

# 基于回归和聚类的青岛市空气质量分析

李雨昕, 董冉, 吴国丽

南京信息工程大学数学与统计学院, 江苏 南京

收稿日期: 2023年3月15日; 录用日期: 2023年4月21日; 发布日期: 2023年4月28日

## 摘要

本文通过研究山东省青岛市空气质量指数和空气质量相关指标, 分析得到青岛市主要污染物并对该污染物于2020年在青岛市的空分布情况进行分析。首先通过描述性统计对2016~2020年青岛市空气质量情况进行分析, 随后利用回归分析进行进一步的判断, 综合分析后得出五年间青岛市空气主要污染物类别。而后利用聚类分析以及方差分析的方法, 对主要污染物空分布进行分析。本文最终得出的结论: 近五年来, 2018年青岛市空气质量最优; 主要污染物为PM<sub>2.5</sub>。全年空气污染较严重时期为冬季; 青岛市空气污染较为严重的区域是黄岛区、市北区以及城阳区。

## 关键词

方差分析, 多元回归, 聚类分析, 空气主要污染物

# Analysis of Air Quality in Qingdao Based on Regression and Cluster

Yuxin Li, Ran Dong, Guoli Wu

School of Mathematics and Statistics, Nanjing University of Information Science & Technology, Nanjing Jiangsu

Received: Mar. 15<sup>th</sup>, 2023; accepted: Apr. 21<sup>st</sup>, 2023; published: Apr. 28<sup>th</sup>, 2023

## Abstract

This paper analyzes the main pollutants in Qingdao and the spatial distribution of the pollutants in Qingdao in 2020 by studying the air quality index and air quality related indicators. Firstly, the air quality of Qingdao from 2016 to 2020 was analyzed through descriptive statistics, and then the regression analysis was used to make further judgment. After comprehensive analysis, the main air pollutant categories in Qingdao in five years were obtained. Then, cluster analysis and variance analysis are used to analyze the spatial distribution of main pollutants. The final conclusion

of this paper is that the air quality of Qingdao in 2018 is the best in the past five years. The main pollutant is PM<sub>2.5</sub> and the serious period of air pollution throughout the year is winter. The areas with heavy air pollution in Qingdao are Huangdao District, Shibei District and Chengyang District.

## Keywords

Analysis of Variance, Multiple Regression, Cluster Analysis, Main Air Pollutants

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

随着人类社会的发展,近年来人类对环境的过度干涉急剧增长,已经使得全球大气环境发生了较为严重的变化,对人类的生存和发展带来了现实性的危害和潜在的威胁。空气质量相关问题受到世界各国专家学者的关注。

刘丽等[1]认为春节期间烟花爆竹燃放等人为活动使空气质量变差;郭保林等[2]利用 Mann-Kendall 趋势检验法分析时间序列数据的可信度和变化趋势,认为 2015~2019 年五年期间东北地区各区域的空气质量指数以及空气中颗粒物(PM<sub>2.5</sub>、PM<sub>10</sub>)、SO<sub>2</sub>、CO、NO<sub>2</sub> 浓度呈现出下降趋势;陈美如等[3]对秦皇岛市环境空气中主要污染物通过绘制折线图等描述性分析,认为近年以来秦皇岛市空气环境污染物以 PM<sub>2.5</sub> 和 O<sub>3</sub> 最为突出。周作明等[4]利用空气污染物负荷系数作为评价指标,认为湘潭市大气污染属于煤烟型污染,并且空气污染物的浓度呈现出逐年上升的趋势。Grange 和 David [5]指出时间序列的明显特征可以被气象正态化程序所阐明,使得该技术具有灵活性,适合于空气质量干预勘探的一系列潜在应用。

