

# 목차

1. 배경 분석

2. 데이터 확정

3. 실험 계획

4. 실험 결과

5. 실험 한계

# 배경 분석

# 배경분석

## 사회적 문제 '흡연'



동아일보

### 사망률 1위 폐암... 흡연-미세먼지-돌연변이 등 원인

흡연자는 비흡연자에 비해 폐암 발생 위험이 10~30배가량 높다고 알려져 있다. 폐암 발생 위험은 흡연 시작 연령이 낮을수록, 흡연 기간이 길수록

부산일보

### [SNS:판] '층간흡연, 뽕집서 흥기 들고 찾아왔어요'... 법적 책임은?

글쓴이는 그동안 뽕집과 층간 흡연 때문에 크고 작은 분쟁이 있었다고 했다. 그러면서 최근 일어난 일화를 언급했다. 그는 "뽕집 주인이 엘리베이터 ...

헬스조선

### "흡연, 뇌혈관 터뜨려"... 美 심장협회 연구

그 결과, 조사 기간 중 120명이 지주막하 출혈로 사망했고 사망 시 연령은 평균 61세였다. 또한 담배를 피우지 않는 대상군과 비교했을 때 담배를 ...



중앙일보

### 금연 늘었지만...흡연으로 한해 6만1723명 숨져 - 중앙일보

성인 남성 흡연율은 감소하고 있는데 왜 흡연 사망자가 늘어날까. 20~25년 전 흡연 이력이 쌓여서 질병 사망으로 나타나기 때문이다. 정금지 교수는 " ...

2019. 8. 28.



한국일보

### 길에 버린 공초...1년 8억 개비 바다로 흘러든다

해마다 흡연자가 무단투기한 담배공초 8억개비가 거리 우수관(빗물 ... 에서 수집한 3,879개의 쓰레기를 분석한 결과, 가장 많이 수거된 쓰레기도 ...



헬스조선

### 간접흡연의 심각한 폐해... "코로나19 감염시킬 수도"

흡연자는 담배를 태우기 위해 흡연장소로 가야 한다. 외부에 마련돼 있는 곳도 있지만, 대다수가 실내에서 밀폐돼 있거나, 많은 사람이 모여서 피우는 ...



# 서울의 흡연율

행정구역	2019 흡연율
충청북도	22.2
충청남도	21.5
서울특별시	21.1
인천광역시	20.7
제주특별자치도	20.6
경기도	19.9
대구광역시	19.8
광주광역시	19.6
경상북도	19.2
울산광역시	19.0
부산광역시	19.0
전라남도	18.9
대전광역시	18.9
경상남도	18.9
전라북도	17.8
강원도	17.0
세종특별자치시	15.9



흡연율이 전국 3위로 높은 순위

흡연의 다양한 요인 중 가장 큰 영향을 주는 것은 무엇일까?

# 데이터 확정

# 데이터 확정

## 1. 성별 인구수

자치구	세대	계	남자	여자
강남구	231,612	561,052	268,941	292,111
강동구	177,407	440,359	218,699	221,660
강북구	142,533	328,002	160,252	167,750
강서구	254,257	608,255	296,175	312,080
관악구	255,352	520,929	262,006	258,923
광진구	160,798	372,298	180,645	191,653
구로구	171,570	441,559	221,386	220,173
금천구	106,066	253,491	129,775	123,716
노원구	217,619	558,075	271,025	287,050
도봉구	137,378	346,234	169,553	176,681
동대문구	159,938	366,011	181,185	184,826
동작구	172,995	408,493	198,827	209,666
마포구	169,408	385,783	183,248	202,535
서대문구	137,266	325,028	155,250	169,778
서초구	173,594	445,401	213,422	231,979
성동구	132,902	312,711	154,077	158,634
성북구	187,112	455,407	221,103	234,304

