경영통계 G조 보고서

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 과목명 | 경영통계 | | | |
| 담당교수 | 곽기영 교수님 | | | |
| 주제 | 다중회귀분석을 이용한 초미세먼지와 대기오염물질에 대한 분석 | | | |
| 팀 명 | G조 | | | |
| 팀원 | 학부 | 학번 | 성명 | 비고 |
| 경영정보학부 | 20185257 | 황윤재 | 팀장 |
| 경영정보학부 | 20175284 | 윤홍은 | 팀원 |
| 경영정보학부 | 20202702 | 손채형 | 팀원 |
| 경영정보학부 | 20202703 | 신정호 | 팀원 |

**서론**

2019년 글로벌 대기오염 조사기관 에어 비주얼이 발표한 ‘2019년 세계 대기질 보고서’에 따르면 우리나라의 작년 연평균 미세먼지 농도는 24.8㎍/㎥으로 OECD 회원국 중에서 가장 나빴다. 또한 최근 기사에 따르면 초미세먼지가 인류의 기대수명을 평균 1.03년가량 단축하고 있다고 밝혔다. 최근 코로나문제로 인해 이전보다 주목을 덜 받고 있는 문제이지만 우리나라의 초미세먼지 증가에 따른 피해도 심각하다고 판단하여 이번 분석을 통해 어떤 요소들이 초미세먼지와 연관성이 있고 이를 토대로 초미세먼지 감소 정책 수립에 근거가 될 수 있는 자료를 만들어보고자 한다.

**분석범위 및 방법**

**2.1 분석범위 및 요인 설정**

지역 크기 대비 인구밀집도가 매우 높아 대기오염이 심각할 것으로 예상되는 서울시를 중심으로 데이터를 수집하였다. 종속변수를 초미세먼지(PM2.5)로 설정하고 독립변수에는 대표적인 대기오염물질로 알려져 있는 SO2, CO, O3, NO2 4가지와 조원들의 의견에 따라 강수량과 온도 2가지를 더 추가하여 총 6가지의 독립변수를 설정하였다.

**2.2 분석 방법**

다중회귀분석은 2개 이상의 독립변수를 이용한 회귀분석으로 독립변수가 종속변수에 미치는 영향의 크기를 파악하고 이를 통해 독립변수의 일정 값에 대응하는 종속변수 값을 예측하는 방법이다.

본 분석에서는 특정한 독립변수 값에 해당하는 종속변수 값들은 정규분포를 해야 하고 정규분포의 분산은 동일해야 하며 종속변수 값들은 통계적으로 서로 독립적이어야 한다는 전제조건을 만족하고 있다. 회귀분석 결과 다중 공선성이 존재할 가능성이 높은 변수가 검출되어 이를 제거하기 전의 회귀분석 결과와 제거 후의 회귀분석 결과를 비교하고 마지막으로 regsubsets 함수를 통해 최적의 회귀모델을 찾는 과정까지 진행해보도록 하겠다.

**다중회귀분석 결과**

회귀분석결과에 따르면 Temp, SO2, Prec의 개별 회귀계수는 유의수준 0.05에서 통계적으로 유의하지 않으며 모집단에서 0일 가능성을 배제할 수 없었다. CO, O3, NO2의 개별 회귀계수는 통계적으로 유의하며 이 세 변수의 회귀계수는 0이라고 볼 수 없다. F값에 대응되는 p-값은 유의수준 0.05에서 통계적으로 유의했다. 하지만 통계적으로 유의할 것이라고 예상되던 Prec의 회귀계수가 유의하지 않았다.

**다중 공선성 변수 제거 전**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 종속변수 | 독립변수 | 비표준화회귀계수 | 표준오차 | 베타 | t | 유의확률 | VIF |
| 초미세먼지  농도(PM2.5) | (상수) | -2.725e+01 | 7.948e+00 |  | -3.429 | <0.001 |  |
| Temp | -1.858e-01 | 1.076e-01 | -0.2774 | -1.727 | 0.089 | 6.302 |
| SO2 | -1.201e+03 | 7.744e+02 | -0.1564 | -1.551 | 0.126 | 2.483 |
| CO | 5.270e+01 | 1.398e+01 | 0.8399 | 3.771 | <0.001 | 12.105 |
| O3 | 5.998e+02 | 9.008e01 | 0.7520 | 6.658 | <0.001 | 3.113 |
| NO2 | 4.764e+02 | 1.779e+02 | 0.4251 | 2.678 | 0.009 | 6.152 |
| Prec | 1.016e-02 | 5.662e-03 | 0.1486 | 1.795 | 0.078 | 1.673 |
| F=30.84(p<0.001) | | | | | | | |

