
기상자료와 의학자료를 이용한 천식악화 방지



조이름: 3실

조원: 문석주 김기현 최혜원

목차 / List

- | | |
|----|--------------------|
| 1장 | 공모배경 |
| 2장 | 활용데이터 정의 |
| 3장 | 데이터 처리방안 및 활용분석 기법 |
| 4장 | 분석 결과 |
| 5장 | 서비스 활용 방안 |
| 6장 | 서비스 기대 효과 |

1

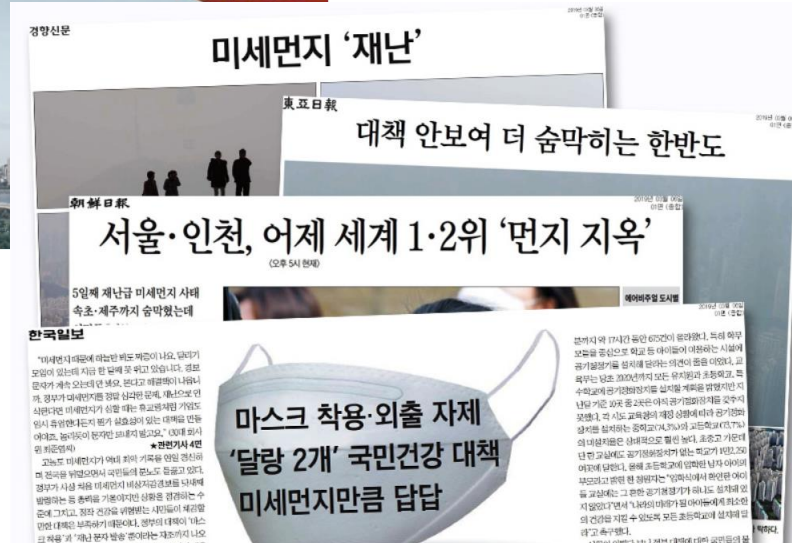
공모배경

1. 공모배경



폭염과 강추위의 발생

2018년, 과거 경험하지 못한 강한한파로 관측 이래 가장 낮은 최저기온을 보인 반면, 장기간 지속된 폭염으로 일 최고 기온 최고치를 경신하는 등 **극한의 기온 변화**를 보였다.



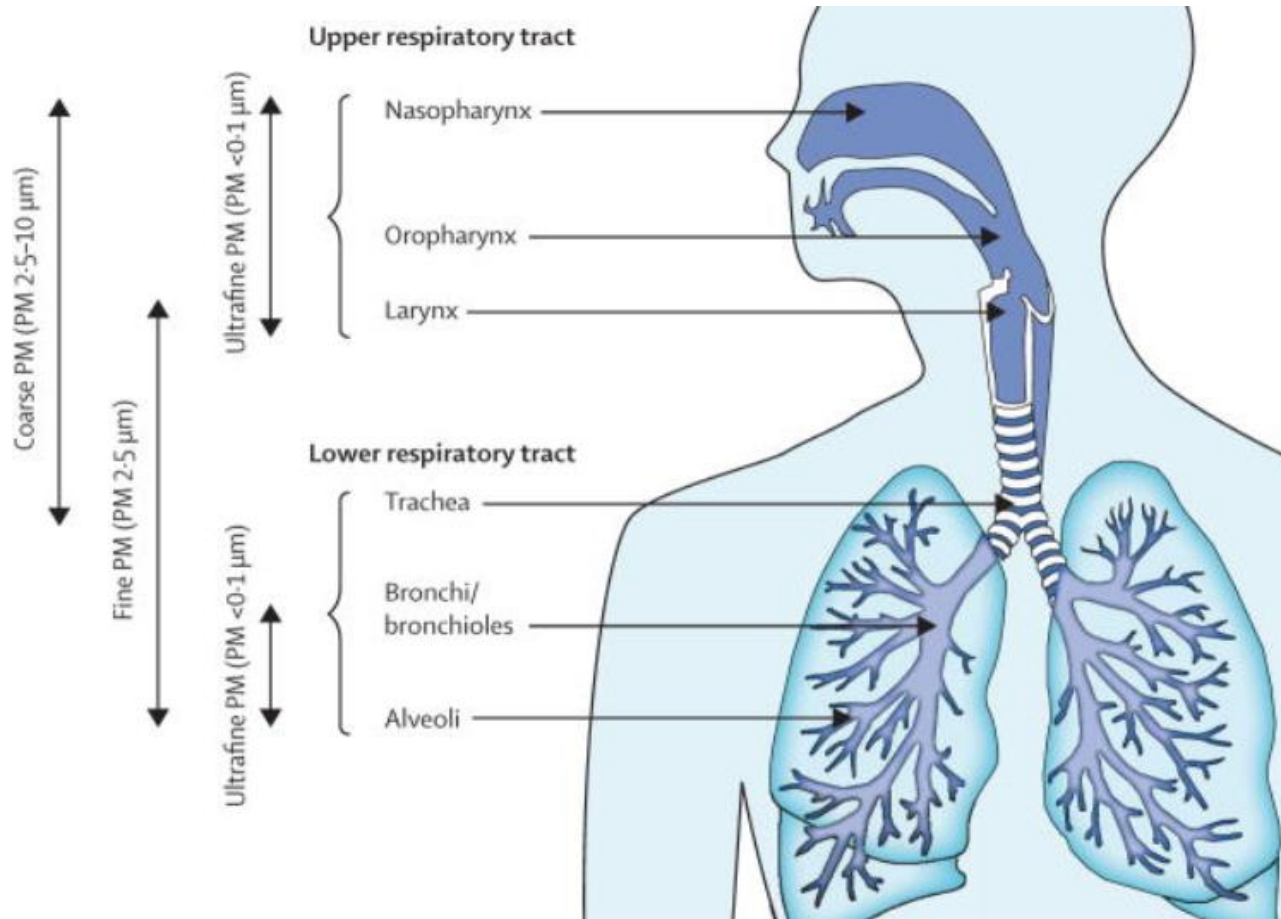
대기오염, 미세먼지의 심각성

발암물질인 미세먼지가 사회적 이슈로 부각되면서, 국민들의 불안과 우려의 증가로 인해 현재 이에 대응하는 많은 조사, 분석이 이루어지고 있다.

1. 공모배경

Outdoor air pollution and asthma

[Michael Guarnieri](#), MD and [John R. Balmes](#), MD

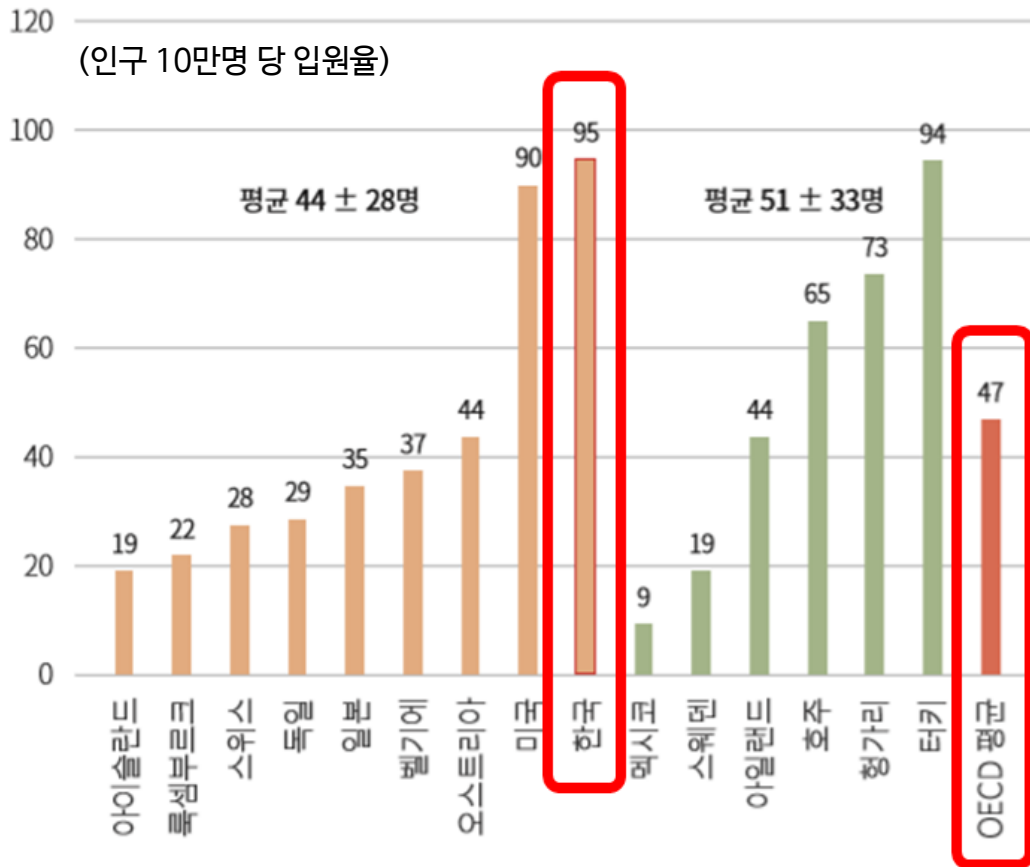


체내 섬모에서 여과되지 않고
혈관까지 침투하면
신체 여러 부분에서 염증을 통해
다양한 호흡기 질환을 발생시킴.

이 중 가장 흔하게 발생하는 천식의 경우,
기후와 대기환경에 큰 관계가 있다

1. 공모배경

OECD 국가 천식 입원율



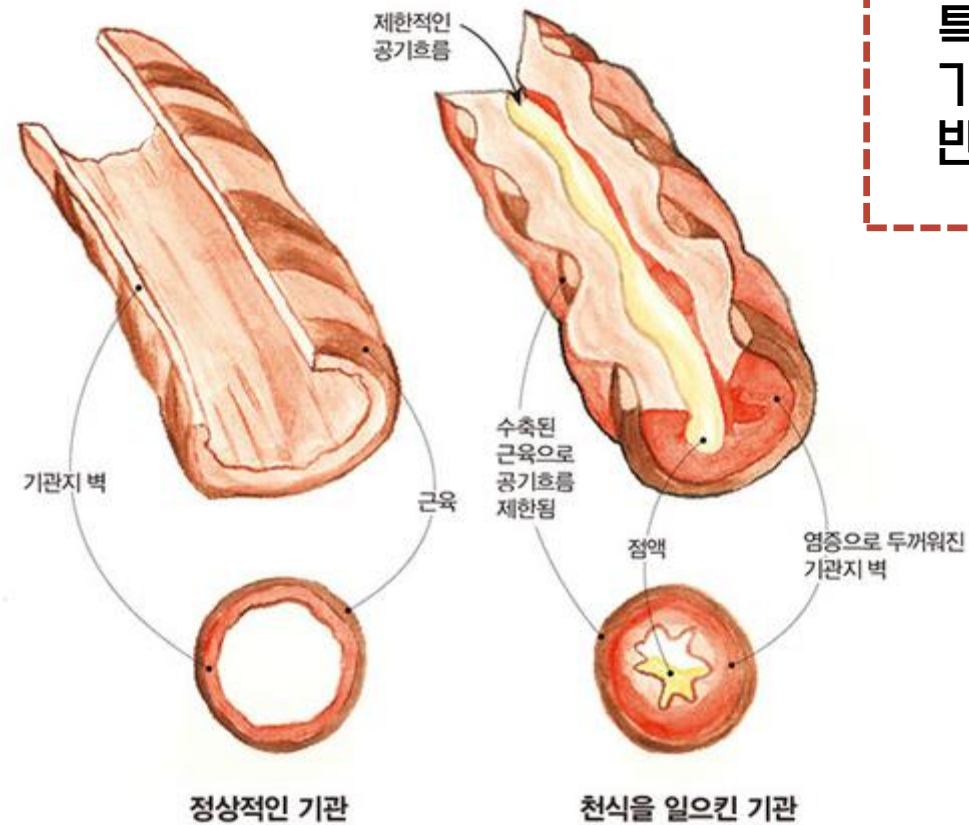
천식은 적절한 대응으로 증상 악화 및 입원을 예방할 수 있는 대표적인 민감성 질환이다.

하지만 한국에서는

'꾸준하게 치료를 해야하는 만성질환'이라는 인식이 상대적으로 부족한 상황이다.

