

# 2014-2020 年长沙地区空气质量与气象因素的相关性分析

唐亚周 519021910804

电子信息与电气工程学院 计算机科学与技术专业

**摘 要** 一直以来, 空气质量问题都是我国城市化发展过程中亟待解决的难题。长沙市作为一座快速发展的新一线城市, 空气质量问题自然十分关键。本文以天为单位, 通过分析以 PM2.5 为代表的空气质量指数和当天的天气状况、气温、风速等指标的关系, 得出一系列结论。

**关键词** 空气质量, 长沙, 天气, 量化分析

## 目录

|                              |          |
|------------------------------|----------|
| <b>1 研究背景</b>                | <b>3</b> |
| <b>2 前期准备</b>                | <b>3</b> |
| 2.1 文献调研 . . . . .           | 3        |
| 2.2 数据搜集及处理 . . . . .        | 3        |
| <b>3 数据分析</b>                | <b>3</b> |
| 3.1 空气质量指数与温度的关系 . . . . .   | 3        |
| 3.2 空气质量指数与天气状况的关系 . . . . . | 4        |
| 3.3 空气质量指数与风速的关系 . . . . .   | 5        |
| <b>4 结论与展望</b>               | <b>5</b> |
| <b>5 公式</b>                  | <b>5</b> |
| <b>6 结束</b>                  | <b>5</b> |
| <b>A Python 爬虫代码</b>         | <b>6</b> |
| <b>B Python 数据分析代码</b>       | <b>6</b> |

## 1 研究背景

近期召开的十九届五中全会审议通过的《中共中央关于制定国民经济和社会发展第十四个五年规划和二〇三五年远景目标的建议》(以下简称《建议》),从多个方面对生态文明建设和生态环境保护作出重要部署、提出明确要求,为做好“十四五”生态环境保护工作指明了前进方向、提供了根本遵循。2021 年是我国现代化建设进程中具有特殊重要性的一年,也是“十四五”时期的开局之年,要立足新发展阶段,坚持新发展理念,构建新发展格局,把生态文明建设和生态环境保护作为推动高质量发展的题中应有之义,进一步凸显其重要地位和关键作用,抓紧谋划“十四五”生态环境保护工作,坚持稳中求进工作总基调,对标对表 2035 年远景目标,牢固树立落实绿水青山就是金山银山的理念,促进经济社会发展全面绿色转型,加快推动绿色低碳发展,持续改善环境质量,提升生态系统质量和稳定性,全面提高资源利用效率,更加注重系统观念在生态环境保护工作中的科学运用和实践深化,突出精准治污、科学治污、依法治污,更好发挥生态环境保护对高质量发展、构建新发展格局的支撑服务保障作用,为开启全面建设社会主义现代化国家新征程、向第二个百年奋斗目标进军奠定坚实基础。<sup>[1]</sup>

在这样的大背景下,借这次概率统计课程大作业的机会,本人决定对家乡长沙的空气质量进行一次研究。

## 2 前期准备

### 2.1 文献调研

经过查阅资料<sup>[2]</sup>得知,空气质量的主要影响因素有局地污染和西北地区沙尘传输造成的自然降尘、气象要素(降水量、风速、逆温等)、地形的空间差异和人类活动等多个方面。考虑到长沙市的地理特征、产业结构以及数据收集的难易程度,我选择对气象因素和空气质量的关系进行研究。

### 2.2 数据搜集及处理

本人在网络<sup>[3]</sup>上搜集到 2014 年到 2020 年长沙市的空气质量指数,再利用 python 爬虫在网络<sup>[4]</sup>上获取 2014 年到 2020 年长沙市的气象要素数据。然后去除数据集中的无效数据,得到 2531 个有效数据。

## 3 数据分析

### 3.1 空气质量指数与温度的关系

收集到的数据中,有当日最高温度和最低温度两个数据,将这两个数据求平均值处理,得到的值作为一天的温度。以年份为单位,将温度及 PM2.5 和日期的关系进行可视化,如图 1 所示

