**비모수통계학을 활용한**

**미세먼지 및 초미세먼지 동향 분석**

**2019110461 최대한**

**2019110477 김부겸**

**목 차**

**1. 서 론**

**제1절 연구의 목적**

**제2절 연구의 범위 및 방법**

**2. 본 론**

**제1-1절 : 미세먼지와 초미세먼지의 상관관계**

**1. 데이터 추출**

**2. 데이터 분석**

**제1-2절 : 미세먼지와 초미세먼지의 연도별 추세 분석**

**1. 데이터 추출**

**2. 데이터 분석**

**제2절 : 코로나 전후 미세먼지 농도 변화 분석**

**1. 데이터 추출**

**2. 데이터 분석**

**제3절 : 미세먼지 비상저감조치 효과 확인**

**1. 데이터 추출**

**2. 데이터 분석**

**3. 결 론**

**4. 참 고 문 헌**

**1. 서 론**

**제1절 미세먼지의 정의 및 원인**

미세먼지란 대기 중에 떠다니면 눈에 보이지 않을 정도로 작은 먼지이며, 질산염(NO3-), 암모늄 이온(NH4+), 황산염(SO42-) 등의 이온 성분과 탄소 화합물과 금속 화합물 등으로 이루어져 있다. 먼지의 지름이 10마이크로미터 이하인 먼지(PM10)를 미세먼지로 부르고, 지름이 2.5마이크로미터 이하인 먼지(PM2.5)를 초미세먼지로 부른다.

미세먼지는 공기 중 고체상태와 액체상태의 입자의 혼합물로 배출되며 화학반응 또는 자연적으로 생성된다. 또한 문명이 산업화됨에 따라 여러 가지 배출원이 생겨나게 되었고, 이에 따라 사업장 연소, 자동차 연료 연소, 생물성 연소 과정등 특정 배출원으로부터 직접 배출된다.

한국은 중국에서 불어오는 편서풍의 영향을 받고 있으며, 한국에서의미세먼지는 중국에서 배출하는 오염물질이 직접적인 원인으로 작용하고 있다. 또한 한국의 미세먼지 원인 조기사망률은 30.5%로 세계에서 네번째로 높았다는 결과가 나오고 있다. 미세먼지에 지속적으로 노출될 경우, 다양한 신체적 문제가 발생할 수 있다. 일반적으로 먼지는 체내에서 배출되는데 하루에서 이틀이 걸리지만, 미세먼지는 입자가 매우 작기 때문에 배출하는데 일주일 이상 걸린다는 연구 결과가 있다.[[1]](#footnote-1) 그리고 미세먼지로 인해 피부병, 눈병, 호흡기 질환, 암 등 치명적인 병에 걸릴 확률이 늘어나게 된다.

**제2절 연구 자료 및 방법**

본 조는 공공데이터포털에 있는 서울특별시\_시간별 (초)미세먼지와 서울특별시 대기환경정보 사이트에서 월별 미세먼지 및 초미세먼지 평균값을 수집해서 데이터로 활용하였다.

**[그림 1] 서울특별시\_시간별 (초)미세먼지 2022년도 자료**

텍스트, 스크린샷, 번호, 폰트이(가) 표시된 사진

자동 생성된 설명

**※ 연구의 방법**

본 조가 가지고 있는 데이터는 2012년부터 2022년까지의 모든 시간별(1시간 단위)로 모든 구의 미세먼지와 초미세먼지의 측정치이다. 가지고 있는 자료가 매우 방대하고 결측값 또한 매우 적기 때문에 자료의 신뢰성은 높다. 데이터를 R 프로그램으로 불러온 다음, 각 주제에 맞게 전처리를 한 다음, 다양한 비모수적 방법을 통해 본 조가 정한 주제에 대해 검정해 볼 것이다.

