



# 생활데이터 분석: 안구건조

20121332 국어국문학과 김정연

목차



## 주제의 선정 및 개요

주제의 선정 및 조사 개요  
설문 문항 소개

## 자료 가공 및 분석

문항별 기초 자료 요약  
서로 다른 두 문항을 이용한 분석  
분석을 통해 밝혀진 사실 요약



주제의 선정 및 조사개요



최근 1년 동안 눈이 자주 아프고 시린데, 정확한 이유를 알 수가 없어  
“생활데이터”라는 이름에 걸맞게 데이터를 수집하여  
안구건조의 이유를 파악하고자 함



주제의 선정 및 조사개요



1. 조사 기간 : 2016년 10월 2일 - 11월 20일

2. 입력된 자료의 수 : 38개의 데이터

3. 입력 방법 : 눈이 아플 때 노트북이 눈 앞에 있으면 바로 작성,  
이외에는 수기로 작성하였다가 하루 마무리를 하면서 기록



## 설문 문항 소개



문항	값	사용의도
1. 오늘 안구건조(눈뻑뻑/시림/아픔)가 일어났습니까?	1.예 2.아니오	안구건조 증상이 일어났는지의 여부를 알고자 함. 안구건조 증상이 일어나지 않았어도 그 날의 데이터는 수집하는 것이 좋다고 판단했기 때문.
2. 어떤 종류의 증상이었습니까?	1. 눈뻑뻑 2. 시림 3. 아픔	어느 증상이 어느 요인과 관계가 있는지 알아보고 싶어 작성한 문항.
3. 오늘 렌즈를 착용하였습니까?	1.예 2.아니오	유독 렌즈를 낄 때 눈이 힘든 것 같아서 작성한 문항.
3-1. 렌즈를 낀 지 몇 시간만에 안구건조를 경험하였습니까?	몇 시간만에 눈이 불편하였는지 반올림하여 직접 숫자 입력	렌즈를 낀 뒤 몇 시간만에 눈이 불편했는지 알기 위해 작성한 문항. 렌즈를 밖에 나갈 때만 끼기 때문에 밖에 나가는 것이 전제되어 있음. 밖에 나가는 것을 전제로 한 것은 미세먼지와 증상을 관련지을 수 있을지 여부를 알아보기 위함임.
3-2. 렌즈를 끼지 않았다면, 집에서 나온 지 몇 시간만에 안구건조를 경험하였습니까?	몇 시간만에 눈이 불편하였는지 반올림하여 직접 숫자 입력	렌즈를 끼지 않고서도 눈이 불편할 때가 있어서 작성한 문항. 이번 학기는 특히 렌즈를 끼지 않은 적이 많아 바로 위 문항과 더불어 꼭 있어야 했던 문항이었다고 생각.



## 설문 문항 소개



문항	값	사용의도
4. 안구건조 증상을 극복하는 데 걸린 시간은 어느 정도입니까?	1.눈 깜빡 몇번만에 나옴 2.1-3분 3. 5분	증세에 따라서 어느 정도 극복 시간이 걸리는지 알아보고자 작성한 문항.
5. 안구건조 증상 극복 이후 렌즈가 뿌여지거나 계속해서 눈이 시리는 등 일상생활에 불편이 있었습니까?	1.예 2.아니오	최근 1년 안구건조 증상이 심해지면서 증상이 나아도 계속해서 눈이 불편한 경우가 왕왕 있게 되어 작성한 문항.
6. 전날 잠은 몇 시간 잤습니까?	몇 시간 동안 잤는지 반올림하여 직접 숫자 입력	전날의 수면시간이 안구건조에 영향을 미치는지 알아보기 위해 작성한 문항.
7. 오늘 전자기기는 도합 대략 몇 시간 사용하였습니까? (노트북 및 핸드폰 화면 등)	몇 시간 동안 이용하였는지 반올림하여 직접 숫자 입력	전자기기 모니터를 바라본 시간이 안구건조에 영향을 미치는지 알아보기 위해 작성한 문항.
8. 오늘 안약(인공눈물 포함)을 점안 하였습니다?	1.예 2.아니오	안구건조에 안약 점안이 효과적인지 알아보기 위해 넣은 문항인데(5번과 엮어 분석하고 싶었음), 실험 기간 동안 안약을 잘 넣지 않아서 유의미한 데이터를 이끌어내기는 어려울 것으로 보임.



설문 문항 소개



## 문항

## 값

## 사용의도

9. 오늘의 미세먼지 농도는 어느 정도였습니까? (아시아의 대기오염 홈페이지 기준)

1.0-50  
2.51-100  
3.101-150  
4.151-200  
5.201-300  
6.300+

미세먼지 농도가 안구건조에 영향을 미치는지 알아보기 위해 작성한 문항. 국내 포털 사이트에서 제공하는 미세먼지 농도는 국제 표준과 다르다는 이야기가 많아 국제 표준을 사용하는 사이트의 기준을 사용하였고, 매일 미세먼지 농도를 확인하고 작성함.

10. (7번 데이터부터 추가) 컴퓨터 모니터는 몇 시간 보았습니까?