## 2. 지역 내 총생산량(소득수준)

자치구	지역내총생산(당해년가격)	1인당 지역내총생산
강남구	62,295,492	115,245
강동구	7,519,077	16,722
강북구	2,759,140	8,571
강서구	10,149,098	17,802
관악구	5,597,548	10,734
광진구	5,696,485	15,417
구로구	12,017,187	26,919
금천구	14,953,994	59,434
노원구	5,606,176	9,903
도봉구	2,906,723	8,506
동대문구	6,589,147	17,996
동작구	5,521,181	13,391
마포구	17,955,655	47,012
서대문구	7,507,777	24,235
서초구	30,390,265	72,294
성동구	10,231,864	34,653
성북구	5,407,042	11,762

### 3. 음주율

지역	월간음주율	연간음주자의 고위험음주율
강남구	63.6	11.1
강동구	60.3	14.7
강북구	59.4	20.6
강서구	58.2	17.5
관악구	60.9	15.8
광진구	61.7	13.3
구로구	58.5	17.4
금천구	59.7	24.6
노원구	56.5	19.5
도봉구	61.3	18.1
동대문구	63.4	22.3
동작구	59.1	14.9
마포구	61.2	15.7
서대문구	59.3	19.7
서초구	60.5	11.5
성동구	65.4	17.8
성북구	63	

### 4. 이혼 가정 수

자치구	혼인	이혼
강남구	2,677	805
강동구	2,427	784
강북구	1,436	742
강서구	4,069	1,120
관악구	3,205	817
광진구	2,082	612
구로구	2,460	775
금천구	1,439	525
노원구	2,420	1,003
도봉구	1,543	666
동대문구	1,657	624
동작구	2,282	598
마포구	2,294	576
서대문구	1,546	485
서초구	2,288	561
성동구	2,096	509
성북구	1,917	732

### 5. 스트레스율

분류	전혀 느끼지 않았다	느끼지 않은 편이다	보통이다	느낀 편이다	매우 많이 느꼈다	평균(10점)
강남구	0.6	21.4	23.9	31.9	22.2	6.34
강동구	0.7	19.5	18.3	40	21.5	6.55
강북구	0.5	12.6	24.8	46.8	15.3	6.59
강서구	1.2	24.2	24	24.4	26.1	6.25
관악구	2.4	16.7	25.2	36.7	19.1	6.33
광진구	0.8	17.8	26.4	37	17.9	6.33
구로구	3.4	8.7	30.4	39.2	18.2	6.5
금천구	0.6	13.2	27.3	37.9	21	6.63
노원구	0.7	15.2	33.8	27.7	22.6	6.41
도봉구	1.2	11.2	30.2	38.6	18.8	6.57
동대문구	1.9	19.1	21.7	33.2	24.2	6.47
동작구	0.9	27.6	21.3	24.1	26.2	6.17
마포구	2.2	12.3	40.8	24.7	20.1	6.2
서대문구	1.3	11.5	36.2	30.7	20.4	6.44
서초구	0.7	26.7	15.2	33.4	24.1	6.33
성동구	0.3	8.7	32	44.4	14.6	6.61
성북구	0.9	11.9	35.9	35.9	15.4	6.32

## 6. 평균 연령

지역	전체평균연령	남자	여자
합계	42.9	42.1	43.7
종로구	44.4	43.9	44.9
중구	44.8	44.2	45.4
용산구	43.7	42.8	44.6
성동구	42.7	42	43.4
광진구	42.1	41.6	42.6
동대문구	43.4	42.8	44.1
중랑구	44.6	43.7	45.4
성북구	42.9	42.1	43.7
강북구	45.9	44.7	47
도봉구	45.2	44.1	46.3
노원구	43.1	41.7	44.4
은평구	44	43	44.9
서대문구	43.1	42.3	43.9
마포구	41.7	41	42.2
양천구	42.1	41.1	42.9
강서구	42.5	41.7	43.2

## 7. 우울증을 앓고 있는 사람 수

지역	현재흡연율	우울증 경험율
강남구	19.1	5.2
강동구	23	12.3
강북구	26.1	1.9
강서구	26.1	8.7
관악구	27.7	7.6
광진구	28.6	8.8
구로구	27.2	11.2
금천구	27	7.7
노원구	24.2	10.6
도봉구	24	6.5
동대문구	26.5	7.9
동작구	21	5.5
마포구	25.3	4.7
서대문구	26.3	12.9
서초구	19.2	8.4
성동구	26.4	7.9
성북구	25.4	7.6
송파구	18	11
양천구	22.2	7.5
영등포구	20.6	13.2
용산구	26.5	8
은평구	27.7	5.3
종로구	22.2	7.9
중구	25.2	2.8
중랑구	25.5	6.8