多元回归分析,是基于空气质量监测站点数据与其它影响因素变量,借助最小二乘法建立的多变量回归建模方法[6]。尹杰[7]利用线性回归模型模拟 PM<sub>2.5</sub> 的空间分布特征。方差分析是英国统计学家兼遗传学家 Gitbert 提出的一种分析方法,在科学试验、医学研究等众多领域有极其广泛的应用[8]。聚类分析同样广泛应用于多种领域,聚类可以分为层次聚类方法及系统聚类方法[9],它是数据挖掘中的一种经典算法,将数据分别划分为多个类,具有较高相似度的对象被划分为一类,不同类之间具有较大的差别。其中,系统聚类法是较为常用的聚类分析方法。系统聚类的中心思想是确定点和点之间的距离,根据距离确定具体分类,距离测算可采用欧式距离、平方欧式距离等;类间距离还可使用最短距离法、最长距离法等[10]。

基于上述原因,深入了解、认识空气污染情况,探索青岛市主要空气污染物,准确、可靠的预测空气污染物情况是具有十分重要的意义的。本文最终所得到的结论可为青岛市大气污染防治提供科学依据,对进一步预测、控制及治理大气污染提供参考。

## 2. 方法

### 2.1. 线性回归

多元线性回归模型  $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \epsilon, \epsilon \sim N(0, \sigma^2)$ , 定义模型的总离差平方和、回归平方和与残差平方和为:

$$SS_T = \sum_i (y_i - \bar{y})^2, SS_R = \sum_i (\hat{y}_i - \bar{y})^2, SS_E = \sum_i (y_i - \hat{y}_i)^2.$$

$$SS_T = SS_R + SS_E \quad (1)$$

其中  $\hat{y}_i$  为样本拟合值,  $\bar{y}$  为样本均值[9]。易证, 即回归模型的方差解。

## 2.2. 单因素方差分析

单因素方差分析的模型为其中  $u_i$  代表  $Y$  在因素  $A$  (共  $r$  个)

$$\begin{cases} y_{ij} = u_{ij} + \varepsilon_{ij}, i = 1, \dots, r; j = 1, \dots, n_j, \\ \varepsilon_{ij} \sim N(0, \sigma^2) \text{ 且相互独立} \end{cases}$$

水平第  $i$  个水平下的总体均值. 引入  $r-1$  个虚拟变量得到方差分析的虚拟变量回归模型[10]:

$$A_i = \begin{cases} 1, \text{水平 } i \\ 0, \text{其他} \end{cases}, i = 1, \dots, r-1,$$

$$Y = \beta_0 + \beta_1 A_1 + \dots + \beta_{r-1} A_{r-1} + \varepsilon, \varepsilon \sim N(0, \delta^2). \quad (2)$$

其中  $\beta_0 = \mu_r, \beta_i = \mu_i - \mu_r (i = 1, \dots, r-1)$ 。因素显著性可以用原假设  $H_0: \mu_1 = \mu_2 = \dots = \mu_r$  或  $H_0: \beta_1 = \dots = \beta_{r-1} = 0$  进行检验。由于因素  $A$  将数据  $Y$  分组, 定义  $Y$  的总离差平方和、组间离差平方和、组内离差平方和分别为:

$$SS_T(Y) = \sum_i \sum_j (y_{ij} - \bar{y})^2, SS_A(Y) = \sum_i \sum_j (y_i - \bar{y})^2, SS_E(Y) = \sum_i \sum_j (y_{ij} - \bar{y}_i)^2,$$

其中  $\bar{y}_i$  为水平  $i$  下的样本均值. 易证在方差分析的虚拟变量回归模型中  $\hat{y}_{ij} = \bar{y}_i$ , 故模型的回归平方和与残差平方和为:

$$SS_R = SS_A(Y), SS_E = SS_E(Y). \quad (3)$$

## 2.3. 聚类分析

聚类分析(cluster analysis)是数据挖掘中的一种经典算法, 将数据分别划分为多个类。具有较高相似度的对象被划分为同一类, 不同类之间具有较大的差别。聚类分析广泛应用于多种领域, 聚类可以分为层次聚类方法及系统聚类方法[11]。

## 2.4. Spearman 秩相关

Spearman 秩相关系数计算公式如式(4)所示。

$$r_s = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (X_j - Y_j)^2. \quad (4)$$

