

2024년 1학기 경영통계와 의사결정

기말 리포트 안내

한국외국어대학교 이순희 교수

학번: 202368020

이름: 김재환

다음 안내사항 필독하기 바랍니다.

- 프로젝트의 목표: 실제 데이터를 이용해서, 가설검정 또는 추정을 적용하여 의사결정을 내리는 과정을 이해한다.
- 프로젝트에 포함되어야 하는 내용: (1) 가설검정/추정을 적용할 수 있는 문제를 정의함 (2) 수집된 데이터에 대한 설명 (3) 통계적인 추론 과정(가설검정 또는 추정)에 대한 설명 (4) 결론
- 리포트는 총 3-4페이지 권장. (넘을 경우라도, 패널티 적용하지는 않음)
- 제출할 것: 리포트 (워드 또는 pdf), 데이터 파일 (엑셀 또는 텍스트 파일); 데이터 파일을 따로 제출하므로 리포트에는 데이터가 들어가지 않아도 됨.
- 가설검정은 엑셀/R을 이용함.
- 평가 기준: 주제가 가설검정/추정하기에 적합한 주제인가? 흥미로운 주제인가? 통계적 추론 과정이 맞는지? 결론이 맞는지? 의 기준으로 평가할 계획입니다.

예제: 아래 예제는 가설검정으로 풀수 있는 문제의 예시이고, 대략적인 내용만을 참고하기 바랍니다 (아래 주제는 이미 예시로 설명하였으므로, 프로젝트 주제로 선정하지 않기 바랍니다)

-가설검정문제: 코로나 전에 비해서 코로나 이후에 서울의 미세먼지량이 훨씬 줄었들었는가?

-데이터 수집: 기상데이터 중 2019년 3월-5월 날짜별 미세먼지량(코로나전) 수집하고, 2020년 3월-5월 날짜별 미세먼지량(코로나후) 수집.

-통계적 추론과정: t-test 중 대응표본을 이용해서 분석함.

-결론: p-value 값이 xxx 이므로 코로나 전에 비해서 코로나 이후의 서울 미세먼지량이 훨씬 줄어들었다. 추가적으로 평균미세먼지량은 00 ± 00 로 추정됨.

2023년 2학기 경영통계와 의사결정 기말 리포트

학번: 202368020 이름: 김재환

프로젝트 요약

1. 프로젝트 주제 & 선정배경:

프로젝트의 주제는 중국에서의 황사 영향이 코로나 기간 전과 코로나 기간 동안 유의미한 차이가 발생하였는지에 관한 연구입니다.

선정배경은 교수님의 주제 추천에 대한 도전의식과 평소 황사와 관련된 분석을 자세히 해보고 싶다는 생각이 있었습니다.

2. 데이터에 대한 설명: (수집 방법 포함, 다른 웹사이트 참고할 경우, 인용)

데이터는 국가 공인기관들(경기도 보건환경연구원, 대전광역시 보건환경연구원, 서울특별시 보건환경연구원, 인천광역시 보건환경연구원)들의 홈페이지에서 공개된 자료를 직접 다운로드 받아서 사용하였습니다.

그 외에 구글 스콜라에서 황사에 관련된 논문 "황사출현에 영향을 미치는 요인에 관한 연구(2009)" 와 "황사의 대기환경영향(2011)" 을 참고하였습니다.

3. 적용한 통계적 방법론:

적용한 통계적 방법론은 shapiro 정규성 검정과 leven의 등분산 검정, ANOVA 검정(3개집단으로 구분시), 독립표본 T검정(2개의 집단 비교시), wilcox 비모수 검정 등을 적용하여 분석하였습니다.

4. 결론

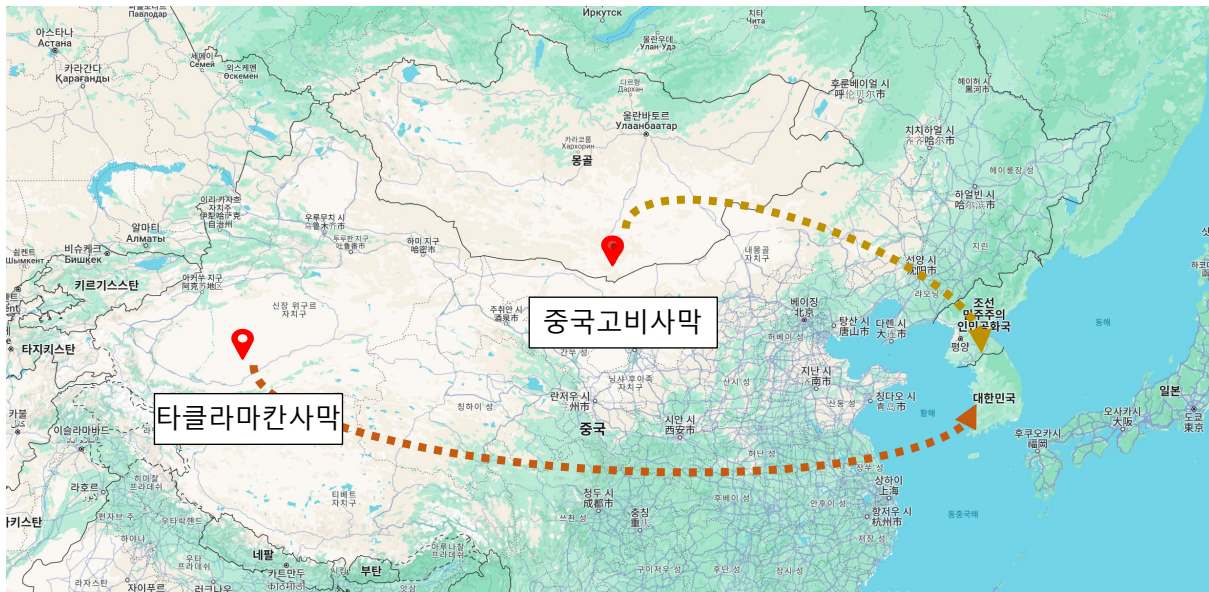
인천의 데이터를 세부적으로 검정한 결과 황사기간 동안 코로나로 인한 미세먼지와 초미세먼지의 농도 감소 현상은 나타났으며, 그 영향은 지역별로 다르게 나타났습니다. 또한 서울의 경우 황사에 영향을 많이 받았을 것이라 여겨지는 지역들만을 분석하였을 때 그 효과가 두드러지게 나타났으며, 경기도와 대전에서는 황사기간 동안 코로나의 영향으로 인한 특정 중금속의 대기중 농도 감소 현상이 유의적으로 나타났습니다.

결론은 황사기간 동안의 COVID-19로 인한 황사피해의 감소 현상의 발생이 유의적이었음을 확인하였습니다.

프로젝트 개요

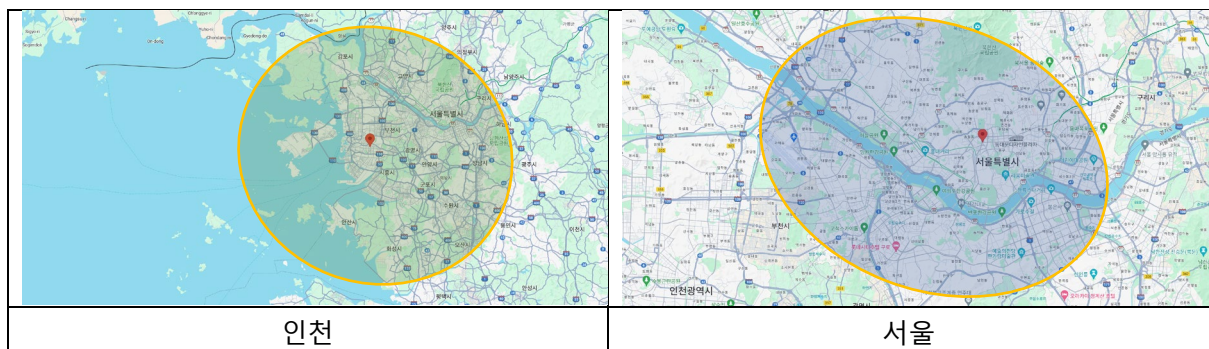
1. 조사분석 범위 설정

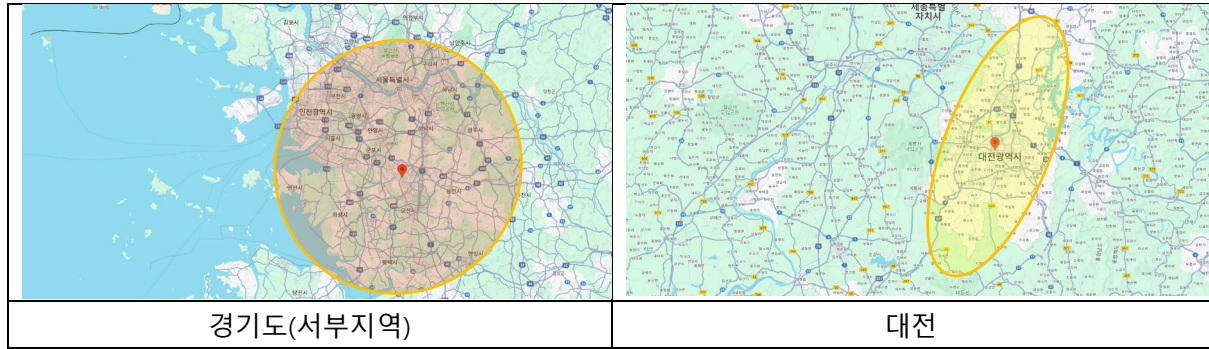
황사란 중국의 고비사막과 타클라마칸 사막에서 출발하여 대한민국의 수도권과 서부해안 지역, 중부 내륙지역의 대기오염에 영향을 끼치는 자연현상으로 주로 3월, 4월, 5월에 집중적으로 발생하는 것으로 알려져 있습니다.



주로 봄철의 서풍과 북서풍이 대한민국의 영토에 영향을 주는 것으로 알려져 있습니다. 따라서 전국의 모든 지역을 선정하기 보다, 지리적으로 중국에서 불어오는 편서풍에 영향을 많이 받을 것 같은 4가지 지역을 선정하여 조사를 진행하였습니다.

각 지역은 인천과 서울, 경기도와 대전 지역을 선정하였습니다.





2. 황사의 측정 단위 및 황사에 포함된 중금속 함유량에 대한 의미

국가연구기관 및 기상특보기관 모두 전통적으로 $PM_{10}(\mu g/m^3)$ 의 대기중 농도를 기준으로 황사연구 및 황사 경보를 발령하고 있고, 그에 따라 저의 연구도 $PM_{10}(\mu g/m^3)$ 을 기준으로 진행하였습니다. 하지만 $PM_{2.5}(\mu g/m^3)$ 의 농도 또한 연구기관에서 활용하고 있기 때문에 그에 대한 분석을 같이 진행하였습니다. (PM_{10} 은 지름이 10마이크로미터 이하인 입자를 말하며, $PM_{2.5}$ 는 지름이 2.5마이크로미터 이하인 입자를 의미합니다. 보통 PM_{10} 을 미세먼지, $PM_{2.5}$ 를 초미세먼지라고 부르고 있습니다.)

PM_{10} 의 구성요소는 다시 여러 부분으로 나눌 수 있는데, 납(Pb), 카드뮴(Cd), 크롬(Cr), 구리(Cu), 망간(Mn), 철(Fe), 니켈(Ni), 비소(As), 베릴륨(Be), 알루미늄(Al), 칼슘(Ca), 마그네슘(Mg) 등이 포함되어 있습니다. 각 조사기관에서는 매달 1~2주 동안 공기중에 포함된 중금속을 측정하고 있습니다.

“황사의 대기환경영향(2011)”이라는 논문에 의하면 황사에는 다음과 같은 미량원소, 중금속 및 오염물질이 포함될 수 있습니다.

미량원소: Mg, Ca, Na, Ti, Co, Mn, Sr, Ba 등

중금속: Zn, Pb, Cr, Cd, Cu, Ni, V, As, Co, Sb, Sc 등

무기이온: Na^+ , K^+ , Mg^{2+} , Ca^{2+} , 질산염, 황산염 등

잔류성 유기오염물질(POPs): PAHs(다환방향족탄화수소), PCBs(폴리염화비페닐), 다이옥신 등

미량원소와 무기이온 중 일부는 지각에서 기원하지만, 나머지 미량원소와 중금속, POPs 등은 황사 발원지 토양의 오염이나 이동 경로에서 인위적 오염물질과의 혼합에 의해 황사에 포함될 수 있습니다. 따라서 오염지역을 지나지 않은 황사에서는 이들 오염물질의 농도가 낮은 반면, 오염지역을

통과한 황사는 유해 중금속이나 POPs 농도가 높게 나타납니다.

직접적으로 관련된 원소 (지각 기원): Al (알루미늄), Ca (칼슘), Mg (마그네슘), Fe (철), Mn (망간)

간접적으로 관련된 원소 (인위적 오염 또는 토양 오염): Pb (납), Cd (카드뮴), Cr (크롬), Cu (구리), Ni (니켈), As (비소)

황사와의 관련성이 명확하지 않은 원소: Be (베릴륨)

위의 분류는 일반적인 경향을 나타내며, 황사의 발원지와 이동 경로에 따라 원소의 구성과 농도가 달라질 수 있습니다. 예를 들어, 오염된 지역을 통과한 황사는 Pb, Cd, Cr, Cu, Ni, As 등의 농도가 높아질 수 있습니다. 반면, 오염되지 않은 지역에서 발원한 황사는 이들 원소의 농도가 상대적으로 낮을 것입니다.

이 금속 성분들을 토양에서 자연적으로 발생하는 금속 성분과 인공적으로 발생하는(주로 공장에서 배출되는) 금속 성분으로 나눌 수 있습니다. 그러나 몇몇 금속은 두 범주에 모두 속할 수 있습니다. 예를 들어, 철(Fe)이나 알루미늄(Al)은 자연적으로도 존재하지만 공업 활동에 의해 대기 중으로 배출될 수도 있습니다.

3. 코로나 발병의 원인과 그 출현 시기, 그리고 비상사태 종료 시기

코로나 바이러스의 최초 발병시기는 2019년 말로 추정됩니다. 최초사례는 2019년 12월 중국 후베이성 우한시에서 원인불명의 폐렴 사례가 보고 되었으며, 같은 해 12월 31일 중국 보건당국은 WHO에 이러한 사례를 보고하였습니다.

초기확산은 2020년 1월, 중국에서 최초사망 사례를 보고하였으며, 같은 달 태국, 일본, 한국 등 아시아 여러 나라에서 초기 감염사례를 발견하였습니다.

2020년 1월 초, 중국 당국은 신종 코로나 바이러스로 인산 것임을 확인하였으며, 1월 7일 유전자의 염기서열이 공개되었습니다.

2020년 2월 , 유럽과 미국 등 전세계로 확산 되었으며, 3월 11일, WHO는 코로나19를 팬데믹으로 선언하였습니다.

코로나19(COVID-19) 팬데믹은 아직 공식적으로는 종식되지 않았습니다. 다만 2023년 5월 5일 세계보건기구(WHO)는 코로나19 팬데믹에 대한 공중보건 비상사태(PHEIC, Public Health Emergency of International Concern) 선언을 해제하였습니다. 이는 공식적인 종식선언이라고 할 수는 없지만, 전 세계적으로 코로나19의 심각성이 감소했음을 의미합니다.

따라서 현상황은 일부 지역에서는 여전히 변이 바이러스에 의한 감염이 발생함에도 불구하고, 많

은 국가에서 백신 접종과 자연 면역 형성으로 인해 심각한 확산은 억제된 상태라고 볼 수 있습니다. 코로나19는 독감처럼 계절성 질환으로 관리되는 방향으로 전환 중 이라고 볼 수 있습니다.

4. 검증하고자 하는 가설, 구체적 분석범위 및 분석 방법

검증하고자 하는 가설은 코로나 기간 동안 중국의 공장들의 가동이 많이 중단되었다는 뉴스기사가 많았으며 그에 따른 우리나라의 황사피해에 관한 영향이 직접적으로 유의미한 영향이 있었는지에 대한 여부이며, 구체적으로는 PM10과 PM2.5의 농도의 수치변화와 PM10에 포함된 중금속의 농도 수치의 변화를 분석할 계획입니다.

구체적 조사방법은 위에서 선택한 지역들의 연구원들의 자료를 활용하여 PM10과 PM2.5의 농도 대한 코로나 기간 전과 후의 데이터와 PM10에 포함된 중금속의 농도에 대한 자료를 수집하고 각 데이터들을 1. 코로나 발생 전, 2. 코로나 기간, 3. 코로나 발생 후 기간으로 나누어서 T검정 혹은 ANOVA, 혹은 비모수 검정을 실시할 계획입니다. 황사기간은 주로 황사가 발생하는 기간인 3월, 4월, 5월에 집중이 되므로 분석 할 때 각 기간들의 3월 4월 5월의 데이터만 추출하여 분석할 계획입니다.

분석실시 및 결과

1. 인천광역시 보건환경연구원 데이터 분석

(1) 코로나 발생전과 후, 두개의 구간으로 나눈 PM10 농도의 분석

인천광역시 보건환경연구원의 데이터에는 대기오염(PM10, PM2.5, SO₂(ppm), NO₂(ppm), O₃(ppm), CO(ppm)) 데이터를 구할 수 있었습니다. SO₂(아황산가스), NO₂(이산화질소), O₃(오존), CO(일산화탄소)는 황사와 직접적인 관련이 없으며 국가기관에서도 황사경보 혹은 황사일수 판단을 할 때, 위의 원소에 대한 농도는 사용하지 않습니다. 따라서 직접적인 관련이 있는 PM10과 PM2.5에 대한 데이터만을 사용하였습니다.

우선 각 연도와 월별의 PM10과 PM2.5 평균 농도 데이터에 대한 파일을 "data_1.csv" 파일로 만들었으며 사전 테스트로 황사기간인 3월과 4월, 5월의 데이터가 다른 달에 비해서 유의미하게 다른지(농도가 높은지 여부) t-test를 실시하였습니다. (pretest_data_1.R)

아래는 t-test 결과입니다.

```
Welch Two Sample t-test

data:  spring_data$PM10 and other_data$PM10
t = 6.7407, df = 50.86, p-value = 7.076e-09
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 12.19877      Inf
sample estimates:
mean of x mean of y
 58.67873  42.44521

> pm2.5_test

Welch Two Sample t-test

data:  spring_data$PM2.5 and other_data$PM2.5
t = 2.6175, df = 45.212, p-value = 0.006003
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 1.302464      Inf
sample estimates:
mean of x mean of y
 25.33739  21.70369
```


위의 데이터는 PM10 농도에 대한 데이터이며 t값은 6.7407로 매우 크며, p-value는 7.06e-09로 0.05 보다 매우 작은 것으로 보아 3월, 4월, 5월의 PM10의 농도가 다른 달의 PM10 농도보다 유의미하게 높다는 것을 의미합니다. 3, 4, 5월의 PM10 농도의 평균은 58.67873이며, 3, 4, 5월을 제외한 다른 달의 평균은 42.44521 입니다.

두번째 데이터의 결과는 PM2.5의 농도에 대한 결과이며, t값은 2.6175, p-value값은 0.05 보다 작은 0.006003 이므로 3, 4, 5월달의 PM2.5 농도가 다른 달의 PM2.5 농도보다 유의미하게 높다는 것을 의미합니다. 3, 4, 5월의 PM2.5 농도의 평균은 25.33739이며, 다른달의 평균은 21.70369입니다.

같은 데이터에 PM10의 농도 데이터만을 구분하여 정규성 검정(Shapiro test)를 실시하였습니다.(data_1_PM10)

```
> # 정규성 검정
> shapiro.test(spring_data$PM10[spring_data$period == "Before COVID"])

      Shapiro-Wilk normality test

data:  spring_data$PM10[spring_data$period == "Before COVID"]
W = 0.92325, p-value = 0.2159

> shapiro.test(spring_data$PM10[spring_data$period == "During COVID"])

      Shapiro-Wilk normality test

data:  spring_data$PM10[spring_data$period == "During COVID"]
W = 0.91929, p-value = 0.1879
```

2020년 1월을 기준으로 "Before COVID" 와 "During COVID"를 구분하였으며 두 데이터 모두 p-value가 0.05 보다 크기 때문에 데이터가 정규분포를 따른다는 귀무가설을 기각할 수 없음을 의미합니다. 즉, "Before COVID" 와 "During COVID" 기간의 PM10 농도 데이터는 **모두 정규성을 만족**합니다. 따라서 이 데이터를 사용한 t-test는 신뢰할 수 있습니다.

이후 분산의 동질성을 판단하기 위하여 leven test를 실시하였습니다.

```
> leveneTest(PM10 ~ period, data = spring_data)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value  Pr(>F)
group  1   3.123 0.08809 .
      28
---
```



```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

p-value가 0.08809로 0.05보다 높습니다. 따라서 이는 두 그룹 (Before COVID와 During COVID)의 분산이 동일하다는 귀무 가설을 기각할 수 없음을 의미합니다. 즉, Before COVID와 During COVID 기간의 PM10 농도의 분산은 통계적으로 유의미한 차이가 없습니다. 따라서, 두 그룹 간의 평균 차이를 비교할 때 분산의 차이에 따른 영향을 고려할 필요가 없습니다. 따라서 이 데이터를 사용한 t-test는 신뢰할 수 있습니다

이후 같은 데이터로 t-test를 실시하였습니다..

