,扩大绿化面积等。对于空气质量较好的城市,应加强管理与保护。

参考文献:

- [1] 魏臻,林芳.基于聚类分析和主成分分析城市空气质量评价[J].淮阴工学院学报,2018,27(3):86-96.
- [2] 胡锐玲.基于因子分析的各省市环境综合水平研究[J].中国经贸导刊(中),2019(3):133-134.
- [3] 孙晓玲,朱家明,贺晓剑.基于模糊综合评价及因子分析对城市空气质量的研究[J].哈尔滨师范大学自然科学学报,2017,33(6):5-10.
- [4] 宋加梅.我国主要城市空气质量情况及影响因素分析[J].中国传媒大学学报(自然科学版),2019,26(4):50-53+60.
- [5] 许海柱,祝佳佳,张棚,等.基于聚类分析和因子分析的肺小结节患者中医证候特点研究[J].中国中医药信息杂志,2020,27(2):84-87.