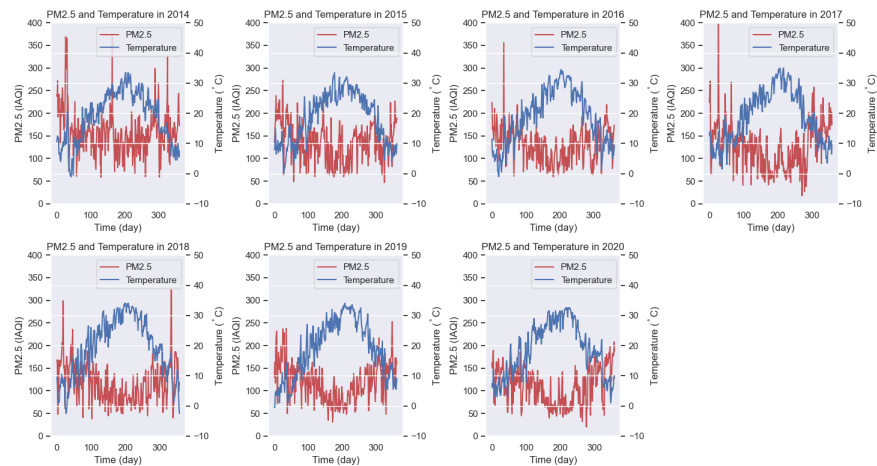


图 1: 2014-2020 年长沙地区各年 PM2.5 指数及气温与日期的关系图

可以看出，总体上 PM2.5 指数与气温呈负相关。然后对于 PM2.5 指数和气温绘制二维直方图，如图 2所示，并使用 Python 内置函数计算相关系数。

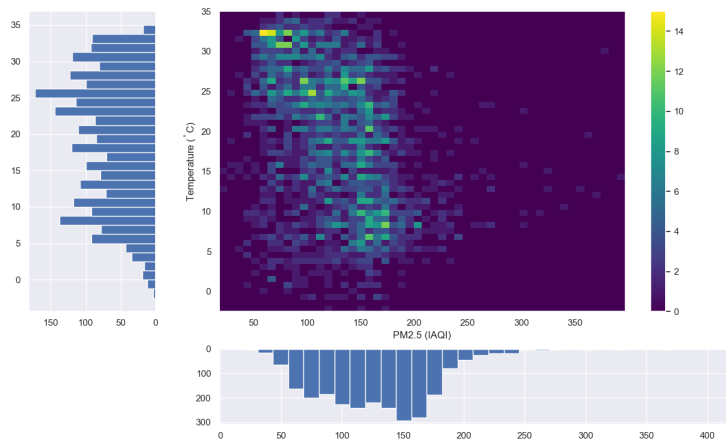


图 2: 2014-2020 年长沙地区 PM2.5 指数与气温的分布图

$$\rho_{XY} = \frac{\text{cov}(X,Y)}{\sqrt{D(X) \cdot D(Y)}} = -0.35$$

(1)

由此看出，PM2.5 指数与温度呈一定的负相关，但相关性并不明显。

3.2 空气质量指数与天气状况的关系

通过遍历数据，我们得到了 7 年来长沙地区出现过的所有天气情况，并按照降雨量/降雪量的大小将其分级，如表??所示。

将数据中每天的两个天气信息及其对应的级别均记录下来，也就是说一个空气质量的信息对应着两个天气状况信息。得到 1 级天气 2568 次、2 级天气 544 次、3 级天气 1592 次、四级天气 358 次。