```
> t.test(PM10 ~ period, data = spring_data)

Welch Two Sample t-test

data:  PM10 by period
t = 1.9116, df = 20.535, p-value = 0.06999
alternative hypothesis: true difference in means between group Before COVID and group During COVID is not equal to 0
95 percent confidence interval:
 -0.6772867 15.8290327
sample estimates:
mean in group Before COVID mean in group During COVID
                62.46667                54.89079
```

t-value 가 1.9116으로 Before COVID 기간의 평균이 During COVID 기간의 평균보다 높음을 나타냅니다. p-value가 0.06999으로 0.05보다 큼니다. 이는 두 그룹 간의 평균 차이가 통계적으로 유의미하지 않음을 의미합니다. Before COVID 기간의 PM10 농도 평균은 62.46667 이며, During COVID 기간의 PM10 농도 평균은 54.89079 입니다.

따라서 t-test 결과 Before COVID와 During COVID 라는 두 기간으로 나눈 PM10 농도 평균 차이는 통계적으로 유의미하지 않습니다.

동일한 데이터로 Wilcoxon rank sum test를 실시하였습니다.

```
> wilcox.test(PM10 ~ period, data = spring_data)

Wilcoxon rank sum test with continuity correction

data:  PM10 by period
W = 162, p-value = 0.04209
alternative hypothesis: true location shift is not equal to 0
```

p-value 가 0.04209이므로 0.05보다 작습니다. 이는 두 그룹 간의 위치 차이가 통계적으로 유의미함을 의미합니다. Wilcoxon rank sum test 결과, Before COVID와 During COVID 기간의 PM10 농도의 위치 차이가 통계적으로 유의미합니다.

Wilcoxon rank sum test 해석

Wilcoxon rank sum test는 비모수 검정으로, 데이터가 정규 분포를 따르지 않거나 분산이 동일하지 않을 때도 사용할 수 있는 검정 방법입니다. 그러나, 이 데이터셋의 경우 정규성 검정과 분산 동질성 검정을 통해 데이터가 정규성을 만족하고, 분산도 동일함을 확인했습니다.

t-검정 결과는 p-value가 0.06999이며 0.05 보다 크기 때문에 평균 차이는 통계적으로 유의미하지 않다고 판단되어 집니다.. 하지만 Wilcoxon rank sum test 결과는 p-value가 0.04209로 0.05보다 작으므로 순위 기반으로 인한 위치 차이가 유의미 하다는 것을 의미합니다..

정규성 test 를 활용하여 데이터가 정규 분포를 따른다는 것을 이미 확인하였고, leven test 를 실시하여 두 그룹의 분산이 동일함을 확인하였습니다. 따라서 t-검정을 실시하여 평균 차이가 유의미하지 않음을 밝혀 냈지만, Wilcoxon rank sum test 의 결과는 위치 차이가 유의미함을 나타내었습니다.

따라서 t-검정 결과는 평균의 차이가 유의미하지 않다고 나타내지만, Wilcoxon rank sum test 결과는 위치 차이가 유의미하다고 판단되어 집니다. 이는 데이터의 중앙값이나 분포의 형태에서 차이가 있을 수 있음을 시사합니다. Wilcoxon rank sum test 의 결과를 고려하면, COVID-19 전후의 PM10 농도에 중앙값이나 분포의 형태에서 유의미한 차이가 있음을 나타낼 수 있습니다. 따라서, 평균 차이는 유의미하지 않지만, PM10 농도의 분포 형태나 위치에서 COVID-19 전후의 변화가 있었다고 결론지을 수 있습니다.

(2) 코로나 발생전과 후, 두개의 구간으로 나눈 PM2.5 농도의 분석

위와 같은 "data_1.csv" 파일을 사용하여 PM2.5 농도 데이터만을 구분하여 정규성 검정 (Shapiro test)를 실시하였습니다.(data_1_PM2.5)

```
> shapiro.test(spring_data$PM2.5[spring_data$period == "Before COVID"])
```

```
Shapiro-Wilk normality test
```

```
data: spring_data$PM2.5[spring_data$period == "Before COVID"]
```

```
W = 0.93913, p-value = 0.4869
```

```
> shapiro.test(spring_data$PM2.5[spring_data$period == "During COVID"])
```

```
Shapiro-Wilk normality test
```

```
data: spring_data$PM2.5[spring_data$period == "During COVID"]
```

```
W = 0.88585, p-value = 0.05806
```

첫번째 결과는 Before COVID 기간의 PM2.5 농도로 Shapiro-Wilk 검정 통계량 (W)이 0.93913입니다. p-value는 0.4869 이므로 p-value가 0.05보다 큼니다. 이는 데이터가 정규 분포를 따른다는 귀무 가설을 기각할 수 없음을 의미합니다. 즉, Before COVID 기간의 PM2.5 농도 데이터는 정규성을 만족합니다.

두번째 결과는 During COVID 기간의 PM2.5 농도로 Shapiro-Wilk 검정 통계량 (W)은 0.88585 입니다. p-value가 0.05806 이므로 p-value가 0.05보다 큼니다. 이는 데이터가 정규 분포를 따른다는 귀무 가설을 기각할 수 없음을 의미합니다. 하지만 p-value가 0.05에 매우 가깝기 때문에 정규성에 약간의 의심이 있을 수 있습니다.

Before COVID와 During COVID 기간의 PM2.5 농도 모두 정규성을 만족한다고 할 수 있습니다.

하지만 During COVID 기간의 PM2.5 데이터는 정규성을 강하게 만족한다고 확신하기는 어렵습니다.

동일한 데이터를 사용하여 leven test를 실시하였습니다.

```
> leveneTest(PM2.5 ~ period, data = spring_data)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1    0.1 0.7545
      25
```

p-value가 0.7545이므로 0.05보다 큼니다. 이는 두 그룹 (Before COVID와 During COVID)의 분산이 동질하다는 귀무 가설을 기각할 수 없음을 의미합니다. 즉, Levene's Test 결과, Before COVID와 During COVID 기간의 PM2.5 농도의 분산이 동질하다고 할 수 있습니다. 따라서, 두 그룹 간의 평균 차이를 비교할 때 분산의 차이에 따른 영향을 고려할 필요가 없습니다.

같은 데이터로 t-test를 실시하였습니다.

```
> t.test(PM2.5 ~ period, data = spring_data)

Welch Two Sample t-test

data: PM2.5 by period
t = 1.3004, df = 21.884, p-value = 0.207
alternative hypothesis: true difference in means between group Before COVID and group During COVID is not equal to 0
95 percent confidence interval:
 -1.870659  8.156056
sample estimates:
mean in group Before COVID mean in group During COVID
```

27.08333

23.94063

t-value는 1.3004이고 양수이므로 Before COVID 기간의 평균이 During COVID 기간의 평균보다 높음을 나타냅니다. p-value는 0.207 이므로 0.05보다 큼니다. 이는 두 그룹 간의 평균 차이가 통계적으로 유의미하지 않음을 의미합니다.

동일한 데이터로 Wilcoxon rank sum test를 실시하였습니다

```
> wilcox.test(PM2.5 ~ period, data = spring_data)

Wilcoxon rank sum test with continuity correction

data: PM2.5 by period
W = 116, p-value = 0.213
alternative hypothesis: true location shift is not equal to 0
```

p-value가 0.213 이므로 0.05보다 큼니다. 이는 두 그룹 간의 위치 차이가 통계적으로 유의미하지 않음을 의미합니다.

따라서 PM2.5의 농도는 코로나 이전과 이후로 나누어서 분석하였을 때, Before COVID와 During COVID 기간 모두 정규성을 만족하고, 두 그룹 간의 분산이 동질하였습니다. 하지만 t-검정 결과, 평균 차이가 통계적으로 유의미하지 않았으며(p-value = 0.207), Wilcoxon rank sum test 결과 또한 위치 차이가 통계적으로 유의미하지 않았습니다.(p-value = 0.213).

결론

PM2.5 농도는 COVID-19 전후에 유의미한 변화가 없는 것으로 해석됩니다.

PM10 농도는 Wilcoxon rank sum test 결과에 따라 COVID-19 전후에 유의미한 변화를 보입니다.

따라서, PM10 농도는 COVID-19의 영향으로 평균차이에는 변화가 없지만 위치에 기반한 순위변화가 있었습니다. 하지만, PM2.5 농도는 그 어떠한 변화가 없었다고 결론지을 수 있습니다.

추가분석(기간 삼등분)

위의 결과를 얻고 나서 좀 더 세부적으로 분석을 해봐야 되겠다는 생각이 들어서 구간을 세부분으로 나누고(코로나이전(~19년까지), 코로나기간(20년~23년5월까지), 코로나 이후(~현재까지)), 또 다른 방법으로는 분석할 데이터가 충분한 관측소 별로 나누어서 세부적인 추가 분석을 진행하였습니다.

우선 위와 동일하게 각 연도와 월별의 PM10과 PM2.5 평균 농도 데이터에 대한 파일 "data_1.csv" 파일을 사용하며 기간에 대한 구간을 세부분으로 나누어("Before COVID", "During COVID", "After

COVID") 분석을 진행하였습니다. 우선 PM10 농도 만을 대상으로 Shapiro test를 실시하였습니다.(data_1_PM10_3divide.R)

```
> shapiro.test(spring_data$PM10[spring_data$period == "Before COVID"])

        Shapiro-Wilk normality test

data:  spring_data$PM10[spring_data$period == "Before COVID"]
W = 0.92325, p-value = 0.2159

> shapiro.test(spring_data$PM10[spring_data$period == "During COVID"])

        Shapiro-Wilk normality test

data:  spring_data$PM10[spring_data$period == "During COVID"]
W = 0.87257, p-value = 0.07045

> shapiro.test(spring_data$PM10[spring_data$period == "After COVID"])

        Shapiro-Wilk normality test

data:  spring_data$PM10[spring_data$period == "After COVID"]
W = 0.79339, p-value = 0.09882
```

첫번째 결과는 Before COVID 기간의 PM10 농도이며 Shapiro-Wilk 검정 통계량 (W)은 0.92325 이고 p-value는 0.2159 이므로 0.05보다 큽니다. 이는 데이터가 정규 분포를 따른다는 귀무 가설을 기각할 수 없음을 의미합니다. 즉, Before COVID 기간의 PM10 농도 데이터는 정규성을 만족합니다.

두번째 결과는 During COVID 기간의 PM10 농도이며, Shapiro-Wilk 검정 통계량 (W)은 0.87257 이며, p-value는 0.07045 입니다. p-value가 0.05보다 크지만 매우 근접합니다. 이는 데이터가 정규 분포를 따른다는 귀무 가설을 기각할 수 없음을 의미하지만, 정규성에 약간의 의심이 있을 수 있습니다.

세번째 결과는 After COVID 기간의 PM10 농도이며 Shapiro-Wilk 검정 통계량 (W)은 0.79339 이고, p-value는 0.09882 입니다. p-value가 0.05보다 크기 때문에 데이터가 정규 분포를 따른다는 귀무 가설을 기각할 수 없음을 의미합니다. 즉, After COVID 기간의 PM10 농도 데이터도 정규성을 만족합니다.

Before COVID, During COVID, After COVID 기간의 PM10 농도 모두 정규성을 만족한다고 볼 수 있습니다. 하지만 During COVID 기간의 PM10 데이터는 p-value가 0.05에 매우 근접하여 정규성에 약간의 의심이 있을 수 있습니다

다음으로 leven test를 실시하였습니다.

```
> leveneTest(PM10 ~ period, data = spring_data)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  1.2251 0.3095
      27
```

p-value가 0.3095 이므로 0.05보다 큼니다. 이는 세 그룹 (Before COVID, During COVID, After COVID)의 분산이 동질하다는 귀무 가설을 기각할 수 없음을 의미합니다. 즉, 세 기간 간의 PM10 농도의 분산은 통계적으로 유의미한 차이가 없습니다.

Levene's Test 결과, Before COVID, During COVID, After COVID 기간의 PM10 농도의 분산이 동질하다고 할 수 있습니다. 따라서, 세 그룹 간의 평균 차이를 비교할 때 분산의 차이에 따른 영향을 고려할 필요가 없습니다.

t-test결과

```
> t.test(PM10 ~ period, data = spring_data[spring_data$period != "After COVID", ])

Welch Two Sample t-test

data: PM10 by period
t = 1.4383, df = 14.979, p-value = 0.1709
alternative hypothesis: true difference in means between group Before COVID and group
During COVID is not equal to 0
95 percent confidence interval:
 -3.120556 16.065000
sample estimates:
mean in group Before COVID mean in group During COVID
      62.46667             55.99444
```

p-value는 0.1709 이므로, 0.05보다 큼니다. 이는 두 그룹 간의 평균 차이가 통계적으로 유의미하지 않음을 의미합니다. 95% 신뢰구간은 -3.120556 ~ 16.065000 이므로 신뢰구간이 0을 포함하여 평균 차이가 유의미하지 않음을 나타냅니다. Before COVID 기간의 PM10 농도 평균은 62.46667 이고, During COVID 기간의 PM10 농도 평균은 55.99444 입니다.

결론적으로, Before COVID와 During COVID 간의 PM10 농도의 평균 차이가 통계적으로 유의미하지 않습니다.(p-value = 0.1709). 이는 위의 결과와 동일하게 Before COVID와 During COVID 기간 동안의 PM10 농도 차이가 유의미하지 않다는 것을 나타냅니다.

다음은 세그룹간의 ANOVA 분석결과 입니다.

```
> summary(anova_result)
      Df Sum Sq Mean Sq F value Pr(>F)
period  2    504    251.8   2.108  0.141
```

```

Residuals    27    3225    119.5
> print(anova_result)
Call:
aov(formula = PM10 ~ period, data = spring_data)

Terms:
              period Residuals
Sum of Squares    503.537  3225.264
Deg. of Freedom         2        27

Residual standard error: 10.92951
Estimated effects may be unbalanced

```

Pr(>F)는 p-value를 의미합니다. 0.141은 0.05보다 큽니다. 이는 세 그룹 간의 평균 차이가 통계적으로 유의미하지 않음을 의미합니다.

ANOVA 결과: 세 그룹 (Before COVID, During COVID, After COVID) 간의 PM10 농도 평균 차이는 통계적으로 유의미하지 않음 (p-value = 0.141). 이는 COVID-19 발생 전, 발생 중, 발생 후의 PM10 농도 차이가 유의미하지 않다는 것을 나타냅니다.

다음은 Wilcoxon rank sum test를 실시하였습니다

```

> wilcox.test(PM10 ~ period, data = spring_data[spring_data$period != "After COVID", ])

Wilcoxon rank sum test with continuity correction

data: PM10 by period
W = 125, p-value = 0.09224
alternative hypothesis: true location shift is not equal to 0

```

p-value:는 0.09224이고, 0.05보다 큽니다. 이는 두 그룹 간의 위치 차이가 통계적으로 유의미하지 않음을 의미합니다.

Wilcoxon rank sum test 결과 Before COVID와 During COVID 간의 PM10 농도 위치 차이는 통계적으로 유의미하지 않습니다. (p-value = 0.09224).

정규성 검정은 세 기간 모두 정규성을 만족하였으며, 분산 동질성 검정으로 세 그룹 간 분산이 동질하다는 것을 확인했습니다. t-검정으로 Before COVID와 During COVID 간의 PM10 평균 차이가 유의미하지 않다는 것을 확인했으며, ANOVA 검정으로 Before COVID, During COVID, After COVID 간의 PM10 평균 차이가 유의미하지 않다는 것을 확인했습니다. Wilcoxon rank sum test로 Before COVID와 During COVID 간의 PM10 위치 차이가 유의미하지 않다는 것을 확인했습니다.

이 결과를 바탕으로, COVID-19 발생 전, 발생 중, 발생 후의 PM10 농도는 통계적으로 유의미한 변화를 보이지 않는다고 결론지을 수 있습니다. 이는 COVID-19가 PM10 농도에 미치는 영향이

통계적으로 유의미하지 않음을 시사합니다.

위와 동일한 분석 방법을 사용하여 세부분으로 나누어("Before COVID", "During COVID", "After COVID") PM2.5의 분석을 실시하였습니다.

다음은 Shapiro test 결과입니다.

```
> shapiro.test(spring_data$PM2.5[spring_data$period == "Before COVID"])

      Shapiro-Wilk normality test

data:  spring_data$PM2.5[spring_data$period == "Before COVID"]
W = 0.92746, p-value = 0.2499

> shapiro.test(spring_data$PM2.5[spring_data$period == "During COVID"])

      Shapiro-Wilk normality test

data:  spring_data$PM2.5[spring_data$period == "During COVID"]
W = 0.8769, p-value = 0.09501

> shapiro.test(spring_data$PM2.5[spring_data$period == "After COVID"])

      Shapiro-Wilk normality test

data:  spring_data$PM2.5[spring_data$period == "After COVID"]
W = 0.87583, p-value = 0.3122
```

첫번째 결과는 Before COVID 기간의 PM2.5 농도이며 Shapiro-Wilk 검정 통계량 (W)은 0.92746 이고 p-value는 0.2499 이므로 p-value가 0.05보다 큼니다. 이는 데이터가 정규 분포를 따른다는 귀무 가설을 기각할 수 없음을 의미합니다. 즉, Before COVID 기간의 PM2.5 농도 데이터는 정규성을 만족합니다.

두번째 결과는 During COVID 기간의 PM2.5 농도이며, Shapiro-Wilk 검정 통계량 (W)은 0.8769 이고 p-value는 0.09501 이므로 0.05보다 크지만 0.1보다 작습니다. 이는 데이터가 정규 분포를 따른다는 귀무 가설을 기각할 수 없음을 의미하지만, 정규성에 약간의 의심이 있을 수 있습니다.

세번째 결과는 After COVID 기간의 PM2.5 농도이며, Shapiro-Wilk 검정 통계량 (W)은 0.87583 이며, p-value는 0.3122 입니다. p-value가 0.05보다 큼니다. 이는 데이터가 정규 분포를 따른다는 귀무 가설을 기각할 수 없음을 의미합니다. 즉, After COVID 기간의 PM2.5 농도 데이터는 정규성을 만족합니다.

Before COVID, During COVID, After COVID 기간의 PM2.5 농도 모두 정규성을 만족한다고 볼 수 있습니다. 하지만 During COVID 기간의 PM2.5 데이터는 p-value가 0.1보다 작아 정규성에 약간의

의심이 있을 수 있습니다.

세 기간 모두에서 PM2.5 농도 데이터는 정규성을 만족합니다. During COVID 기간의 PM2.5 데이터의 정규성에 약간의 의심이 있으므로, 비모수 검정을 병행하여 결과를 비교하겠습니다.

다음은 leven 검정입니다.

```
> leveneTest(PM2.5 ~ period, data = spring_data)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2   1.395 0.2658
      26
```

p-value는 0.2658 이고 0.05보다 큽니다. 이는 세 그룹 (Before COVID, During COVID, After COVID)의 분산이 동질하다는 귀무 가설을 기각할 수 없음을 의미합니다. 즉, 세 기간 간의 PM2.5 농도의 분산은 통계적으로 유의미한 차이가 없습니다.

Levene's Test 결과, Before COVID, During COVID, After COVID 기간의 PM2.5 농도의 분산이 동질하다고 할 수 있습니다. 따라서, 세 그룹 간의 평균 차이를 비교할 때 분산의 차이에 따른 영향을 고려할 필요가 없습니다.

다음은 세구간으로 나누었을 때 After COVID를 제외한 두집단(Before COVID, During COVID) 간의 PM2.5 평균 농도의 t-test 결과입니다.

```
> t.test(PM2.5 ~ period, data = spring_data[spring_data$period != "After COVID", ])

Welch Two Sample t-test

data: PM2.5 by period
t = 4.0982, df = 22.754, p-value = 0.0004486
alternative hypothesis: true difference in means between group Before COVID and group
During COVID is not equal to 0
95 percent confidence interval:
 4.839232 14.716057
sample estimates:
mean in group Before COVID mean in group During COVID
          31.51825          21.74061
```

t-value는 4.0982 이고 상당히 큰 양수이므로 Before COVID 기간의 평균이 During COVID 기간의 평균보다 높음을 나타냅니다. p-value는 0.0004486 이고, 0.05보다 훨씬 작습니다. 이는 두 그룹 간의 평균 차이가 통계적으로 유의미함을 강하게 지지합니다. 95% 신뢰구간은 4.839232 ~ 14.716057 이며 신뢰구간이 0을 포함하지 않으며, 이는 두 그룹 간의 평균 차이가 유의미함을 나타냅니다. Before COVID 기간의 PM2.5 농도 평균은 31.51825 이고, During COVID 기간의 PM2.5 농도 평균: 21.74061 입니다.

결론적으로 Before COVID와 During COVID 간의 PM2.5 농도의 평균 차이가 통계적으로 유의미함을 의미합니다.(p-value = 0.0004486). 이는 Before COVID 기간의 PM2.5 농도가 During COVID 기간의 PM2.5 농도보다 유의미하게 높음을 나타냅니다.

다음은 ANOVA test를 실시하였습니다.