## 8. 실업률

지역	실업률
강남구	6.3
강동구	5.2
강북구	8.3
강서구	8.6
관악구	9.2
광진구	9.3
구로구	9.3
금천구	8.4
노원구	6.5
도봉구	5.2
동대문구	6.5
동작구	4.3
마포구	8.9
서대문구	8.3
서초구	6.5
성동구	8.3
성북구	5.6
송파구	5.7
양천구	8.9
영등포구	8.3
용산구	4.2
은평구	10.1
종로구	5.3
중구	9.8
중랑구	8.7



# 실험계획

## 가설설정

- 지역 내 총생산량(소득수준)이 낮을수록 흡연율이 높을 것이다
- 지역 내 남자 성비가 높을수록 흡연율이 높을 것이다.
- 음주율이 높을수록 흡연율이 높을 것이다
- 우울증 환자 수의 비율이 높을수록 흡연율이 높을 것이다
- 이혼 가정수가 많을수록 흡연율이 높을 것이다
- 스트레스를 받는 사람이 많을수록 흡연율이 높을 것이다
- 평균연령이 높을수록 흡연율이 높을 것이다
- 실업률이 높을수록 흡연율이 높을 것이다

## 변수설정

독립변수	의미
gdp2	지역 내 총생산량/1000 (소득수준을 나타내는 지표로 설정)
malerate	남자 수/지역 내 전체 인구 수
drink	지역 내 음주율
depression	우울증을 겪고 있는 사람의 비율
divorce	이혼 가정수
stresshigh	지역 내 스트레스가 높은 사람의 비율
lostjob2	실업율 <sup>2</sup> 값

종속변수	의미
smokerate	흡연율



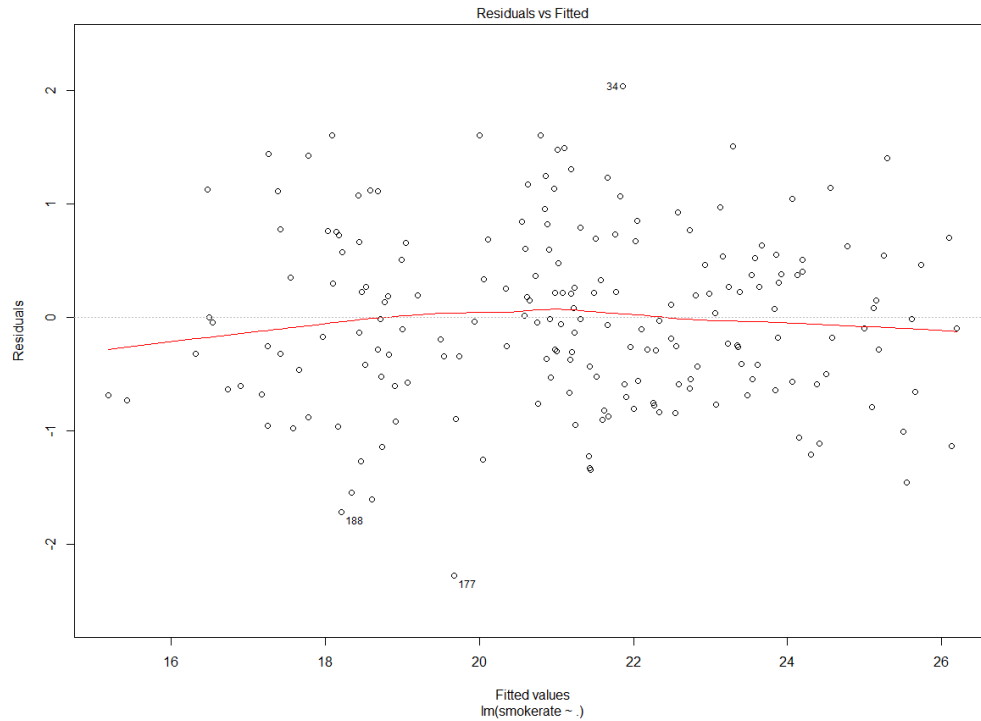
# 실험결과

## 회귀분석의 검증 (독립성)

```
> vif(smoke_model6)
      gdp2  malerate    drink depression  divorce stresshigh  avr_age  lostjob2
1.067479  1.313157  2.332338  2.178692  2.590573  1.224564  3.885419  3.069697
```

