

我国主要城市空气质量指标的聚类分析

李文青, 史雨梅

(滁州学院数学与金融学院, 安徽 滁州 239000)

摘要:我国经济的飞速发展导致环境污染不断加重,其中空气质量问题直接影响着人们的生活与健康。本文采用聚类分析(K均值)方法,按照不同的影响指标对全国31个主要城市的空气污染浓度数据进行分类,并对主要城市空气质量问题提出合理建议。

关键词:聚类分析;空气污染;空气质量

【DOI】10.12231/j.issn.1000-8772.2020.35.303

1 前言

随着我国经济的高速发展,工业化进程的不断加快,城市逐渐成为人口聚集之地,进而成为环境污染的聚集之地。工业的发展、人口的聚集、化石燃料的过度消耗以及废气的排放导致各地区空气质量越来越差。十九大报告明确指出:要执行严格的生态环境保护制度,建设美好中国。环境问题已成为我国政府正在关注以及未来重点关注的重要问题之一。因此有必要对全国各大城市的空气质量进行研究。为了深入研究,更好的把握防治空气污染的力度,本文借助SPSS软件对来源于2020年《中国统计年鉴》的全国31个主要城市的空气污染浓度数据进行统计分析,希望研究结果能为相关部门治理大气污染提供一定的参考价值。

2 K均值聚类

K均值聚类分析的基本原理是将数据中每个城市当成一个样品,每个城市的空气质量指标则为样品的要素,具体操作步骤如下:

(1)将所有样品分成K个初始类,K代表所要聚类的类数,事先已经设定好。

(2)计算样品与各类中心的距离,并将样品归类到距离最小的一类中,重新组成新的类。

(3)计算各类新的类中心,即均值。

(4)重复步骤(2)、(3),直到各类中的样品不再发生变化为止。

3 实证分析

通过查阅《中国统计年鉴》,本文选取空气污染指标中的SO₂浓度、NO₂浓度和PM₁₀浓度三个指标进行统计研究,见下表1。

运用SPSS中K均值法对全国31个城市空气质量的标准化数据进行聚类,通过实验分为5类分析结果如下。

表2为初始聚类中心,由标准化的数据可知,5个类别分别对应的是哈尔滨、郑州、海口、太原、西宁的数据为初始位置,如下所示。

表3展示了5个类中心每次迭代时的偏移情况。从表中可知,第3次迭代后,5个类的中心点的偏移均小于指定的判定标准(0.02),聚类分析结束。

表4为聚类成员表,给出了31个城市的聚类信息和每个成员到聚类中心的距离,如下所示。

由上述结果可知,2019年全国31个城市空气质量情况分为5类,具体为:第一类有14个城市,分别为北京、长春、上海、南京、杭州、合肥、南昌、武汉、长沙、广州、南宁、重庆、成都、昆明;第二类城市有4个,分别是天津、郑州、西安、乌鲁木齐;第3类城市有4个,为福州、海口、贵阳、拉萨;第4类城市也有4个,分别是石家庄、太

表1 城市的空气质量指标

城市	二氧化硫	二氧化氮	可吸入颗粒物
	年平均浓度 (ug/m ³)	年平均浓度 (ug/m ³)	(PM ₁₀)年平均 浓度 (ug/m ³)
北京	4	37	68
天津	11	42	76
石家庄	16	46	118
太原	22	50	107
呼和浩特	15	39	71
沈阳	21	36	77
长春	11	34	64
哈尔滨	17	32	67
上海	7	42	45
南京	10	42	69
杭州	7	41	66
合肥	6	42	68
福州	5	22	42
南昌	9	34	68
济南	15	45	106
郑州	9	45	99
武汉	9	44	71
长沙	7	33	57
广州	7	45	53
南宁	9	29	53
海口	5	13	32
重庆	7	40	60
成都	6	42	68
贵阳	10	21	47
昆明	11	31	45
拉萨	6	19	31
西安	9	48	96
兰州	18	50	79
西宁	17	37	59
银川	15	37	68
乌鲁木齐	8	42	84

表2 初始聚类中心

	聚类				
	1	2	3	4	5
Zscore: 二氧化硫年平均浓度 (ug/m ³)	-0.73884	-0.32984	-1.14784	2.32866	1.30616
Zscore: 二氧化氮年平均浓度 (ug/m ³)	0.50223	0.83116	-2.67739	1.37937	-0.04598
Zscore: 可吸入颗粒物 (PM ₁₀)年平均浓度 (ug/m ³)	-1.10043	1.46163	-1.71723	1.84120	-0.43619

表 3 迭代历史记录

迭代	聚类中心中的变动				
	1	2	3	4	5
1	0.848	0.436	0.757	0.747	0.477
2	0.062	0.287	0.000	0.220	0.211
3	0.000	0.000	0.000	0.000	0.000

表 4 聚类成员

城市	聚类	距离	城市	聚类	距离	城市	聚类	距离
北京	1	0.866	杭州	1	0.417	海口	3	0.757
天津	2	0.745	合肥	1	0.647	重庆	1	0.262
石家庄	4	0.840	福州	3	0.507	成都	1	0.647
太原	4	0.928	南昌	1	0.619	贵阳	3	0.869
呼和浩特	5	0.526	济南	4	0.659	昆明	1	1.278
沈阳	5	0.914	郑州	2	0.496	拉萨	3	0.349
长春	1	0.808	武汉	1	0.818	西安	2	0.539
哈尔滨	5	0.465	长沙	1	0.636	兰州	4	1.143
上海	1	0.882	广州	1	0.848	西宁	5	0.455
南京	1	0.707	南宁	1	1.113	银川	5	0.419
						乌鲁木齐	2	0.421

原和济南;第五类的 5 个城市分别是呼和浩特、沈阳、哈尔滨、西宁、银川。

4 结束语

城市空气质量的分析对于大气环境保护有一定的重要性,本文

主要对全国 31 个城市的空气质量进行聚类分析,分成 5 个类别。由分类结果,对于空气质量较差的城市,应找出影响空气质量的关键因素,积极采取措施从而有效防治空气污染,如转移燃煤和化工产业,产业结构调整 and 能源结构调整以减少二氧化硫、二氧化氮等有毒气体的排放,扩大绿化等。

参考文献

[1]中华人民共和国国家统计局.中国统计年鉴[M].中国统计出版社,2020.
 [2]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(03):86-96.
 [3]罗国梁.我国主要城市空气质量面板数据聚类分析[J].现代商贸工业,2014,26(07):8-9.
 [4]薛薇.统计分析与 SPSS 的应用[M].中国人民大学出版社,2011.
 [5]陈颖,张仲伍.基于聚类分析和主成分分析的城市空气质量评价——以山西省 11 个地级市为例[J].山西师范大学学报(自然科学版),2020,34(04):72-78.