该式中:  $n$ ——时间周期的数量;  $X_j$ ——周期  $j$  按时间序列排序的序号;  $Y_j$ ——周期  $j$  内污染物浓度值按升序排序的序号。在 Spearman 秩相关系数分析中, 如果得到  $r_s$  的绝对值小于条件  $n$  下的临界值, 则认为所得变化趋势不具有统计意义, 反之则具有统计意义。若  $r_s$  大于 0 则认为所研究数据具有上升趋势, 反之则具有下降趋势。

## 3. 实证分析

### 3.1. 数据说明

根据我国相关政府文件, 一个城市各点位 24 小时平均浓度的算数平均值作为该城市主要污染物当日

的平均浓度值；各监测点位臭氧日最大 8 小时平均浓度值的算术平均值为该城市当日最大八小时平均臭氧浓度值。

本文数据来源为青岛市空气质量实时发布系统，数值单位为  $\mu\text{g}/\text{m}^3$  (CO 的数值单位为  $\text{mg}/\text{m}^3$ )。其中每日 AQI 数据和主要污染物浓度数据为当日各观测点 24 小时全部数据计算平均值的结果，符合国家相关文件统计方法要求并且保证数据准确性。本文研究所用数据在数据预处理过程中发现少量缺失值，因缺失值极少故采取剔除缺失值的处理方式，取均值后最终得到 1806 条有效数据。

## 3.2. 结果分析

### 3.2.1. 空气污染物变化趋势分析

为大致了解 2016 年~2020 年青岛市空气质量变化情况，求得青岛市 2016~2020 年各年 AQI 值及主要污染物浓度值均值，绘制折线图如图 1。可以看出，5 年来青岛市 AQI 值及各个主要污染物浓度的年均值均在 2018 年达到最低，一定程度上可以认为五年中 2018 年青岛市的空气质量最优；2020 年六类污染物浓度及 AQI 值均低于 2016 年，可一定程度认为 2020 年青岛市空气质量较 5 年前有所改善；五年间空气中  $\text{SO}_2$  浓度呈现逐年降低的较好趋势，2018 年后下降速度变缓；AQI 值、 $\text{NO}_2$ 、 $\text{O}_3$ -8h、 $\text{PM}_{10}$ 、 $\text{PM}_{2.5}$  的浓度总体呈现稳定，虽然有一定的下降趋势，但是其下降趋势不如空气中  $\text{SO}_2$  的浓度下降的趋势明显。

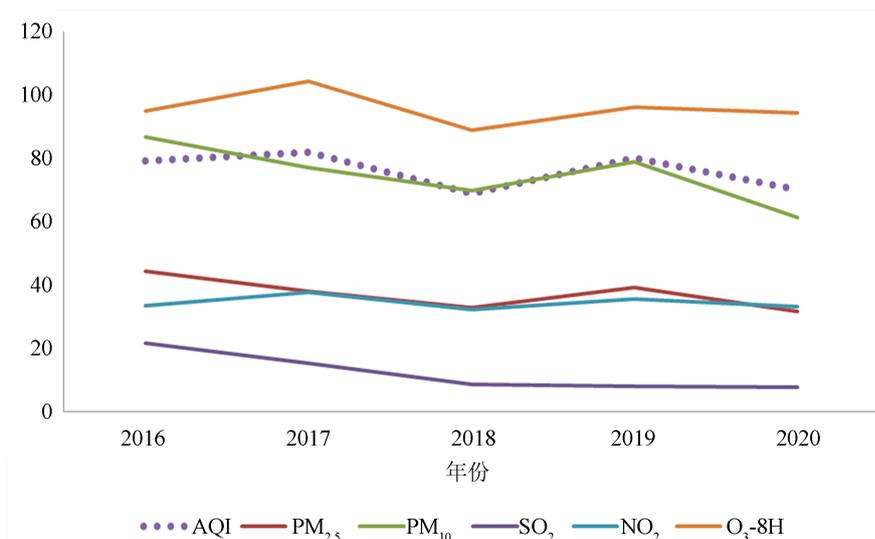


Figure 1. Change trend of AQI and major air pollutant concentrations in Qingdao from 2016 to 2020  
图 1. 2016~2020 年青岛市 AQI 及主要空气污染物浓度变化趋势图