따라서 천식이 악화되기 전 미리 예방 할 필요가 있다

1. 공모배경



천식이란, 폐로 연결되는 통로인 '기관지'의 질환으로, 특정한 유발 원인 물질에 노출되었을 때 기관지의 염증에 의해 기관지가 심하게 좁아져 기침, 천명, 호흡곤란, 가슴 답답함이 반복적으로 발생하는 질환

천식의 구분

유전적 인자	환경 인자
① 비만	① 알레르기 항원
② 성별	② 다양한 바이러스의 감염
	③ 작업성 감작물질
	④ 흡연
	⑤ 실내/실외 공기 오염
	⑥ 음식

2

활용 데이터 정의

2. 활용 데이터 정의



항목	세부내용	날짜
천식 진료 정보	<ul style="list-style-type: none">시도지역: 서울, 대구, 광주, 부산, 인천발생건수: 천식 발생 건수(월별)	2014.01.01. ~ 2018.06.31
기상관측 정보	<ul style="list-style-type: none">평균기온(°C), 평균일교차(°C), 평균상대습도(%) 월합강수량(mm)	2014.01.01. ~ 2018.06.31
대기환경 정보	<ul style="list-style-type: none">SO2(ppm): 이황산가스, NO2(ppm): 이산화질소 CO(ppm): 일산화탄소, PM-10(ug/m3): 미세먼지	2014.01.01. ~ 2018.06.31

2. 활용 데이터 정의



국민건강영양조사
Korea National Health & Nutrition Examination Survey

항목	세부내용	날짜
시군구별 주민등록인구	<ul style="list-style-type: none">총 인구수 (명)	2014.01.01. ~ 2018.06.31
국민건강 영양 조사	<ul style="list-style-type: none">천식, 고혈압, 당뇨병유무, 나이, 음주량, 흡연여부, BMI지수	2014 ~ 2017

3

데이터 처리 방안 및 활용 분석 기법

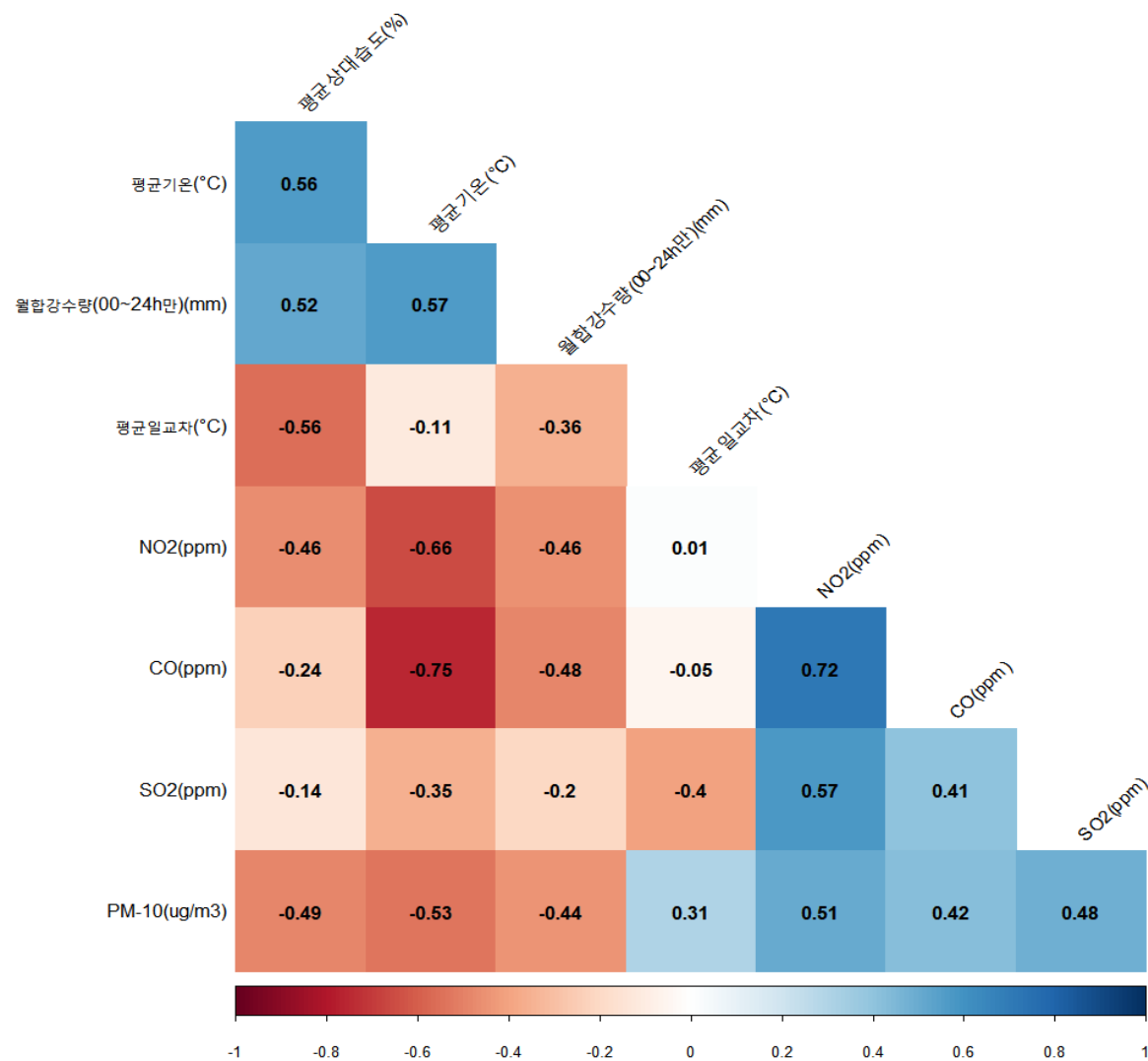
3.1 분석 계획

3.2 데이터 전처리

3.1 데이터 처리 방안 및 활용 분석 기법 - 분석계획

- ① 다중 회귀 분석을 통한 천식과 기상환경과의 관련성 분석
- ② 시계열 모형과 머신 러닝 기법을 이용하여 천식 환자 수 예측
- ③ 독립성 검정을 통한 내부적 특성과 천식의 관련성 분석
- ④ 로지스틱 회귀 분석을 통한 내부적 특성 파악

3.2 데이터 처리 방안 및 활용 분석 기법 - 데이터 전처리



① 상관관계 (Correlation plot)

높은 상관관계를 띄는 변수 (0.60이상)

평균기온과 CO (0.75), CO와 NO2 (0.72)
평균기온과 NO2 (0.66)

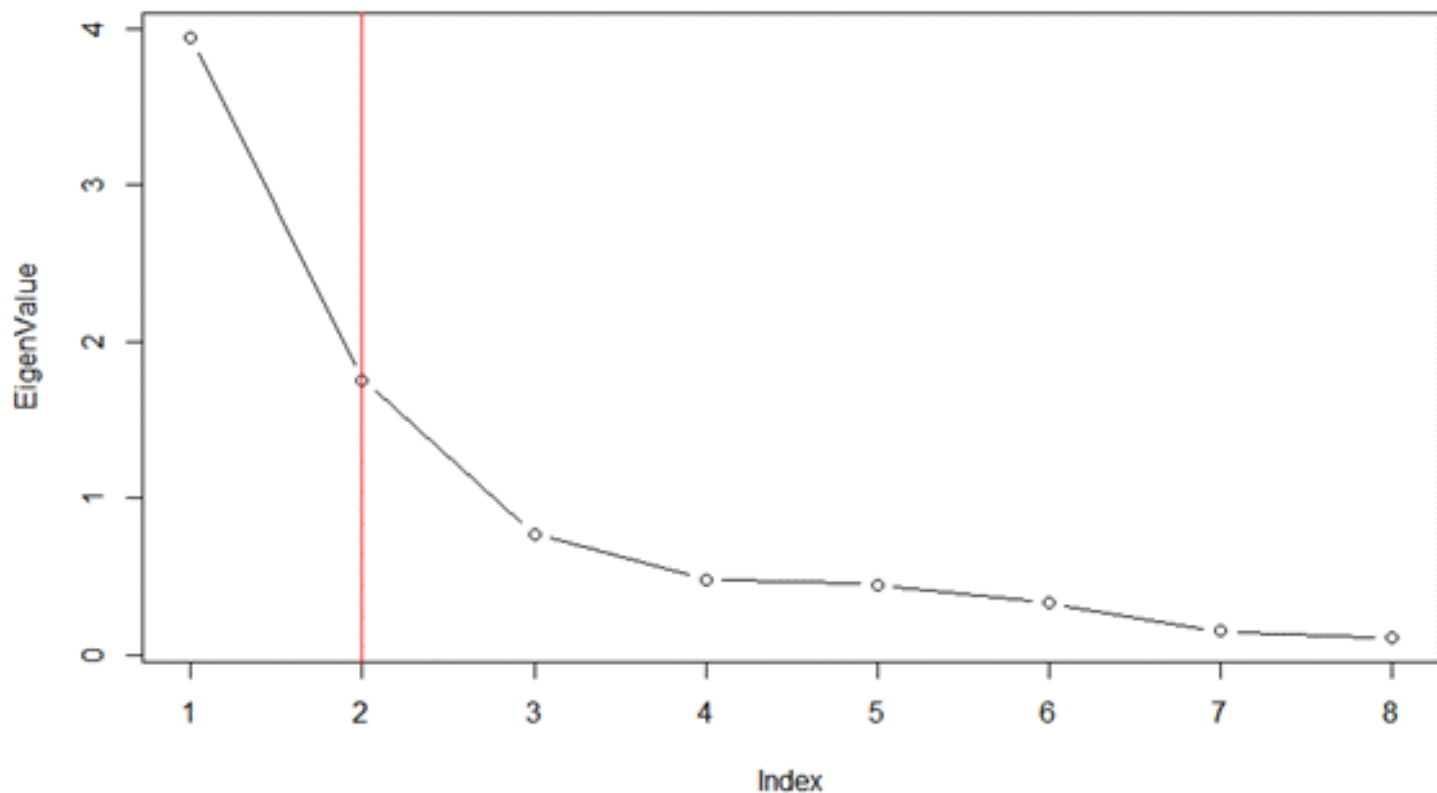
다중공선성 우려

기상변수 간 상관관계가 높으므로 회귀분석 시 다중공선성이 우려된다.

인자분석 고려

다중공선성의 문제를 해결하기 위해 차원을 축소한다.

3.2 데이터 처리 방안 및 활용 분석 기법 - 데이터 전처리



```
> df_all_weather_factor$values
```

```
[1] 3.9404963 1.7554692 0.7725439 0.4811832 0.4514339 0.3341012 0.1515123 0.1132601
```

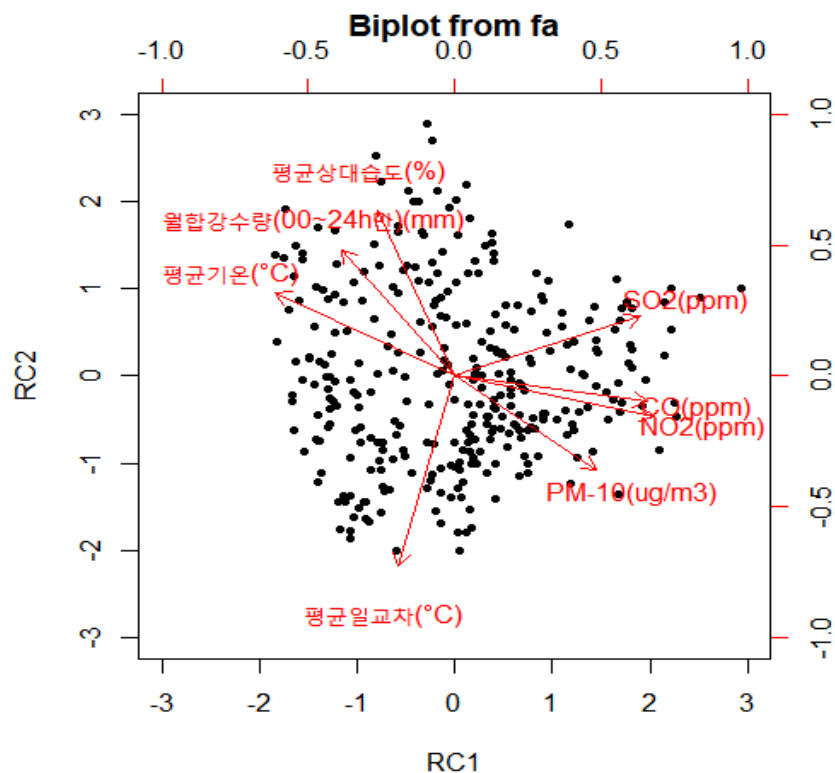
② 인자분석 (Scree plot)

주성분 방법을 통한
인자 수 결정

고유값이 1이상
주성분까지 선택

3.2 데이터 처리 방안 및 활용 분석 기법 - 데이터 전처리

② 인자분석 (Scree plot)



- **첫번째 인자 (RC1)**
: 미세먼지가 높고 기온이 낮을 때 높은 점수를 갖는 변수
- **두번째 인자 (RC2)**
: 일교차가 작고 습도와 강수량이 많을 때 높은 점수를 갖는 변수

```
> df_all_weather_varimax
Principal Components Analysis
Call: principal(r = df_all_weather, nfactors = 2, rotate = "varimax")
Standardized loadings (pattern matrix) based upon correlation matrix
```

	RC1	RC2	h2	u2	com
평균기온(°C)	-0.77	0.40	0.75	0.25	1.5
평균일교차(°C)	-0.24	-0.91	0.88	0.12	1.1
평균상대습도(%)	-0.33	0.79	0.73	0.27	1.3
월합강수량(00~24h만)(mm)	-0.48	0.60	0.60	0.40	1.9
SO2(ppm)	0.79	0.29	0.71	0.29	1.3
NO2(ppm)	0.85	-0.20	0.77	0.23	1.1
CO(ppm)	0.83	-0.12	0.70	0.30	1.0
PM-10(ug/m3)	0.60	-0.45	0.56	0.44	1.8

	RC1	RC2
SS loadings	3.39	2.30
Proportion Var	0.42	0.29
Cumulative Var	0.42	0.71
Proportion Explained	0.60	0.40
Cumulative Proportion	0.60	1.00

선택된 인자가 전체 분산 중
71%를 설명

Mean item complexity = 1.4
Test of the hypothesis that 2 components are sufficient.