```
> anova_result <- aov(PM2.5 ~ period, data = spring_data)
> print(anova_result)
Call:
  aov(formula = PM2.5 ~ period, data = spring_data)

Terms:
              period Residuals
Sum of Squares  677.8098  896.8714
Deg. of Freedom      2      26

Residual standard error: 5.873249
Estimated effects may be unbalanced
> summary(anova_result)
              Df Sum Sq Mean Sq F value    Pr(>F)
period         2  677.8   338.9    9.825 0.000664 ***
Residuals     26  896.9    34.5
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Pr(>F)는 0.000664 이므로 p-value가 0.05보다 훨씬 작습니다. 이는 세 그룹 간의 평균 차이가 통계적으로 유의미함을 강하게 지지합니다.

ANOVA 결과는 세 그룹 (Before COVID, During COVID, After COVID) 간의 PM2.5 농도 평균 차이는 통계적으로 유의미함을 입증하고 있습니다. (p-value = 0.000664). 이는 세 기간 간의 PM2.5 농도에 유의미한 차이가 있음을 나타냅니다

정규성 검정으로 세 기간 모두 정규성을 만족하였습니다. (약간의 의심이 있는 During COVID 기간 존재). 분산 동질성 검정으로 세 그룹 간 분산이 동질하다는 것을 확인하였습니다. t-검정으로 Before COVID와 During COVID 간의 PM2.5 평균 차이가 유의미함을 확인하였습니다. (p-value = 0.0004486). ANOVA 검정으로 Before COVID, During COVID, After COVID 간의 PM2.5 평균 차이가 유의미함을 확인하였습니다. (p-value = 0.000664).

COVID-19 발생 전, 발생 중, 발생 후의 PM2.5 농도는 통계적으로 유의미한 차이를 보입니다.

특히, Before COVID와 During COVID 간의 PM2.5 농도 차이가 유의미하게 큰 것으로 나타났습니다.

이를 통해 COVID-19 발생이 PM2.5 농도에 미친 영향이 있음을 시사할 수 있습니다.

결론

2개의 구간으로 나누었을 때 유의미한 차이가 없던 PM10의 농도는 3개의 구간으로 나누었을 때에도 여전히 t-test 결과 유의적이지 않음으로 해석할 수 있습니다. 또한 2개의 구간에 유의적이었던 Wilcoxon rank sum test 결과에서 3개 구간으로 나누어 코로나 전과 코로나 기간을 비교한 결과 유의적인 결과가 사라진 것을 확인할 수 있습니다. 그리고 wilcoxon rank sum test를 나머지 구간의 그룹들에서 모두 적용한 결과 유의성이 존재하지 않았습니다. 따라서 처음에 검정하였던 2개 구간에서 비모수검정의 PM10 농도의 유의적인 결과가 그 다지 신뢰성이 없었다는 것을 확인할 수 있습니다.

하지만, t-test 결과 PM2.5의 농도가, 조금 더 세밀하게 3개의 기간으로 구분하여 분석하였을 때 유의미한 차이가 발생하였습니다. 이는 PM2.5의 농도에 있어서 앞서 두개의 기간으로 나눈 것이 코로나 진행 시기와 코로나 이후 시기를 합하여 계산하여 코로나 진행 시기의 데이터를 희석하였던 것으로 볼 수 있습니다.

추가분석(관측소별)

따라서 분석을 좀 더 세밀하게 하기 위하여 지금부터 모두 3개의 구간(코로나이전, 코로나기간, 코로나이후)으로 나누어서 진행하도록 하겠습니다.

다음은 각 관측소(지역) 별로 나누어서 세부적인 분석을 진행하도록 하겠습니다. 현재 분석한 데이터는 도로변측정소의 PM10과 PM2.5의 각 7개의 관측소들(서해, 송현, 석바위, 남동, 부평역, 중봉, 임학)의 평균 데이터들을 사용하였으며 조금 더 정확하게 분석하려면 관측소 별로 구분하여 분석해야 올바른 데이터를 얻을 수 있다고 생각하여 위의 분석을 진행하기 위한 데이터를 가공하기 전 데이터로 분석할 수 있도록 "data_2_PM10.csv" 파일로 데이터를 가공하였습니다.

우선 PM10의 농도 데이터를 분석하였습니다. 다만, 송현(b)과 석바위(c), 부평역(e)의 데이터를 제외한 나머지 데이터는 코로나 이전 데이터가 없기 때문에 비교하여 분석을 할 수 없어서 제외하였습니다.

우선 각각의 지역의 PM10 농도 데이터에 대한 Shapiro test를 실시 하였습니다.(data_2_PM10_3divide.R)

```
> shapiro.test(pre_covid$b)
```

```
Shapiro-Wilk normality test
```

```
data: pre_covid$b
```

```
W = 0.9457, p-value = 0.4595
```

```
> shapiro.test(during_covid$b)

      Shapiro-Wilk normality test

data:  during_covid$b
W = 0.88761, p-value = 0.1885

> shapiro.test(post_covid$b)

      Shapiro-Wilk normality test

data:  post_covid$b
W = 0.75, p-value < 2.2e-16

> shapiro.test(pre_covid$c)

      Shapiro-Wilk normality test

data:  pre_covid$c
W = 0.96961, p-value = 0.8522

> shapiro.test(during_covid$c)

      Shapiro-Wilk normality test

data:  during_covid$c
W = 0.82593, p-value = 0.04019

> shapiro.test(post_covid$c)

      Shapiro-Wilk normality test

data:  post_covid$c
W = 0.88107, p-value = 0.3275

> shapiro.test(pre_covid$e)

      Shapiro-Wilk normality test

data:  pre_covid$e
W = 0.97782, p-value = 0.9525

> shapiro.test(during_covid$e)

      Shapiro-Wilk normality test
```

```
data: during_covid$e
W = 0.86779, p-value = 0.1164
```

```
> shapiro.test(post_covid$e)
```

Shapiro-Wilk normality test

```
data: post_covid$e
W = 0.97197, p-value = 0.6788
```

위의 결과에 따르면 송현(b)의 post_covid 데이터가 p-value값이 $2.2e-16$ 이므로 정규성을 충족하지 못하고, 석바위(c)의 during_covid 데이터가 p-value가 0.04019 이므로, 정규성을 충족하지 못한다. 나머지 데이터는 모두 p-value 값이 0.05 이상이므로 정규성을 충족합니다.

각각의 데이터 집합에 대하여 leven test를 실시하였습니다.

```
> print(levene_b)
```

Levene's Test for Homogeneity of Variance (center = median)

	Df	F value	Pr(>F)
--	----	---------	--------

group	2	0.4871	0.6197
-------	---	--------	--------

	27		
--	----	--	--

```
> print(levene_c)
```

Levene's Test for Homogeneity of Variance (center = median)

	Df	F value	Pr(>F)
--	----	---------	--------

group	2	0.9939	0.3833
-------	---	--------	--------

	27		
--	----	--	--

```
> print(levene_e)
```

Levene's Test for Homogeneity of Variance (center = median)

	Df	F value	Pr(>F)
--	----	---------	--------

group	2	0.4937	0.6158
-------	---	--------	--------

	27		
--	----	--	--

Leven test결과 세개의 데이터 집합(송현(b)과 석바위(c), 부평역(e))의 p-value는 모두 0.05 보다 높으며, 각 데이터 집합간의 그룹간 분산이 서로 동질하다는 것을 확인할 수 있습니다.

다음은 일원분산분석(ANOVA)를 실시한 결과입니다.

```
> anova_b <- aov(b ~ group, data = data)
```

```
> print(anova_b)
```

Call:

```
aov(formula = b ~ group, data = data)
```

Terms:

```

              group Residuals
Sum of Squares 2392.311 4103.989
Deg. of Freedom      2      27

Residual standard error: 12.32881
Estimated effects may be unbalanced
> summary(anova_b)
              Df Sum Sq Mean Sq F value Pr(>F)
group          2  2392    1196   7.869 0.00203 **
Residuals     27  4104     152
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> anova_c <- aov(c ~ group, data = data)
> print(anova_c)
Call:
  aov(formula = c ~ group, data = data)

Terms:
              group Residuals
Sum of Squares   586.211 3853.156
Deg. of Freedom      2      27

Residual standard error: 11.94611
Estimated effects may be unbalanced
> summary(anova_c)
              Df Sum Sq Mean Sq F value Pr(>F)
group          2   586    293.1   2.054  0.148
Residuals     27  3853    142.7
> anova_e <- aov(e ~ group, data = data)
> print(anova_e)
Call:
  aov(formula = e ~ group, data = data)

Terms:
              group Residuals
Sum of Squares   418.711 4172.656
Deg. of Freedom      2      27

Residual standard error: 12.43152
Estimated effects may be unbalanced
> summary(anova_e)
              Df Sum Sq Mean Sq F value Pr(>F)
group          2   419    209.4   1.355  0.275
Residuals     27  4173    154.5

```

송현(b)의 ANOVA 결과 $Pr(>F)$ 는 0.00203 이므로 p-value가 0.05보다 작습니다. 이는 세 기간 간

의 평균 차이가 통계적으로 유의미함을 의미합니다.

석바위(c)의 ANOVA 결과 $\Pr(>F)$ 는 0.148이므로 p-value가 0.05보다 큼니다. 이는 세 기간 간의 평균 차이가 통계적으로 유의미하지 않음을 의미합니다.

부평역(e) ANOVA 결과 $\Pr(>F)$ 는 0.275이므로 p-value가 0.05보다 큼니다. 이는 세 기간 간의 평균 차이가 통계적으로 유의미하지 않음을 의미합니다.

따라서 ANOVA 검정결과 송현은 세 기간 간의 평균 차이가 유의미하고, 석바위와 부평역은 각 기간 간의 평균 차이가 유의미하지 않다는 것을 의미합니다.

다음은 Wilcoxon Rank Sum Test 결과입니다.

```
> # 비모수 검정 (Wilcoxon Rank Sum Test) - coin 패키지 사용
> wilcox_test_b <- wilcox_test(b ~ factor(group), data = subset(data, group %in%
c("pre_covid", "during_covid")))
> wilcox_test_c <- wilcox_test(c ~ factor(group), data = subset(data, group %in%
c("pre_covid", "during_covid")))
> wilcox_test_e <- wilcox_test(e ~ factor(group), data = subset(data, group %in%
c("pre_covid", "during_covid")))
> # 결과 출력
> print(wilcox_test_b)

      Asymptotic Wilcoxon-Mann-Whitney Test

data:  b by
      factor(group) (during_covid, pre_covid)
Z = -3.2817, p-value = 0.001032
alternative hypothesis: true mu is not equal to 0

> print(wilcox_test_c)

      Asymptotic Wilcoxon-Mann-Whitney Test

data:  c by
      factor(group) (during_covid, pre_covid)
Z = -2.0594, p-value = 0.03945
alternative hypothesis: true mu is not equal to 0

> print(wilcox_test_e)

      Asymptotic Wilcoxon-Mann-Whitney Test

data:  e by
```

```
factor(group) (during_covid, pre_covid)
Z = -1.7606, p-value = 0.07831
alternative hypothesis: true mu is not equal to 0
```

송현(b)의 Wilcoxon Rank Sum Test 결과 p-value는 0.001032 이므로 p-value가 0.05보다 작습니다.

이는 pre_covid와 during_covid 간의 b의 위치 차이가 통계적으로 유의미함을 의미합니다.

석바위(c)의 Wilcoxon Rank Sum Test 결과 p-value는 0.03945 이므로 p-value가 0.05보다 작습니다.

이는 pre_covid와 during_covid 간의 c의 위치 차이가 통계적으로 유의미함을 의미합니다.

부평역(e) Wilcoxon Rank Sum Test 결과 p-value는 0.07831 이므로 p-value가 0.05보다 큽니다.

이는 pre_covid와 during_covid 간의 변수 e의 위치 차이가 통계적으로 유의미하지 않음을 의미합니다.

송현(b)은 COVID-19 발생 전과 발생 중의 PM10 농도 차이가 유의미하게 다릅니다.

석바위(c)는 COVID-19 발생 전과 발생 중의 PM10 농도 차이가 유의미하게 다릅니다.

부평역(e)은 COVID-19 발생 전과 발생 중의 PM10 농도 차이가 유의미하지 다르지 않습니다.

결론

COVID-19의 영향으로 인해 송현과 석바위 지역의 PM10 농도에서 유의미한 변화가 있었습니다. 반면에 부평역의 PM10 농도는 COVID-19 발생 전과 발생 중에 유의미한 차이가 없었습니다.

위의 결과를 보충하기 위하여 t-test도 진행하였습니다.

```
> # t-검정
> t.test(pre_covid$b, during_covid$b, var.equal = TRUE)

Two Sample t-test

data: pre_covid$b and during_covid$b
t = 4.1195, df = 22, p-value = 0.0004506
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 10.24045 31.00399
sample estimates:
mean of x mean of y
 61.06667  40.44444

> t.test(pre_covid$c, during_covid$c, var.equal = TRUE)
```

```

Two Sample t-test

data:  pre_covid$c and during_covid$c
t = 2.2595, df = 22, p-value = 0.03409
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.8345207 19.4765904
sample estimates:
mean of x mean of y
 58.60000  48.44444

> t.test(pre_covid$e, during_covid$e, var.equal = TRUE)

```

```

Two Sample t-test

data:  pre_covid$e and during_covid$e
t = 1.6103, df = 22, p-value = 0.1216
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-2.48206 19.72650
sample estimates:
mean of x mean of y
 67.73333  59.11111

```

송현(b) t-검정 결과 p-value는 0.0004506 으로 0.05보다 훨씬 작습니다.

이는 pre_covid와 during_covid 간의 변수 b의 평균 차이가 통계적으로 유의미함을 의미합니다.

석바위(c) t-검정 결과 p-value는 0.03409이고 0.05보다 작습니다. 이는 pre_covid와 during_covid 간의 석바위 PM10 농도의 평균 차이가 통계적으로 유의미함을 의미합니다.

부평역(e) t-검정 결과 p-value는 0.1216이고 0.05보다 큼니다. 이는 pre_covid와 during_covid 간의 변수 부평역의 PM10 농도의 평균 차이가 통계적으로 유의미하지 않음을 의미합니다.

결론

위의 Wilcoxon Rank Sum Test와 동일하게 송현과 석바위는 COVID-19의 영향으로 PM10 농도가 유의미하게 감소하였음을 알 수 있습니다.

부평역은 COVID-19가 PM10 농도에 유의미한 영향을 미치지 않았음을 알 수 있습니다.



해석

인천의 서해관측소(a)와 남동관측소(d)는 코로나 기간인 20년 3월에 처음으로 관측을 시작하였습니다. 그리고 중봉관측소(f)는 코로나 기간인 21년 1월에 처음으로 관측을 시작하였으며, 임학관측소(g)는 코로나가 끝난 다음인 23년 9월에 처음 관측을 시작하였습니다. 그리고 서해는 비교적 해안가 이지만 나머지 관측소 들은 비교적 내륙에 있다는 점도 영향을 끼쳤을 것이며 관측이 포함된 시기 또한 전체 데이터에 영향을 미쳤을 것으로 생각되어집니다. 따라서 처음 분석한 인천 전체의 PM10의 농도에 대한 평균 데이터에는 이러한 데이터들의 교란이 발생했을 것이고, 농도 데이터를 서로 간의 상호 희석효과로 인하여 기존의 결과에서 유의미한 차이를 발견할 수 없다고 결론이 나왔던 것으로 추측할 수 있습니다.

또한, 상대적으로 내륙보다는 해안가에 혹은 교외지역에서 황사현상을 잘 발견할 수 있는바, 비교적 서해안 지역에 가까운 송현 지역에서 코로나 기간 동안의 PM10 농도 감소현상을 훨씬 더 잘 발견할 수 있습니다.

다음은 위에 했던 것과 동일하게 관측소 별로 PM2.5의 농도 데이터를 분석하도록 하겠습니다.

PM10과 동일하게 송현(b)과 석바위(c), 부평역(e)의 데이터를 제외한 나머지 데이터는 코로나 이전 데이터가 없기 때문에 비교하여 분석을 할 수 없어서 제외하였습니다. 관련 데이터를 R로 분석할 수 있도록 "data_2_PM2.5.csv" 로 저장하였습니다.

다음은 Shapiro test 결과입니다. (data_2_PM2.5_3divide.R)

```
> shapiro.test(pre_covid$b)
```

Shapiro-Wilk normality test

```
data: pre_covid$b
W = 0.82865, p-value = 0.1047

> shapiro.test(during_covid$b)

      Shapiro-Wilk normality test

data: during_covid$b
W = 0.8783, p-value = 0.1507

> shapiro.test(post_covid$b)

      Shapiro-Wilk normality test

data: post_covid$b
W = 0.85465, p-value = 0.253

> shapiro.test(pre_covid$c)

      Shapiro-Wilk normality test

data: pre_covid$c
W = 0.85118, p-value = 0.03797

> shapiro.test(during_covid$c)

      Shapiro-Wilk normality test

data: during_covid$c
W = 0.91837, p-value = 0.3789

> shapiro.test(post_covid$c)

      Shapiro-Wilk normality test

data: post_covid$c
W = 0.92308, p-value = 0.4633

> shapiro.test(pre_covid$e)

      Shapiro-Wilk normality test

data: pre_covid$e
W = 0.83754, p-value = 0.1244
```

```
> shapiro.test(during_covid$e)

      Shapiro-Wilk normality test

data:  during_covid$e
W = 0.85766, p-value = 0.09041

> shapiro.test(post_covid$e)

      Shapiro-Wilk normality test

data:  post_covid$e
W = 0.9959, p-value = 0.8777
```

위의 결과에 따르면 석바위(c)의 pre_covid 데이터가 p-value가 0.03797 이므로, 정규성을 충족하지 못합니다. 나머지 데이터는 모두 p-value 값이 0.05 이상이므로 정규성을 충족합니다.

Leven test

```
leveneTest.default(y = y, group = group, ...)에서: group coerced to factor.
> print(levene_b)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  0.1819 0.8352
      18
> print(levene_c)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  0.5251 0.5981
      24
> print(levene_e)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  0.1523 0.8598
      18
```

송현(b)과 석바위(c), 부평역(e)의 p-value 가 모두 0.05 보다 크기 때문에 분산이 통계적으로 유의미한 차이가 없음을 의미합니다.

송현(b)과 석바위(c), 부평역(e)의 세 기간 (pre_covid, during_covid, post_covid) 간의 분산이 통계적으로 유의미한 차이가 없습니다.

이는 분산 동질성 가정을 충족하므로, ANOVA를 통해 그룹 간의 평균 차이를 비교하는 것이 적절

합니다.

송현(b)과 석바위(c)는 정규성을 만족하고, 분산 동질성을 만족하므로 ANOVA와 t-검정을 사용할 수 있습니다.

부평역(e)의 경우 pre_covid 기간에서 정규성을 만족하지 않으므로, 비모수 검정 (Wilcoxon Rank Sum Test)을 사용하여 비교하는 것이 적절합니다.

```
> # 일원 분산 분석 (ANOVA)
> anova_b <- aov(b ~ group, data = data)
> print(anova_b)
Call:
  aov(formula = b ~ group, data = data)

Terms:
              group Residuals
Sum of Squares  174.6825  661.8889
Deg. of Freedom      2      18

Residual standard error: 6.06396
Estimated effects may be unbalanced
결측으로 인하여 9 개의 관측치가 삭제되었습니다.
> summary(anova_b)
      Df Sum Sq Mean Sq F value Pr(>F)
group    2  174.7   87.34   2.375  0.121
Residuals 18  661.9   36.77
결측으로 인하여 9 개의 관측치가 삭제되었습니다.
> anova_c <- aov(c ~ group, data = data)
> print(anova_c)
Call:
  aov(formula = c ~ group, data = data)

Terms:
              group Residuals
Sum of Squares  389.3519 1087.3889
Deg. of Freedom      2      24

Residual standard error: 6.73112
Estimated effects may be unbalanced
결측으로 인하여 3 개의 관측치가 삭제되었습니다.
> summary(anova_c)
      Df Sum Sq Mean Sq F value Pr(>F)
group    2  389.4  194.68   4.297 0.0254 *
Residuals 24 1087.4   45.31
```



```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

결측으로 인하여 3 개의 관측치가 삭제되었습니다.

```
> anova_e <- aov(e ~ group, data = data)
> print(anova_e)
```

Call:

```
aov(formula = e ~ group, data = data)
```

Terms:

	group	Residuals
Sum of Squares	165.1429	670.0000
Deg. of Freedom	2	18

Residual standard error: 6.101002

Estimated effects may be unbalanced

결측으로 인하여 9 개의 관측치가 삭제되었습니다.

