

주제 선정 배경

배경

- 현재 서울특별시에서는 통합대기환경지수(Comprehensive Air-quality Index)를 활용하여 서울 대기의 오염 정도를 수치화 및 등급화하여 시민들에게 그 정보를 알리고 있다.
- 하지만, 서울에서 사용하는 통합대기환경지수의 지수산출 방법에는 각 오염 물질들 간 통계적 상관관계의 반영이 미흡한 점이 확인된다.
- 이에 우리 조에서는 수업시간에 배운 다변량 관리도를 활용해 좀 더 개선된 대기오염 측정 시스템을 구축해보자 한다.

목적

- 통합대기환경지수(Comprehensive Air-quality Index)는 계산 방식이 일반 사람이 이해하기에는 수식에 따른 무리가 있다.
- 따라서, 관리도를 통한 대기오염 모니터링으로 누구나 쉽게 이해 할 수 있도록 한다.

수행 방법

- 팀 구성은 3명으로 구성하였으며, 주제 범위는 2012~13년 3월, 4월 서울특별시 대기 정보를 이용하여 실시 하였다. 프로젝트 일정은 아래와 같다.

날짜	07/15	07/22	07/29	08/09	08/16
조사계획	[Progress Bar]				
본조사	[Progress Bar]				
자료분석	[Progress Bar]				
보고서작성	[Progress Bar]				

- 자료수집 : 서울특별시 대기 환경 정보 홈페이지(<http://cleanair.seoul.go.kr/main.htm>) 2012~13년 3월, 4월 일 평균 데이터를 기록 및 저장
- 수행방법 : 2012~13년 3월, 4월 일 평균 데이터를 이용하여 일변량 및 다변량 관리도 작성

통계 분석

1. 주요 대기오염물질

- 아래 5개의 대기오염물질은 인체에 악영향을 미치는 주요 물질로 기존 대기오염측정 시스템에서 수치화되어 사용되고 있으며, 이후 우리가 적용할 새로운 시스템(다변량 관리도)에서도 이 변수들을 사용한다.

변수	설명	그림
미세먼지 (PM-10)	대기 중 고체입자와 액적입자의 혼합물로 공기역학직경이 10μm 이하인 부분을 의미하며 호흡계 질환 및 심혈관계 질환을 일으킨다.	대기중미세먼지
오존 (O ₃)	광화학 옥시던트의 하나로 2차 오염물질에 속하며 무색, 무미의 기체로 산화력이 강하다. 폐에 피해를 주며, 기슴통증, 기침 등을 유발한다.	식물의 오존 피해
이산화질소 (NO ₂)	질소산화물 중 대기오염에 가장 영향이 많은 물질로 적갈색의 자극성 냄새가 나는 유독한 기체. 눈과 호흡기 등에 자극을 주며 폐수중, 페인트 등 유발한다.	식물의 이산화질소 피해
일산화탄소 (CO)	무색, 무취의 맹독성 기체로 연료의 불완전 연소로 발생한다. 헤모글로빈을 카르복시헤모글로빈으로 변형시켜 혈액의 산소운반 능력 저하시킨다.	산물 등의 자연재해
아황산가스 (SO ₂)	불쾌한 자극적인 냄새가 나는 무색의 불연성기체. 고농도에서 비강과 인후에 흡수되고 점막염과 반응하여 황산염 형성해 염증 및 호흡장애를 일으킨다.	식물의 아황산가스 피해

2. 기준시스템

2-1) 통합대기 환경지수 지수 (Comprehensive Air-quality Index)

대기오염도 측정치를 국민이 쉽게 알 수 있도록 하고 대기오염으로부터 피해를 예방하기 위한 행동지침을 국민에게 제시하기 위하여 대기오염도에 따른 인체 위험성과 대기환경기준을 고려하여 개발된 대기오염도 표현방식. (PM-10, O₃, NO₂, CO, SO₂의 대표적인 5가지 오염물질에 대한 대기질의 상태지수를 제곱)