**2. 본 론**

**제1절 : 미세먼지와 초미세먼지의 상관관계**

**1. 데이터 추출**

각 연도별 미세먼지와 초미세먼지의 변화를 살펴보기 위해 서울특별시 대기환경정보에서 제공된 대기질통계 자료를 참고하여 수집하였다. 미세먼지는 2012년 1월부터 2022년 12월 , 초미세먼지는 2013년 10월부터 2022년 12월까지의 서울의 각 구별 평균과 총 평균을 수집하였다.

**[그림 2-1] 월별 미세먼지와 초미세먼지의 평균**

텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명

**2. 데이터 분석**

2013년 10월부터의 미세먼지와 초미세먼지의 spearman 상관계수를 구해본 결과, 약 0.624 정도가 나옴을 확인할 수 있었다. 상관계수가 1에 가깝기 때문에 미세먼지와 초미세먼지가 양의 상관관계가 있다고 생각할 수 있다. 각 연도별을 기준으로 Spearman 상관계수를 했을 때 0.80~ 0.92 값이 나온 것으로 보아 이 둘의 상관관계는 매우 강한 것으로 생각된다. 그리고 Spearman 상관계수를 permutation 방법으로도 검정해볼 것이다. 이 방법을 사용하면, 가설 검정을 하는 것처럼 귀무가설과 대립가설을 세워야 되는데 이 경우에서 귀무가설은 상관관계가 없다이고, 대립가설은 상관관계가 있다이다. 검정을 진행해본 결과, pvalue가 0.0001으로 일반적인 기각역인 0.05보다 매우 작기 때문에 귀무가설을 기각하게 된다. 따라서 상관관계가 있다는 결론이 나오는데, 이는 처음에 구했던 spearman 상관계수의 결과와 어느 정도 일치하는 것을 확인할 수 있다.

**[그림 2-2] spearman 상관계수를 permutation 방법으로 구한 결과**

텍스트, 도표, 스크린샷, 라인이(가) 표시된 사진

자동 생성된 설명

**제1-2절 미세먼지와 초미세먼지 추세분석**

1. **데이터 추출**

미세먼지와 초미세먼지의 상관관계는 회귀선 분석을 통해 비슷한 추세면 상관관계가 높다고 추정할 수 있을 것이다. 월별 서울시 미세먼지와 초미세먼지의 서울시 평균을 추출하여 미세먼지농도의 변동특성을 파악하기 위해서 LOESS 분석을 실행하였다.

1. **데이터 분석**

R 에서 제공되는 loess 함수를 이용하여 각 연도 월별 농도에 따른 회귀선을 그렸다. 분석결과 대상기간동안 미세먼지와 초미세먼지의 증감변동이 파악되고 둘의 변동이 매우 유사함이 나타난다. 미세먼지는 2016년까지 증가 추세를 보이다가 이후 감소하는 모습이고 초미세먼지는 그래프 형태는 유사하나 2018년까지 증가 추세를 보이다가 감소하는 모습이다. 미세먼지와 초미세먼지의 추세가 비슷하기 때문에 상관관계가 있다는 주장을 뒷받침하는 자료로 볼 수 있을 것이다.

**[그림 2-4] 미세먼지와 초미세먼지의 분포에서 LOESS를 통해 회귀선을 그린 결과**

스크린샷, 도표이(가) 표시된 사진

자동 생성된 설명

그리고 이 추세선을 loess 방법 말고 Spline 방법으로도 진행해봤을때도 비슷한 결과가 나오는 것을 확인할 수 있었다.

**[그림 2-5] 미세먼지와 초미세먼지의 분포에서 Spline을 통해 회귀선을 그린 결과**



**제2절 : 코로나 전후 미세먼지 농도 변화 분석**

**1. 데이터 추출**

연도별 평균 미세먼지 농도를 데이터에서 추출한 다음 코로나 발생년도인 2020년도를 기준으로 해서 그룹 2개(before, after)를 생성했다.