```
> summary(anova_e)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	165.1	82.57	2.218	0.138
Residuals	18	670.0	37.22		

결측으로 인하여 9 개의 관측치가 삭제되었습니다.

송현(b)의 $Pr(>F)$ 는 0.121이므로 p-value가 0.05보다 큽니다. 이는 세 기간 (pre_covid, during_covid, post_covid) 간의 변수 b의 평균 차이가 통계적으로 유의미하지 않음을 의미합니다.

석바위(c) $Pr(>F)$ 는 0.0254 이므로 p-value가 0.05보다 작습니다. 이는 세 그룹 (pre_covid, during_covid, post_covid) 간의 변수 c의 평균 차이가 통계적으로 유의미함을 의미합니다.

부평역(e)의 $Pr(>F)$ 는 0.138 이므로 p-value가 0.05보다 큽니다. 이는 세 그룹 (pre_covid, during_covid, post_covid) 간의 변수 e의 평균 차이가 통계적으로 유의미하지 않음을 의미합니다.

송현(b)과 석바위(c)는 COVID-19 발생 전, 중, 후의 PM2.5 농도 차이가 유의미하지 않음을 의미하는 반면에 부평역(e)는 COVID-19 발생 전, 중, 후의 PM2.5 농도 차이가 유의미합니다.

다음은 Wilcoxon Rank Sum Test 결과입니다.

```
library(coin)
> # 비모수 검정 (Wilcoxon Rank Sum Test) - coin 패키지 사용
> wilcox_test_b <- wilcox_test(b ~ factor(group), data = subset(data, group %in%
c("pre_covid", "during_covid")))
> wilcox_test_c <- wilcox_test(c ~ factor(group), data = subset(data, group %in%
c("pre_covid", "during_covid")))
> wilcox_test_e <- wilcox_test(e ~ factor(group), data = subset(data, group %in%
c("pre_covid", "during_covid")))
> # 결과 출력
```

```

> print(wilcox_test_b)

      Asymptotic Wilcoxon-Mann-Whitney Test

data:  b by
      factor(group) (during_covid, pre_covid)
Z = -2.0717, p-value = 0.0383
alternative hypothesis: true mu is not equal to 0

> print(wilcox_test_c)

      Asymptotic Wilcoxon-Mann-Whitney Test

data:  c by
      factor(group) (during_covid, pre_covid)
Z = -2.8576, p-value = 0.004269
alternative hypothesis: true mu is not equal to 0

> print(wilcox_test_e)

      Asymptotic Wilcoxon-Mann-Whitney Test

data:  e by
      factor(group) (during_covid, pre_covid)
Z = -1.7773, p-value = 0.07552
alternative hypothesis: true mu is not equal to 0

```

송현(b)의 Wilcoxon Rank Sum Test 결과 p-value는 0.0383 이므로 0.05보다 작습니다.

이는 pre_covid와 during_covid 간의 변수 b의 위치 차이가 통계적으로 유의미함을 의미합니다.

석바위(c) Wilcoxon Rank Sum Test 결과 p-value는 0.004269 이므로 0.05보다 작습니다.

이는 pre_covid와 during_covid 간의 변수 c의 위치 차이가 통계적으로 유의미함을 의미합니다.

부평역(e) Wilcoxon Rank Sum Test 결과 p-value는 0.07552 이므로 0.05보다 큼니다.

이는 pre_covid와 during_covid 간의 변수 e의 위치 차이가 통계적으로 유의미하지 않음을 의미합니다.

COVID-19의 영향으로 인해 송현(b)와 석바위(c)의 PM2.5 농도에서 유의미한 변화가 있었습니다.

부평역(e)의 PM2.5 농도는 COVID-19 발생 전과 발생 중에 유의미한 차이가 없었습니다.

t-test

```

> # t-검정

```

```
> t.test(pre_covid$b, during_covid$b, var.equal = TRUE)

Two Sample t-test

data: pre_covid$b and during_covid$b
t = 1.938, df = 13, p-value = 0.07465
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.7201629 13.2757184
sample estimates:
mean of x mean of y
 25.83333  19.55556
```

```
> t.test(pre_covid$c, during_covid$c, var.equal = TRUE)

Two Sample t-test

data: pre_covid$c and during_covid$c
t = 2.7824, df = 19, p-value = 0.01187
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  2.133612 15.088610
sample estimates:
mean of x mean of y
 28.50000  19.88889
```

```
> t.test(pre_covid$e, during_covid$e, var.equal = TRUE)

Two Sample t-test

data: pre_covid$e and during_covid$e
t = 1.6839, df = 13, p-value = 0.116
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.603475 12.936808
sample estimates:
mean of x mean of y
 31.00000  25.33333
```

송현(b)의 t-검정 결과 p-value가 0.07465 이므로 0.05보다 큼니다. 이는 pre_covid와 during_covid 간의 변수 b의 평균 차이가 통계적으로 유의미하지 않음을 의미합니다.

석바위(c)의 t-검정 결과 p-value가 0.01187 이므로 0.05보다 작습니다. 이는 pre_covid와 during_covid 간의 변수 c의 평균 차이가 통계적으로 유의미함을 의미합니다.

부평역(e)의 t-검정 결과 p-value은 0.116 이므로 0.05보다 큼니다. 이는 pre_covid와 during_covid

간의 변수 e의 평균 차이가 통계적으로 유의미하지 않음을 의미합니다.

결론

Wilcoxon Rank Sum Test 와 t-검정 결과 송현(b)와 부평역(e)는 통계적으로 유의미한 차이가 없는 반면, 석바위(c)는 유의미한 차이가 있습니다.

석바위(c)는 COVID-19 발생 전과 발생 중의 PM2.5 농도에서 유의미한 차이가 있음을 확인할 수 있습니다.

송현(b)와 부평역(e)은 COVID-19 발생이 PM2.5 농도에 유의미한 영향을 미치지 않았음을 알 수 있습니다.



해석

2.5 마이크로미터 이하의 초미세먼지(PM2.5)의 경우에는 우리의 예상과 달리 해안가에 가까운 송현이 아니라, 석바위 지역에서 좀 더 명확하게 코로나 기간 동안의 농도 감소 현상이 발생하고 있습니다. 미세먼지의 농도와 다르게 초미세먼지의 농도에서 왜 송현이 석바위 지역보다 코로나로 인한 농도감소 효과가 적은지에 대한 정확한 원인을 파악하기 위해서는 위의 조사방법 이외의 다른 종류의 추가적인 조사와 분석이 필요하다고 생각합니다.

다만, 송현 지역은 비록 t-test 에서는 유의적이지 않았지만, Wilcoxon Rank Sum Test 에서는 유의적으로 나왔기 때문에 평균의 차이는 없지만 순위 기반으로 인한 위치차이는 존재하는 것이라고 추측할 수 있습니다.

내륙이며 도심지역인 부평역은 미세먼지와 초미세먼지에서 코로나의 유의미한 영향은 다른 지역에 비하여 상대적으로 없었던 것으로 파악이 되고 있습니다.

2. 서울특별시 보건환경연구원 데이터 분석

서울특별시 보건환경연구원의 데이터 역시 대기오염에 관한 데이터를 구할 수 있었습니다. 따라서 PM10과 PM2.5에 대한 데이터만을 사용하여 분석을 하였습니다.

위와 동일한 방법을 사용하되 처음부터 3개 구간으로 나누어서 진행하겠습니다. 먼저 서울 전체 PM10과 PM2.5 데이터의 월 평균 데이터를 사용하여 코로나이전, 코로나기간, 코로나 이후로 나누어서 진행 하겠습니다. 분석을 위하여 "서울특별시 보건환경연구원" 폴더에 "data_1.csv"라는 파일을 생성하였습니다.

우선 PM10의 데이터로 각 기간에 대한 정규성 검정을 진행하겠습니다. (data_1_seoul.R)

```
# 정규성 검정
> shapiro.test(spring_data$PM10[spring_data$period == "Before COVID"])

        Shapiro-Wilk normality test

data:  spring_data$PM10[spring_data$period == "Before COVID"]
W = 0.93632, p-value = 0.3384

> shapiro.test(spring_data$PM10[spring_data$period == "During COVID"])

        Shapiro-Wilk normality test

data:  spring_data$PM10[spring_data$period == "During COVID"]
W = 0.87398, p-value = 0.07342

> shapiro.test(spring_data$PM10[spring_data$period == "After COVID"])

        Shapiro-Wilk normality test

data:  spring_data$PM10[spring_data$period == "After COVID"]
W = 0.7874, p-value = 0.08476
```

Before COVID는 PM10 농도 데이터는 정규성을 만족합니다.

During COVID는 PM10 농도 데이터는 p-value가 0.05보다 크지만 0.1보다 작아 정규성에 약간의 의심이 있습니다.

After COVID는 PM10 농도 데이터는 p-value가 0.05보다 크지만 0.1보다 작아 정규성에 약간의 의심이 있습니다.

세 기간 모두에서 PM10 농도 데이터는 대체로 정규성을 만족하지만, During COVID와 After COVID 기간에서는 정규성에 약간의 의심이 있을 수 있습니다.

이러한 경우, 정규성 가정을 사용하는 검정 (예: t-검정)과 비모수 검정 (예: Wilcoxon Rank Sum Test)을 병행하여 결과를 비교하는 것이 좋습니다.

Leven test

```
leveneTest(PM10 ~ period, data = spring_data)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  0.2241 0.8007
      27
```

p-value가 0.8007 이므로 0.05보다 큽니다. 이는 세 그룹 (Before COVID, During COVID, After COVID) 간의 분산이 통계적으로 유의미한 차이가 없음을 의미합니다.

t-test

```
# 독립 표본 t-검정
> t.test(PM10 ~ period, data = spring_data[spring_data$period != "After COVID", ])

Welch Two Sample t-test

data: PM10 by period
t = 1.7196, df = 21.143, p-value = 0.1001
alternative hypothesis: true difference in means between group Before COVID and group
During COVID is not equal to 0
95 percent confidence interval:
 -1.604866 16.971532
sample estimates:
mean in group Before COVID mean in group During COVID
           55.93333           48.25000
```

PM10 농도의 Before COVID와 During COVID 기간 비교

p-value가 0.1001 이므로 0.05보다 큽니다. 이는 두 그룹 간의 평균 차이가 통계적으로 유의미하지 않음을 의미합니다.

t-검정 결과 Before COVID와 During COVID 간의 PM10 농도의 평균 차이가 통계적으로 유의미하지 않습니다. (p-value = 0.1001).

이는 COVID-19 발생 전과 발생 중의 PM10 농도 차이가 유의미하지 않음을 의미합니다.

ANOVA test 수행

```
> # ANOVA 수행
```



```

> anova_result <- aov(PM10 ~ period, data = spring_data)
> print(anova_result)
Call:
  aov(formula = PM10 ~ period, data = spring_data)

Terms:
              period Residuals
Sum of Squares   638.183   3431.183
Deg. of Freedom         2         27

Residual standard error: 11.27301
Estimated effects may be unbalanced
> summary(anova_result)
              Df Sum Sq Mean Sq F value Pr(>F)
period         2    638   319.1    2.511   0.1 .
Residuals     27   3431   127.1
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

세 그룹 간의 PM10 농도 비교 (Before COVID, During COVID, After COVID)

Pr(>F)가 0.1이므로 0.05보다 큼니다. 이는 세 그룹 간의 평균 차이가 통계적으로 유의미하지 않음을 의미합니다.

Wilcoxon Rank Sum Test

```

library(coin)
> # Before COVID 와 During COVID 비교
> wilcox_test_before_during <- wilcox_test(PM10 ~ factor(period), data =
subset(spring_data, period != "After COVID"))
> print(wilcox_test_before_during)

Asymptotic Wilcoxon-Mann-Whitney Test

data:  PM10 by
      factor(period) (Before COVID, During COVID)
Z = 1.8115, p-value = 0.07006
alternative hypothesis: true mu is not equal to 0

> # Before COVID 와 After COVID 비교
> wilcox_test_before_after <- wilcox_test(PM10 ~ factor(period), data =
subset(spring_data, period %in% c("Before COVID", "After COVID")))
> print(wilcox_test_before_after)

```

Asymptotic Wilcoxon-Mann-Whitney Test

```
data: PM10 by
      factor(period) (After COVID, Before COVID)
Z = -1.7241, p-value = 0.0847
alternative hypothesis: true mu is not equal to 0

> # During COVID와 After COVID 비교
> wilcox_test_during_after <- wilcox_test(PM10 ~ factor(period), data =
subset(spring_data, period %in% c("During COVID", "After COVID")))
> print(wilcox_test_during_after)
```

Asymptotic Wilcoxon-Mann-Whitney Test

```
data: PM10 by
      factor(period) (After COVID, During COVID)
Z = -0.28919, p-value = 0.7724
alternative hypothesis: true mu is not equal to 0
```

Before COVID와 During COVID 간의 비교

p-value가 0.07006 이므로 0.05보다 큽니다. 이는 Before COVID와 During COVID 간의 PM10 농도의 위치 차이가 통계적으로 유의미하지 않음을 의미합니다.

Before COVID와 After COVID 간의 비교

p-value가 0.0847 이므로 0.05보다 큽니다. 이는 Before COVID와 After COVID 간의 PM10 농도의 위치 차이가 통계적으로 유의미하지 않음을 의미합니다.

During COVID와 After COVID 간의 비교

p-value가 0.7724 이므로 0.05보다 큽니다. 이는 During COVID와 After COVID 간의 PM10 농도의 위치 차이가 통계적으로 유의미하지 않음을 의미합니다.

Wilcoxon Rank Sum Test (coin 패키지 사용) 결과, Before COVID, During COVID, After COVID 기간 간의 PM10 농도의 위치 차이는 통계적으로 유의미하지 않습니다.

이는 COVID-19 발생 전, 발생 중, 발생 후의 PM10 농도 차이가 통계적으로 유의미하지 않음을 시사합니다.

결론

ANOVA, t-검정, Wilcoxon Rank Sum Test 결과: 모두 COVID-19 발생 전, 중, 후의 PM10 농도는 통계적으로 유의미한 변화를 보이지 않습니다.

따라서, COVID-19가 서울시 전체 월별 PM10 평균 농도에는 유의미한 영향을 미치지 않았다고 결

론지을 수 있습니다.

위의 분석과 동일하게 서울시 전체 PM2.5의 월별 평균 농도 데이터를 이용하여 분석을 하겠습니다.

우선 PM2.5의 데이터로 각 기간에 대한 정규성 검정을 진행하겠습니다. (data_1_seoul2.R)

```
# 정규성 검정
> shapiro.test(spring_data$PM2.5[spring_data$period == "Before COVID"])

        Shapiro-Wilk normality test

data:  spring_data$PM2.5[spring_data$period == "Before COVID"]
W = 0.90694, p-value = 0.1216

> shapiro.test(spring_data$PM2.5[spring_data$period == "During COVID"])

        Shapiro-Wilk normality test

data:  spring_data$PM2.5[spring_data$period == "During COVID"]
W = 0.82364, p-value = 0.01762

> shapiro.test(spring_data$PM2.5[spring_data$period == "After COVID"])

        Shapiro-Wilk normality test

data:  spring_data$PM2.5[spring_data$period == "After COVID"]
W = 0.87097, p-value = 0.2983
```

Before COVID 기간의 PM2.5 농도

p-value는 0.1216 이므로 0.05보다 큽니다. 이는 데이터가 정규 분포를 따른다는 귀무 가설을 기각할 수 없음을 의미합니다. 즉, Before COVID 기간의 PM2.5 농도 데이터는 정규성을 만족합니다.

During COVID 기간의 PM2.5 농도

p-value가 0.01762 이므로 0.05보다 작습니다. 이는 데이터가 정규 분포를 따르지 않는다는 것을 의미합니다. 즉, During COVID 기간의 PM2.5 농도 데이터는 정규성을 만족하지 않습니다.

After COVID 기간의 PM2.5 농도

p-value가 0.2983 이므로 0.05보다 큽니다. 이는 데이터가 정규 분포를 따른다는 귀무 가설을 기각할 수 없음을 의미합니다. 즉, After COVID 기간의 PM2.5 농도 데이터는 정규성을 만족합니다.

Levene's Test

```
> leveneTest(PM2.5 ~ period, data = spring_data)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  1.2156 0.3122
      27
```

PM2.5 농도의 분산 동질성 검정

p-value가 0.3122 이므로 0.05보다 큽니다. 이는 세 그룹 (Before COVID, During COVID, After COVID) 간의 분산이 통계적으로 유의미한 차이가 없음을 의미합니다.

t-test

```
> # 독립 표본 t-검정
> t.test(PM2.5 ~ period, data = spring_data[spring_data$period != "After COVID", ])

Welch Two Sample t-test

data:  PM2.5 by period
t = 2.9316, df = 24.62, p-value = 0.00718
alternative hypothesis: true difference in means between group Before COVID and group
During COVID is not equal to 0
95 percent confidence interval:
 1.969603 11.297064
sample estimates:
mean in group Before COVID mean in group During COVID
          28.80000          22.16667
```

PM2.5 농도의 Before COVID와 During COVID 기간 비교

p-value가 0.00718 이므로 0.05보다 작습니다. 이는 두 그룹 간의 평균 차이가 통계적으로 유의미함을 의미합니다. 95% 신뢰구간이 1.969603 ~ 11.297064 이고 신뢰구간이 0을 포함하지 않으므로 평균 차이가 유의미함을 나타냅니다. Before COVID 기간의 PM2.5 농도 평균은 28.80000 이며, During COVID 기간의 PM2.5 농도 평균은 22.16667 입니다.

t-검정 결과, Before COVID와 During COVID 간의 PM2.5 농도의 평균 차이가 통계적으로 유의미함을 의미합니다. (p-value = 0.00718). 이는 COVID-19 발생 전과 발생 중의 PM2.5 농도 차이가 유의미하다는 것을 의미합니다.

Wilcoxon Rank Sum Test

```
Wilcoxon rank sum test with continuity correction
```

```
data: PM2.5 by period
W = 147, p-value = 0.005641
alternative hypothesis: true location shift is not equal to 0
```

PM2.5 농도의 Before COVID와 During COVID 기간 비교

p-value가 0.005641 이므로 0.05보다 작습니다. 이는 Before COVID와 During COVID 간의 PM2.5 농도의 위치 차이가 통계적으로 유의미함을 의미합니다.

Wilcoxon Rank Sum Test 결과, Before COVID와 During COVID 간의 PM2.5 농도의 위치 차이가 통계적으로 유의미함을 의미합니다. (p-value = 0.005641). 이는 COVID-19 발생 전과 발생 중의 PM2.5 농도 차이가 유의미하다는 것을 의미합니다.

결론

t-검정 결과, Before COVID와 During COVID 간의 PM2.5 농도의 평균 차이가 유의미함 (p-value = 0.00718).

Wilcoxon Rank Sum Test 결과, Before COVID와 During COVID 간의 PM2.5 농도의 위치 차이가 유의미함 (p-value = 0.005641).

따라서 서울시 전체 초미세먼지인 PM2.5의 농도는 COVID-19의 영향을 받아 해당 기간동안 유의적으로 감소했음을 확인할 수 있습니다.

추가분석(전체 서울지역이 아닌 일부지역)

위의 전체 분석이 아닌 황사의 영향을 많이 받을 것이라 생각되는 서울의 지역구를 7 개를 선정하였다. 선정된 지역구는 다음과 같다. (강서구와 마포구, 영등포구, 용산구, 서대문구, 강남구, 동대문구) 이들 지역구만을 선정하여 코로나 전과 코로나 기간, 코로나 후로 나누어서 통계적 분석을 진행하려 한다. 그렇게 하기 위하여 위의 종합된 데이터가 아닌 원래의 모든 지역구가 표시된 데이터를 csv 파일로 만들었습니다.(data_1_PM10.csv)

위의 조정된 데이터를 가지고 다시 정규성 검정을 실시하였습니다.(data_1_PM10_seoul.R)

```
> # 정규성 검정 (Shapiro-Wilk test)
> shapiro.test(pre_covid$PM10)

Shapiro-Wilk normality test

data: pre_covid$PM10
W = 0.98335, p-value = 9.355e-05
```

```

> shapiro.test(during_covid$PM10)

      Shapiro-Wilk normality test

data:  during_covid$PM10
W = 0.97276, p-value = 9.419e-05

> shapiro.test(post_covid$PM10)

      Shapiro-Wilk normality test

data:  post_covid$PM10
W = 0.90111, p-value = 4.081e-06

```

Before COVID 기간의 PM10 농도

p-value 가 9.355e-05 이므로 0.05 보다 작습니다. 이는 데이터가 정규 분포를 따르지 않는다는 것을 의미합니다. 즉, Before COVID 기간의 PM10 농도 데이터는 정규성을 만족하지 않습니다.

During COVID 기간의 PM10 농도

p-value 는 9.419e-05 이므로 0.05 보다 작습니다. 이는 데이터가 정규 분포를 따르지 않는다는 것을 의미합니다. 즉, During COVID 기간의 PM10 농도 데이터는 정규성을 만족하지 않습니다.

After COVID 기간의 PM10 농도

p-value 는 4.081e-06 이므로 p-value 가 0.05 보다 작습니다. 이는 데이터가 정규 분포를 따르지 않는다는 것을 의미합니다. 즉, After COVID 기간의 PM10 농도 데이터는 정규성을 만족하지 않습니다.

Levene's Test

```

> levene_test <- leveneTest(PM10 ~ group, data = data_long)
경고메시지(들):
leveneTest.default(y = y, group = group, ...)에서: group coerced to factor.
> print(levene_test)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value    Pr(>F)
group  2  11.098 1.763e-05 ***
      795
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

PM10 농도의 분산 동질성 검정

p-value 가 1.763e-05 이므로 0.05 보다 작습니다. 이는 세 그룹 (Before COVID, During COVID, After COVID) 간의 분산이 통계적으로 유의미하게 차이가 있음을 의미합니다.

t-검정을 사용할 수 없으므로, 비모수 검정 (Wilcoxon Rank Sum Test)을 사용하여 그룹 간의 위치 차이를 비교해야 합니다.