The root mean square of the residuals (RMSR) is 0.09
with the empirical chi square 145.15 with prob < 2e-24

Fit based upon off diagonal values = 0.96

```
> KMO(cor(df[,c(6,9:15)]))
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = cor(df[, c(6, 9:15)]))
Overall MSA = 0.67
MSA for each item =
```

MSA=0.67로 인자분석이
적합하다고 할 수 있음

평균기온(°C)	평균일교차(°C)	평균상대습도(%)	월합강수량(00~24h만)(mm)	
0.72	0.40	0.63		0.91
SO2(ppm)	NO2(ppm)	CO(ppm)		PM-10(ug/m3)
0.54	0.81	0.68		0.71

4

분석결과

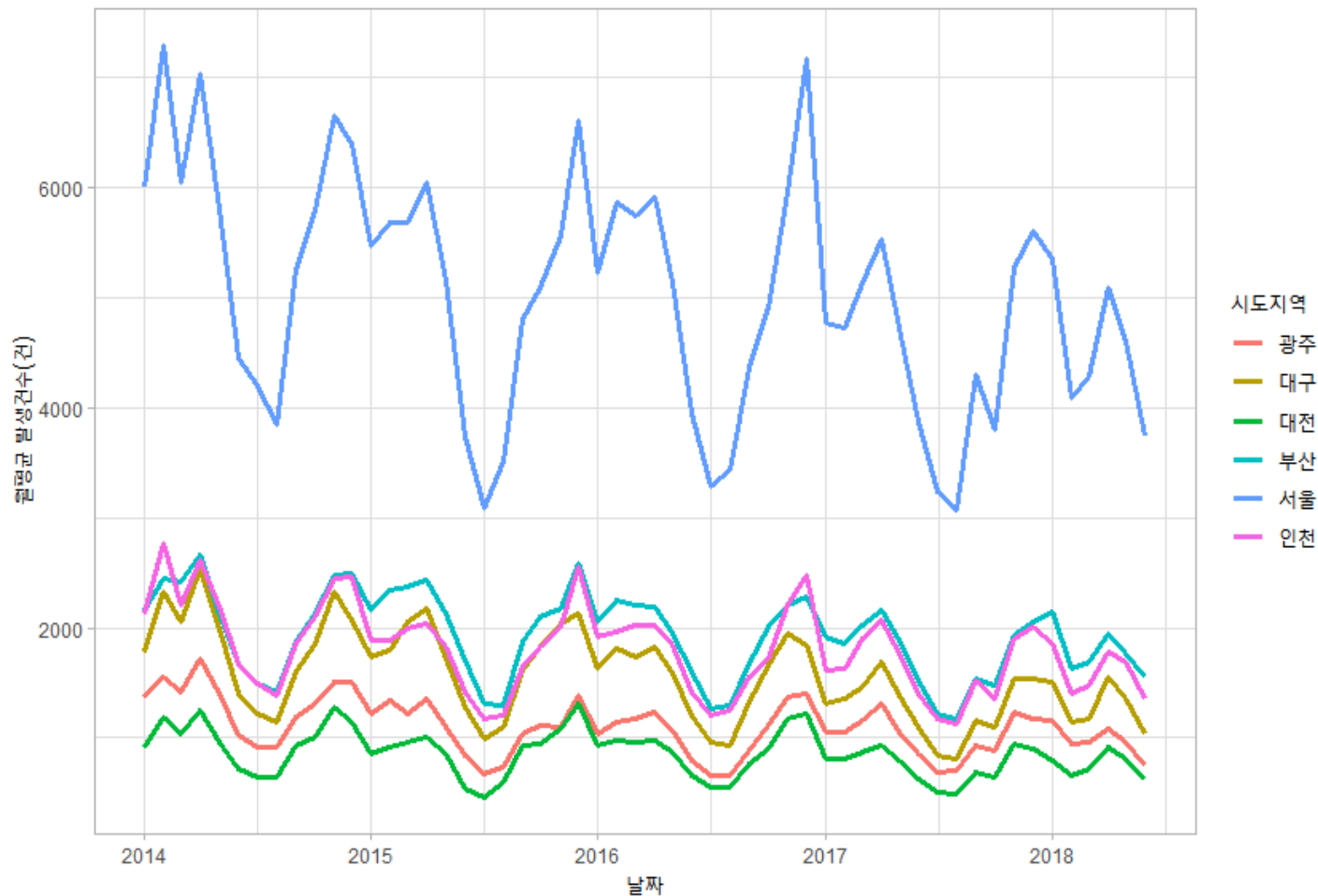
4.1 환경 요인

4.2 내부적 요인

4.3 결과

4.1 분석결과 - 환경요인

① 탐색적 자료 분석 (천식 환자 수)

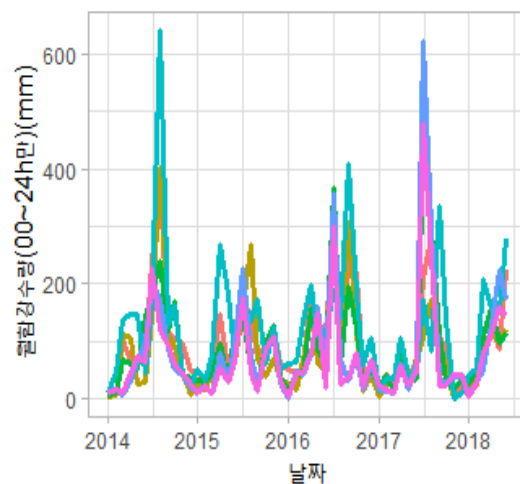
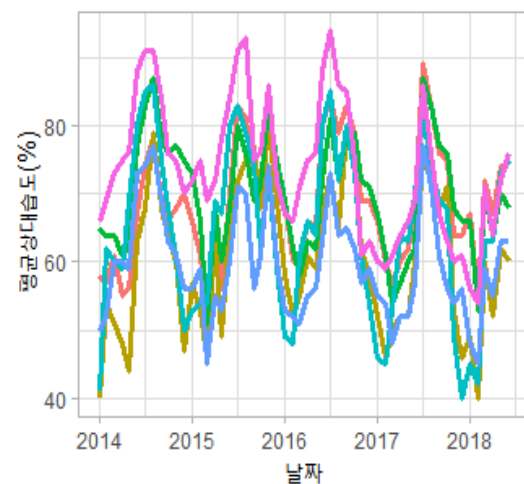
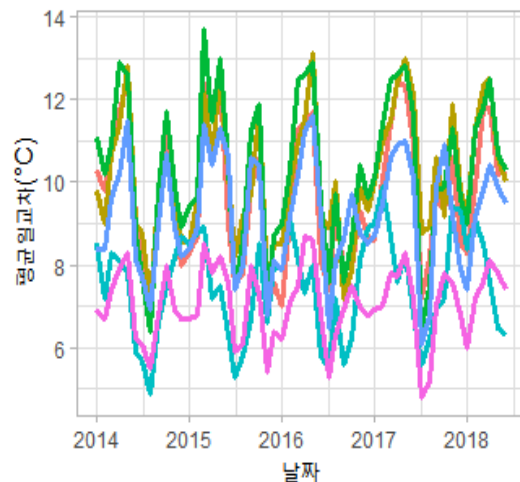
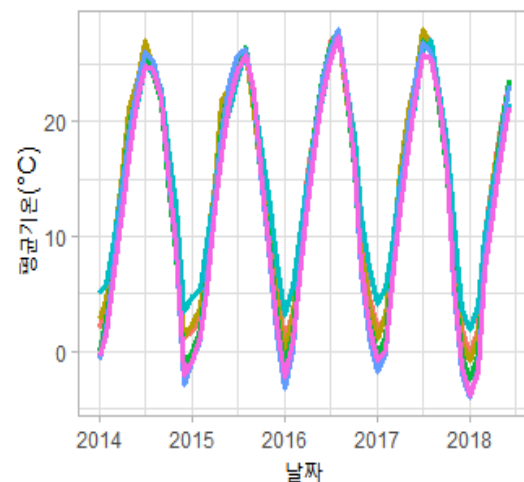


시도지역 별, 계절 별 차이

- ✓ 시도지역 별
: 지역의 따라 환자 수가 다름
지역에 따라 분석 및 예측 필요
- ✓ 계절 별
: 겨울에 증가하고
여름에 감소하는 패턴

4.1 분석결과 - 환경요인

② 탐색적 자료 분석 (날씨 환경)



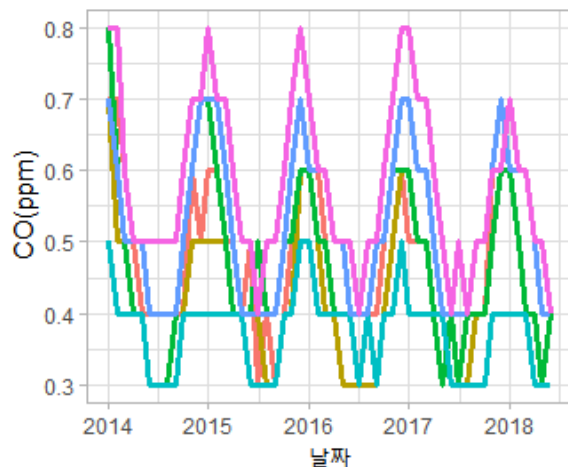
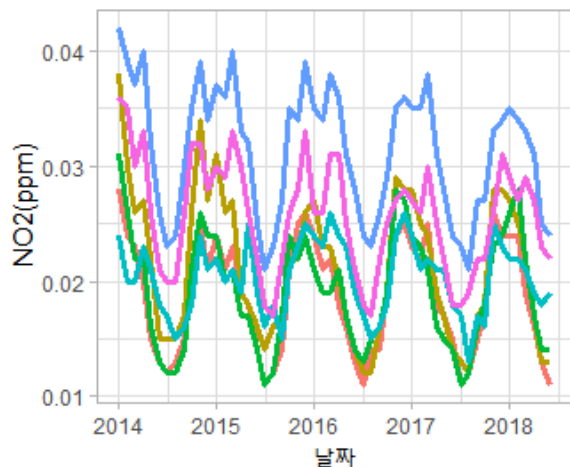
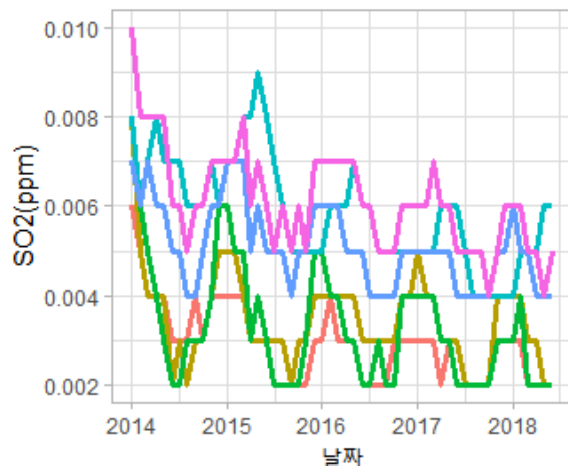
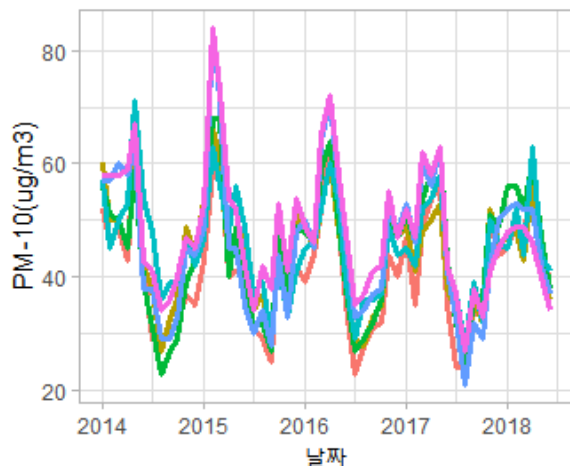
지역 별 날씨 차이