**[그림 3-1] before, after 그룹 수치**

텍스트, 스크린샷, 폰트, 라인이(가) 표시된 사진

자동 생성된 설명

**2. 데이터 분석**

본 조는 코로나 이전보다 코로나 발생 후의 평균 미세먼지 농도가 더 낮을 것이다라는 가설을 세웠다. 이 가설을 세운 근거로는 코로나 발생 이후 약 2년간 전세계적으로 제조업 산업에 타격이 갔고, 이로 인해 세계 여러 곳곳에서 공장 가동이 중지되었었다. 우리나라의 미세먼지 농도에 가장 큰 영향을 주는 국가가 바로 중국인데, 중국 공장 역시 이 기간동안 공장가동이 중지되었었다. 미세먼지 발생 원인에 공장 매연 역시 큰 요인으로 작용하는데, 공장 가동이 중지됨에 따라 공장 매연이 발생하지 않기 때문에 이러한 근거를 가지고 가설을 세웠다. 그리고 이를 검정하기 위해 wilcoxon ranksum test를 진행하였고, 이를 permutation 방법으로도 검정하였다.

첫 번째로, wilcoxon ranksum test를 진행한 결과, pvalue가 약 0.006이 나왔다. 이 경우에는, 귀무가설은 코로나 전후 미세먼지 농도의 차이가 없다이고 대립가설은 코로나 이전의 미세먼지 농도가 코로나 이후 미세먼지 농도보다 높다이다. 분석 진행 결과, pvalue가 0.05보다 매우 작기 때문에(약 0.006) 귀무가설을 기각하게 된다. 따라서 코로나 이후 미세먼지 농도가 더 낮음을 확인할 수 있다.

**[그림 3-2] wilcoxon ranksum test 결과**

텍스트, 스크린샷, 폰트, 라인이(가) 표시된 사진

자동 생성된 설명

두 번째로, 위의 검정을 permutation 방법으로도 진행하였는데, 역시 pvalue가 0.006이 나옴을 확인할 수 있었다.

그리고 접근방식을 연도별 전체 평균이 아닌 연도별 구별 평균을 각각 계산해서 paired data로 만든 후 관련 분석을 진행해봤다. Wilcoxon Signed-Ranks Test와 Sign test를 진행하였느데 이때의 귀무가설은 코로나 이전 연도와 이후 연도의 구별 미세먼지 값들의 중앙값이 같다이고, 대립가설은 코로나 이전 연도의 미세먼지 값들의 중앙값이 이후 연도의 코로나 미세먼지 값들의 중앙값보다 더 크다이다. 각각 진행해본 결과, 둘다 모두 pvalue가 0에 매우 근접한 작은 숫자가 나왔다. 이를 통해 귀무가설을 기각하게 되고, 대립가설을 채택하게 된다.

데이터를 paired와 unpaired 2가지 방식으로 접근해서 분석을 모두 진행해본 결과, 둘 다 코로나 이전 연도의 미세먼지 농도가 더 높음을 확인할 수 있었다.

**제3절 미세먼지 비상저감조치 효과 확인**

**1. 데이터 추출**

미세먼지 비상저감조치란 고농도 미세먼지(PM-2.5)가 일정기간 지속시 단기간에 대기질을 개선하기 위한 비상조치를 말한다. 서울특별시 대기환경정보 사이트에 들어가면 미세먼지 비상저감조치 시행 일자를 확인해볼 수 있다. 해당 일자의 데이터와 다음날 데이터를 엑셀 파일에서 확인해서 분석을 진행해 볼 것이다.