ANOVA 결과

```
> # 일원 분산 분석 (ANOVA)
> anova_result <- aov(PM10 ~ group, data = data_long)
> print(anova_result)
Call:
  aov(formula = PM10 ~ group, data = data_long)

Terms:
              group Residuals
Sum of Squares  15242.02 148115.80
Deg. of Freedom      2      795

Residual standard error: 13.64951
Estimated effects may be unbalanced
> summary(anova_result)
              Df Sum Sq Mean Sq F value Pr(>F)
group          2  15242    7621   40.91 <2e-16 ***
Residuals     795 148116     186
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

PM10 농도의 세 그룹 간 비교 (Before COVID, During COVID, After COVID)

Pr(>F)가 < 2e-16 이므로 0.05 보다 작습니다. 이는 세 그룹 간의 평균 차이가 통계적으로 유의미함을 의미합니다.

ANOVA 결과는 세 그룹 간의 PM10 농도의 평균 차이가 통계적으로 유의미함을 의미합니다. (p-value < 2e-16). 이는 COVID-19 발생 전, 발생 중, 발생 후의 PM10 농도에 유의미한 차이가 있음을 의미합니다.

Wilcoxon Rank Sum Test

```
> # 비모수 검정 (Wilcoxon Rank Sum Test)
> library(coin)
> wilcox_test_before_during <- wilcox_test(PM10 ~ factor(group), data = subset(data_long,
group %in% c("pre_covid", "during_covid")))
```

```

> print(wilcox_test_before_during)

      Asymptotic Wilcoxon-Mann-Whitney Test

data:  PM10 by
      factor(group) (during_covid, pre_covid)
Z = -8.1402, p-value = 3.945e-16
alternative hypothesis: true mu is not equal to 0

> wilcox_test_before_after <- wilcox_test(PM10 ~ factor(group), data = subset(data_long,
group %in% c("pre_covid", "post_covid")))
> print(wilcox_test_before_after)

      Asymptotic Wilcoxon-Mann-Whitney Test

data:  PM10 by
      factor(group) (post_covid, pre_covid)
Z = -4.4534, p-value = 8.45e-06
alternative hypothesis: true mu is not equal to 0

> wilcox_test_during_after <- wilcox_test(PM10 ~ factor(group), data = subset(data_long,
group %in% c("during_covid", "post_covid")))
> print(wilcox_test_during_after)

      Asymptotic Wilcoxon-Mann-Whitney Test

data:  PM10 by
      factor(group) (during_covid, post_covid)
Z = -1.5954, p-value = 0.1106
alternative hypothesis: true mu is not equal to 0

```

Before COVID 와 During COVID 간의 비교

p-value 가 3.945e-16 이므로 p-value 가 0.05 보다 작습니다. 이는 Before COVID 와 During COVID 간의 PM10 농도의 위치 차이가 통계적으로 유의미함을 의미합니다.

Before COVID 와 After COVID 간의 비교

p-value 가 8.45e-06 이므로 0.05 보다 작습니다. 이는 Before COVID 와 After COVID 간의 PM10 농도의 위치 차이가 통계적으로 유의미함을 의미합니다.

During COVID 와 After COVID 간의 비교

p-value 가 0.1106 이므로 0.05 보다 큼니다. 이는 During COVID 와 After COVID 간의 PM10 농도의 위치 차이가 통계적으로 유의미하지 않음을 의미합니다.

이는 COVID-19 발생 전후로 PM10 농도에 변화가 있었으나, 발생 중과 발생 후의 변화는 유의미하지 않았음을 나타냅니다.

Kruskal-Wallis 테스트 결과

```
# 비모수 검정 (Kruskal-Wallis test)
> kruskal_test <- kruskal.test(PM10 ~ group, data = data_long)
> print(kruskal_test)
```

Kruskal-Wallis rank sum test

```
data: PM10 by group
Kruskal-Wallis chi-squared = 70.587, df = 2, p-value =
4.701e-16
```

PM10 농도의 세 그룹 간 비교 (Before COVID, During COVID, After COVID)

p-value 가 4.701e-16 이므로 0.05 보다 작습니다. 이는 COVID-19 발생 전, 발생 중, 발생 후의 PM10 농도에 유의미한 차이가 있음을 의미합니다.

Dunn's Test 결과

```
> library(FSA)
> dunn_result <- dunnTest(PM10 ~ group, data = data_long, method = "bonferroni")
경고메시지(들):
group was coerced to a factor.
> print(dunn_result)
Dunn (1964) Kruskal-Wallis multiple comparison
p-values adjusted with the Bonferroni method.
```

	Comparison	Z	P.unadj	P.adj
1	during_covid - post_covid	-1.702597	8.864345e-02	2.659303e-01
2	during_covid - pre_covid	-8.097409	5.614212e-16	1.684264e-15
3	post_covid - pre_covid	-4.523247	6.089804e-06	1.826941e-05

During COVID 와 Post COVID 간의 비교

Padj (adjusted p-value)가 2.659303e-01 이므로 0.05 보다 큼니다. 이는 During COVID 와 Post COVID 간의 PM10 농도의 위치 차이가 통계적으로 유의미하지 않음을 의미합니다.

During COVID 와 Pre COVID 간의 비교

Padj (adjusted p-value) 가 1.684264e-15 이므로 0.05 보다 작습니다. 이는 During COVID 와 Pre COVID 간의 PM10 농도의 위치 차이가 통계적으로 유의미함을 의미합니다.

Post COVID 와 Pre COVID 간의 비교

Padj (adjusted p-value)가 1.826941e-05 이므로 0.05 보다 작습니다. 이는 Post COVID 와 Pre COVID 간의 PM10 농도의 위치 차이가 통계적으로 유의미함을 의미합니다.

이 결론은 COVID-19 발생 전후로 PM10 농도에 변화가 있었으나, 발생 중과 발생 후의 변화는 유의미하지 않았음을 나타냅니다. 이를 바탕으로 COVID-19 가 PM10 농도에 미친 영향을 분석할 수 있습니다.

t-검정 결과

```
> # t-검정
> t_test_before_during <- t.test(pre_covid$PM10, during_covid$PM10, var.equal = TRUE)
> print(t_test_before_during)

      Two Sample t-test

data:  pre_covid$PM10 and during_covid$PM10
t = 8.7556, df = 670, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 7.442825 11.746064
sample estimates:
mean of x mean of y
44.03095  34.43651

> t_test_before_after <- t.test(pre_covid$PM10, post_covid$PM10, var.equal = TRUE)
> print(t_test_before_after)

      Two Sample t-test

data:  pre_covid$PM10 and post_covid$PM10
t = 7.2104, df = 509, p-value = 2.031e-12
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 8.576944 15.001444
sample estimates:
mean of x mean of y
44.03095  32.24176

> t_test_during_after <- t.test(during_covid$PM10, post_covid$PM10, var.equal = TRUE)
> print(t_test_during_after)
```

Two Sample t-test

```
data: during_covid$PM10 and post_covid$PM10
t = 1.6894, df = 341, p-value = 0.09206
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.3606048  4.7501041
sample estimates:
mean of x mean of y
 34.43651  32.24176
```

Before COVID 와 During COVID 간의 비교

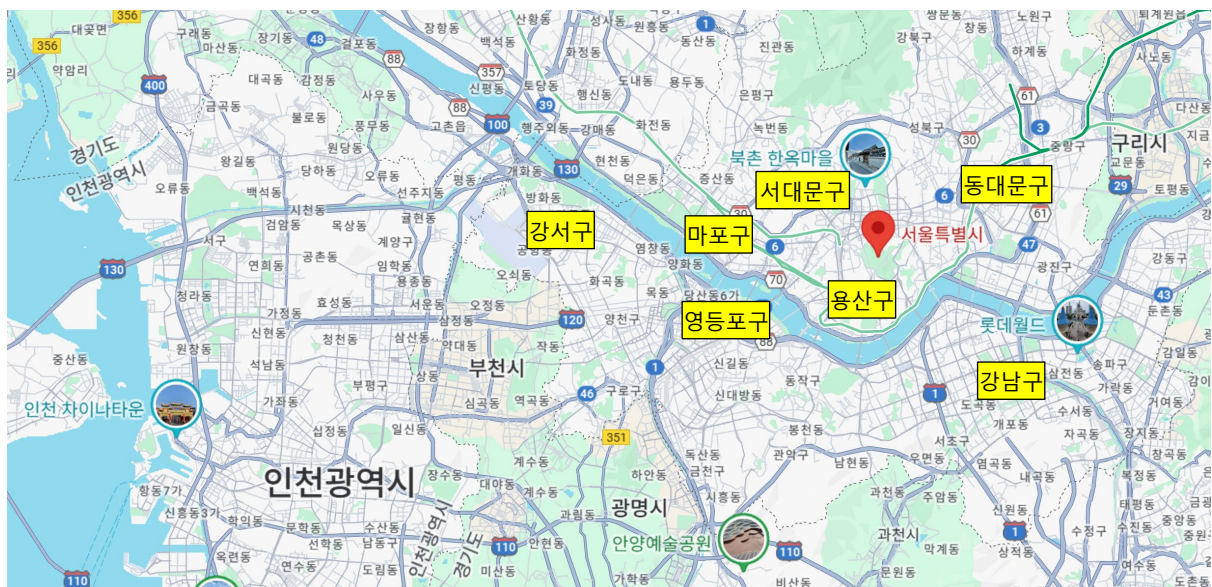
p-value 가 $< 2.2e-16$ 이므로 0.05 보다 작습니다. 이는 Before COVID 와 During COVID 간의 PM10 농도의 평균 차이가 통계적으로 유의미함을 의미합니다.

Before COVID 와 After COVID 간의 비교

p-value 가 $2.031e-12$ 이므로 0.05 보다 작습니다. 이는 Before COVID 와 After COVID 간의 PM10 농도의 평균 차이가 통계적으로 유의미함을 의미합니다

During COVID 와 After COVID 간의 비교

p-value 가 0.09206 이므로 0.05 보다 큼니다. 이는 During COVID 와 After COVID 간의 PM10 농도의 평균 차이가 통계적으로 유의미하지 않음을 의미합니다.



결론

서울 전체 데이터가 아닌 황사에 제일 영향을 받을 만한 지역 7 개를 선별하여 조정된 데이터에서는 ANOVA 결과, Wilcoxon Rank Sum Test, Dunn's Test 결과, t-검정 결과, Kruskal-Wallis 테스트 결과 등 모든 데이터에서 강력한 값으로 같은 결론이 나오고 있습니다. 그것은 코로나 이전과 코로나 기간 이후는 PM10 에 대하여 유의미한 농도의 감소가 있었으며, 코로나 기간과 그 이후에는 유의미한 차이(농도의 변화)는 없었다는 의미입니다.

다음은 위와 동일한 방법으로 동일한 7 개의 지역구 데이터를 선정하여 PM2.5(초미세먼지)의 농도에 대한 분석을 실시하였습니다.(data_1_PM2.5.csv)

Shapiro-Wilk 정규성 검정(data_1_PM2.5_seoul.R)

```
> # 정규성 검정 (Shapiro-Wilk test)
> shapiro.test(pre_covid$PM2.5)

        Shapiro-Wilk normality test

data:  pre_covid$PM2.5
W = 0.98325, p-value = 8.841e-05

> shapiro.test(during_covid$PM2.5)

        Shapiro-Wilk normality test

data:  during_covid$PM2.5
W = 0.98681, p-value = 0.02034

> shapiro.test(post_covid$PM2.5)

        Shapiro-Wilk normality test

data:  post_covid$PM2.5
W = 0.96606, p-value = 0.01783
```

Before COVID 기간의 PM2.5 농도

p-value 은 8.841e-05 이므로 0.05 보다 작습니다. 이는 데이터가 정규 분포를 따르지 않는다는 것을 의미합니다. 즉, Before COVID 기간의 PM2.5 농도 데이터는 정규성을 만족하지 않습니다.

During COVID 기간의 PM2.5 농도

p-value 은 0.02034 이므로 0.05 보다 작습니다. 이는 데이터가 정규 분포를 따르지 않는다는 것을 의미합니다. 즉, During COVID 기간의 PM2.5 농도 데이터는 정규성을 만족하지 않습니다.

After COVID 기간의 PM2.5 농도

p-value 는 0.01783 이므로 0.05 보다 작습니다. 이는 데이터가 정규 분포를 따르지 않는다는 것을 의미합니다. 즉, After COVID 기간의 PM2.5 농도 데이터는 정규성을 만족하지 않습니다.

Levene's Test 결과

```
levene_test <- leveneTest(PM2.5 ~ group, data = data_long)
경고메시지(들):
leveneTest.default(y = y, group = group, ...)에서: group coerced to factor.
> print(levene_test)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value    Pr(>F)
group  2  10.546 3.017e-05 ***
      795
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

PM2.5 농도의 분산 동질성 검정

p-value 가 3.017e-05 이므로 0.05 보다 작습니다. 이는 세 그룹 (Before COVID, During COVID, After COVID) 간의 분산이 통계적으로 유의미하게 차이가 있음을 의미합니다.

ANOVA 결과

```
# 일원 분산 분석 (ANOVA)
> anova_result <- aov(PM2.5 ~ group, data = data_long)
> print(anova_result)
Call:
aov(formula = PM2.5 ~ group, data = data_long)

Terms:
              group Residuals
Sum of Squares  5262.14  37632.41
Deg. of Freedom      2      795

Residual standard error: 6.880143
Estimated effects may be unbalanced
> summary(anova_result)
      Df Sum Sq Mean Sq F value Pr(>F)
```

```
group      2   5262  2631.1   55.58 <2e-16 ***
Residuals 795  37632   47.3
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

PM2.5 농도의 세 그룹 간 비교 (Before COVID, During COVID, After COVID)

Pr(>F)는 < 2e-16 이므로 p-value 가 0.05 보다 작습니다. 이는 COVID-19 발생 전, 발생 중, 발생 후의 PM2.5 농도에 유의미한 차이가 있음을 의미합니다.

Wilcoxon Rank Sum Test

```
> # 비모수 검정 (Wilcoxon Rank Sum Test)
> library(coin)
> wilcox_test_before_during <- wilcox_test(PM2.5 ~ factor(group), data = subset(data_long,
group %in% c("pre_covid", "during_covid")))
> print(wilcox_test_before_during)
```

Asymptotic Wilcoxon-Mann-Whitney Test

```
data: PM2.5 by
      factor(group) (during_covid, pre_covid)
Z = -8.666, p-value < 2.2e-16
alternative hypothesis: true mu is not equal to 0
```

```
> wilcox_test_before_after <- wilcox_test(PM2.5 ~ factor(group), data = subset(data_long,
group %in% c("pre_covid", "post_covid")))
> print(wilcox_test_before_after)
```

Asymptotic Wilcoxon-Mann-Whitney Test

```
data: PM2.5 by
      factor(group) (post_covid, pre_covid)
Z = -6.6806, p-value = 2.38e-11
alternative hypothesis: true mu is not equal to 0
```

```
> wilcox_test_during_after <- wilcox_test(PM2.5 ~ factor(group), data = subset(data_long,
group %in% c("during_covid", "post_covid")))
> print(wilcox_test_during_after)
```

Asymptotic Wilcoxon-Mann-Whitney Test

```
data: PM2.5 by
      factor(group) (during_covid, post_covid)
Z = -0.71599, p-value = 0.474
```

```
alternative hypothesis: true mu is not equal to 0
```

Before COVID 와 During COVID 간의 비교

p-value 가 $< 2.2e-16$ 이므로 0.05 보다 작습니다. 이는 Before COVID 와 During COVID 간의 PM2.5 농도의 위치 차이가 통계적으로 유의미함을 의미합니다.

Before COVID 와 After COVID 간의 비교

p-value 가 $2.38e-11$ 이므로 0.05 보다 작습니다. 이는 Before COVID 와 After COVID 간의 PM2.5 농도의 위치 차이가 통계적으로 유의미함을 의미합니다.

During COVID 와 After COVID 간의 비교

p-value 가 0.474 이므로 0.05 보다 큼니다. 이는 During COVID 와 After COVID 간의 PM2.5 농도의 위치 차이가 통계적으로 유의미하지 않음을 의미합니다.

이는 COVID-19 발생 전후로 PM2.5 농도에 유의미한 변화(감소)가 있었으나, 발생 중과 발생 후의 변화는 유의미하지 않았음을 나타냅니다. 이를 바탕으로 COVID-19 가 PM2.5 농도에 미친 영향을 분석할 수 있습니다.

Kruskal-Wallis 테스트

```
# 비모수 검정 (Kruskal-Wallis test)
> kruskal_test <- kruskal.test(PM2.5 ~ group, data = data_long)
> print(kruskal_test)
```

```
Kruskal-Wallis rank sum test
```

```
data: PM2.5 by group
```

```
Kruskal-Wallis chi-squared = 93.966, df = 2, p-value < 2.2e-16
```

PM2.5 농도의 세 그룹 간 비교 (Before COVID, During COVID, After COVID)

p-value 가 $< 2.2e-16$ 이므로 0.05 보다 작습니다. 이는 세 그룹 간의 PM2.5 농도에 통계적으로 유의미한 차이가 있음을 의미합니다.

이는 COVID-19 발생 전, 발생 중, 발생 후의 PM2.5 농도에 유의미한 차이가 있음을 의미합니다.

Dunn's Test 결과

```
library(FSA)
> dunn_result <- dunnTest(PM2.5 ~ group, data = data_long, method =
"bonferroni")
```

경고메시지(들):

```
group was coerced to a factor.
```

```
> print(dunn_result)
```

```
Dunn (1964) Kruskal-Wallis multiple comparison
```

```
p-values adjusted with the Bonferroni method.
```

	Comparison	Z	P.unadj	P.adj
1	during_covid - post_covid	-0.3967374	6.915611e-01	1.000000e+00
2	during_covid - pre_covid	-8.7987850	1.383054e-18	4.149162e-18
3	post_covid - pre_covid	-6.4761705	9.407973e-11	2.822392e-10

```
>
```

During COVID 와 Post COVID 간의 비교

Padj (adjusted p-value)가 1.000000e+00 이므로 0.05 보다 큽니다. 이는 During COVID 와 Post COVID 간의 PM2.5 농도의 위치 차이가 통계적으로 유의미하지 않음을 의미합니다.

During COVID 와 Pre COVID 간의 비교

Padj (adjusted p-value)는 4.149162e-18 이므로 0.05 보다 작습니다. 이는 During COVID 와 Pre COVID 간의 PM2.5 농도의 위치 차이가 통계적으로 유의미함을 의미합니다.

Post COVID 와 Pre COVID 간의 비교

Padj (adjusted p-value)는 2.822392e-10 이므로 0.05 보다 작습니다. 이는 Post COVID 와 Pre COVID 간의 PM2.5 농도의 위치 차이가 통계적으로 유의미함을 의미합니다.

이 결론은 위의 PM10 농도의 분석결과와 동일하게 COVID-19 발생 전후로 PM2.5 농도에 유의미한 변화(감소)가 있었으나, 발생 중과 발생 후의 변화는 유의미하지 않았음을 나타냅니다.

t-검정

```
> # t-검정
> t_test_before_during <- t.test(pre_covid$PM2.5, during_covid$PM2.5, var.equal = TRUE)
> print(t_test_before_during)
```

Two Sample t-test

data: pre_covid\$PM2.5 and during_covid\$PM2.5

t = 9.2936, df = 670, p-value < 2.2e-16

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

4.183370 6.424566

sample estimates:


```

mean of x mean of y
24.74048 19.43651

> t_test_before_after <- t.test(pre_covid$PM2.5, post_covid$PM2.5, var.equal = TRUE)
> print(t_test_before_after)

Two Sample t-test

data: pre_covid$PM2.5 and post_covid$PM2.5
t = 7.9147, df = 509, p-value = 1.56e-14
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 4.926880 8.180446
sample estimates:
mean of x mean of y
24.74048 18.18681

> t_test_during_after <- t.test(during_covid$PM2.5, post_covid$PM2.5, var.equal = TRUE)
> print(t_test_during_after)