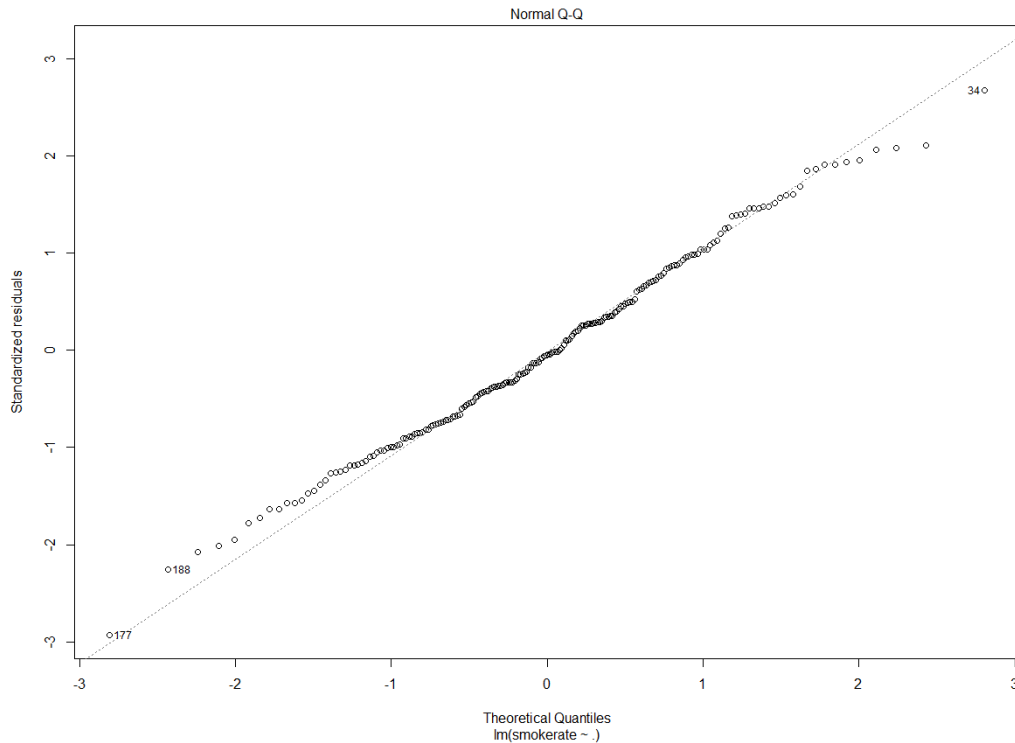
모든 변수에 해당하는 VIF 값이 4 미만이므로 다중공선성이 일어나지 않는다.  
( 변수들 사이에 높은 상관관계가 존재하지 않는다 )