날씨는 지역별로 차이가 나타남

- ✓ 부산, 인천 (해안가)
: 겨울 철 온도와 습도가 높음
- ✓ 서울 등 내륙
: 겨울 철 온도와 습도가 낮음

4.1 분석결과 - 환경요인

③ 탐색적 자료 분석 (대기오염)



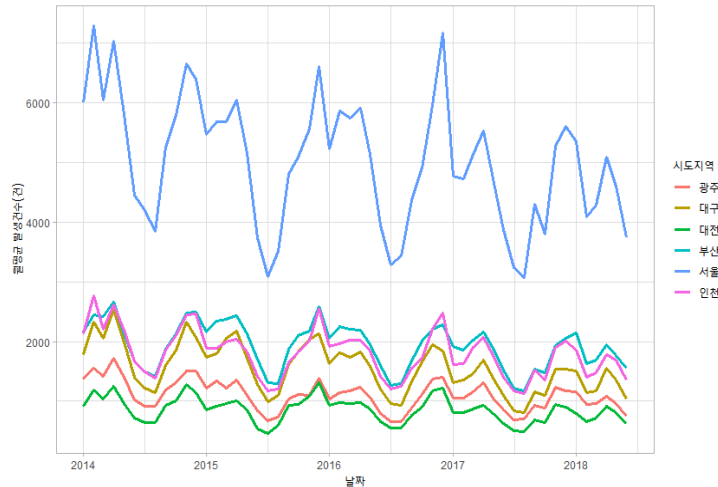
지역 별 대기오염 차이

대기오염은 지역별로 차이가 나타남

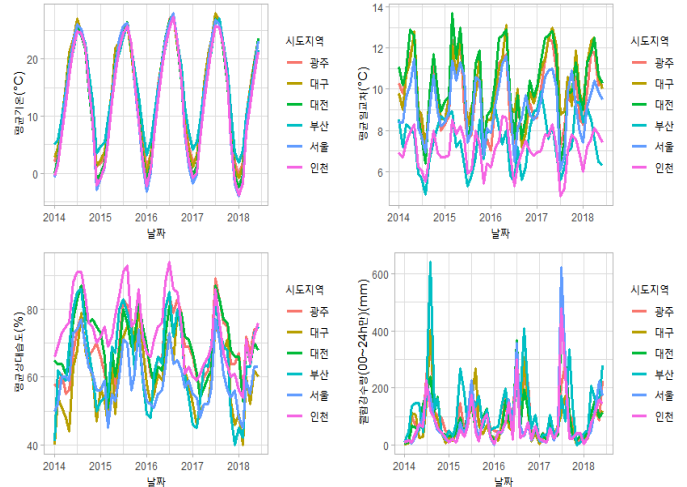
- ✓ NO2(이산화질소) : 서울 높음
- ✓ SO2(아황산가스) : 서울, 인천 높음
- ✓ CO(일산화탄소) : 서울, 인천 높음

4.1 분석결과 - 환경요인

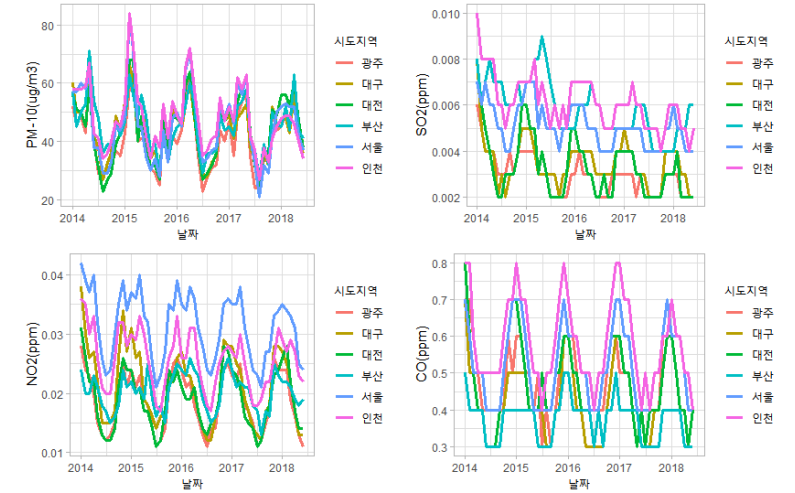
④ 탐색적 자료 분석 (종합)



천식 환자 수



날씨 환경



대기오염

천식은 계절성이 있어 날씨, 대기오염과 밀접한 관련성이 있다.
지역별로 날씨 환경, 대기오염, 지역 인구 수에 따라
천식 환자 수가 다르기 때문에 지역에 따른 분석과 예측 필요

4.1 분석결과 - 환경요인

⑤ 다중 회귀분석

```
> summary(df_model)
```

Call:

```
lm(formula = log(`월평균 발생건수(건)`) ~ pc1 + pc2 + pc1 * pc2 +  
    시도지역, data = df)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.39401	-0.11083	0.01211	0.11424	0.44507

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.07441	0.02316	305.428	< 2e-16 ***
pc1	0.17807	0.01154	15.434	< 2e-16 ***
pc2	-0.03266	0.01332	-2.452	0.0147 *
시도지역대구	0.30171	0.03077	9.804	< 2e-16 ***
시도지역대전	-0.27899	0.03050	-9.146	< 2e-16 ***
시도지역부산	0.49595	0.03354	14.786	< 2e-16 ***
시도지역서울	1.34975	0.03321	40.639	< 2e-16 ***
시도지역인천	0.21878	0.04342	5.039	7.90e-07 ***
pc1:pc2	0.07631	0.01303	5.856	1.19e-08 ***

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

Residual standard error: 0.1582 on 315 degrees of freedom
Multiple R-squared: 0.9339, Adjusted R-squared: 0.9322
F-statistic: 556.3 on 8 and 315 DF, p-value: < 2.2e-16

정규성 만족을 위해 로그변환을 하여 분석 진행

모든 인자
인자간 교호 작용
지역 변수
모두 유의함

R-square : 0.93

첫번째 변수 (pc1)

: 미세먼지 농도가 높고
온도가 낮을 때 높은 점수를 갖는 변수
→ 발생수를 증가시킨다. $\text{Exp}(0.178)$

두번째 변수 (pc2)

: 일교차가 낮고 습도, 강수량이 높을 때
높은 점수를 갖는 변수
→ 발생수를 감소시킴 $\text{Exp}(-0.033)$

교호작용 (pc1:pc2)

→ 환자수를 증가시킴

지역

→ 광주를 기준으로 광주보다 인구가
많은 지역이 발생 수가 더 증가

4.1 분석결과 - 환경요인

⑤ 다중 회귀분석

```
> summary(df_model)
```

Call:

```
lm(formula = log(`월평균 발생건수(건)`) ~ pc1 + pc2 + pc1 * pc2 +  
    시도지역, data = df)
```

정규성 만족을 위해 로그변환을 하여 분석 진행

Residuals:

Min	1Q	Median	3Q	Max
-0.39401	-0.11083	0.01211	0.11424	0.44507

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.07441	0.02316	305.428	< 2e-16 ***
pc1	0.17807	0.01154	15.434	< 2e-16 ***
pc2	-0.03266	0.01332	-2.452	0.0147 *
시도지역대구	0.30171	0.03077	9.804	< 2e-16 ***
시도지역대전	-0.27899	0.03050	-9.146	< 2e-16 ***
시도지역부산	0.49595	0.03354	14.786	< 2e-16 ***
시도지역서울	1.34975	0.03321	40.639	< 2e-16 ***
시도지역인천	0.21878	0.04342	5.039	7.90e-07 ***
pc1:pc2	0.07631	0.01303	5.856	1.19e-08 ***

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

Residual standard error: 0.1582 on 315 degrees of freedom
Multiple R-squared: 0.9339, Adjusted R-squared: 0.9322
F-statistic: 556.3 on 8 and 315 DF, p-value: < 2.2e-16

모든 인자
인자간 교호 작용
지역 변수
모두 유의함

R-square : 0.93

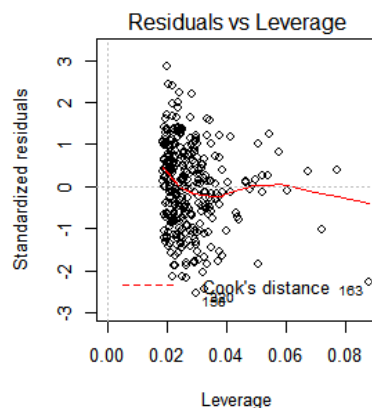
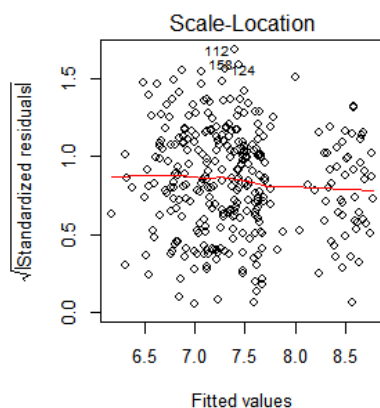
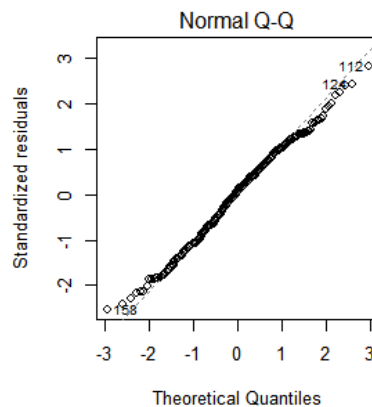
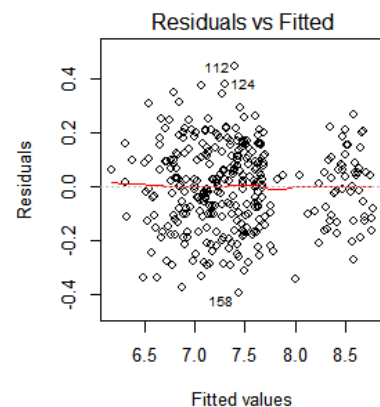
해 석

- ✓ 온도와 습도가 높을 경우
천식 환자 수 감소
- ✓ 미세먼지가 높을 경우
천식 환자 수 증가
- ✓ 교호작용을 통하여
습도가 높고 미세먼지가 많을 경우,

습도가 높더라도
천식 환자 수가 오히려 증가

4.1 분석결과 - 환경요인

⑤ 다중 회귀분석



정규성, 등분산성, 왜도, 첨도
회귀분석의 가정을 만족