进一步分析 2016~2020 年青岛市空气质量的变化趋势，当时间周期为 5 时，统计量的临界值  $\gamma = 0.9$ 。根据公式计算所得近五年青岛市 6 类环境评价监测指标的 Spearman 秩相关系数如下表 1 所示：

Table 1. Calculated Spearman Rank correlation coefficients for six monitoring indicators

表 1. 计算所得 6 类监测指标的 Spearman 秩相关系数表

指标	AQI	$\text{NO}_2$	$\text{O}_3$ -8h	$\text{PM}_{10}$	$\text{PM}_{2.5}$	CO	$\text{SO}_2$
$r_s$	-0.3	-0.3	-0.3	-0.7	-0.7	-0.7	-1

计算所得青岛市近 5 年来 6 类指标的 Spearman 秩相关系数如表 1 所示，可以看出， $\text{SO}_2$  的  $r_s = -1$ ， $|r_s| > \gamma$  (0.9)；AQI、 $\text{NO}_2$ 、 $\text{O}_3$ -8h 的  $r_s = -0.3$ ， $|r_s| < \gamma$  (0.9)；CO、 $\text{PM}_{10}$ 、 $\text{PM}_{2.5}$  的  $r_s = -0.7$ ， $|r_s| < \gamma$  (0.9)。

故得到以下结论：2016~2020年青岛市空气中SO<sub>2</sub>浓度呈显著下降趋势；空气中NO<sub>2</sub>、O<sub>3</sub>-8h的浓度有下降趋势，但并不显著；CO、PM<sub>10</sub>、PM<sub>2.5</sub>的浓度有下降趋势同样不显著，但其下降趋势相较于NO<sub>2</sub>、O<sub>3</sub>-8h而言，是显著的；另外，从AQI角度来看，5年来青岛市空气质量有提升趋势但并不显著。

综上所述，通过计算Spearman秩相关系数的方法对青岛市2016~2020年5年来空气质量变化趋势进行量化分析得到的结果与上节描述性分析所得的结论大致相同，可以得到以下结论：2016~2020年间青岛市空气质量得到较大幅度的改善；空气中SO<sub>2</sub>的浓度呈现出显著的下降趋势；除SO<sub>2</sub>以外的五种主要监测空气污染物的浓度均有所下降，但该趋势并不显著。

### 3.2.2. 空气主要污染物分析

因为空气质量状况与六类污染物的浓度状况全部相关，故本节利用Eviews软件构建多元线性回归模型，目的在于判断哪一种空气污染物的浓度变化更能影响当日的空气质量，影响越大，则认为该污染物为主要污染物。此处利用AQI值刻画每日空气质量状况。另外，由于研究过程中CO的浓度与其余五类污染物的单位不同，且AQI是无单位的，故本节构建多元线性回归模型的过程中将全部数据进行标准化处理，变成无量纲形式进行分析。将六类污染物全部引入作为自变量，AQI值作为因变量，依据最小二乘原理构建多元线性回归模型，输出结果如下表2：

Table 2. Multiple linear regression results under least squares

表 2. 最小二乘下多元线性回归结果

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.75E-09	0.007878	-6.03E-07	1.0000
PM <sub>10</sub>	0.415423	0.019781	21.00160	0.0000
PM <sub>2.5</sub>	0.445338	0.026041	17.10143	0.0000
SO <sub>2</sub>	-0.075324	0.010772	-6.992737	0.0000
NO <sub>2</sub>	-0.016023	0.013153	-1.218201	0.2233
CO	0.155178	0.023431	6.622656	0.0000
O <sub>3</sub>	0.338091	0.008276	40.85060	0.0000
R-squared	0.888297	Mean dependent var		-9.97E-08
Adjusted R-squared	0.887924	S.D. dependent var		1.000000
S.E. of regression	0.334778	Akaike info criterion		0.653168
Sum squared resid	201.6248	Schwarz criterion		0.674481
Log likelihood	-582.8104	Hannan-Quinn criter.		0.661034
F-statistic	2384.357	Durbin-Watson stat		1.120094
Prob (F-statistic)	0.000000			