## 회귀분석의 검증 (등분산성)



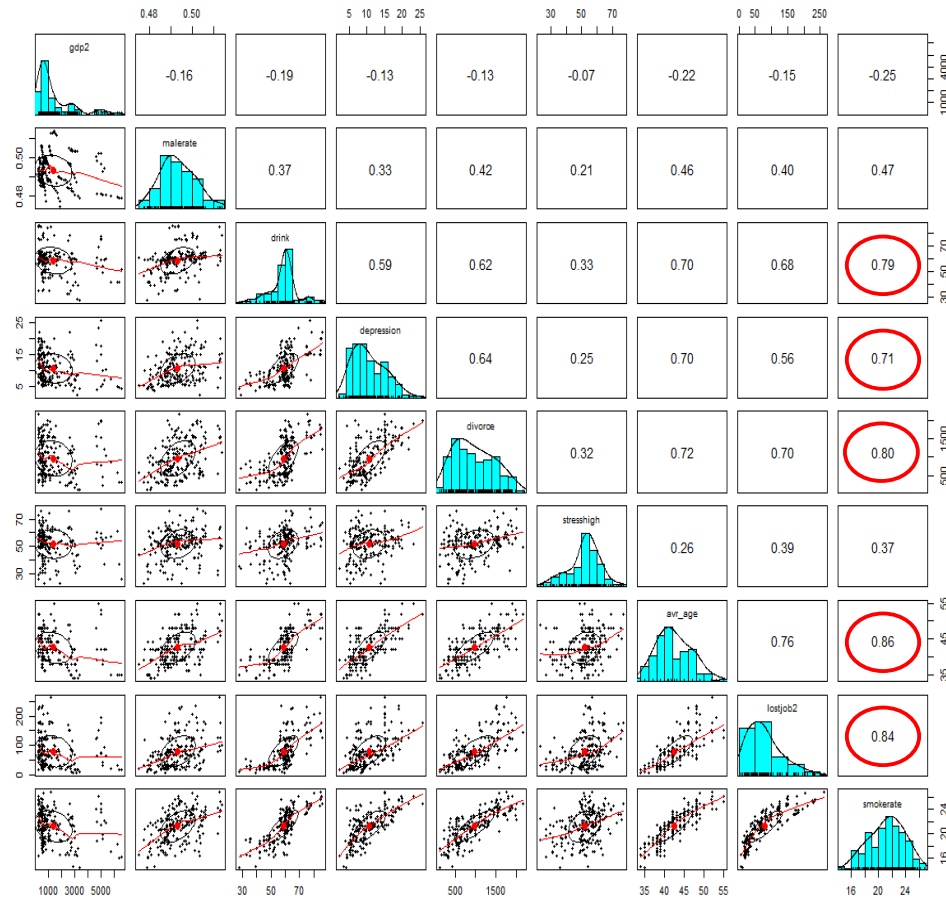
- Fitted Value 와 Residual 의 Plot을 확인
- 잔차가 전체적으로 고른 분산을 가진다
- 등분산성을 만족한다고 볼 수 있다.

## 회귀분석의 검증 (정규성)



- Normal Q-Q Plot으로 Plot을 확인
- 45도 line으로 이상치 1개를 제외하고 전체적인 잔차가 직선에 분포하는 것을 확인할 수 있다.
- 정규성을 만족한다고 볼 수 있다.

## 상관분석 결과



- 변수 간 상관관계 파악 ( Pairs.panels 모형 )
- 상관계수의 값이 -1에 가까우면 음(-)의 상관관계  
1에 가까우면 양(+)의 상관관계를 나타냄
- 양의 관계: malerate, drink, depression, divorce, stresshigh, avr\_age, lostjob
- 음의 관계: gdp





## 회귀분석 결과

```
> summary(smoke_model6)
```

Call:

```
lm(formula = smokerate ~ gdp2 + malerate + drink + depression +  
    divorce + stresshigh + avr_age + lostjob2, data = smoke_data3)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-2.32152	-0.51431	0.03234	0.57684	2.19238

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	4.443e+00	4.302e+00	1.033	0.30303	
gdp2	-1.302e-04	4.682e-05	-2.780	0.00598	**
malerate	9.804e+00	8.978e+00	1.092	0.27618	
drink	4.858e-02	9.288e-03	5.230	4.42e-07	***
depression	6.718e-02	2.039e-02	3.295	0.00117	**
divorce	1.031e-03	2.113e-04	4.877	2.26e-06	***
stresshigh	5.761e-03	6.688e-03	0.861	0.39017	
avr_age	1.454e-01	2.834e-02	5.131	7.06e-07	***
lostjob2	1.362e-02	2.093e-03	6.510	6.48e-10	***

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

Residual standard error: 0.8983 on 191 degrees of freedom

Multiple R-squared: 0.8842, Adjusted R-squared: 0.8793

F-statistic: 182.2 on 8 and 191 DF, p-value: < 2.2e-16

## 1. Multiple R-squared / Adjusted R-squared