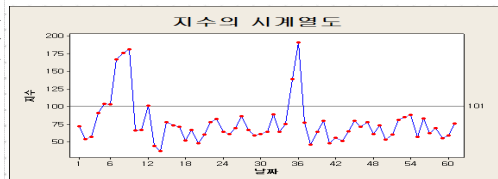
5개 대기오염 물질별로 통합대기환경지수 점수를 산정하여 가장 높은 점수를 통합 지수값으로 사용. 산출된 각각의 오염 물질별 지수점수가 '민감군 영향(3단계)' 이상의 등급이 2개 물질 이상일 경우 통합지수값에 가산점 부여. 통합대기환경지수는 0에서 500까지의 지수를 6단계로 나누어 점수가 커질수록 대기상태가 좋지 않음을 나타냄. 실제 서울시에서는 '민감군 영향(3단계)'부터 유의한 수준이라 판단한다.

$$I_p = \frac{I_{HI} - I_{LO}}{BP_{HI} - BP_{LO}} \times (C_p - BP_{LO}) + I_{LO}$$

- I_p : 오염물질의 대기지수점수
- C_p : 오염물질의 대기중 농도
- BP_{HI} : 오염물질의 오염도 해당 구간 최대 오염도
- BP_{LO} : 오염물질의 오염도 해당 구간 최소 오염도
- I_{HI} : BP_{HI} 에 해당하는 지수값(구간 최대 지수값)
- I_{LO} : BP_{LO} 에 해당하는 지수값(구간 최소 지수값)

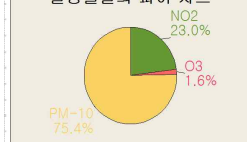
지수 산출 방법	지수 구분	지수 범위	구간의미
좋음	0~50	대기오염 관련 환자군에서도 영향이 유발되지 않을 수준	
	보통	51~100	환자군에게 만성 노출시 경미한 영향이 유발될 수 있는 수준
민감군 영향	101~150	환자군 및 민감군에게 유해한 영향이 유발될 수 있는 수준	
	나쁨	151~251	환자군 및 민감군(어린이, 노약자 등)에게 유해한 영향 유발, 일반인도 건강상 불쾌감을 경험할 수 있는 수준
매우 나쁨	251~350	환자군 및 민감군에게 급성 노출시 심각한 영향 유발, 일반인도 약한 영향이 유발될 수 있는 수준	
	위험	351~500	환자군 및 민감군에게 응급 조치가 발생되거나, 일반인에게 유해한 영향이 유발될 수 있는 수준

2-2) 2013년 3월~4월 데이터 그래프 분석

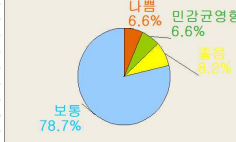


- 위 통합대기환경지수 시계열도를 확인해보면, 3월 5, 6, 7, 8, 9, 12일 4월 4, 5일이 '민감군 영향(3단계)' 이상의 지수구분값을 가지는 것으로 보인다.
- 실제 확인결과, 뉴스를 통해 이 날짜 기간 동안 약한 황사가 온 것을 확인했다.

결정물질의 파이 차트



등급의 파이 차트



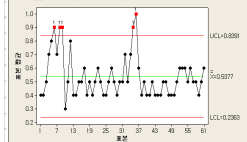
- 빈도분석의 결과 대기 중 오염결정물질은 미세먼지(PM-10)가 75.4%로 가장 많이 차지하고 있으며, 다음으로 이산화질소(NO₂)와 오존(O₃)이 따른다.
- 대기 오염 등급 파이 차트를 보면 서울시에서 유의하게 판단하는 '민감군 영향(3단계)'이상의 날이 전체 일수에 13.2%(1주일에 하루 정도)를 차지하는 것을 확인할 수 있다.

3. 새로운 시스템

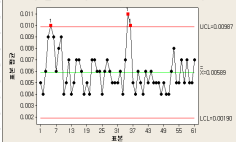
1) 일변량 관리도 (Xbar 관리도)

- 다변량 관리도 적용하기 이전 먼저 일변량 관리도를 적용해보자.
- 통계적 품질관리 중의 중요한 목적 중의 하나는 공정의 이상을 조기에 발견하고 그 원인을 파악하여 공정을 안정화 시키는 것이다. 이러한 목적으로 사용되는 기법이 관리도이며, 하나의 품질특성치를 다루는 것을 일변량 관리도라 한다.
- 관리도는 2012년 데이터를 기준으로 2013년 모니터링 실시한다

