# 对西安近3年空气质量分析

## 概述

水、空气、食物和环境是生命存活的四大条件。其中尤为重要的是空气。成年人平均每天需吸入8立方米，大概10公斤的空气，是食物和水需求量的几倍。空气通过肺部与人体组织进行交换，空气中的污染物可以直接进入肺部、然后进入血液，导致呼吸、心血管、神经等系统疾病或其他疾病，严重时还可引发肺心疾病、癌症等。

我们每时每刻都在呼吸，一个人每天要呼吸两万多次，每天至少要与环境交换一万多升气体，可见空气质量的好坏与人的健康息息相关。本篇文章通过对西安近五年的空气质量指数（AQI）进行分析，找到在不同的四个季度西安空气质量的联系，并进一步对于西安空气质量指数进行参数估计，从而得到西安近5年来空气质量的发展趋势和大致水平。

## 研究背景

空气是地球表面一切有生命的物质赖以生存的基本条件。如果没有空气，人类的生存及其社会活动就无法维持下去，植物的光合作用不能进行，其它生物也不复存在。所以，当大气遭受污染之后， 其成分、性质都发生了改变，这势必会对人体健康、动植物生长生活以及生态平衡乃至各种器物的存放产生有害的影响。近年来，随着城市工业的发展，大气污染日益严重，空气质量进一步恶化，不仅危害到人们的正常生活，而且威胁着人们的身心健康。我国11个最大城市中，空气中的烟尘和细颗粒物每年使40万人感染上慢性支气管炎。在一定程度上，城市生活正在背离人们所追求的健康目标。呼吸道疾病，温室效应，臭氧层破坏，酸雨，PM2.5等等这些名词频繁地出现在我们的日常生活中。当下对大气污染的深刻认识和保护对策的思考变得刻不容缓。如今，我们有了量化空气质量的标准----空气质量指数AQI，并且也能够对于空气中的每种污染物进行量化检测，从这些数据的检测中能够判断出西安目前空气质量的变化趋势，分析出西安空气质量指数的置信区间和空气质量指数分布的各个参数，在这篇文章中采用西安近年的空气质量指数，来分析西安空气质量的变化趋势。

## 问题研究

首先对于空气质量指数进行研究，设随机变量为第i年第j个月第k天的空气质量指数（2019=<i<=2021，1=<j<=12，1=<k<=31），近似地认定一个月内每天的空气质量指数服从同一分布，再在下文中证明相邻两天的空气质量指数互相独立。

说明相邻两天空气质量指数互相独立，只需要说明即可。在这里采用2021年9月作为例子。两者的样本方差作为方差的估计值。

取k从1到30，则：

则根据公式可以计算得出这里的相关系数.

由计算得，相邻两天的空气质量指数是正相关的，但并不是相互独立的，而且两者之间是弱相关。但是为了更方便的统计，近似认为相邻两天的空气质量指数是互相独立的，则可以推理出每一天的空气质量指数都可以看做是互相独立的。

设，则在一个月内，随机变量互相独立。

根据独立同分布的中心极限定理（n取决于每个月的天数）

设，，为每个月的空气质量指数的平均值，得出。

的方差是不一样的，但是当我们得出了服从正态分布这个结果后，就不用去关心，因为一天空气质量的好与坏的分布准确性不高，而每个月的空气质量指数平均值更有统计意义。

接下来将一年分为4个季度进行估计。

采用矩估计法进行点估计，得到以下表格。

对于，可证明这个估计量具有无偏性，一致性和有效性，对于这个估计量，可以证明这个估计量不具有无偏性，也不具有有效性，但这个估计量具有一致性。

|  |  |  |
| --- | --- | --- |
| 2019年 |  |  |
| 第一季度 | 73.9 | 183.6 |
| 第二季度 | 78.6 | *40.9* |
| 第三季度 | 65.8 | *198.4* |
| 第四季度 | 75.6 | *271.4* |
| 2020年 |  |  |
| 第一季度 | 42.3 | 4.2 |
| 第二季度 | 76.2 | *11.5* |
| 第三季度 | 83.6 | *287.4* |
| 第四季度 | 75.1 | *320.1* |
| 2021年 |  |  |
| 第一季度 | 74.5 | 98.4 |
| 第二季度 | 71.2 | *5.4* |
| 第三季度 | 42.5 | *72.3* |
| 第四季度 | 41.2 | *8.2* |