```
> summary(gvmodel)
```

	Value	p-value	Decision
Global Stat	4.197e+00	0.3801	Assumptions acceptable.
Skewness	4.375e-01	0.5083	Assumptions acceptable.
Kurtosis	2.469e+00	0.1161	Assumptions acceptable.
Link Function	4.335e-05	0.9947	Assumptions acceptable.
Heteroscedasticity	1.290e+00	0.2561	Assumptions acceptable.

인자분석을 통해 다중 공선성($VIF < 5$) 문제 해결

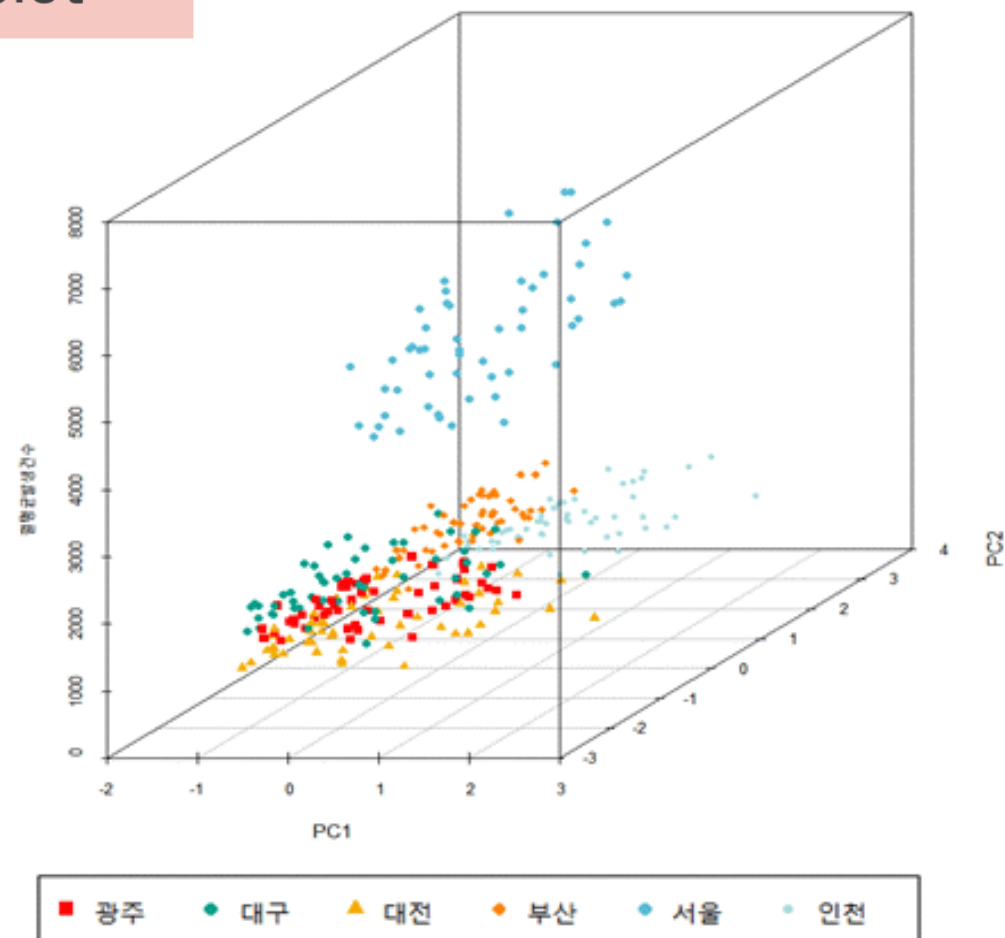
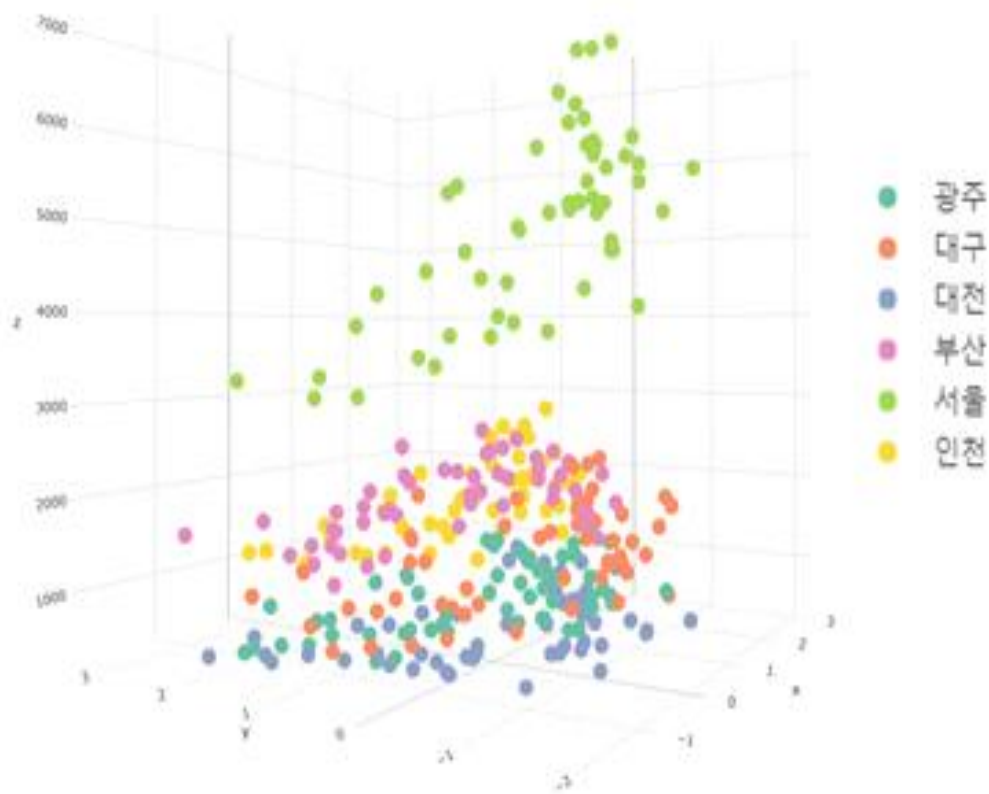
```
> vif(df_model)
```

	GVIF	Df	$GVIF^{1/(2*Df)}$
pc1	1.717122	1	1.310390
pc2	2.287630	1	1.512491
시도지역	3.414404	5	1.130659
pc1:pc2	1.615757	1	1.271124

4.1 분석결과 - 환경요인

⑤ 다중 회귀분석

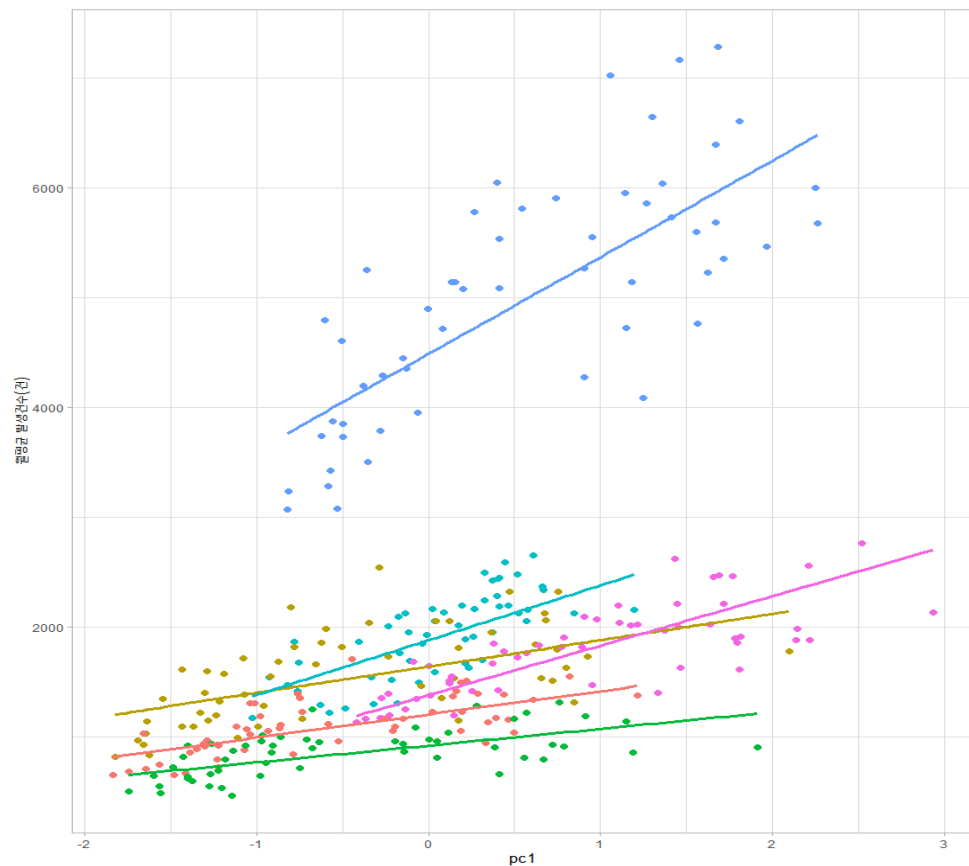
3D scatter plot



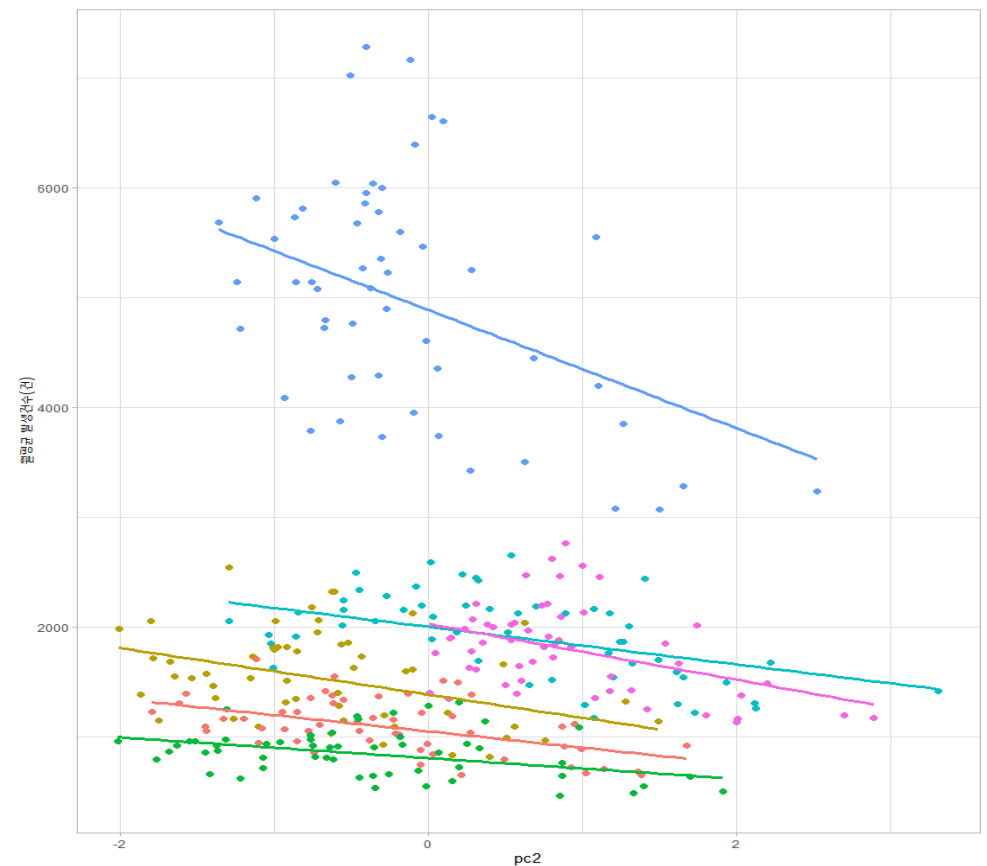
4.1 분석결과 - 환경요인

⑤ 다중 회귀분석

Regression plot



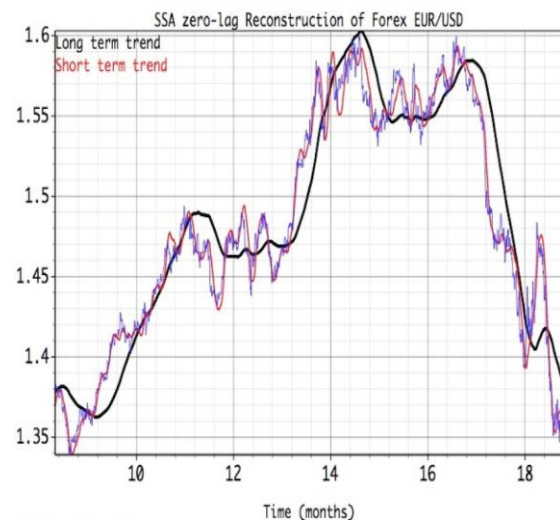
시도지역
광주
대구
대전
부산
서울
인천



4.1 분석결과 - 환경요인

⑥ 천식 환자 수 예측 기법

Seasonal ARIMA 모형



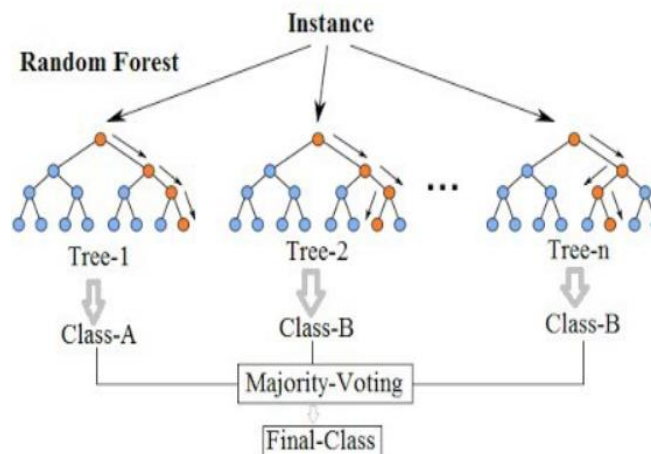
시계열분석

: 시간의 흐름에 따라 종속변수를 예측하는 것

Seasonal ARIMA 모형

: 계절성이 띄고 종속변수의 흐름이 정상성을 띌 때 사용 하며, 예측 시 바로 전 시점을 사용.