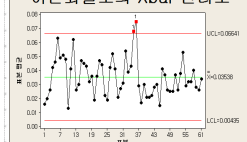
일산화탄소의 Xbar 관리도



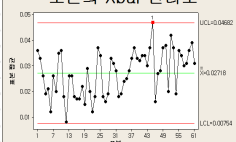
아황산가스의 Xbar 관리도



이산화질소의 Xbar 관리도



오존의 Xbar 관리도



미세먼지의 Xbar 관리도



✓ 관리상한선 유의수준: 0.0027%

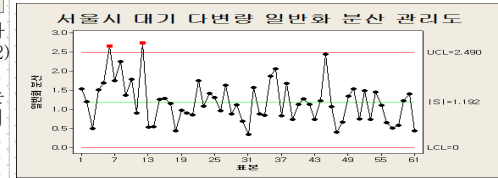
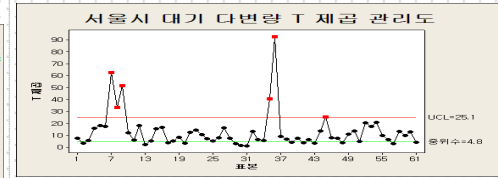
- 하지만 변수들간의 상관관계를 반영할 수 없기 때문에 더 정확한 결과를 도출하는데 부적절하다. 따라서, 변수들간의 상관관계를 반영할 뿐만 아니라 변수들을 동시에 관리할 수 있는 다변량 관리도가 적절하다.

2) 다변량 관리도

- 품질 특성치들 간의 상호 연관성 등의 문제로 인하여 기존의 일변량 관리도를 통하여 관리하는 것은 올바르지 않으며, 이러한 특성치를 동시에 관리하는 다변량 관리도의 중요성이 강조되고 있다.
- 호텔링 T² 관리도는 Xbar 관리도를 다변량 개념으로 확장한 것으로 T² 통계량을 이용하여 서로 상관관계가 없는 여러 개의 변수들을 동시에 관리하는 관리도이다. 적용을 위해서는 데이터가 다변량 정규분포를 따라야 하며 통계량 계산은 다음과 같다.

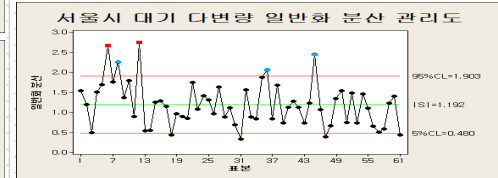
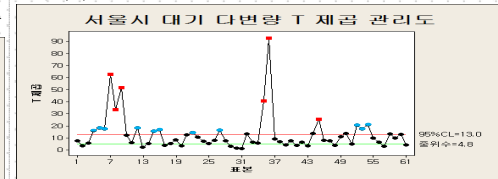
$$T^2 = (x - \bar{x})' S^{-1} (x - \bar{x})$$

- 일반화 분산 관리도는 S 관리도에 해당하는 다변량 관리도로서, 서로 관련된 두 개 이상의 공정 특성으로 인한 공정 변동들 동시에 관측할 수 있다.



✓ 관리상한선 유의수준: 0.27%

- 관리상한선을 벗어난 날짜 : 3월 7,8,9일 4월 4,5,14일
- 실제 대기 환경지수를 통한 등급과 비교해본 결과 등급이 '나쁨'인 날짜와 일치한다.
- 관리 상한선의 기준이 높은 점으로 인하여 '민감군 영향' 등급인 날짜들을 발견하지 못하였지만, 유의수준을 낮춤으로써 발견 가능하다.



✓ 관리상한선 유의수준: 5%

- 관리상한선을 벗어난 날짜 : 3월 4,5,6,12,15,16,22,27일 4월 21,22,23일
- 실제 대기 환경지수를 통한 등급과 비교해본 결과 등급이 '민감군 영향'인 날짜를 모두 찾아냈으며, 등급이 '보통'이지만, 지수가 높은 날짜도 관리상한선 유의수준 5%안에서 모두 찾아 낼 수 있다.

결론

- 현재 서울시에서 제공하는 통합대기 환경지수는 수식 및 계산방식이 쉽게 이해하는데 어려움이 있다.
- 이에 따라, 관리도를 통하여 새로운 모니터링 방법을 제안하였고, 변수들간의 상관관계를 반영하는 다변량 관리도가 적합하다.
- 다변량 관리도를 적용해본 결과, 현재 시스템보다 더 뛰어난 모니터링 성능을 확인할 수 있다.