表2中可以看出F统计量的值为2384.357，P值远小于0.05，所以认为模型通过检验，显著有效；观察每个变量的t检验结果发现，NO<sub>2</sub>的t统计量值为-1.218，P值大于0.05，故该自变量不显著。调整后的R<sup>2</sup>约为0.888，大于0.7，可以认为模型拟合效果比较好。接下来检验模型是否符合模型设立的各个假设。首先判断是否符合各自变量独立的假设，检验多重共线性。若存在多重共线性，则有可能导致自变量的t检验失去意义、变量的系数值的含义不合理等情况，应考虑消除多重共线性。

多重共线性检验结果如下，此处根据方差膨胀系数VIF作为判断标准。由表3可以看出，自变量PM<sub>2.5</sub>的VIF值大于10，所以是存在多重共线性的。所以该模型需要进行进一步的完善处理消除多重共线性。

**Table 3.** Results of multicollinearity test**表 3.** 多重共线性检验结果

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	6.21E-05	1.000000	NA
PM <sub>10</sub>	0.000391	6.301448	6.301448
PM <sub>2.5</sub>	0.000678	10.92140	10.92140
SO <sub>2</sub>	0.000116	1.868684	1.868684
NO <sub>2</sub>	0.000173	2.786391	2.786391
CO	0.000549	8.842256	8.842256
O <sub>3</sub>	6.85E-05	1.103148	1.103148

首先利用向前逐步回归的方法再次构建模型，试图消除多重共线性，然而结果并不理想，仍然存在 VIF 大于 10 的自变量，所以又考虑到删除变量消除共线性。删除哪一个变量是当下面对的问题，经过上面的表可以看出 PM<sub>2.5</sub> 的 VIF 较大，可以删除 PM<sub>2.5</sub>。然而，根据 PM<sub>2.5</sub> 和 PM<sub>10</sub> 的定义我们可知，PM<sub>10</sub> 包含 PM<sub>2.5</sub> 且 PM<sub>10</sub> 中接近 70% 为 PM<sub>2.5</sub>，故此处删除变量 PM<sub>10</sub>，再次拟合回归模型。

**Table 4.** Regression model after reducing variables**表 4.** 减少变量后的回归模型

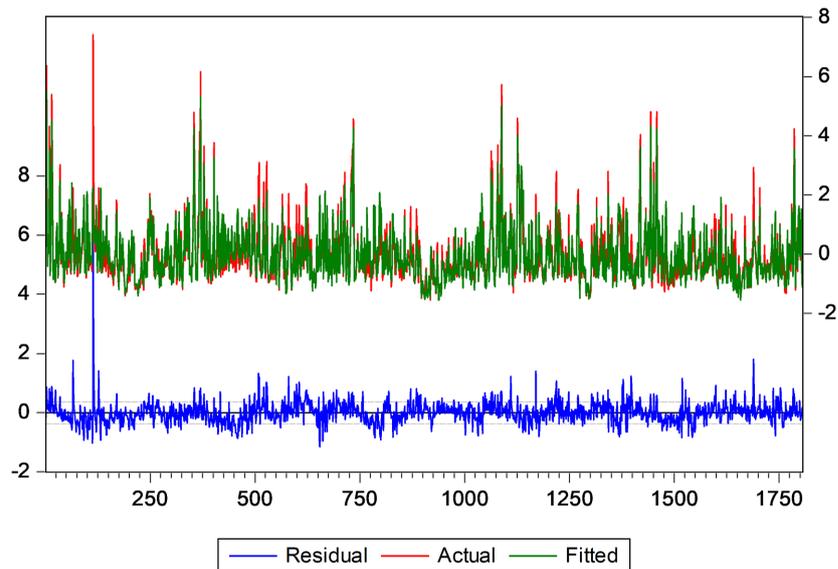
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.39E-08	0.008788	6.13E-06	1.0000
PM <sub>2.5</sub>	0.818610	0.021232	38.55573	0.0000
O <sub>3</sub>	0.368693	0.009088	40.56706	0.0000
NO <sub>2</sub>	0.082787	0.013703	6.041668	0.0000
SO <sub>2</sub>	-0.049559	0.011938	-4.151274	0.0000
CO	0.070929	0.025753	2.754168	0.0059
R-squared	0.860910	Mean dependent var		-9.97E-08
Adjusted R-squared	0.860523	S.D. dependent var		1.000000
S.E. of regression	0.373466	Akaike info criterion		0.871335
Sum squared resid	251.0578	Schwarz criterion		0.889604
Log likelihood	-780.8157	Hannan-Quinn criter.		0.878078
F-statistic	2228.249	Durbin-Watson stat		1.129242
Prob (F-statistic)	0.000000			