表 1: 天气状况及人为分级

| 天气情况  | 级别 | 分级依据          |
|-------|----|---------------|
| 晴     | 1  | 一般为干燥         |
| 多云    | 1  | 一般为干燥天气       |
| 阴     | 2  | 一般湿度比较大，但没有降水 |
| 小雨    | 3  |               |
| 小到中雨  | 3  | 一般降水量较少       |
| 阵雨    | 3  | 一般降水量较少       |
| 雷阵雨   | 3  | 一般降水量较少       |
| 雨     | 3  | 一般降水量较少       |
| 中雨    | 4  | 一般降水量较大       |
| 中到大雨  | 4  | 一般降水量较大       |
| 大雨    | 4  | 一般降水量较大       |
| 大到暴雨  | 4  | 一般降水量较大       |
| 暴雨    | 4  | 一般降水量较大       |
| 冻雨    | 3  | 一般降水量较少       |
| 雨夹雪   | 3  | 一般降水量较少       |
| 小雪    | 3  | 一般降水量较少       |
| 小雪-中雪 | 3  | 一般降水量较少       |
| 中雪    | 4  | 一般降水量较大       |
| 中到大雪  | 4  | 一般降水量较大       |
| 大雪    | 4  | 一般降水量较大       |

对四种天气下的 PM2.5 指数做统计，得到直方图如图??所示。

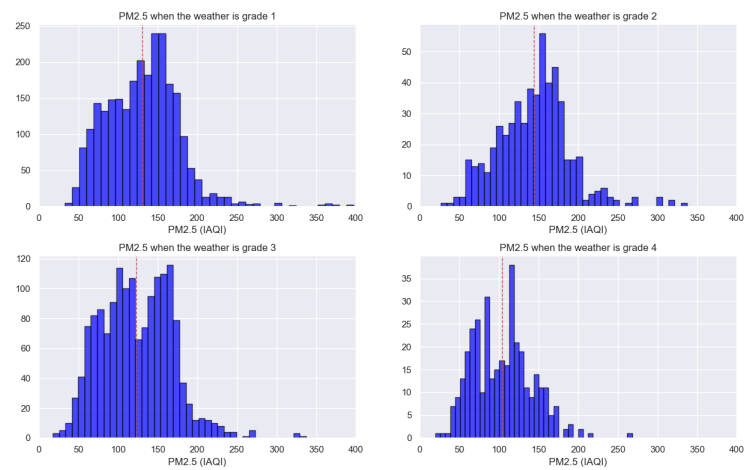


图 3: 2014-2020 年间长沙地区 4 种天气情况下的 PM2.5 指数分布直方图

由  $E(X) = \sum_{i=1}^n x_i p_i$  和  $D(X) = \sum_{i=1}^n [x_i - E(X)]^2 p_i$  可计算出四种天气下 PM2.5 指数的期望和方差，如表??所示。

表 2: 各级别天气下 PM2.5 指数的样本期望和方差

| 天气级别 | 期望     | 方差      |
|------|--------|---------|
| 级别 1 | 130.87 | 2032.00 |
| 级别 2 | 143.81 | 2051.42 |
| 级别 3 | 123.16 | 1993.46 |
| 级别 4 | 104.01 | 1392.35 |

表 3: 各级别天气下中度污染及以上天数的样本期望和数量

| 天气级别 | 期望     | 出现次数 | 占总数比例  |
|------|--------|------|--------|
| 级别 1 | 177.79 | 867  | 33.67% |
| 级别 2 | 180.49 | 253  | 46.51% |
| 级别 3 | 175.75 | 469  | 29.46% |
| 级别 4 | 169.62 | 43   | 12.01% |

可以看到, 总体上来说, 当天气的级别越高, 也就是降水量越大、空气越湿润时, PM2.5 指数总体上越小。另外, 我们可以按照国家标准<sup>[7]</sup>里所规定的”中度污染“以上的标准, 也就是 PM2.5 的 IAQI 值大于等于 151 的情况下, 对四种级别的天气进行统计, 如图??和表??所示。可以看到, 总体上来说, 当天气的级别越高, 出现”中度污染“及以上的比例要减小。但是在级别 2 的天气(阴天)时, PM2.5 指数的期望要比级别 1(晴天和多云)要大, 出现”中度污染“及以上的次数比例也要高于级别 1。综上所述, PM2.5 指数在不同天气下从高到低排序应该是  $2 > 1 > 3 > 4$ 。