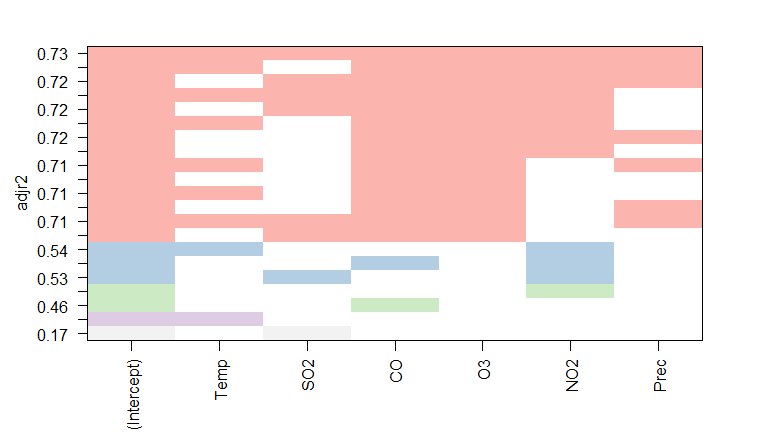
이에 따라 다중 공선성의 존재가 의심되어 점검해보았더니 CO변수의 VIF가 10을 초과하였고 다중 공선성이 존재할 가능성이 높다고 판단하여 제거하고 다시 회귀분석을 실시하였다. 그 결과 Temp, SO2, Prec의 유의확률에 변화가 있었다. 다중 공선성 제거 전 개별회귀계수가 유의하지 않았던 결과와 달리 세 변수의 개별회귀계수가 통계적으로 유의하다는 결과가 나왔다. 하지만 회귀식의 설명력은 다중 공선성 변수 제거 전에 비해 감소한 것으로 밝혀졌다.

**다중 공선성 변수 제거 후**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 종속변수 | 독립변수 | 비표준화회귀계수 | 표준오차 | 베타 | t | 유의확률 | VIF |
| 초미세먼지  농도(PM2.5) | (상수) | -2.167e+00 | 4.806e+00 |  | -0.451 | 0.653 |  |
| Temp | -4.474e-01 | 9.088e-02 | -0.6681 | -4.923 | <0.001 | 3.683 |
| SO2 | -1.798e+03 | 8.374e+02 | -0.2341 | -2.146 | 0.036 | 2.379 |
| O3 | 4.155e+02 | 8.358e+01 | 0.5209 | 4.971 | <0.001 | 2.196 |
| NO2 | 8.845e+02 | 1.560e+02 | 0.7893 | 5.668 | <0.001 | 3.877 |
| Prec | 1.275e-02 | 6.208e-03 | 0.1864 | 2.054 | 0.044 | 1.648 |
| F=27.99(p<0.001) | | | | | | | |

Regsubsets 함수를 이용하여 서브셋회귀분석 그래프를 나타낸 결과 그래프 가장 위의 행에 있는 모든 변수를 포함한 모델의 수정된 이 0.73으로 가장 높아 최적의 모델인 것을 알 수 있었다.

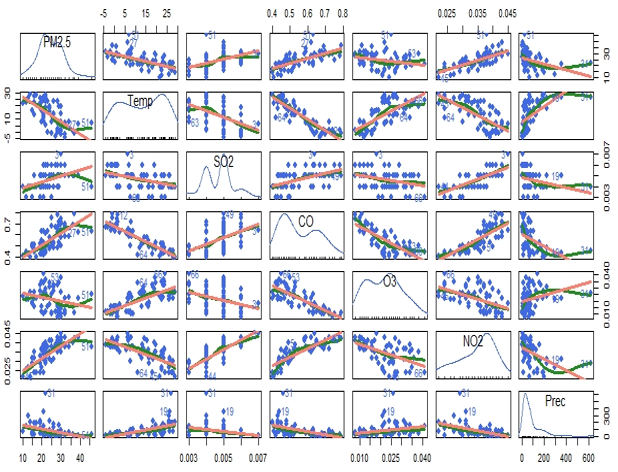
**Regsubsets 그래프**



**결론**

회귀분석 결과 SO2, CO, NO2 세 가지는 초미세먼지와 정의 상관관계를 갖는 물질로 나타났다. 따라서 초미세먼지의 농도를 낮추기 위해서는 위의 SO2, CO, NO2에 초점을 맞춰 정책을 수립하는 것이 바람직할 것으로 보인다. 그런데 초미세먼지와 정의 상관관계를 갖을 것으로 예상되던 O3는 분석결과 초미세먼지와 부의 상관관계를 갖는 것으로 나타났다. 이 결과에 의아하여 이와 관련된 자료를 찾아보니 최근 코로나19로 인해 이동이 제한되어 화석연료 소비량이 크게 줄어 대부분의 대기오염물질의 수치는 감소했지만 O3는 증가했다는 기사를 볼 수 있었다. 이 기사를 보고 이번 분석결과에서 O3가 초미세먼지와 부의 상관관계를 갖는다는 것이 우연한 결과나 분석오류가 아닐 수도 있다는 생각이 들었고 이에 대해서는 추가적인 연구가 필요할 것으로 보인다.

**scatterplotMatrix 함수를 이용한 산점도 행렬**



**참고 문헌**

이선희, 윤동현, and 고준환. "다중회귀분석을 이용한 서울시 1 인 가구의 연령별 분포요인에 관한 연구." *한국공간정보학회지* 23.3 (2015): 11-21.

Pierre Sicard, Alessandra De Marco, Evgenios Agathokleous, Zhaozhong Feng, Xiaobin Xu, Elena Paoletti, José Jaime Diéguez Rodriguez, Vicent Calatayud, Amplified ozone pollution in cities during the COVID-19 lockdown, Science of The Total Environment, Volume 735, 2020, 139542, ISSN 0048-9697, <https://doi.org/10.1016/j.scitotenv.2020.139542>.

조선일보(김효인 기자), 한국 미세먼지, OECD 중 가장 나쁘다, 2020.02,

<https://www.chosun.com/site/data/html_dir/2020/02/27/2020022700301.html?utm_source=naver&utm_medium=original&utm_campaign=news>