Two Sample t-test

data: during_covid$PM2.5 and post_covid$PM2.5
t = 1.8044, df = 341, p-value = 0.07205
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.1125859 2.6119754
sample estimates:
mean of x mean of y
19.43651 18.18681

```

Before COVID 와 During COVID 간의 비교

p-value 은 $< 2.2e-16$ 이므로 0.05 보다 작습니다. 이는 Before COVID 와 During COVID 간의 PM2.5 농도의 평균 차이가 통계적으로 유의미함을 의미합니다.

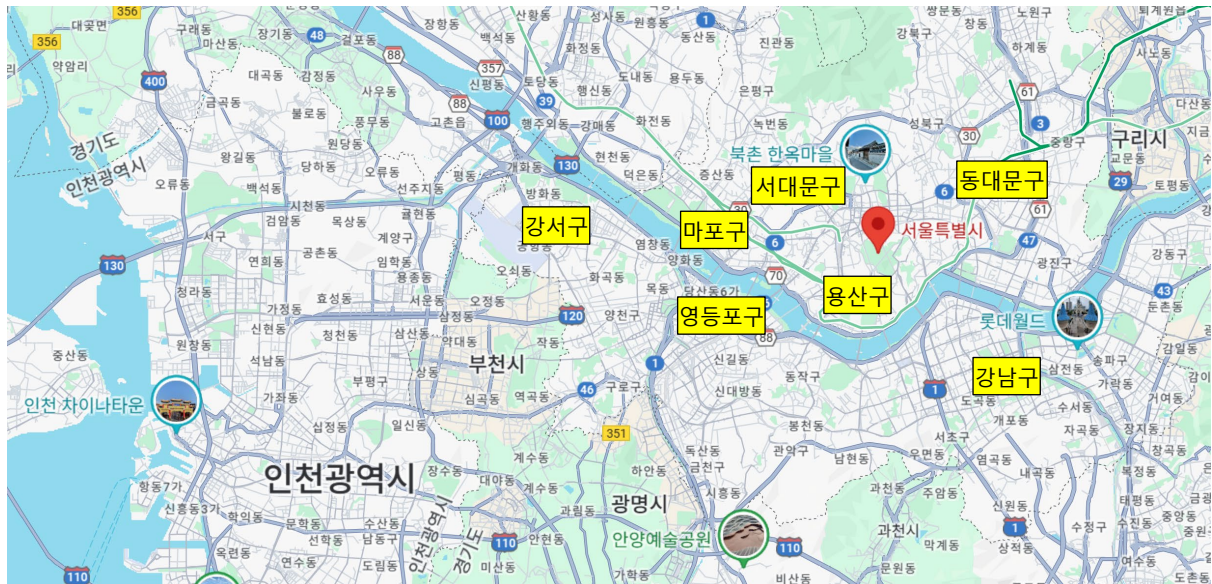
Before COVID 와 After COVID 간의 비교

p-value 는 $1.56e-14$ 이므로 0.05 보다 작습니다. 이는 Before COVID 와 After COVID 간의 PM2.5 농도의 평균 차이가 통계적으로 유의미함을 의미합니다.

During COVID 와 After COVID 간의 비교

p-value 은 0.07205 이므로 0.05 보다 큼니다. 이는 During COVID 와 After COVID 간의 PM2.5 농도의 평균 차이가 통계적으로 유의미하지 않음을 의미합니다.

이 결론은 위의 PM10 농도의 분석결과와 동일하게 진행한 모든 검정결과(정규성 검정 (Shapiro-Wilk), Levene's Test, ANOVA, Kruskal-Wallis 테스트, Dunn's Test, t-검정)에서 코로나 이전과 코로나 기간, 혹은 코로나 이전과 코로나 이후 사이의 변화에서 동일한 결과(유의미한 감소 존재)가 도출됩니다. COVID-19 발생 전후로 PM2.5 농도에 유의미한 변화(감소)가 있었으나, 코로나발생 기간과 코로나 이후 간의 변화는 유의미하지 않았음을 나타냅니다.



해석

서울의 전체 미세먼지의 농도에 대한 코로나 이전과 코로나 이후에 대한 유의미한 변화는 없었지만 초미세먼지의 경우 코로나 이전과 이후 유의미한 감소가 있는 것으로 확인되었습니다.

또한 이러한 변화는 전체 25 개의 지역구가 아닌 지리적으로 황사의 영향을 많이 받을 것이라 생각되어지는 7 개의 지역구를 선별하여 분석하였을 때 미세먼지 혹은 초미세먼지 두 대기오염 농도의 모든 검정결과에서 코로나 이후에 코로나 이전과 비교하여 상당한 감소가 있었던 것으로 확인 할 수 있습니다.

3. 경기도 보건환경연구원 데이터 분석

경기도 보건환경연구원의 데이터는 대기오염과 PM10에 포함된 중금속에 관한 데이터를 구할 수 있었습니다. 따라서 PM10과 PM2.5의 농도 데이터 분석 이외의 PM10에 포함된 중금속 함유량에 대한 분석을 모두 실시 할 수 있었습니다.

위와 동일한 분석을 하기 위하여 경기도 전체의 PM10,과 PM2.5 데이터를 "data_1.csv" 파일로 종합하였습니다.

우선 R프로그램을 사용하여 정규성 검정을 진행하였습니다.(data_1_PM10_KKD.R)

```
# 정규성 검정
> shapiro.test(spring_data$PM10[spring_data$period == "Before COVID"])

        Shapiro-Wilk normality test

data:  spring_data$PM10[spring_data$period == "Before COVID"]
W = 0.63598, p-value = 0.0002528

> shapiro.test(spring_data$PM10[spring_data$period == "During COVID"])

        Shapiro-Wilk normality test

data:  spring_data$PM10[spring_data$period == "During COVID"]
W = 0.49403, p-value = 1.285e-06

> shapiro.test(spring_data$PM10[spring_data$period == "After COVID"])

        Shapiro-Wilk normality test

data:  spring_data$PM10[spring_data$period == "After COVID"]
W = 0.81203, p-value = 0.1436
```

Before COVID 기간의 PM10 농도

p-value가 0.0002528 이므로 0.05보다 작습니다. 이는 데이터가 정규 분포를 따르지 않는다는 것을 의미합니다. 즉, Before COVID 기간의 PM10 농도 데이터는 정규성을 만족하지 않습니다.

During COVID 기간의 PM10 농도

p-value가 1.285e-06 이므로 0.05보다 작습니다. 이는 데이터가 정규 분포를 따르지 않는다는 것을 의미합니다. 즉, During COVID 기간의 PM10 농도 데이터는 정규성을 만족하지 않습니다.

After COVID 기간의 PM10 농도

p-value는 0.1436 이므로 0.05보다 큽니다. 이는 데이터가 정규 분포를 따른다는 것을 의미합니다. 즉, After COVID 기간의 PM10 농도 데이터는 정규성을 만족합니다.

Levene's Test

```
> # 분산 동질성 검정
> library(car)
> leveneTest(PM10 ~ period, data = spring_data)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  0.5417  0.59
      20
```

PM10 농도의 세 그룹 간 분산 동질성 검정

p-value는 0.59이므로 0.05보다 큽니다. 이는 세 그룹 간의 분산 차이가 통계적으로 유의미하지 않음을 의미합니다. 다른 말로 세 그룹 간의 분산이 동질함을 의미합니다.

t-검정

```
> # 독립 표본 t-검정
> t.test(PM10 ~ period, data = spring_data[spring_data$period != "After COVID", ])

Welch Two Sample t-test

data: PM10 by period
t = 0.69687, df = 15.644, p-value = 0.4961
alternative hypothesis: true difference in means between group Before COVID and group
During COVID is not equal to 0
95 percent confidence interval:
 -20.37836  40.28236
sample estimates:
mean in group Before COVID mean in group During COVID
                22.801                12.849
```

Before COVID와 During COVID 간의 비교

p-value는 0.4961 이므로 0.05보다 큽니다. 이는 Before COVID와 During COVID 간의 PM10 농도의 평균 차이가 통계적으로 유의미하지 않음을 의미합니다.

ANOVA

```
> # ANOVA 수행
> anova_result <- aov(PM10 ~ period, data = spring_data)
```

```

> print(anova_result)
Call:
  aov(formula = PM10 ~ period, data = spring_data)

Terms:
              period Residuals
Sum of Squares   3052.389 17583.652
Deg. of Freedom         2         20

Residual standard error: 29.65101
Estimated effects may be unbalanced
> summary(anova_result)
      Df Sum Sq Mean Sq F value Pr(>F)
period    2   3052   1526.2    1.736  0.202
Residuals 20  17584    879.2

```

PM10 농도의 세 그룹 간 비교 (Before COVID, During COVID, After COVID)

Pr(>F)는 0.202 이므로 p-value가 0.05보다 큼니다. 이는 세 기간 간의 PM10 농도의 평균 차이가 통계적으로 유의미하지 않음을 의미합니다.

Wilcoxon Rank Sum Test

```

> # # 비모수 검정 (Mann-Whitney U 테스트)
> wilcox.test(PM10 ~ period, data = spring_data[spring_data$period != "After COVID", ])

      Wilcoxon rank sum test with continuity correction

data:  PM10 by period
W = 48.5, p-value = 0.9696
alternative hypothesis: true location shift is not equal to 0

```

p-value는 0.9696 이므로 0.05보다 큼니다. 이는 Before COVID와 During COVID 간의 PM10 농도의 위치 차이가 통계적으로 유의미하지 않음을 의미합니다.

결론

경기도 전체의 PM10 데이터 분석 결과 코로나 이전과 이후의 유의미한 변화(감소)는 발생하지 않았습니다.

다음은 PM2.5에 대한 동일한 분석을 진행하였습니다.

우선 정규성 검정을 진행하였습니다.(data_1_PN2.5_KKD.R)

```
# 정규성 검정
> shapiro.test(spring_data$PM2.5[spring_data$period == "Before COVID"])

      Shapiro-Wilk normality test

data:  spring_data$PM2.5[spring_data$period == "Before COVID"]
W = 0.6909, p-value = 0.001111

> shapiro.test(spring_data$PM2.5[spring_data$period == "During COVID"])

      Shapiro-Wilk normality test

data:  spring_data$PM2.5[spring_data$period == "During COVID"]
W = 0.51406, p-value = 2.242e-06

> shapiro.test(spring_data$PM2.5[spring_data$period == "After COVID"])

      Shapiro-Wilk normality test

data:  spring_data$PM2.5[spring_data$period == "After COVID"]
W = 0.75, p-value < 2.2e-16
```

Before COVID 기간의 PM2.5 농도

p-value 가 0.001111 이므로 0.05 보다 작습니다. 이는 데이터가 정규 분포를 따르지 않는다는 것을 의미합니다. 즉, Before COVID 기간의 PM2.5 농도 데이터는 정규성을 만족하지 않습니다.

During COVID 기간의 PM2.5 농도

p-value 가 2.242e-06 이므로 0.05 보다 작습니다. 이는 데이터가 정규 분포를 따르지 않는다는 것을 의미합니다. 즉, During COVID 기간의 PM2.5 농도 데이터는 정규성을 만족하지 않습니다.

After COVID 기간의 PM2.5 농도

p-value 는 < 2.2e-16 이므로 0.05 보다 작습니다. 이는 데이터가 정규 분포를 따르지 않는다는 것을 의미합니다. 즉, After COVID 기간의 PM2.5 농도 데이터는 정규성을 만족하지 않습니다.

모두 정규성을 만족하지 않으므로 비모수 검정 (Wilcoxon Rank Sum Test 및 Kruskal-Wallis 테스트)을 사용하여 검정을 실시한다.

Levene's Test

```
분산 동질성 검정
> library(car)
> leveneTest(PM2.5 ~ period, data = spring_data)
```

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  0.8031 0.4619
      20
```

p-value 는 0.4619 이므로 0.05 보다 큽니다. 이는 세 그룹 간의 분산 차이가 통계적으로 유의미하지 않음을 의미합니다. 이는 세 그룹 간의 분산이 동질함을 의미합니다.

Wilcoxon Rank Sum Test

```
> # 비모수 검정 (Mann-Whitney U 테스트)
> wilcox.test(PM2.5 ~ period, data = spring_data[spring_data$period != "After COVID", ])

      Wilcoxon rank sum test with continuity correction

data:  PM2.5 by period
W = 68, p-value = 0.15
alternative hypothesis: true location shift is not equal to 0
```

p-value 가 0.15 이므로 0.05 보다 큽니다 이는 Before COVID 와 During COVID 간의 PM2.5 농도의 위치 차이가 통계적으로 유의미하지 않음을 의미합니다.

결론

경기도 전체의 PM2.5 의 농도는 코로나 기간전과 코로나 기간후의 영향에 유의적인 변화(감소)가 발생하지 않았습니다.

추가연구제안

경기도는 상당히 넓은 지역이기 때문에 황사의 영향을 받는 서부(서북부, 서남부)지역 이외의 전지역을 포함하여 평균 분석하면 당연히 황사의 효과가 희석이 될 것이라 생각되어 집니다. 따라서 안산, 혹은 시흥, 화성 같은 단일 지역을 분석하는 것이 황사효과를 연구하기에는 좀 더 적합하다고 제안되어 집니다.

추가분석(대기 중에 포함된 중금속)

경기도보건환경연구원은 매달 미세먼지에 포함된 중금속의 함유량을 분석한 데이터를 대기환경월보에 포함해서 홈페이지에 PDF 파일로 게시하고 있습니다. 황사의 구성요소에 대한 추가적인 분석을 위하여 해당파일을 종류별로 데이터를 모두 종합하였습니다. (data_2.csv)

측정된 원소는 순서대로 Pb 납 Cd 카드뮴 Cr 크롬 Cu 구리 Mn 망간 Fe 철 Ni 니켈 As 비소 Be 베릴륨 Al 알루미늄 Ca 칼슘 Mg 마그네슘이며, 이중 일반적으로 알려진 분류는 다음과 같습니다.

직접적으로 관련된 원소 (지각 기원): Al (알루미늄), Ca (칼슘), Mg (마그네슘), Fe (철), Mn (망간)

간접적으로 관련된 원소 (인위적 오염 또는 토양 오염): Pb (납), Cd (카드뮴), Cr (크롬), Cu (구리), Ni (니켈), As (비소)

황사와의 관련성이 명확하지 않은 원소: Be (베릴륨)

베릴륨을 제외한 모든 원소가 황사와 직간접적으로 관련이 되어 있지만 실제로 황사기간인 3 월, 4 월, 5 월 사이에 다른 월보다 유의미하게 증가하는 원소가 어떤 것들인지 각 원소들에 대하여 Wilcoxon Rank Sum Test 를 모두 실시하였습니다. (data_2_pretest_KKD.R)

```
> library(dplyr)
```

다음의 패키지를 부착합니다: 'dplyr'

The following object is masked from 'package:car':

recode

The following objects are masked from 'package:timeSeries':

filter, lag

The following objects are masked from 'package:stats':

filter, lag

The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

```
> library(coin)
```

```
> # 데이터 불러오기
```

```
> data <- read.csv("C:\\Users\\김재환\\OneDrive\\바탕 화면\\과제 참고자료\\경영통계\\2024-기말고사\\경기도보건환경연구원\\date_2.csv")
```

```
> # 데이터 구조 확인
```

```
> str(data)
```

```
'data.frame':    92 obs. of  13 variables:
```

```
$ date: chr  "2016-09-01" "2016-10-01" "2016-11-01" "2016-12-01" ...
```

```
$ Pb : num  0.045 0.0425 0.0267 0.0466 0.0367 ...
```

```
$ Cd : num  0.00237 0.0053 0.00103 0.00255 0.00137 ...
```

```
$ Cr : num  0.00713 0.01505 0.00698 0.0066 0.0065 ...
```



```

$ Cu : num 0.0445 0.061 0.0289 0.0479 0.0364 ...
$ Mn : num 0.0366 0.0592 0.0349 0.0382 0.0359 ...
$ Fe : num 0.764 1.242 0.831 1.056 0.85 ...
$ Ni : num 0.0099 0.00845 0.0039 0.0056 0.00413 ...
$ As : num 0.00513 0.0067 0.0039 0.00887 0.00978 ...
$ Be : num 0 0 0 0 0 0 0 0 0 ...
$ Al : num 0.398 0.799 0.624 0.812 0.568 ...
$ Ca : num 0.566 1.063 0.821 1.048 0.869 ...
$ Mg : num 0.149 0.229 0.219 0.229 0.279 ...
> # 날짜 데이터를 Date 형식으로 변환
> data$date <- as.Date(data$date, format="%Y-%m-%d")
> # 3, 4, 5 월 데이터와 다른 달 데이터 구분
> data <- data %>%
+   mutate(month = format(date, "%m")) %>%
+   mutate(period = ifelse(month %in% c("03", "04", "05"), "DustPeriod", "NonDustPeriod"))
> # 원소별로 비모수 검정 (Wilcoxon Rank Sum Test)
> element_list <- names(data)[!(names(data) %in% c("date", "month", "period"))]
> # 결과 저장할 리스트 초기화
> test_results <- list()
> # 각 원소에 대해 Wilcoxon Rank Sum Test 수행
> for (element in element_list) {
+   test <- wilcox_test(data[[element]] ~ factor(data$period), data = data)
+   test_results[[element]] <- test
+ }
> # 결과 출력
> for (element in element_list) {
+   cat("Wilcoxon Rank Sum Test for", element, ":\n")
+   print(test_results[[element]])
+   cat("\n")
+ }
Wilcoxon Rank Sum Test for Pb :

      Asymptotic Wilcoxon-Mann-Whitney Test

data:  data[[element]] by
      factor(data$period) (DustPeriod, NonDustPeriod)
Z = 0.12026, p-value = 0.9043
alternative hypothesis: true mu is not equal to 0

Wilcoxon Rank Sum Test for Cd :

      Asymptotic Wilcoxon-Mann-Whitney Test

data:  data[[element]] by

```

```
      factor(data$period) (DustPeriod, NonDustPeriod)
Z = 1.3414, p-value = 0.1798
alternative hypothesis: true mu is not equal to 0
```

Wilcoxon Rank Sum Test for Cr :

Asymptotic Wilcoxon-Mann-Whitney Test

```
data: data[[element]] by
      factor(data$period) (DustPeriod, NonDustPeriod)
Z = -0.0092506, p-value = 0.9926
alternative hypothesis: true mu is not equal to 0
```

Wilcoxon Rank Sum Test for Cu :

Asymptotic Wilcoxon-Mann-Whitney Test

```
data: data[[element]] by
      factor(data$period) (DustPeriod, NonDustPeriod)
Z = 2.5393, p-value = 0.01111
alternative hypothesis: true mu is not equal to 0
```

Wilcoxon Rank Sum Test for Mn :

Asymptotic Wilcoxon-Mann-Whitney Test

```
data: data[[element]] by
      factor(data$period) (DustPeriod, NonDustPeriod)
Z = 0.4024, p-value = 0.6874
alternative hypothesis: true mu is not equal to 0
```

Wilcoxon Rank Sum Test for Fe :

Asymptotic Wilcoxon-Mann-Whitney Test

```
data: data[[element]] by
      factor(data$period) (DustPeriod, NonDustPeriod)
Z = 1.9935, p-value = 0.04621
alternative hypothesis: true mu is not equal to 0
```

Wilcoxon Rank Sum Test for Ni :

Asymptotic Wilcoxon-Mann-Whitney Test

```
data: data[[element]] by
      factor(data$period) (DustPeriod, NonDustPeriod)
Z = 1.1749, p-value = 0.24
alternative hypothesis: true mu is not equal to 0
```

Wilcoxon Rank Sum Test for As :

Asymptotic Wilcoxon-Mann-Whitney Test

```
data: data[[element]] by
      factor(data$period) (DustPeriod, NonDustPeriod)
Z = -0.24052, p-value = 0.8099
alternative hypothesis: true mu is not equal to 0
```

Wilcoxon Rank Sum Test for Be :

Asymptotic Wilcoxon-Mann-Whitney Test

```
data: data[[element]] by
      factor(data$period) (DustPeriod, NonDustPeriod)
Z = 0.48051, p-value = 0.6309
alternative hypothesis: true mu is not equal to 0
```

Wilcoxon Rank Sum Test for Al :

Asymptotic Wilcoxon-Mann-Whitney Test

```
data: data[[element]] by
      factor(data$period) (DustPeriod, NonDustPeriod)
Z = 2.6688, p-value = 0.007612
alternative hypothesis: true mu is not equal to 0
```

Wilcoxon Rank Sum Test for Ca :

Asymptotic Wilcoxon-Mann-Whitney Test

```
data: data[[element]] by
```

```

      factor(data$period) (DustPeriod, NonDustPeriod)
Z = 2.0768, p-value = 0.03782
alternative hypothesis: true mu is not equal to 0

Wilcoxon Rank Sum Test for Mg :

      Asymptotic Wilcoxon-Mann-Whitney Test

data:  data[[element]] by
      factor(data$period) (DustPeriod, NonDustPeriod)
Z = 2.4375, p-value = 0.01479
alternative hypothesis: true mu is not equal to 0

```

Wilcoxon Rank Sum Test 결과 해석 (중금속 원소)

황사발생기간(3, 4, 5 월)과 비황사기간의 중금속 원소 농도 비교

Pb (납)의 p-value 는 0.9043 이므로 Pb 의 농도는 황사발생기간과 비황사기간 간에 유의미한 차이가 없습니다.