Random Forest

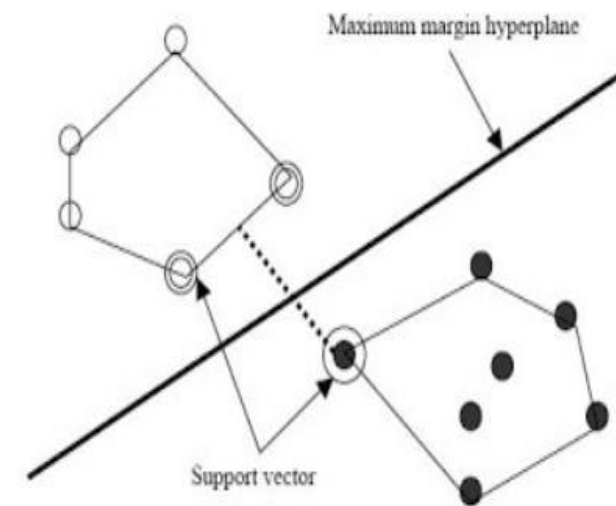


랜덤 포레스트

: 앙상블 기법 중 하나로 의사결정나무의 알고리즘을 사용. 배깅을 통하여 예측력을 높임.

예측 시 기상변수와 대기오염 변수 인구변수 사용 .

Support Vector Machine



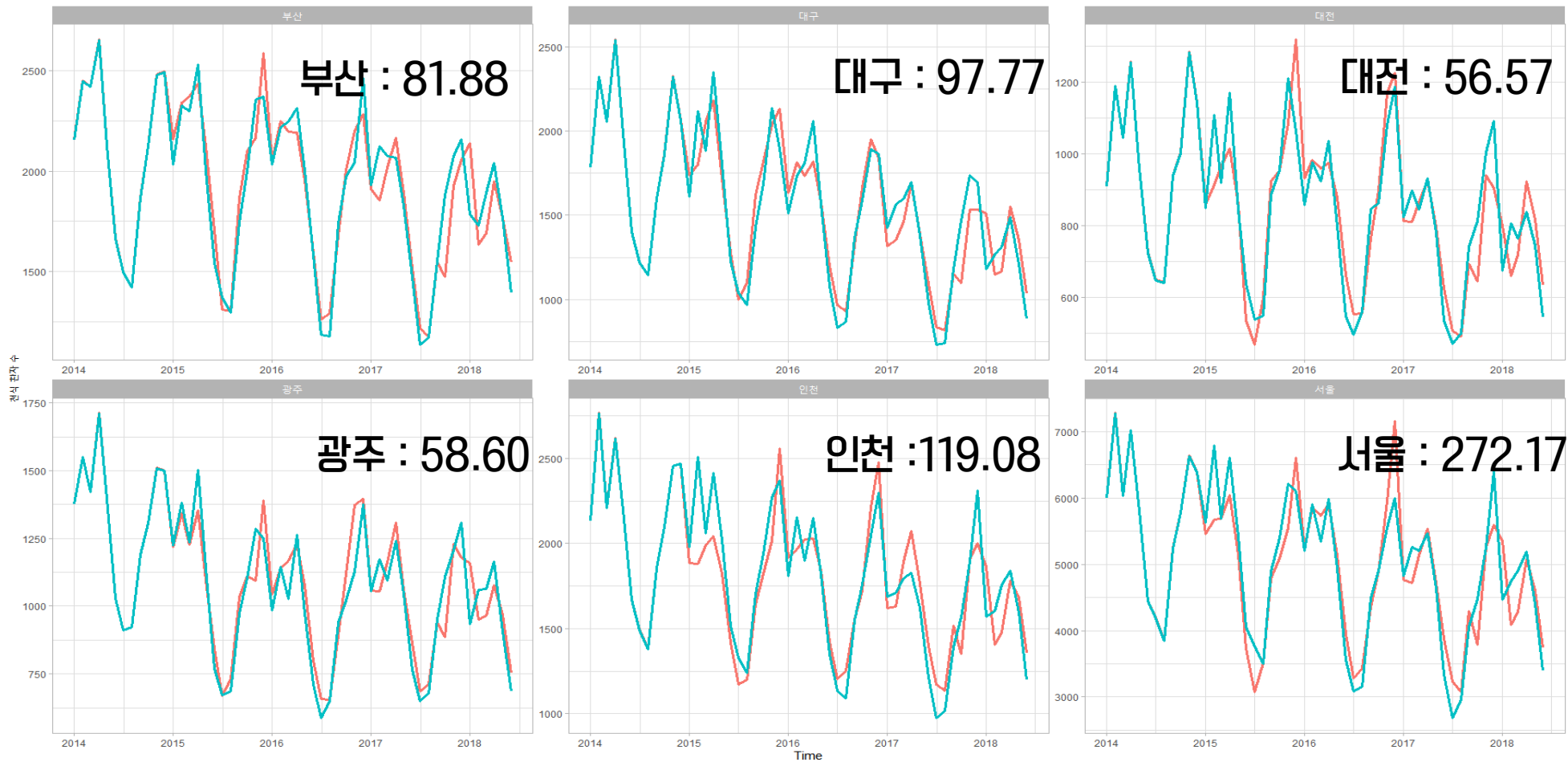
Support Vector Machine

: 자료 분석을 위한 지도 학습이며 분류, 회귀에 이용. 두 데이터를 가장 잘 나눌 수 있는 초평면을 찾아 예측함.

예측 시 기상변수와 대기오염 변수 인구변수 사용 .

4.1 분석결과 - 환경요인

⑥ 천식 환자 수 예측 기법 (Seasonal ARIMA 모형 예측)



MAE (평균절대오차)

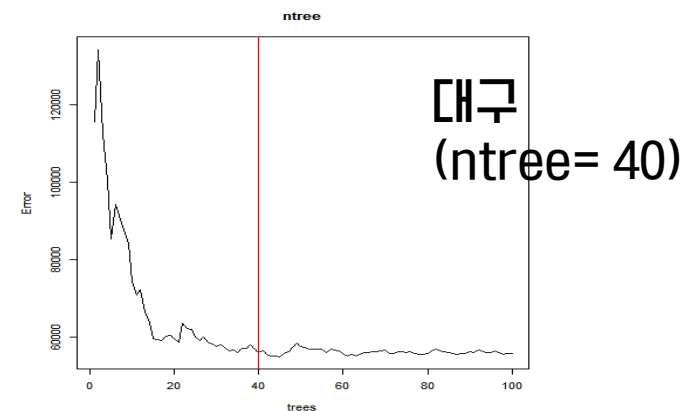
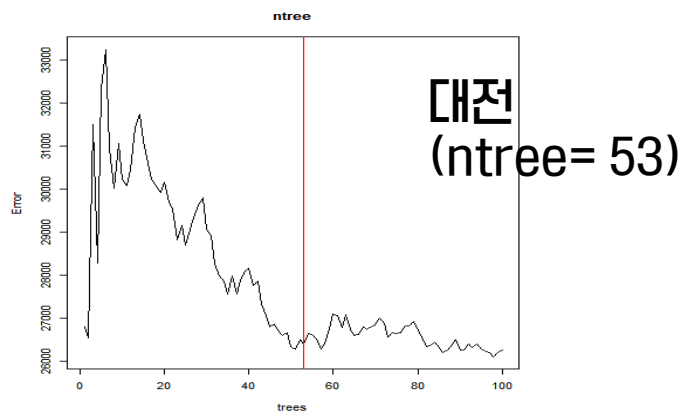
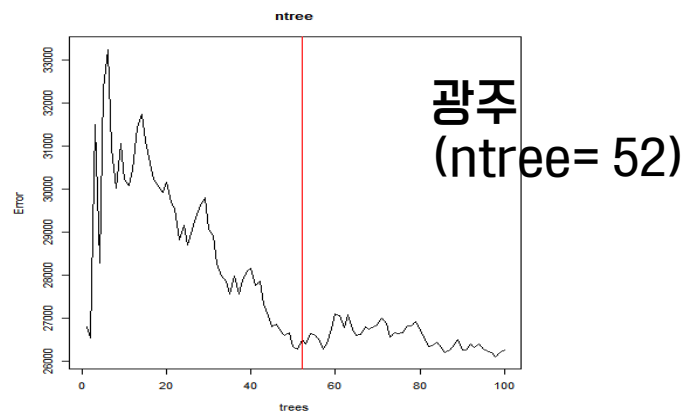
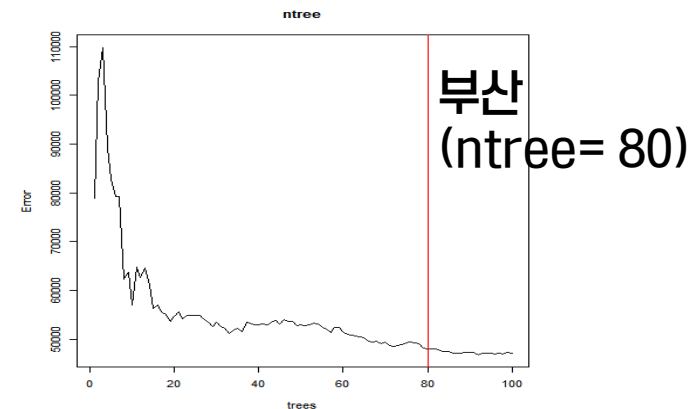
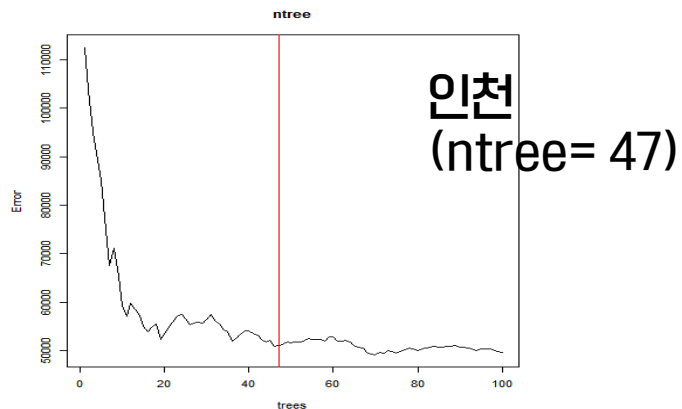
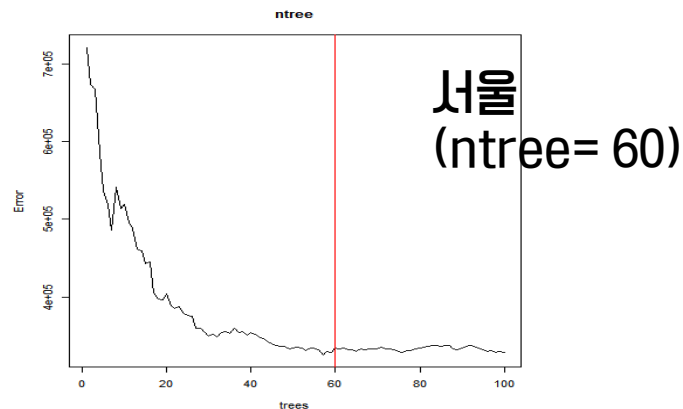
평균 : 114.35

4.1 분석결과 - 환경요인

⑥ 천식 환자 수 예측 기법 (Random Forest)

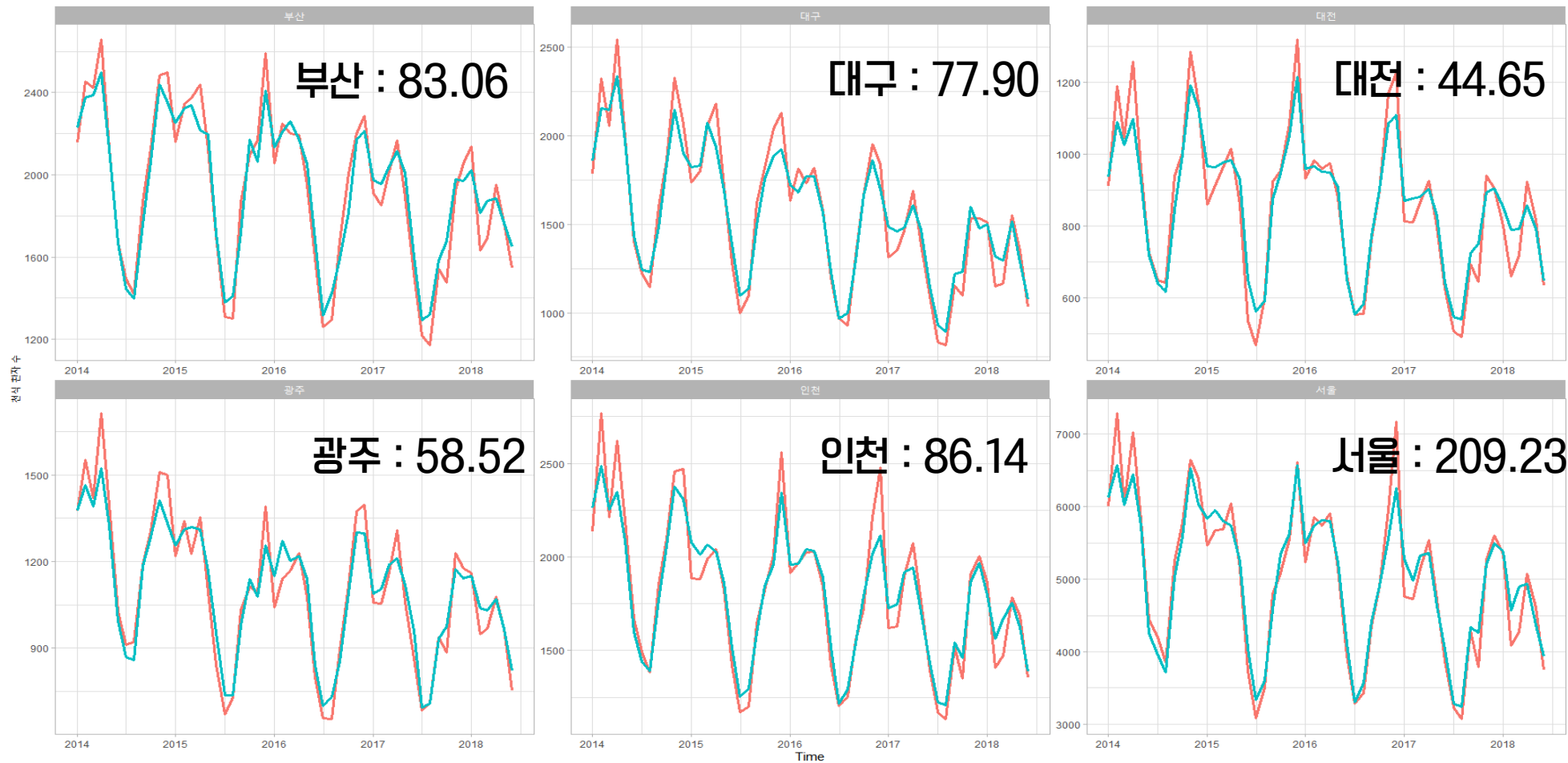
Hyperparameter 결정

: 오차를 최소화 하는 tree수 결정



4.1 분석결과 - 환경요인

⑥ 천식 환자 수 예측 기법 (Random Forest 예측)



MAE (평균절대오차)

평균 : 93.25

model

— Actual

— RandomForest

4.1 분석결과 - 환경요인

⑥ 천식 환자 수 예측 기법 (Support Vector Machine)

Hyperparameter 결정

: 오차를 최소화 하는 epsilon, cost 결정