通过统计表格可以得知，空气质量指数在第三季度和第四季度比较低，在第第二季度和第一季度较高。

观察方差的情况可以观察出一个现象，就是在第二季度空气质量指数相对稳定，我认为可能是因为西安在夏天属于暖温带半湿润大陆性季风气候，夏天降水相较于其他几个季节降水量比较多，由于降雨量相对稳定，导致第二季度空气质量相对稳定。

各个季度的空气质量指数和年份的关系不大，基本都在85上下浮动，而85属于良以上级别，对极少数敏感人群有影响，所以近些年上西安的空气质量较好且相对稳定。

对和进行区间估计。因为我们更加关心空气质量指数的置信区间，所以主要是计算的置信区间。

先计算置信度为0.05的双侧置信区间。

根据公式，查表知，所以四个季度的区间估计如下（不考虑每一年每个季度的参数，将同一季度放在一起估算参数）：

|  |  |
| --- | --- |
| 季度 | 双侧置信区间 |
| 第一季度 | (13.48,177.80) |
| 第二季度 | (42.50,100.52) |
| 第三季度 | (1.55,208.21) |
| 第四季度 | (3.49,211.16) |

计算的置信度为0.05的单侧置信上限，利用公式。

|  |  |
| --- | --- |
| 季度 | 单侧置信上限 |
| 第一季度 | 153.14 |
| 第二季度 | 103.66 |
| 第三季度 | 195.72 |
| 第四季度 | 168.93 |

从这个单侧置信上限可以看出，的上限最高不超过200，而空气质量指数>200时为重度污染，这也就是说西安每季度的空气质量指数只有小于5%的概率有几天空气质量为重度污染，因此西安的空气质量相对较好。

1. **研究结论**

本篇文章通过对于西安2019年到2021年的空气质量指数的分析，得出以下结论：

1. 相邻两天的空气质量指数之间为弱相关，近似为相互独立。

2.西安近年来空气质量较为稳定，与年份几乎没有关系，尤其第二季度的空气质量最为稳定，可能是因为西安在夏天属于暖温带半湿润大陆性季风气候，夏天降水相较于其他几个季节降水量比较多，由于降雨量相对稳定且多，导致第二季度空气质量相较于其他季度更为稳定。

3.西安的空气质量总体来说较好，全年天数出现重度污染的概率小于百分之五，故出现重度污染的天数占比很小。

1. **研究不足与反思**

本篇文章只用了空气质量指数的数据，没有用到PM2.5等其他评估天气质量的数据对西安空气质量进行综合分析。但我认为若将各类数据综合起来分析西安的天气质量，再在纵向时间上进行对比分析，数据处理，感觉难度较大，不好操作。

其次我只选取了西安近三年的空气数据进行分析，在结束论文的时候我意识到数据跨度应该再取大一点，比如取到近五年或者近十年的空气数据进行分析，可能会更加的客观准确，也更能反应出由于近代西安工业的发展对西安空气等环境的影响。（因为本篇论文通过对数据的分析得出的结论是空气质量与年限没有什么太大关联）

在数据处理方面，我认为取一年的空气质量进行分析可能会导致对空气质量的分析数据不够充分，不能很好地反应最终的结果。而选取月份来进行分析，数据量又很大，过于冗杂。故最后深思熟虑后选取了以季度作为横向跨度量进行分析参考。

1. **参考文献及数据来源**
2. 《概率论与数理统计》 西安电子科技大学 张卓奎，陈慧婵著.

[2]《概率论与数理统计》第四版 浙江大学 盛骤，谢式千，潘承毅著.

[3]2019年--2021年西安空气质量指数统计历史数据.

**六．关键词**

空气质量指数，概率论中的正态分布，参数估计（点估计），中心极限定理，置信区间，