Cd (카드뮴)의 p-value 는 0.1798 이므로 Cd 의 농도는 황사발생기간과 비황사기간 간에 유의미한 차이가 없습니다.

Cr (크롬)의 p-value 는 0.9926 이므로 Cr 의 농도는 황사발생기간과 비황사기간 간에 유의미한 차이가 없습니다.

Cu (구리)의 p-value 는 0.01111 이므로 Cu 의 농도는 황사발생기간에 비황사기간보다 유의미하게 높습니다.

Mn (망간)의 p-value 는 0.6874 이므로 Mn 의 농도는 황사발생기간과 비황사기간 간에 유의미한 차이가 없습니다.

Fe (철)의 p-value 는 0.04621 이므로 Fe 의 농도는 황사발생기간에 비황사기간보다 유의미하게 높습니다.

Ni (니켈)의 p-value 는 0.24 이므로 Ni 의 농도는 황사발생기간과 비황사기간 간에 유의미한 차이가 없습니다.

As (비소)의 p-value 는 0.8099 이므로 As 의 농도는 황사발생기간과 비황사기간 간에 유의미한 차이가 없습니다.

Be (베릴륨)의 p-value 는 0.6309 이므로 Be 의 농도는 황사발생기간과 비황사기간 간에 유의미한 차이가 없습니다.

Al (알루미늄)의 p-value 는 0.007612 이므로 Al 의 농도는 황사발생기간에 비황사기간보다 유의미하게 높습니다.

Ca (칼슘)의 p-value 가 0.03782 이므로 Ca 의 농도는 황사발생기간에 비황사기간보다 유의미하게 높습니다.

Mg (마그네슘)의 p-value 는 0.01479 이므로 Mg 의 농도는 황사발생기간에 비황사기간보다 유의미하게 높습니다.

위의 데이터를 종합하면 비황사기간 보다 황사기간 동안 유의미한 차이가 없는 원소는 Pb (납), Cd (카드뮴), Cr (크롬), Mn (망간), Ni (니켈), As (비소), Be (베릴륨) 이고, 유의미한 차이가 있는 원소는 . Cu (구리), Fe (철), Al (알루미늄), Ca (칼슘), Mg (마그네슘) 입니다. 유의미한 차이가 있는 원소 중에서 Cu(구리)를 제외한 다른 모든 원소들은 지각에서 기원한 원소 입니다. (Fe(철)과 Al(알루미늄)은 지각에서 얻을 수도 있지만 공장에서도 방출될 수 있음)

해석

위의 결과는 코로나의 농도 감소 효과를 반영하지 않았기 때문에 정확한 결과를 반영한다고 볼수는 없습니다. 위에 데이터는 코로나 시기를 구분하지 않고 황사기간과 비황사 기간으로만 구분하여 분석을 실시하였기 때문에 코로나 기간 동안의 감소효과가 중금속 농도에 대한 일반적 황사효과를 찾는 분석에 영향을 주었을 가능성이 있습니다. 때문에 일반적인 황사효과를 찾기 위해서는 코로나 전 기간만을 구분하여 추가로 분석해야 한다고 생각합니다. 또한 황사효과와 코로나의 효과를 찾는 정확한 분석을 하기위해서 기간을 나누고 3, 4, 5 월의 데이터 만을 비교해야 됩니다.

다음은 동일한 데이터를 사용하여, 황사기간 내에서(3 월, 4 월, 5 월의 자료만을 이용) 코로나 발생전과 코로나 발생기간 이후에 증가 혹은 감소의 변화폭이 유의미한 차이가 존재하는 원소가 있는지에 대한 분석을 진행 하였습니다. (data_2_metal_KKD.R)

다음은 원소별로 Kruskal-Wallis 를 실시한 결과입니다.

```
print(test_results)
$Pb

      Asymptotic Kruskal-Wallis Test

data:  data[[element]] by
      factor(data$period) (After COVID, Before COVID, During COVID)
```

```
chi-squared = 6.2345, df = 2, p-value = 0.04428
```

```
$Cd
```

```
Asymptotic Kruskal-Wallis Test
```

```
data: data[[element]] by
      factor(data$period) (After COVID, Before COVID, During COVID)
chi-squared = 0.69449, df = 2, p-value = 0.7066
```

```
$Cr
```

```
Asymptotic Kruskal-Wallis Test
```

```
data: data[[element]] by
      factor(data$period) (After COVID, Before COVID, During COVID)
chi-squared = 12.899, df = 2, p-value = 0.001582
```

```
$Cu
```

```
Asymptotic Kruskal-Wallis Test
```

```
data: data[[element]] by
      factor(data$period) (After COVID, Before COVID, During COVID)
chi-squared = 5.8779, df = 2, p-value = 0.05292
```

```
$Mn
```

```
Asymptotic Kruskal-Wallis Test
```

```
data: data[[element]] by
      factor(data$period) (After COVID, Before COVID, During COVID)
chi-squared = 14.855, df = 2, p-value = 0.0005947
```

```
$Fe
```

```
Asymptotic Kruskal-Wallis Test
```

```
data: data[[element]] by
      factor(data$period) (After COVID, Before COVID, During COVID)
```

```
chi-squared = 7.4326, df = 2, p-value = 0.02432
```

```
$Ni
```

```
Asymptotic Kruskal-Wallis Test
```

```
data: data[[element]] by
      factor(data$period) (After COVID, Before COVID, During COVID)
chi-squared = 15.871, df = 2, p-value = 0.0003578
```

```
$As
```

```
Asymptotic Kruskal-Wallis Test
```

```
data: data[[element]] by
      factor(data$period) (After COVID, Before COVID, During COVID)
chi-squared = 1.2303, df = 2, p-value = 0.5406
```

```
$Be
```

```
Asymptotic Kruskal-Wallis Test
```

```
data: data[[element]] by
      factor(data$period) (After COVID, Before COVID, During COVID)
chi-squared = 11.801, df = 2, p-value = 0.002738
```

```
$Al
```

```
Asymptotic Kruskal-Wallis Test
```

```
data: data[[element]] by
      factor(data$period) (After COVID, Before COVID, During COVID)
chi-squared = 4.5881, df = 2, p-value = 0.1009
```

```
$Ca
```

```
Asymptotic Kruskal-Wallis Test
```

```
data: data[[element]] by
      factor(data$period) (After COVID, Before COVID, During COVID)
```

```
chi-squared = 7.9543, df = 2, p-value = 0.01874
```

```
$Mg
```

```
Asymptotic Kruskal-Wallis Test
```

```
data: data[[element]] by  
      factor(data$period) (After COVID, Before COVID, During COVID)  
chi-squared = 6.2108, df = 2, p-value = 0.04481
```

Kruskal-Wallis Test 결과

유의미한 차이가 있는 원소 (p-value < 0.05)

Pb (납): p-value = 0.04428

Cr (크롬): p-value = 0.001582

Mn (망간): p-value = 0.0005947

Fe (철): p-value = 0.02432

Ni (니켈): p-value = 0.0003578

Be (베릴륨): p-value = 0.002738

Ca (칼슘): p-value = 0.01874

Mg (마그네슘): p-value = 0.04481

유의미한 차이가 없는 원소 (p-value >= 0.05)

Cd (카드뮴): p-value = 0.7066

Cu (구리): p-value = 0.05292

As (비소): p-value = 0.5406

Al (알루미늄): p-value = 0.1009

Kruskal-Wallis Test 결과 중 코로나 기간에 대한 유의미한 차이가 있는 원소들을 대상으로 변화 방향을 확인하기 위하여 Dunn's test 검정을 실시하였습니다.

Dunn's Test 결과

Dunn's Test for Pb :

Dunn (1964) Kruskal-Wallis multiple comparison

p-values adjusted with the Bonferroni method.

	Comparison	Z	P.unadj	P.adj
1	After COVID - Before COVID	-1.9167454	0.05527028	0.16581084
2	After COVID - During COVID	-0.4458753	0.65568730	1.00000000
3	Before COVID - During COVID	2.1968582	0.02803058	0.08409173

Dunn's Test for Cr :

Dunn (1964) Kruskal-Wallis multiple comparison

p-values adjusted with the Bonferroni method.

	Comparison	Z	P.unadj	P.adj
1	After COVID - Before COVID	-2.35907130	0.0183207354	0.054962206
2	After COVID - During COVID	-0.07545583	0.9398520425	1.00000000
3	Before COVID - During COVID	3.38972232	0.0006996345	0.002098904

Dunn's Test for Mn :

Dunn (1964) Kruskal-Wallis multiple comparison

p-values adjusted with the Bonferroni method.

	Comparison	Z	P.unadj	P.adj
1	After COVID - Before COVID	-1.8430245	0.0653254405	0.1959763214
2	After COVID - During COVID	0.7545583	0.4505140721	1.00000000
3	Before COVID - During COVID	3.8271058	0.0001296588	0.0003889763

Dunn's Test for Fe :

Dunn (1964) Kruskal-Wallis multiple comparison

p-values adjusted with the Bonferroni method.

	Comparison	Z	P.unadj	P.adj
1	After COVID - Before COVID	-1.4989932	0.133875391	0.40162617
2	After COVID - During COVID	0.3086829	0.757562732	1.00000000
3	Before COVID - During COVID	2.6706903	0.007569545	0.02270863

Dunn's Test for Ni :

Dunn (1964) Kruskal-Wallis multiple comparison

p-values adjusted with the Bonferroni method.

	Comparison	Z	P.unadj	P.adj
1	After COVID - Before COVID	-2.2859151	0.0222592263	0.0667776790
2	After COVID - During COVID	0.3362045	0.7367166789	1.00000000
3	Before COVID - During COVID	3.8777666	0.0001054198	0.0003162594

Dunn's Test for Be :

Dunn (1964) Kruskal-Wallis multiple comparison
p-values adjusted with the Bonferroni method.

	Comparison	Z	P.unadj	P.adj
1	After COVID - Before COVID	3.435306	0.000591886	0.001775658
2	After COVID - During COVID	2.637111	0.008361558	0.025084675
3	Before COVID - During COVID	-1.273845	0.202718210	0.608154629

Dunn's Test for Ca :

Dunn (1964) Kruskal-Wallis multiple comparison
p-values adjusted with the Bonferroni method.

	Comparison	Z	P.unadj	P.adj
1	After COVID - Before COVID	-1.3761249	0.168782960	0.50634888
2	After COVID - During COVID	0.5213312	0.602136094	1.00000000
3	Before COVID - During COVID	2.7966037	0.005164283	0.01549285

Dunn's Test for Mg :

Dunn (1964) Kruskal-Wallis multiple comparison
p-values adjusted with the Bonferroni method.

	Comparison	Z	P.unadj	P.adj
1	After COVID - Before COVID	-1.2532566	0.21011233	0.63033698
2	After COVID - During COVID	0.4184369	0.67562774	1.00000000
3	Before COVID - During COVID	2.4652526	0.01369167	0.04107501

>

Pb (납) Before COVID - During COVID: p.adj = 0.08409173 (유의미한 차이 없음)

Cr (크롬) Before COVID - During COVID: p.adj = 0.002098904 (유의미한 차이 있음, 코로나 기간 감소)

Mn (망간) Before COVID - During COVID: p.adj = 0.0003889763 (유의미한 차이 있음, 코로나 기간 감소)

Fe (철) Before COVID - During COVID: p.adj = 0.02270863 (유의미한 차이 있음, 코로나 기간 감소)

Ni (니켈) Before COVID - During COVID: p.adj = 0.0003162594 (유의미한 차이 있음, 코로나 기간 감소)

Be (베릴륨) After COVID - Before COVID: p.adj = 0.001775658 (유의미한 차이 있음, 코로나 이후 증가), After COVID - During COVID: p.adj = 0.025084675 (유의미한 차이 있음, 코로나 이후 증가)

Be 의 농도는 코로나 기간에 비해 코로나 이후에 유의미하게 증가.

Ca (칼슘) Before COVID - During COVID: $p_{adj} = 0.01549285$ (유의미한 차이 있음, 코로나 기간 감소)

Mg (마그네슘) Before COVID - During COVID: $p_{adj} = 0.04107501$ (유의미한 차이 있음, 코로나 기간 감소)

해석

Cr (크롬), Mn (망간), Fe (철), Ni (니켈), Ca (칼슘), Mg (마그네슘)의 원소들은 코로나 기간 전 보다 코로나 기간에 감소하였습니다. 이중에서 Cr(크롬)과 Fe(철), Ni(니켈) 등은 공장에서 배출되는 원소이므로 코로나 기간 공장가동 중단에 영향을 감소된 것이라 생각되어 집니다.

Be(베릴륨)은 코로나 이후 증가하였지만 그 양이 미미하여 무시할 만 합니다.

나머지 Pb (납), Cd (카드뮴), Cu (구리), As (비소), Al (알루미늄) 은 유의미한 변화가 존재하지 않았습니다. 이를 종합하면 공장가동 중단의 여파로 중국의 모든 공장이 닫은 것은 아니며, Cr(크롬)과 Fe(철), Ni(니켈)을 배출하는 금속가공 혹은 화학 공장들 혹은 해당 공장들의 특정 제조라인이 코로나 기간동안 주로 조업중단이 되었을 것이라고 추측할 수 있습니다.

추가적인 분석에 의하면 결과에 의한 원소들이 공교롭게도 자동차 엔진의 배기가스에 정확하게 Cr(크롬)과 Fe(철), Ni(니켈) 원소들이 모두 포함되어 있고 Mn (망간)의 경우 연료첨가제와 관련이 있는 것으로 짐작해볼 때 , 코로나 기간 동안의 중국의 이동봉쇄 정책과 감염위험에 따른 자가격리 혹은 재택근무의 증가로 인한 배기가스 감소도 원인 중에 하나라고 추측 할 수 있습니다.

분명한 것은 특정 중금속의 농도에서 황사기간 동안 코로나의 영향으로 감소되었다는 사실입니다.

더 정확한 분석을 위하여 대기오염 수치(PM10, PM2.5)와 비슷하게 중금속의 농도수치 역시, 황사의 영향을 많이 받을 것이라 예상되는 경기도 서부(안산, 시흥, 화성 등)의 지역을 특정하여 분석한다면 조금 더 정확하고 의미있는 결과를 얻을 수 있을 것이라 예상됩니다.

4. 대전광역시 보건환경연구원 데이터 분석

대전광역시 보건환경연구원의 데이터는 대기오염과 PM10에 포함된 중금속에 관한 데이터를 구할 수 있었습니다. 따라서 PM10과 PM2.5의 농도 데이터 분석 이외의 PM10에 포함된 중금속 함유량에 대한 분석을 모두 실시 할 수 있었습니다.

위의 방법과 동일하게 PM10과 PM2.5에 대하여 분석하기 위하여 대전광역시 보건환경연구원 홈페이지에서 관련 파일을 다운로드 받아 데이터를 종합하였습니다. (data_1.csv)

PM10 분석을 위하여 Shapiro test를 실시하였습니다. (data_1_PM10_DJ.R)

```
# 정규성 검정
> shapiro.test(spring_data$PM10[spring_data$period == "Before COVID"])

        Shapiro-Wilk normality test

data:  spring_data$PM10[spring_data$period == "Before COVID"]
W = 0.94275, p-value = 0.4182

> shapiro.test(spring_data$PM10[spring_data$period == "During COVID"])

        Shapiro-Wilk normality test

data:  spring_data$PM10[spring_data$period == "During COVID"]
W = 0.87813, p-value = 0.0985

> shapiro.test(spring_data$PM10[spring_data$period == "After COVID"])

        Shapiro-Wilk normality test

data:  spring_data$PM10[spring_data$period == "After COVID"]
W = 0.93997, p-value = 0.6541
```

PM10 농도의 정규성 검정

Before COVID의 p-value는 0.4182 이므로 정규성을 만족합니다.

During COVID의 p-value 가 0.0985 이므로 정규성을 만족합니다.

After COVID의 p-value 가 0.6541 이므로 정규성을 만족합니다.

Levene's Test

```
# 분산 동질성 검정
> library(car)
> leveneTest(PM10 ~ period, data = spring_data)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  0.6338 0.5383
      27
```

p-value가 0.5383 이므로 분산이 동질적입니다. ANOVA 혹은 T-test를 진행할 수 있습니다.

독립 표본 t-검정

```
# 독립 표본 t-검정
> t.test(PM10 ~ period, data = spring_data[spring_data$period != "After COVID", ])

Welch Two Sample t-test

data: PM10 by period
t = 1.56, df = 16.938, p-value = 0.1373
alternative hypothesis: true difference in means between group Before COVID and group
During COVID is not equal to 0
95 percent confidence interval:
 -2.575988 17.176290
sample estimates:
mean in group Before COVID mean in group During COVID
          54.83457          47.53442
```

p-value가 0.1373 0.05보다 크므로, Before COVID와 During COVID 간의 PM10 농도 차이는 유의미하지 않습니다.

ANOVA

```
# ANOVA 수행
> anova_result <- aov(PM10 ~ period, data = spring_data)
> print(anova_result)
Call:
aov(formula = PM10 ~ period, data = spring_data)

Terms:
              period Residuals
Sum of Squares 1025.988 3162.193
Deg. of Freedom      2      27
```

```

Residual standard error: 10.82212
Estimated effects may be unbalanced
> summary(anova_result)
      Df Sum Sq Mean Sq F value Pr(>F)
period    2   1026    513.0    4.38 0.0225 *
Residuals 27   3162    117.1
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

p-value가 0.0225 이므로 세 기간 간의 PM10 농도 차이는 유의미합니다.

위의 ANOVA 결과에서 유의미한 차이가 발생한 것으로 나왔습니다. ANOVA는 세 개 이상의 그룹 간의 평균 차이를 검정하는데, 귀무가설이 기각되면 어떤 그룹 간에 차이가 있는지 확인하기 위해 사후 검정이 필요합니다. Tukey's HSD Test는 모든 가능한 그룹들의 쌍(pair)들 간의 평균 차이를 비교하여 이 문제를 해결합니다. 또한 Tukey's HSD Test는 다중 비교로 인한 오류를 최소화합니다. 이는 다중 검정에서 발생할 수 있는 1종 오류(거짓 양성)를 줄이는 데 효과적입니다.

따라서 Tukey's HSD Test를 추가로 진행하였습니다.

Tukey's HSD Test 결과 요약

```

> # 사후 분석 (Tukey's HSD Test)
> tukey_result <- TukeyHSD(anova_result)
> print(tukey_result)
Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = PM10 ~ period, data = spring_data)

$period
              diff            lwr            upr            p adj
Before COVID-After COVID 17.173990    2.074469 32.273511 0.0233883
During COVID-After COVID  9.873839   -5.792997 25.540674 0.2787933
During COVID-Before COVID -7.300151  -17.951551  3.351249 0.2238172

```

Before COVID vs After COVID

p-value (p adj)가 0.0233883 이므로 Before COVID와 After COVID 간의 PM10 농도 차이는 유의미합니다. ($p < 0.05$).

During COVID vs After COVID

p-value (p adj)가 0.2787933 이므로 During COVID와 After COVID 간의 PM10 농도 차이는 유의미하지 않습니다. ($p > 0.05$).

During COVID vs Before COVID

p-value (p adj)가 0.2238172 이므로 During COVID와 Before COVID 간의 PM10 농도 차이는 유의미하지 않습니다. ($p > 0.05$).

Before COVID와 After COVID 간의 PM10 농도는 유의미한 차이가 있으며, Before COVID 기간에 비해 After COVID 기간에 PM10 농도가 증가한 것으로 보입니다. 그러나 During COVID와 다른 두 기간 간의 차이는 유의미하지 않습니다.

위의 Tukey's HSD Test 결과 중 Before COVID와 After COVID 간의 PM10 농도 차이는 유의미한 결과가 나왔기 때문에 Before COVID와 After COVID 간의 t-test를 별도로 진행하였습니다.