```
> print(tune_result)
```

Parameter tuning of 'svm':

- sampling method: 10-fold cross validation

- best parameters:
epsilon cost
0 2

- best performance: 301106.4

서울

```
> print(tune_result)
```

Parameter tuning of 'svm':

- sampling method: 10-fold cross validation

- best parameters:
epsilon cost
0.15 4

- best performance: 41486.81

부산

```
> print(tune_result)
```

Parameter tuning of 'svm':

- sampling method: 10-fold cross validation

- best parameters:
epsilon cost
0 2

- best performance: 61111.16

인천

```
> print(tune_result)
```

Parameter tuning of 'svm':

- sampling method: 10-fold cross validation

- best parameters:
epsilon cost
0.3 2

- best performance: 23360.31

광주

```
> print(tune_result)
```

Parameter tuning of 'svm':

- sampling method: 10-fold cross validation

- best parameters:
epsilon cost
0.35 2

- best performance: 14722.5

대전

```
> print(tune_result)
```

Parameter tuning of 'svm':

- sampling method: 10-fold cross validation

- best parameters:
epsilon cost
0.27 4

- best performance: 43769.16

대구

4.1 분석결과 - 환경요인

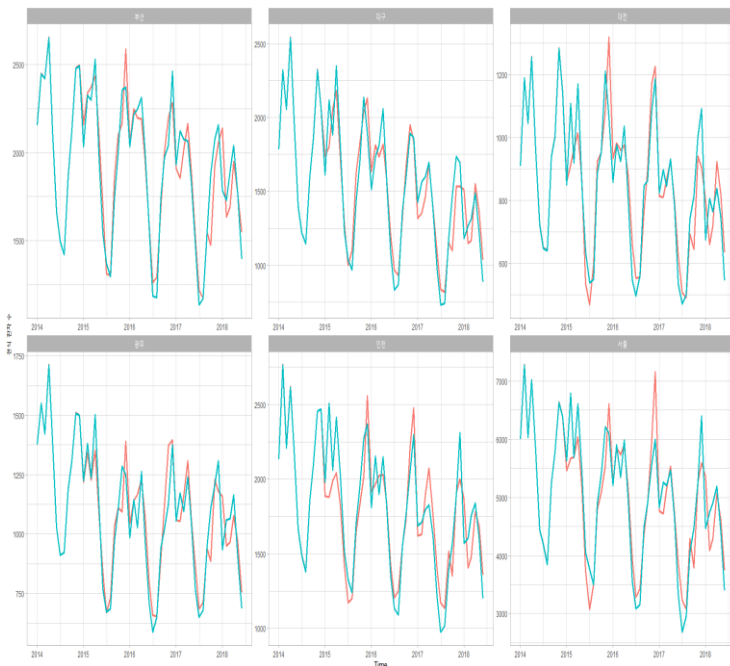
⑥ 천식 환자 수 예측 기법 (Support Vector Machine 예측)



4.1 분석결과 - 환경요인

⑥ 천식 환자 수 예측 기법 (최종 모형)

Seasonal ARIMA 모형



MAE AVERAGE : 114.35

Random Forest



MAE AVERAGE : 93.25

Support Vector Machine



MAE AVERAGE : 91.09

MAE 평균이 가장 나은 **Support Vector machine 모델**을 최종 모형으로 결정.

4.2 분석결과 - 내부적 요인

① 독립성 검정 (천식과 관련성)

성별 (p-value = 0.2923)

```
> chisq.test(data$asthma,data$sex)
```

Pearson's Chi-squared test with Yates' continuity correction

data: data\$asthma and data\$sex
X-squared = 1.1091, df = 1, p-value = 0.2923

음주량 (p-value < 0.01)

```
> chisq.test(data$asthma,data$alcohol)
```

Pearson's Chi-squared test

data: data\$asthma and data\$alcohol
X-squared = 12.601, df = 2, p-value = 0.001836

흡연 여부 (p-value = 0.03584)

```
> chisq.test(data$asthma,data$smoking)
```

Pearson's Chi-squared test

data: data\$asthma and data\$smoking
X-squared = 6.6572, df = 2, p-value = 0.03584

비만 유병 여부 (p-value = 0.01726)

```
> chisq.test(data$asthma,data$obesity)
```

Pearson's Chi-squared test with Yates' continuity correction

data: data\$asthma and data\$obesity
X-squared = 5.6695, df = 1, p-value = 0.01726

고혈압 유병 여부 (p-value < 0.01)

```
> chisq.test(data$asthma,data$hypertension)
```

Pearson's Chi-squared test with Yates' continuity correction

data: data\$asthma and data\$hypertension
X-squared = 33.547, df = 1, p-value = 6.955e-09

음주량, 흡연 여부,
비만 유병 여부,
고혈압 유병 여부와
천식은 관련성이 있음.

4.2 분석결과 - 내부적 요인

② T-test 검증

나이 (p-value <0.01)

```
> t.test(data_y$age,data_n$age,var.equal = F,alternative = 'greater')
```

Welch Two Sample t-test

```
data: data_y$age and data_n$age
t = 9.5779, df = 512.64, p-value < 2.2e-16
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 11.48919      Inf
sample estimates:
mean of x mean of y
 57.38056  43.50397
```

수축기 혈압 (p-value <0.01)

```
> t.test(data_y$sbp,data_n$sbp,var.equal = T,alternative = 'greater')
```

Two Sample t-test

```
data: data_y$sbp and data_n$sbp
t = 5.6656, df = 610, p-value = 1.13e-08
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 5.368797      Inf
sample estimates:
mean of x mean of y
122.6333 115.0635
```

이완기 혈압 (p-value <0.01)

```
> t.test(data_y$dbp,data_n$dbp,var.equal = T,alternative = 'greater')
```

Two Sample t-test

```
data: data_y$dbp and data_n$dbp
t = 3.0642, df = 610, p-value = 0.001139
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 1.149727      Inf
sample estimates:
mean of x mean of y
 75.46667  72.98016
```

**천식 유병 여부에 따라
나이, 수축기 혈압, 이완기 혈압의
차이가 있다.**

4.2 분석결과 - 내부적 요인

③ 로지스틱 회귀 분석

Step: AIC=746.38

asthma ~ sex + age + smoking + obesity + hypertension

Stepwise 방법을 이용하여
변수 선택 및 최종 모형 결정

	Df	Deviance	AIC
<none>		732.38	746.38
- obesity	1	734.77	746.77
- hypertension	1	734.78	746.78
+ alcohol	2	730.02	748.02
- smoking	2	738.48	748.48
- sex	1	736.79	748.79
- age	1	778.71	790.71

독립성 검정과
T-test에서 검정한
변수들을 사용하여
로지스틱 회귀분석 진행

Call: glm(formula = asthma ~ sex + age + smoking + obesity + hypertension,
family = binomial(link = "logit"), data = data)

Coefficients:

(Intercept)	sex2	age	smoking2	smoking3	obesity1	hypertension1
-2.25440	0.47210	0.03749	0.58104	0.49790	0.30030	0.35043

Degrees of Freedom: 611 Total (i.e. Null); 605 Residual

Null Deviance: 829.3

Residual Deviance: 732.4 AIC: 746.4

4.2 분석결과 - 내부적 요인

③ 로지스틱 회귀 분석

```
> summary(model)
```

```
Call:
glm(formula = asthma ~ sex + age + smoking + obesity + hypertension,
     family = binomial(link = "logit"), data = data)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.1064	-1.0298	0.6305	0.9102	1.8134

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-2.254400	0.345548	-6.524	6.84e-11	***
sex2	0.472103	0.226476	2.085	0.0371	*
age	0.037491	0.005738	6.534	6.41e-11	***
smoking2	0.581038	0.258955	2.244	0.0248	*
smoking3	0.497903	0.276346	1.802	0.0716	.
obesity1	0.300297	0.194902	1.541	0.1234	
hypertension1	0.350433	0.226423	1.548	0.1217	

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 829.25 on 611 degrees of freedom
Residual deviance: 732.38 on 605 degrees of freedom
AIC: 746.38

Number of Fisher Scoring iterations: 4

Pr(>|z|)

- ✓ $\Pr(>|z|) < 0.1$
성별, 나이, 흡연 여부(비흡연자, 흡연 경력이 있는 사람, 흡연자)가 유의하다
=> 천식의 위험성을 높임.
- ✓ $\Pr(>|z|) \approx 0.1$
비만 유병 여부, 고혈압 유병 여부 영향을 미치는 변수라고 생각할 수 있다.

Estimate

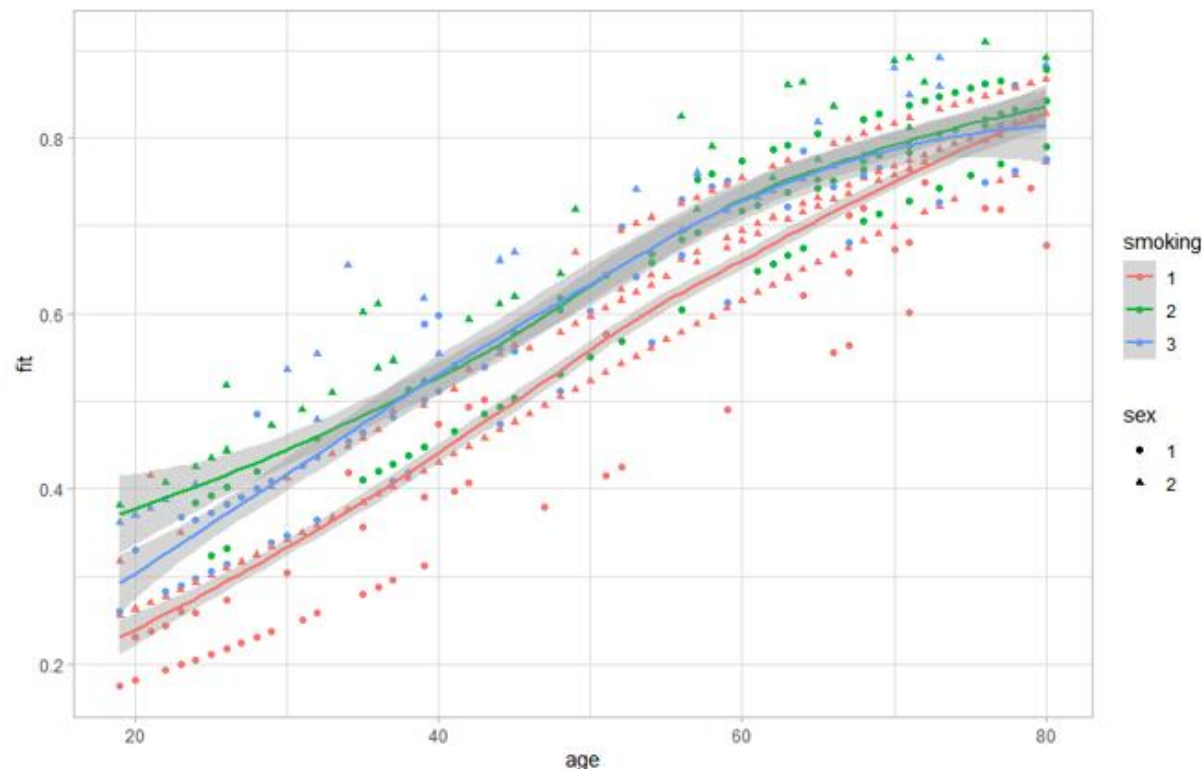
- ✓ 성별: 여성일 경우 $\exp(0.472130) = 1.6$ 배 높음
- ✓ 나이: 나이가 1단위 증가할 때 $\exp(0.037491) = 1.03$ 배 높음
- ✓ 흡연: 흡연자의 경우 $\exp(0.497903) = 1.64$ 배 높음
- ✓ 비만: 비만일 경우 $\exp(0.300297) = 1.35$ 배 높음
- ✓ 고혈압: 고혈압일 경우 $\exp(0.350433) = 1.41$ 배 높음

=> 천식은 **성별, 나이, 흡연 여부, BMI, 혈압**에 따라 위험률이 달라진다.