根据表 4 可以看出模型通过了 F 检验，模型显著有效；五个自变量均通过了 t 检验，对空气质量水平有显著影响；调整后的 R<sup>2</sup> 为 0.86，大于 0.7，故模型拟合程度也比较好。根据表 5 判断是否存在多重共线性，发现所有变量的 VIF 均小于 10，故多重共线性得以消除。

接下来对误差项进行检验，判断误差项是否符合正态分布，即是否符合误差正态分布的假定。根据图 2 残差图中下方的折线图可以看出，残差并无较明显的变化趋势，残差服从正态分布，符合模型假定。

**Table 5.** Multiple collinearity test results after removing variables  
**表 5.** 去除变量后多重共线性检验结果

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	7.72E-05	1.000000	NA
PM <sub>2.5</sub>	0.000451	5.833811	5.833811
O <sub>3</sub>	8.26E-05	1.068953	1.068953
NO <sub>2</sub>	0.000188	2.429886	2.429886
SO <sub>2</sub>	0.000143	1.844446	1.844446
CO	0.000663	8.583081	8.583081



**Figure 2.** Error chart  
**图 2.** 误差图

最后，对模型进行基于怀特检验异方差性检验。结果如表 6，可以看出 P 值均小于 0.05，所以拒绝原假设，认为存在异方差性，不符合异方差假设，所以对模型进行进一步修正。修正后模型表 7。

**Table 6.** Heteroscedasticity test results  
**表 6.** 异方差检验结果

Heteroskedasticity Test: White			
F-statistic	10.71216	Prob. F (20, 1785)	0.0000
Obs*R-squared	193.5348	Prob. Chi-Square (20)	0.0000
Scaled explained SS	3989.166	Prob. Chi-Square (20)	0.0000

根据模型修正后的结果可以看出，五个变量均可以通过 t 检验，且模型通过 F 检验，模型拟合程度较好。表 6 中发现自变量 PM<sub>2.5</sub> 的系数接近于 0.818，大于 0 且是五个变量中系数最大的变量，故可以认为五个自变量中 PM<sub>2.5</sub> 对于 AQI 值变化影响最大，并且其浓度越高，AQI 值越大，空气污染越严重。一定程度可以认为 PM<sub>2.5</sub> 是主要污染物。

**Table 7.** Model heteroscedasticity correction results  
**表 7.** 模型异方差修正结果

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000282	0.000222	1.268840	0.204662
CO	0.071973	0.001326	54.277637	0.000000
NO <sub>2</sub>	0.083084	0.000632	131.436029	0.000000
PM <sub>2.5</sub>	0.817941	0.001666	490.861031	0.000000
SO <sub>2</sub>	-0.049558	0.000355	-139.516537	0.000000
O <sub>3</sub>	0.369095	0.000291	1270.012330	0.000000
R-squared	0.999661		Mean dependent var	0.000229
Adjusted R-squared	0.999660		S.D. dependent var	1.082747
S.E. of regression	0.045037		Akaike info criterion	-3.359353
Sum squared resid	3.650979		Schwarz criterion	-3.341084
Log likelihood	3039.495548		Hannan-Quinn criter.	-3.352610
F-statistic	1062237.174192		Durbin-Watson stat	1.263452
Prob (F-statistic)	0.000000		Weighted mean dep.	-0.046793