```
> # 독립 표본 t-검정
> t.test(PM10 ~ period, data = spring_data[spring_data$period != "During COVID", ])
```

```
Welch Two Sample t-test

data: PM10 by period
t = -3.8202, df = 5.7053, p-value = 0.009637
alternative hypothesis: true difference in means between group After COVID and group
Before COVID is not equal to 0
95 percent confidence interval:
 -28.313426  -6.034553
sample estimates:
mean in group After COVID mean in group Before COVID
                37.66058                54.83457
```

p-value가 0.009637 이므로 0.05보다 작습니다. 따라서 After COVID와 Before COVID 간의 PM10 농도 차이는 유의미합니다. After COVID 기간의 PM10 농도가 Before COVID 기간에 비해 유의미하게 낮다고 할 수 있습니다.

결론

대전광역시 보건환경데이터 분석결과 대전시 전체의 PM10의 농도는 코로나 발생전과 코로나 기간간의 유의미한 변화(감소)가 발생하지 않았지만 코로나 발생전(19년 이전)과 코로나 이후(23년 5월이후)의 PM10 농도 데이터에서는 유의미한 감소가 있었습니다. 또한 코로나 기간과 코로나 이후의 t-test를 추가로 실시한 결과 유의미하지 않았습니다. 하지만 농도의 평균에 있어서 뚜렷하게 시계열적으로 감소하는 경향이 있었습니다.

해석

따라서 대전광역시는 코로나로 인한 감소 효과가 천천히 진행되어 코로나기간 동안은 PM10의 농도가 감소는 하였지만 유의미한 감소는 아니었습니다. 그리고 코로나기간 이후가 되어서야 유의

미한 차이가 발생한 것이라 해석할 수 있습니다. 따라서 대전광역시는 코로나(COVID-19)로 인한 미세먼지 농도의 감소효과가 다른 지역에 비해서 천천히 진행되었다고 볼 수 있습니다.

다음은 위의 동일한 데이터로 PM2.5의 농도를 분석하였습니다.

우선 정규성 검정을 하기위하여 Shapiro test를 실시하였습니다. (data_1_PM2.5_DJ.R)

```
> # 정규성 검정
> shapiro.test(spring_data$PM2.5[spring_data$period == "Before COVID"])

      Shapiro-Wilk normality test

data:  spring_data$PM2.5[spring_data$period == "Before COVID"]
W = 0.94708, p-value = 0.4797

> shapiro.test(spring_data$PM2.5[spring_data$period == "During COVID"])

      Shapiro-Wilk normality test

data:  spring_data$PM2.5[spring_data$period == "During COVID"]
W = 0.91005, p-value = 0.2137

> shapiro.test(spring_data$PM2.5[spring_data$period == "After COVID"])

      Shapiro-Wilk normality test

data:  spring_data$PM2.5[spring_data$period == "After COVID"]
W = 0.94515, p-value = 0.5485
```

Before COVID

p-value는 0.4797 이므로 0.05보다 크므로, 정규성을 만족합니다.

During COVID

p-value는 0.2137 이므로 0.05보다 크므로, 정규성을 만족합니다.

After COVID

p-value는 0.5485 이므로 0.05보다 크므로, 정규성을 만족합니다.

Leven's test

```
> # 분산 동질성 검정
> library(car)
> leveneTest(PM2.5 ~ period, data = spring_data)
```



```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  0.9023 0.4175
      27
```

분산의 동질성 검정결과 p-value가 0.4175 이므로 0.05보다 큽니다. 세 그룹 간의 분산은 동질적입니다. 따라서 ANOVA를 수행하는 데 문제가 없습니다.

독립표본 t-test

```
# 독립 표본 t-검정
> t.test(PM2.5 ~ period, data = spring_data[spring_data$period != "After COVID", ])

Welch Two Sample t-test

data: PM2.5 by period
t = 3.1398, df = 24.748, p-value = 0.004338
alternative hypothesis: true difference in means between group Before COVID and group
During COVID is not equal to 0
95 percent confidence interval:
 2.008402 9.677758
sample estimates:
mean in group Before COVID mean in group During COVID
                25.40160                19.55852

> t.test(PM2.5 ~ period, data = spring_data[spring_data$period != "Before COVID", ])

Welch Two Sample t-test

data: PM2.5 by period
t = -1.4945, df = 4.9426, p-value = 0.1959
alternative hypothesis: true difference in means between group After COVID and group
During COVID is not equal to 0
95 percent confidence interval:
 -7.624745  2.030742
sample estimates:
mean in group After COVID mean in group During COVID
                16.76152                19.55852

> t.test(PM2.5 ~ period, data = spring_data[spring_data$period != "During COVID", ])

Welch Two Sample t-test

data: PM2.5 by period
t = -4.1835, df = 6.9036, p-value = 0.004248
```

```

alternative hypothesis: true difference in means between group After COVID and group
Before COVID is not equal to 0
95 percent confidence interval:
 -13.537508 -3.742657
sample estimates:
 mean in group After COVID mean in group Before COVID
                16.76152                25.40160

```

Before COVID vs During COVID

p-value는 0.004338 이므로 0.05보다 작습니다. Before COVID와 During COVID 간의 PM2.5 농도 차이는 유의미합니다. During COVID 기간의 PM2.5 농도가 Before COVID 기간에 비해 유의미하게 낮습니다.

After COVID vs During COVID

p-value가 0.1959 이므로 0.05보다 큼니다. After COVID와 During COVID 간의 PM2.5 농도 차이는 유의미하지 않습니다.

After COVID vs Before COVID

p-value는 0.004248 이므로 0.05보다 작습니다. After COVID와 Before COVID 간의 PM2.5 농도 차이는 유의미합니다. After COVID 기간의 PM2.5 농도가 Before COVID 기간에 비해 유의미하게 낮습니다.

ANOVA test

```

# ANOVA 수행
> anova_result <- aov(PM2.5 ~ period, data = spring_data)
> summary(anova_result)
              Df Sum Sq Mean Sq F value    Pr(>F)
period          2   326.2   163.11    6.937 0.00371 **
Residuals      27   634.9    23.51
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

p-value가 0.0037 이므로 0.05보다 작습니다. 세 그룹 간의 PM2.5 농도 차이는 유의미합니다. 이를 통해 사후 분석 (Tukey's HSD Test)을 수행할 수 있습니다.

Tukey's HSD Test

```

> # 사후 분석 (Tukey's HSD Test)
> tukey_result_pm25 <- TukeyHSD(anova_result)
> print(tukey_result_pm25)
Tukey multiple comparisons of means
95% family-wise confidence level

```

```
Fit: aov(formula = PM2.5 ~ period, data = spring_data)
```

```
$period
```

	diff	lwr	upr	p adj
Before COVID-After COVID	8.640082	1.036202	16.243962	0.0235376
During COVID-After COVID	2.797002	-4.963676	10.557679	0.6488973
During COVID-Before COVID	-5.843080	-10.499487	-1.186674	0.0117629

Before COVID - After COVID

p-value (조정)는 0.0235376 이므로 유의미한 차이가 있습니다. Before COVID 기간의 PM2.5 농도가 After COVID 기간에 비해 유의미하게 높습니다.(역으로 말하면 After COVID기간에 감소하였습니다.)

During COVID - After COVID

p-value (조정)는 0.6488973 이므로 유의미한 차이가 없습니다.

During COVID - Before COVID

p-value (조정)는 0.0117629 이므로 유의미한 차이 있습니다. During COVID 기간의 PM2.5 농도가 Before COVID 기간에 비해 유의미하게 낮습니다.

결론

대전광역시의 PM2.5의 농도는 황사기간 동안만을 구분하여 분석을 실시한 결과 코로나기간 동안 유의미한 농도의 감소가 있었음을 확인할 수 있습니다.

해석

황사기간 동안의 대전광역시의 전체 PM10과 PM2.5의 농도 데이터를 분석한 결과 PM10의 농도 변화는 코로나로 인한 영향으로부터 천천히 진행되어 유의미한 감소는 코로나 종식기간인 23년 5월 이후의 데이터부터 유의미한 감소 결과가 발생하였으며, PM2.5의 농도의 변화는 코로나의 영향으로 유의미하게 감소하였습니다.

PM10농도가 코로나 기간동안 대전광역시 지역에서 감소효과가 더디게 진행되는 원인은 추가적인 조사와 분석이 필요하다고 생각합니다.

추가분석(대기 중에 포함된 중금속)

대전광역시 보건환경연구원 홈페이지 자료실에는 2004년부터 2024년 5월까지의 중금속 농도 데이터가 엑셀파일로 올라와 있습니다. 해당자료를 종합하여 csv 파일인 "data_2.csv" 로 "대전광역

시보건환경연구원" 폴더에 저장하였습니다.

저장한 "data_2.csv" 파일을 이용하여 Kruskal-Wallis Test를 실시하였습니다.

\$Pb

Asymptotic Kruskal-Wallis Test

data: Pb by
period (After COVID, Before COVID, During COVID)
chi-squared = 25.579, df = 2, p-value = 2.79e-06

\$Cd

Asymptotic Kruskal-Wallis Test

data: Cd by
period (After COVID, Before COVID, During COVID)
chi-squared = 26.906, df = 2, p-value = 1.437e-06

\$Cr

Asymptotic Kruskal-Wallis Test

data: Cr by
period (After COVID, Before COVID, During COVID)
chi-squared = 7.0185, df = 2, p-value = 0.02992

\$Cu

Asymptotic Kruskal-Wallis Test

data: Cu by
period (After COVID, Before COVID, During COVID)
chi-squared = 17.71, df = 2, p-value = 0.0001427

\$Mn

Asymptotic Kruskal-Wallis Test

```
data: Mn by
      period (After COVID, Before COVID, During COVID)
chi-squared = 15.535, df = 2, p-value = 0.0004234
```

\$Fe

Asymptotic Kruskal-Wallis Test

```
data: Fe by
      period (After COVID, Before COVID, During COVID)
chi-squared = 14.509, df = 2, p-value = 0.0007069
```

\$Ni

Asymptotic Kruskal-Wallis Test

```
data: Ni by
      period (After COVID, Before COVID, During COVID)
chi-squared = 15.994, df = 2, p-value = 0.0003365
```

\$As

Asymptotic Kruskal-Wallis Test

```
data: As by
      period (After COVID, Before COVID, During COVID)
chi-squared = 7.3423, df = 2, p-value = 0.02545
```

\$Be

Asymptotic Kruskal-Wallis Test

```
data: Be by
      period (After COVID, Before COVID, During COVID)
chi-squared = 1.1382, df = 2, p-value = 0.566
```

\$Al

Asymptotic Kruskal-Wallis Test

```
data: Al by
      period (After COVID, Before COVID, During COVID)
chi-squared = 8.3563, df = 2, p-value = 0.01533
```

\$Ca

Asymptotic Kruskal-Wallis Test

```
data: Ca by
      period (After COVID, Before COVID, During COVID)
chi-squared = 4.934, df = 2, p-value = 0.08484
```

\$Mg

Asymptotic Kruskal-Wallis Test

```
data: Mg by
      period (After COVID, Before COVID, During COVID)
chi-squared = 6.1658, df = 2, p-value = 0.04583
```

유의미한 차이가 있는 원소 (p-value < 0.05)

Pb (납)

Cd (카드뮴)

Cr (크롬)

Cu (구리)

Mn (망간)

Fe (철)

Ni (니켈)

As (비소)

Al (알루미늄)

Mg (마그네슘)

유의미한 차이가 없는 원소 (p-value >= 0.05)

Be (베릴륨)

Ca (칼슘)

유의미한 차이가 있는 원소에 대한 Dunn's Test를 수행하였습니다.

\$Pb

Dunn (1964) Kruskal-Wallis multiple comparison
p-values adjusted with the Bonferroni method.

	Comparison	Z	P.unadj	P.adj
1	After COVID - Before COVID	-3.4550136	5.502651e-04	0.0016507954
2	After COVID - During COVID	-0.7432444	4.573337e-01	1.0000000000
3	Before COVID - During COVID	4.0806781	4.490451e-05	0.0001347135

\$Cd

Dunn (1964) Kruskal-Wallis multiple comparison
p-values adjusted with the Bonferroni method.

	Comparison	Z	P.unadj	P.adj
1	After COVID - Before COVID	-3.770206	0.0001631129	0.0004893388
2	After COVID - During COVID	-1.077029	0.2814671028	0.8444013085
3	Before COVID - During COVID	3.988368	0.0000665293	0.0001995879

\$Cr

Dunn (1964) Kruskal-Wallis multiple comparison
p-values adjusted with the Bonferroni method.

	Comparison	Z	P.unadj	P.adj
1	After COVID - Before COVID	-2.290534	0.02199035	0.06597105
2	After COVID - During COVID	-1.127933	0.25934827	0.77804482
3	Before COVID - During COVID	1.595856	0.11052098	0.33156294

\$Cu

Dunn (1964) Kruskal-Wallis multiple comparison
p-values adjusted with the Bonferroni method.

	Comparison	Z	P.unadj	P.adj
1	After COVID - Before COVID	-3.419142	0.0006281882	0.001884565
2	After COVID - During COVID	-1.418603	0.1560147498	0.468044249
3	Before COVID - During COVID	2.845206	0.0044382763	0.013314829

\$Mn

Dunn (1964) Kruskal-Wallis multiple comparison
p-values adjusted with the Bonferroni method.

	Comparison	Z	P.unadj	P.adj
1	After COVID - Before COVID	-3.416917	0.0006333449	0.001900035
2	After COVID - During COVID	-1.694658	0.0901402677	0.270420803

3 Before COVID - During COVID 2.359565 0.0182963734 0.054889120

\$Fe

Dunn (1964) Kruskal-Wallis multiple comparison

p-values adjusted with the Bonferroni method.

	Comparison	Z	P.unadj	P.adj
1	After COVID - Before COVID	-3.389454	0.0007003188	0.002100956
2	After COVID - During COVID	-1.800753	0.0717417552	0.215225266
3	Before COVID - During COVID	2.131497	0.0330482511	0.099144753

\$Ni

Dunn (1964) Kruskal-Wallis multiple comparison

p-values adjusted with the Bonferroni method.

	Comparison	Z	P.unadj	P.adj
1	After COVID - Before COVID	-3.039773	0.002367564	0.007102691
2	After COVID - During COVID	-1.023824	0.305918377	0.917755130
3	Before COVID - During COVID	2.944139	0.003238547	0.009715642

\$As

Dunn (1964) Kruskal-Wallis multiple comparison

p-values adjusted with the Bonferroni method.

	Comparison	Z	P.unadj	P.adj
1	After COVID - Before COVID	-2.034957	0.04185524	0.1255657
2	After COVID - During COVID	-0.550882	0.58171456	1.0000000
3	Before COVID - During COVID	2.148752	0.03165403	0.0949621

\$Al

Dunn (1964) Kruskal-Wallis multiple comparison

p-values adjusted with the Bonferroni method.

	Comparison	Z	P.unadj	P.adj
1	After COVID - Before COVID	-2.8186477	0.004822641	0.01446792
2	After COVID - During COVID	-2.5206983	0.011712221	0.03513666
3	Before COVID - During COVID	0.3727061	0.709367231	1.00000000

\$Mg

Dunn (1964) Kruskal-Wallis multiple comparison

p-values adjusted with the Bonferroni method.

	Comparison	Z	P.unadj	P.adj
1	After COVID - Before COVID	-2.4284154	0.01516496	0.04549489
2	After COVID - During COVID	-2.1486441	0.03166263	0.09498788


```
3 Before COVID - During COVID 0.3533783 0.72380482 1.00000000
```

Pb (납)

After COVID - Before COVID

p-adj 가 0.0016508 이므로 0.05보다 작습니다. 따라서 After COVID에서 Before COVID보다 유의미한 농도의 감소가 있습니다.

Before COVID - During COVID

p-adj 가 0.0001347 이므로 0.05보다 작습니다. 따라서 During COVID에서 Before COVID보다 유의미한 농도의 감소가 있습니다.

Cd (카드뮴)

After COVID - Before COVID

p-adj 이 0.0004893 이므로 0.05보다 작습니다. 따라서 After COVID에서 Before COVID보다 유의미한 농도의 감소가 있습니다.

Before COVID - During COVID

p-adj 이 0.0001996 이므로 0.05보다 작습니다. 따라서 During COVID에서 Before COVID보다 유의미한 농도의 감소가 있습니다.

Cr (크롬)

After COVID - Before COVID

p-adj 이 0.0660 이므로 0.05보다 큼니다. After COVID에서 Before COVID보다 유의미한 농도의 감소가 있다고 보기 어렵습니다.

Cu (구리)

After COVID - Before COVID

p-adj 이 0.0018846 이므로 0.05보다 작습니다. After COVID에서 Before COVID보다 유의미한 농도의 감소가 있습니다.

Before COVID - During COVID

p-adj 이 0.0133148 이므로 0.05보다 작습니다. During COVID에서 Before COVID보다 유의미한 농도의 감소가 있습니다.

Mn (망간)

After COVID - Before COVID

p-adj 은 0.0019000 이므로 0.05보다 작습니다. After COVID에서 Before COVID보다 유의미한 농도의 감소가 있습니다.

Before COVID - During COVID

p-adj 는 0.0549 이므로 0.05보다 큼니다. During COVID에서 Before COVID보다 농도의 감소가 유의적이지 않습니다.

Fe (철)

After COVID - Before COVID

p-adj 는 0.0021010 이므로 0.05보다 작습니다. After COVID에서 Before COVID보다 유의미한 농도의 감소가 있습니다.

Before COVID - During COVID

p-adj 는 0.0991 이므로 0.05보다 큼니다. During COVID에서 Before COVID보다 농도의 감소가 유의적이지 않습니다.

Ni (니켈)

After COVID - Before COVID

p-adj 은 0.0071027 이므로 After COVID에서 Before COVID보다 유의미한 농도의 감소가 있습니다.

Before COVID - During COVID

p-adj 은 0.0097156 이므로 0.05보다 작습니다. 따라서 During COVID에서 Before COVID보다 유의적인 농도의 감소가 있습니다.

As (비소)

Before COVID - During COVID

p-adj 는 0.0949621 이므로 0.05보다 큼니다. During COVID에서 Before COVID보다 농도의 감소가 유의적이지 않습니다.

Al (알루미늄)

After COVID - Before COVID

p-adj 는 0.0144679 이므로 0.05보다 작습니다. After COVID에서 Before COVID보다 유의미한 농도의 감소가 있습니다.

After COVID - During COVID

p-adj 는 0.0351367 이므로 0.05보다 작습니다. After COVID에서 During COVID보다 유의미한 농도의 감소가 있습니다.

Mg (마그네슘)

After COVID - Before COVID

p-adj 는 0.0454949 이므로 0.05보다 작습니다. After COVID에서 Before COVID보다 유의미한 농도의 감소가 있습니다.

요약

- 유의미한 감소

납(Pb): After COVID vs Before COVID, Before COVID vs During COVID

카드뮴(Cd): After COVID vs Before COVID, Before COVID vs During COVID

구리(Cu): After COVID vs Before COVID, Before COVID vs During COVID

망간(Mn): After COVID vs Before COVID

철(Fe): After COVID vs Before COVID

니켈(Ni): After COVID vs Before COVID, Before COVID vs During COVID

알루미늄(Al): After COVID vs Before COVID, After COVID vs During COVID

마그네슘(Mg): After COVID vs Before COVID

- 유의미한 차이 없음

크롬(Cr)

비소(As)

베릴륨(Be)

칼슘(Ca)

결론

따라서 황사기간 동안의 대전광역시의 대기중 중금속 농도에서 코로나로 인한 유의미한 감소가 있는 원소는 납(Pb), 구리(Cu), 망간(Mn), 철(Fe), 니켈(Ni), 알루미늄(Al), 마그네슘(Mg) 이고, 유의미한 차이가 없는 원소는 크롬(Cr), 비소(As), 베릴륨(Be), 칼슘(Ca) 입니다.

해석

위의 경기도 전체 지역의 황사기간의 코로나 영향으로 농도가 감소된 원소들은 Cr (크롬), Mn (망

간), Fe (철), Ni (니켈), Ca (칼슘), Mg (마그네슘) 입니다. 그와 대비하여 대전광역시에서 황사기간 동안 코로나의 영향으로 감소한 원소들은 납(Pb), 구리(Cu), 망간(Mn), 철(Fe), 니켈(Ni), 알루미늄(Al), 마그네슘(Mg) 입니다. 이 중 경기도와 대전광역시에서 모두 유의적으로 농도가 감소된 원소는 망간(Mn), 철(Fe), 니켈(Ni), 마그네슘(Mg) 이며, 경기도에서만 유의적으로 감소된 원소는 크롬(Cr)과 칼슘(Ca)이며, 대전광역시에만 유의적으로 감소된 원소는 납(Pb), 구리(Cu), 알루미늄(Al) 입니다.

이는 중국의 타클라마칸 사막에서부터 시작되어 충칭과 스촨성 등 내륙을 지나서 오는 황사풍의 중금속의 함유량과 중국 고비사막에서부터 발원되어 산시성과 허베이성 등을 거쳐서 오는 황사풍의 중금속 함유량이 서로 다르기 때문일 것이라고 생각되어집니다.

경기도와 대전광역시에서 공통으로 감소된 원소는 망간(Mn), 철(Fe), 마그네슘(Mg), 니켈(Ni) 이며, 대부분 철강산업과 합금 산업에서 배출되는 것입니다. 그러므로 만약 코로나 기간에 중국의 공장이 문을 많이 닫았다면 그 업종은 합금 혹은 철강산업일 것이라고 생각됩니다.

분명한 사실은 황사기간 동안 경기도와 대전광역시에서 코로나의 영향으로 특정 중금속 함유량이 유의적으로 감소했다는 사실입니다.

조금 더 정확한 분석을 위하여 황사의 경로로 해당되어지는 특정지역을 선정하여 추가적인 분석을 진행해야 한다고 생각됩니다.