**2. 데이터 분석**

첫 번째로, 2022-02-11 21:00시에 시행한 경우를 분석해보았다. 기본적으로 저 시간대의 구별 초미세먼지 데이터와 다음날인 2022-02-12 21:00시의 구별 초미세먼지를 비교해 paired data를 분석하는 방법인 sign test를 진행해본 결과, pvalue값이 0에 매우 근접한 값이 나왔다. 귀무가설은 두 시간대의 초미세먼지 농도의 중앙값이 같다이고, 대립가설은 2월 11일 21시의 초미세먼지 농도의 중앙값이 2월 12일 21시의 초미세먼지 농도의 중앙값보다 높다인데, pvalue가 매우 작기 때문에 귀무가설을 기각하게 된다. 따라서 다음날 초미세먼지 농도의 중앙값이 낮기 때문에 미세먼지 비상저감조치의 효과가 있었다라고 생각할 수 있다. 실제로 발령했던 날짜에 대해 검정을 진행해 본 결과, 대부분의 경우에서 미세먼지 비상저감조치 효과가 있었음이 확인되었다. 하지만 효과가 없었던 경우도 있었는데 이럴 경우 모든 날짜에서 다음날에도 미세먼지 비상저감조치가 시행되었음을 확인했다. 그래서 초미세먼지 농도가 유의미하게 줄어줄어들 때 까지 비상저감조치를 시행함을 확인할 수 있었다.

**[그림 3-1] 미세먼지 비상저감조치 발령 일자 및 연속 발령 일자 예시**

텍스트, 스크린샷, 폰트, 라인이(가) 표시된 사진

자동 생성된 설명

**3. 결 론**

지금까지 미세먼지와 초미세먼지 데이터를 중심으로 관련 가설을 세워보고 검정해보는 과정을 진행해봤다. 첫 번째로, 미세먼지와 초미세먼지는 상관관계가 있으며 비슷한 시기에 늘고 줄었음을 확인할 수 있었다. 그리고 이러한 유사성을 추세선을 통해 확인하기 위해 LOESS 및 Spline 방법으로 확인해 보았고, 그래프의 기울기와 위치가 유사한 것을 확인할 수 있었다. 두 번째로, 코로나 전후의 미세먼지 농도를 비교해봤는데, 기준을 연도별 전체로 놓은 경우와, 전후 구별로 놓은 경우로 구분해서 검정을 진행했었다. 연도별 전체로 놓았을 때에는 wilcoxin ranksum test 및 permutation 방법을 진행했고, 전후 구별로 할 때에는 sign test 및 wilcoxon sign rank test 방법을 사용했다. 어느 기준으로 검정하든 결과는 동일하게 코로나 이전 시기의 미세먼지 농도가 이후 시기의 미세먼지 농도보다 높았음을 확인할 수 있었다. 마지막으로 미세먼지 비상저감조치가 효과가 있는지 검정해보았는데, 검정 결과 초미세먼지 농도의 중앙값 차이가 0보다 크다는 것이 확인되었기 때문에 비상저감조치가 효과가 있음을 확인할 수 있었다.

아쉬운 점은, 대기와 관련된 다양한 요인들 간의 상관계수가 대부분 낮게 나와서 미세먼지와 초미세먼지 및 두 변수와 관련된 검정밖에 진행하지 못한 점이 있었다.

**참 고 자 료**

**1. 국내문헌**

**1. 동아사이언스 ““해독차 증기 들이마시면 효과?” 미세먼지 둘러싼 거짓정보들”, https://www.dongascience.com/news.php?idx=32847, 동아사이언스**

**2. 서울특별시 대기환경정보, “비상저감조치 발령 내역”,**

**https://cleanair.seoul.go.kr/statistics/emerList?pageNum=2&checkEfcmStep=&pollutantType=&efcmStep=&year=**

**3. 서울특별시 대기환경정보, “대기질 통계 월별 평균”,**

**https://cleanair.seoul.go.kr/statistics/monthAverage**

1. 이정아, “[만연하는 가짜 의학정보] “해독차 증기 들이마시면 효과?” 미세먼지 둘러싼 거짓정보들”, 동아사이언스, 2019.12.08, https://www.dongascience.com/news.php?idx=32847 [↑](#footnote-ref-1)