### 3.2.3. 主要空气污染物空间分布分析

经过上一节的研究可以知道,近些年来青岛市主要空气污染物为颗 PM<sub>2.5</sub>, 本届主要研究 PM<sub>2.5</sub> 在青岛市的时空分布,判断青岛市哪一个区域污染更加严重。本节所研究的数据为 2020 年全年青岛市九个观测点逐时观测数据取平均后得到的每日 PM<sub>2.5</sub> 浓度以及 AQI 值。本文选七个观测站点进行分析,分别是:城阳区子站、李沧区子站、四方区子站、崂山区子站、市南区子站、市北区子站和黄岛区子站。分析过程中,根据国控监测点名称以及位置作为一个行政区的空气污染浓度,且由于老四方区是老工业集中区,具有较强的研究意义,故此处研究时暂不将其归于市北区的空气污染浓度。

利用青岛市七个观测站点 2020 年每日平均 PM<sub>2.5</sub> 浓度的数据做聚类分析,目的是将含有相关关系的站点进行合并,达到数据降维的目的。分类情况整理如表 8 所示。

**Table 8.** System clustering results  
**表 8.** 系统聚类结果

类别	站点名称
第一类	市南区站点、崂山区站点
第二类	市北区站点、四方区站点、黄岛区站点
第三类	城阳区站点
第四类	李沧区站点

为判断上小节分成的四类观测点的数值之间是否存在显著差异,本小节拟采用方差分析的方法进行进一步的判断。为判断数据是否能够进行方差分析,要判断其是否满足方差分析的假定,故首先对数据进行齐次性检验。由于数据原始分布未知,故此处采用的是列文检验的方法。根据上表可以看出,Levene 统计量值为 1.508,显著性为 0.211,显著性大于 0.05,故拒绝原假设,认为数据方差齐次,进行方差分析。

**Table 9.** Information table of analysis of variance results**表 9.** 方差分析结果信息表

	平方和	df	均方	F	显著性
组之间	10124.660	3	3374.887	3.509	0.015
组内	2248961.622	2338	961.917		
总计	2259086.281	2341			

表 9 是方差分析所得到的结果,可以看出方差分析的 P 值为 0.015,小于 0.05,故在显著性为 0.05 的水平下,认为所研究四类别之间存在显著差异,即各个观测站点分组后的新类别之间至少两组之间存在显著差异,但具体差异的组别并不知晓,故进行多重比较,此处采用最小显著差异法进行比较,结果如表 10。

**Table 10.** Multiple comparison results based on LSD method**表 10.** 基于 LSD 方法下的多重比较结果

(I) type	(J) type	平均差(I - J)	标准错误	显著性	下限值	上限值
1	2	4.73627*	1.52429	0.002	1.7472	7.7254
	3	1.21723	1.92810	0.528	-2.5637	4.9982
	4	0.06817	2.11326	0.974	-4.0759	4.2122
2	1	-4.73627*	1.52429	0.002	-7.7254	-1.7472
	3	-3.51904	2.04505	0.085	-7.5294	0.4913
	4	-4.66810*	2.22048	0.036	-9.0224	-0.3138
3	1	-1.21723	1.92810	0.528	-4.9982	2.5637
	2	3.51904	2.04505	0.085	-0.4913	7.5294
	4	-1.14905	2.51488	0.648	-6.0807	3.7826
4	1	-0.06817	2.11326	0.974	-4.2122	4.0759
	2	4.66810*	2.22048	0.036	0.3138	9.0224
	3	1.14905	2.51488	0.648	-3.7826	6.0807

\*表示该值通过 95%显著性检验。

可以清楚地看出每一类别与其余三个类别之间进行两两比较的结果,根据表中标示可以得到结论:在 0.05 显著性水平下第二类别和第一类别、第二类别和第四类别之间存在显著的差异,第三类别与其他类别均无显著差异。

根据表 9 可以看到,第二组是市北区站点、四方区站点、黄岛区站点;第一组是崂山区站点和市南区站点;根据表 10 可以看出第二组三个站点污染程度高于轻度污染的比例相对较高,一定程度认为该三个区域  $PM_{2.5}$  浓度相对较高,污染较为严重,第一组和第四组均与第二组存在显著差异,故认为市南区站点、崂山区站点以及李沧区站点所监测范围污染程度较低;而又因为第三组与第二组无差异,故认为李沧区站点所监测范围污染同样较为严重。