```
Residual standard error: 0.8983 on 191 degrees of freedom  
Multiple R-squared: 0.8842, Adjusted R-squared: 0.8793  
F-statistic: 182.2 on 8 and 191 DF, p-value: < 2.2e-16
```

- $R^2 = 0.8842$ 로 회귀식이 모델에 대해 88.42%의 설명력을 가진다.
- 수정된 결정계수  $R^2 = 0.8793$ 은 독립변수의 수와 표본의 크기를 고려하여 수정된 값으로 87.93%의 설명력을 가진다.
- $R^2$ , Adjusted  $R^2$ 의 값 모두 1에 가깝기 때문에 높은 설명력을 가진다.

## 2. F-statistic

```
Residual standard error: 0.8983 on 191 degrees of freedom  
Multiple R-squared: 0.8842, Adjusted R-squared: 0.8793  
F-statistic: 182.2 on 8 and 191 DF, p-value: < 2.2e-16
```

- F 통계량의 p-value가 작다는 것은 유의한 회귀계수가 있다는 뜻
- 상관관계가 있음을 보이고 싶다면 값이 작을수록 좋다
- 유의확률이 유의수준인 0.05보다 낮은 값을 가지므로 회귀식의 전체 모형이 유의한 것으로 판정

3.  $\Pr(<|t|)$ 

```

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.443e+00  4.302e+00   1.033  0.30303
gdp2         -1.302e-04  4.682e-05  -2.780  0.00598 **
malerate      9.804e+00  8.978e+00   1.092  0.27618
drink         4.858e-02  9.288e-03   5.230  4.42e-07 ***
depression    6.718e-02  2.039e-02   3.295  0.00117 **
divorce       1.031e-03  2.113e-04   4.877  2.26e-06 ***
stresshigh    5.761e-03  6.688e-03   0.861  0.39017
avr_age       1.454e-01  2.834e-02   5.131  7.06e-07 ***
lostjob2      1.362e-02  2.093e-03   6.510  6.48e-10 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

- p-value 값이 0.05보다 작으면 유의한 차이를 지닌다.
- 독립변수 gdp2, drink, depression, divorce, avr\_age, lost job2와 종속 변수 smokerate 사이의 인과관계가 성립한다.

## 4. Estimate

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	4.443e+00	4.302e+00	1.033	0.30303	
gdp2	-1.302e-04	4.682e-05	-2.780	0.00598	**
malerate	9.804e+00	8.978e+00	1.092	0.27618	
drink	4.858e-02	9.288e-03	5.230	4.42e-07	***
depression	6.718e-02	2.039e-02	3.295	0.00117	**
divorce	1.031e-03	2.113e-04	4.877	2.26e-06	***
stresshigh	5.761e-03	6.688e-03	0.861	0.39017	
avr_age	1.454e-01	2.834e-02	5.131	7.06e-07	***
lostjob2	1.362e-02	2.093e-03	6.510	6.48e-10	***

---  
signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

- GRDP가 0.1증가할 때 흡연율이 1.3%만큼 감소한다.
- 음주율이 10% 증가하면 흡연율이 0.46% 증가한다.
- 우울증을 겪는 사람의 비율이 10% 증가하면 흡연율이 0.67% 증가한다.
- 이혼 가정수가 1 증가하면 흡연율이 0.10% 증가한다.
- 평균연령이 0.1세 증가하면 흡연율이 1.45% 증가한다.
- 실업률 값이 0.1 증가하면 흡연율이 0.14% 증가한다.

Regression.Smokerate.com

실험한계

## 분석결과의 한계

1. 사용한 데이터의 개수가 많지 않음
2. 각 개인의 데이터를 사용한 것이 아니라 서울시 자치구별 통계량을 이용함
3. 종속변수와 독립변수가 직접적인 인과관계를 가지는지에 대해서 의문점을 가질 수 있다.
4. 개인의 흡연 여부는 다양한 변수들이 존재한다 (같은 데이터라도 결과가 다를 수 있음)

? Q&A !