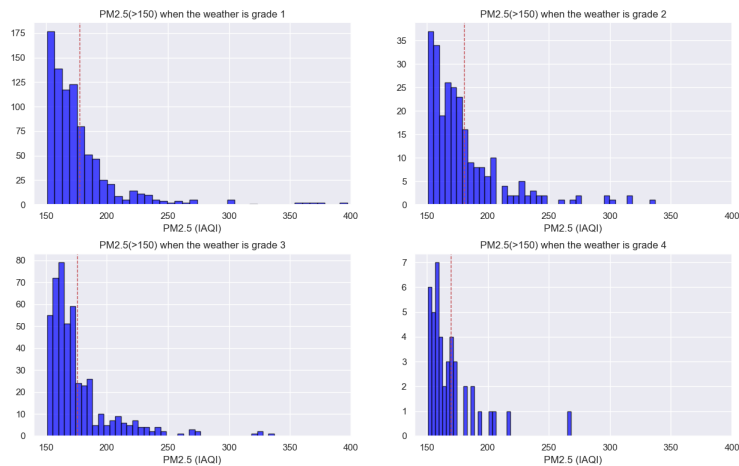


图 4: 2014-2020 年间长沙地区 4 种天气情况下的 PM2.5 指数分布直方图(中度污染及以上)

3.3 空气质量指数与风速的关系

考虑到风力等级与风速并非线性关系, 直接将一天中的两个风力等级数据取平均值作为一天的风力等级并不合理。故同样选择将数据中每天的两个风力等级均记录下来, 也就是说一个空气质量的信息对应着两个风力等级信息。得到 1 级风 779 次, 2 级风 779 次, 3 级风 3110 次, 4 级风 363 次, 5 级风 28 次, 6 级风 3 次。然后按照 3.2 节中相同的处理方式, 如表??和图??所示。



- 菩提本无树，明镜亦非台
- 本来无一物，何处惹尘埃.

| Setting        | A4 size paper |        |
|----------------|---------------|--------|
|                | mm            | inches |
| Top            | 25            | 1.0    |
| Bottom         | 25            | 1.0    |
| Left           | 20            | 0.8    |
| Right          | 20            | 0.8    |
| Column Width   | 82            | 3.2    |
| Column Spacing | 6             | 0.25   |

表 5: a table

$$\alpha \geq \delta + \overline{\eta * \beta}$$

$$\left[\begin{array}{cccc} a_1 & a_2 & \alpha & \beta \\ \chi & \varphi & \gamma & \eta \\ \theta & \zeta_3 & \xi & \omega \end{array}\right]$$

图 6: aaa

参考文献

[1] 中华人民共和国生态环境部. 生态环境部举行党的十九届五中全会精神辅导报告会 [EB/OL].<https://www.mee.gov.cn/xxgk/xxgk/xxgk15/202012/t20201212812690.html>,2020-12-12.

[2] 李小飞, 张明军, 王圣杰, 赵爱芳, 马潜. 中国空气污染指数变化特征及影响因素分析 [J]. 环境科学,2012,33(06):1936-1943.

[3] The World Air Quality Project.Air Quality Historical Data Platform[DB/OL].<https://aqicn.org/data-platform/register/>,2020-12-26.

[4] 天气网. 历史天气查询 [DB/OL].<http://lishi.tianqi.com/>,2020-12-26.

[5] HJ 633—2012, 环境空气质量指数（AQI）技术规定（试行）[S]. 中国: 中华人民共和国生态环境部,2012.

A Python 爬虫代码

some text in Appendix A

B Python 数据分析代码

some text in Appendix B