4.2 분석결과 - 내부적 요인

③ 로지스틱 회귀 분석

Logistic Regression Plot



```
> vif(model)
```

	GVIF	Df	GVIF^(1/(2*Df))
sex	1.531918	1	1.237707
age	1.323887	1	1.150603
smoking	1.594501	2	1.123715
obesity	1.051262	1	1.025310
hypertension	1.315021	1	1.146743

VIF(분산 팽창 지수 <5)
: 다중 공선성이 존재하지 않는다

= Hosmer and Lemeshow 검정 =

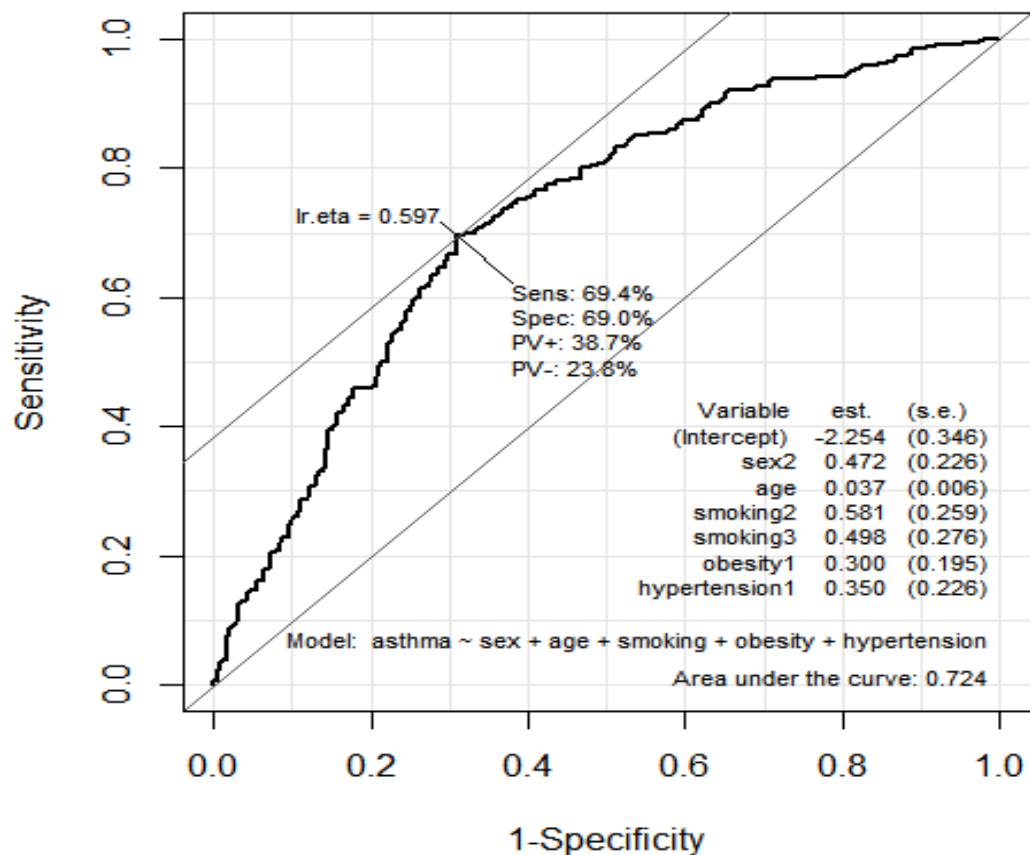
단계	카이제곱	자유도	유의확률
1	7.107	8	.525

Hosmer Lemeshow 검정
(p-value = 0.525) : 모형이 적합하다.

4.2 분석결과 - 내부적 요인

③ 로지스틱 회귀 분석

ROC Curve



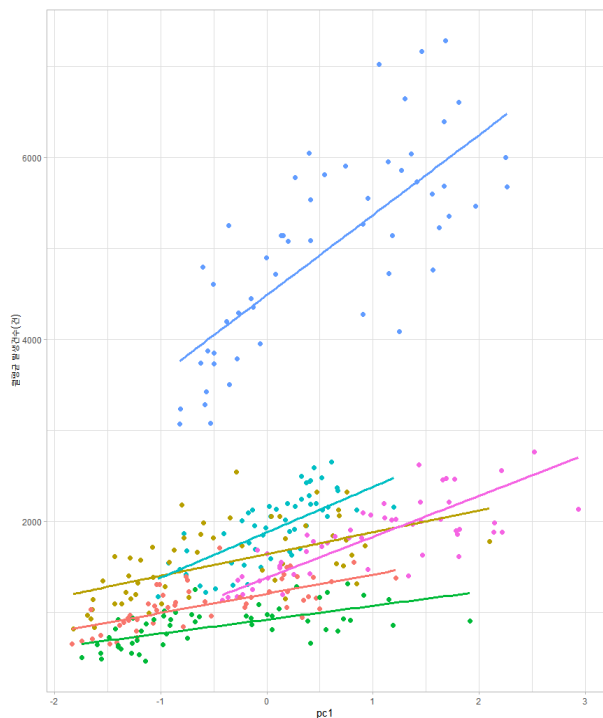
ROC Curve

: 민감도와 특이도가 어떤 관계를 갖고 있는지를 표현한 그래프

- ✓ 민감도(Sensitivity) : 69.4%
: 천식 환자를 천식 환자라고 판단 할 확률
- ✓ 특이도(Specificity) : 69%
: 비 천식환자를 천식환자라고 판단하지 않을 확률
- ✓ AUC(Area Under The Curve) : 0.724
: ROC 곡선 아래 영역

4.3 분석결과 - 결과

Regression plot



- ✓ 기온과 습도가 높으면 천식 발생률 감소
- ✓ 대기 오염이 높으면 발생률이 증가

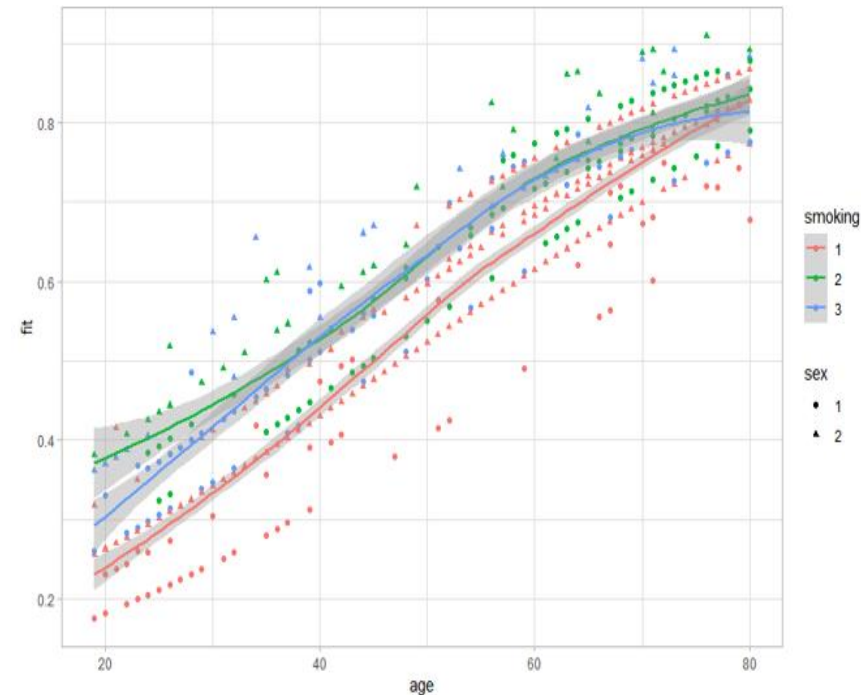
BUT 습도가 높고 대기오염이 같이 높은 경우 오히려 **천식 발생률이 증가**한다.

Support Vector Machine



- ✓ 기상 데이터를 이용하여 정확한 예측이 가능
- ✓ **기상 변수를 이용한** SVM, Random Forest가 예측력이 좋다.

Logistic Regression Plot



- ✓ 유전적 요인(비만, 성별)과 **천식 위험성과 밀접한 관련**이 있음
- ✓ 내부적 요인(술, 담배, 혈압)은 **천식 위험성과 밀접한 관련**이 있음

5

서비스 활용 방안

5. 서비스 활용 방안

① 천식알람 서비스 강화 (어플 활용 예시)



현재 알람 서비스 활용 정보

- ✓ 기상기상 관측정보
- ✓ 대기오염 통계정보

맞춤화 서비스 강화

알람서비스에 개인정보 입력

- ✓ 유전적 요인(비만, 성별)
- ✓ 내부적 환경 요인 (술, 담배, 혈압)

5. 서비스 활용 방안

② 천식 환자를 위한 주거입지 선정에 활용

활용 대상

유전적으로 천식을
가지고 있는 가족



천식 환자를 위한
사회적 공간 필요할 경우



기상환경에 민감한
어린이 또는 노인



제공

- 날씨와 대기오염
정도를 통하여
천식 위험지수를 예측
- 평균적으로
가장 적합한 환경을
제공하는 지역을 선정

➡ 천식 위험성이 낮은 지역으로 선정된 도시는 추후 이를 **도시홍보**에 사용 가능

5. 서비스 활용 방안

③ 천식 악화 방지 환경을 조성하는 IoT기술 개발



에어컨



가습기



공기청정기

에어컨, 가습기, 공기청정기와 같은 제품에 IoT기술을 적용하여 자동적으로 천식 환자에게 **적합한 환경을 제공**하는 기술을 개발 할 수 있다.

6

서비스 기대 효과

6. 서비스 기대효과

개인적 특성을 고려한 알람 서비스

기존의 알람 서비스
기상변수에 따른 지역적
특성만 고려



개인적 특성을 고려



개인의 특성에 맞춰
천식악화방지 방법 고안

천식뿐 아니라 다른 호흡기 질환에도 사용



여러 호흡기 질환들

기상데이터를 이용하여 예측하는 방법과 개인적 특성을 고려하는 방식을 천식 이외의 호흡기 질환과 관련하여 위험성 및 악화 방지 방안을 고안할 수 있다.

참고문헌

- Sang Heon Cho, Yoon Keun Kim, Yoon Seok Chang, Sun Sin Kim, Kyung Up Min, You Young Kim. Original Articles : Asthma insights and reality in Korea. The Korean Journal of Medicine 2006;70(1): 69-77.
- Trokart, Demarche, Schleich, Paquot, Louis. Asthma and obesity. Rev Med Liege 2017 May;72(5):241-245.
- Dunn, Busse, Wechsler. Asthma in the elderly and late-onset adult asthma. Allergy. 2018 Feb;73(2):284-294. doi: 10.1111/all.13258. Epub 2017 Sep 7.
- Guarnieri, Balmes. Outdoor air pollution and asthma. Lancet. 2014 May 3;383(9928):1581-92. doi: 10.1016/S0140-6736(14)60617-6.
- Nasreen S1, Nessa A, Islam F, Khanam A, Sultana R, Alam, Naznin, Tajkia, Rahman, Hossain, Sultana, Kamal. Interaction between Hypertension and Asthma in Adult. Mymensingh Med J. 2018 Jan;27(1):34-40.
- Underner, Perriot, Peiffer, Meurice. Influence of tobacco smoking on the risk of developing asthma. Rev Mal Respir. 2015 Feb;32(2):110-37. doi: 10.1016/j.rmr.2014.07.014. Epub 2014 Oct 18.
- 기후과학국 기후정책과 보도자료 , 기상청 기후정보포털 - 2018년 이상기후 보고서
- 지난 3월 6일자 경향신문, 조선일보, 동아일보, 한국일보 1면에 실린 기사 갈무리. 디자인=안혜나기자
- [출처] 미세먼지가 건강에 미치는 영향과 우리나라 정부 R&D 동향. 한국보건산업진흥원 이명성. NICE. 제 3권 제 4호
- [출처] 건강보험심사평가원. 2015년 기준 OECD 보건통계 해설서

—
감사합니다