#### 4. 结论

2016~2020 年青岛市总体空气质量较好,空气质量水平高于山东省平均水平。六类监测污染物中,空气中  $SO_2$  的浓度连续 5 年下降,下降趋势最为显著;其余五类空气污染物浓度总体呈现下降趋势,但

趋势并不明显。对该现象进行分析认为空气中  $\text{SO}_2$  浓度的持续走低与青岛市政府近年来大力推动清洁能源采暖实施密不可分。5年中2018年空气质量最优,对原因进行分析认为除了长期的防治举措以外,2018年大型会议举办期间,部分地区停工停产措施对当年空气质量改善起到十分重要的作用。

通过2016~2020年青岛市六类监测空气污染物的日均浓度以及每日AQI值进行可视化分析,得到如下结论:主要空气污染物和AQI值都呈现出来了明显的季节特征, $\text{SO}_2$ 、 $\text{NO}_2$ 、CO、颗粒物( $\text{PM}_{10}$ 、 $\text{PM}_{2.5}$ )日平均浓度的变化趋势与AQI值变化趋势相同,在冬季较高,夏季较低; $\text{O}_3$ 的日平均浓度的变化特征与上述五种污染物相反,呈现出夏季和秋季明显高于冬季和春季的特点。

本文通过描述性分析、回归分析的分析方法最终得到2016~2020年青岛市最主要污染物为颗粒物( $\text{PM}_{10}$ 、 $\text{PM}_{2.5}$ )。对青岛市主要污染物即颗粒物进行空间分析,最终认为四方区、市北区、城阳区、黄岛区站点所监测范围的污染程度较严重,市南区、崂山区站点以及李沧区站点所监测范围污染相对不严重。一定程度反映各个行政区域受颗粒物污染程度。

## 参考文献

- [1] 刘丽,唐晓婵,马东卓,等. 2013-2019年青岛市空气质量变化特征分析[J]. 中国环境监测, 2020, 36(6): 66-71.
- [2] 郭保林,周开斌,孙娅,等. 东北空气质量时空变化特征及变化趋势研究[J]. 资源节约与环保, 2021(3): 20-23.
- [3] 陈美如,张文婧,王爽,等. 秦皇岛市2015~2019年空气污染物变化趋势与防治对策[J]. 河北科技师范学院学报, 2020, 34(3): 68-73.
- [4] 周作明,荆国华,徐欣. 湘潭市环境空气质量变化趋势分析及对策[J]. 四川环境, 2005(5): 27-29.
- [5] Grange, S.K. and Carslaw, D.C. (2019) Using Meteorological Normalisation to Detect Interventions in Air Quality Time Series. *The Science of the Total Environment*, **653**, 578-588. <https://doi.org/10.1016/j.scitotenv.2018.10.344>
- [6] Gilliland, F., Avol, E., Kinney, P., et al. (2005) Air Pollution Exposure Assessment for Epidemiologic Studies of Pregnant Women and Children: Lessons Learned from the Centers for Children Environmental Health and Disease Prevention Research. *Environmental Health Perspectives*, **113**, 1447-1454. <https://doi.org/10.1289/ehp.7673>
- [7] 尹杰. 重庆市主城区PM2.5的时空格局及影响因素分析[D]: [硕士学位论文]. 重庆: 重庆师范大学, 2019.
- [8] 薛定宇,陈阳泉. 基于MATLAB/Simulink的系统仿真技术与应用[M]. 北京: 清华大学出版社, 2002: 50-79.
- [9] 唐振,于文. 基于系统聚类的环境质量对城镇居民医疗支出的影响分析[J]. 统计学与应用, 2014, 3(2): 76-84.
- [10] 夏帆. 虚拟变量回归与t检验等效性的探讨[J]. 统计研究, 2008, 25(10): 109-110.
- [11] 甘伦知. 虚拟变量回归与方差分析的联系[J]. 统计与决策, 2011(8): 159-160.