몇 시간동안 이용하였는지  
반올림하여 직접 숫자 입력

이번 학기 코딩을 시작하고 평소보다 노트북 모니터를 뚫어지게 볼 기회가 많아지면서 눈이 더 건조해지는 것 같아 추가한 문항.

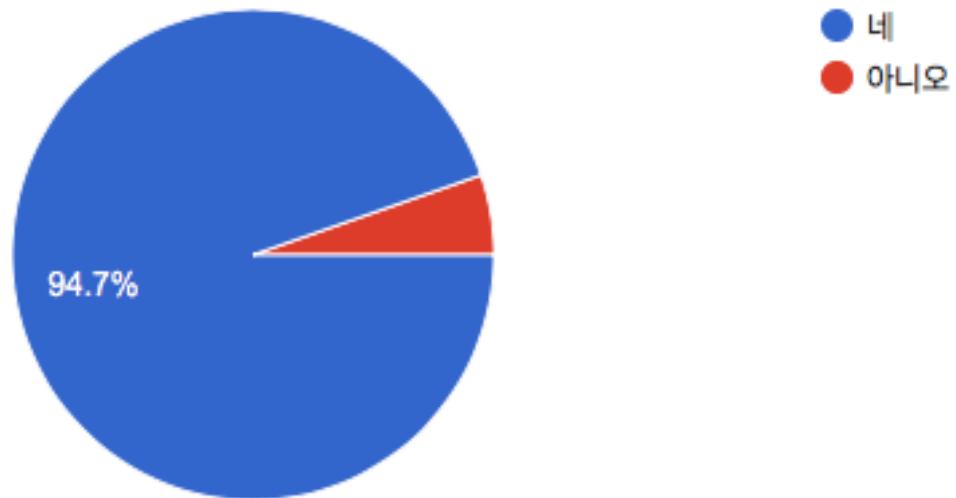


자료 가공 및 분석



## 문항별 기초자료 요약

1. 오늘 안구건조(눈뻑뻑/시림/아픔)가 일어났습니까? (응답 38개)



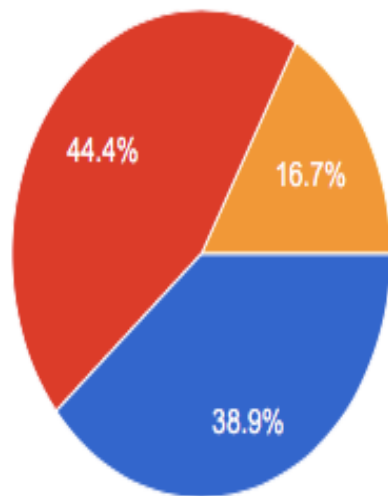
- 수집한 38개의 데이터 중 안구건조가 일어난 날이 총 36개로 94.7%를 차지함
- 안구건조가 거의 매일 일어났다고 볼 수 있음





## 문항별 기초자료 요약

### 2. 어떤 종류의 증상이었습니까? (응답 36개)

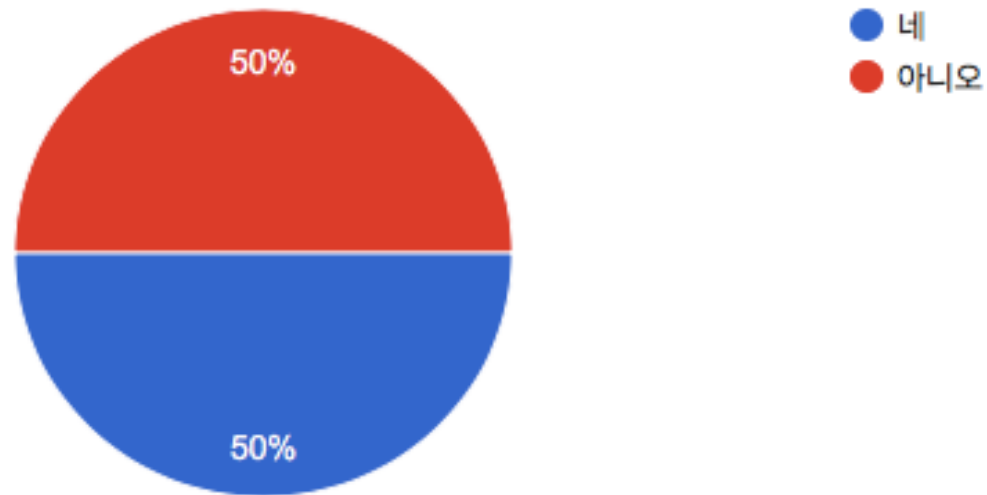


- 눈백백 (렌즈 꺾을 때 시야 가려지는 정도)
- 시림
- 아픔

- 눈이 시리거나 백백한 증상이 증상의 대다수를 차지함 (83.3%)
- 눈이 아픈 경우는 상대적으로 적은 것을 확인할 수 있음

## 문항별 기초자료 요약

### 3. 오늘 렌즈를 착용하였습니까? (응답 38개)



- 수집한 38개의 데이터에서  
렌즈 낀 날의 빈도수와  
렌즈 끼지 않은 날의 빈도수는  
각각 19개로 같은 비율을 보임

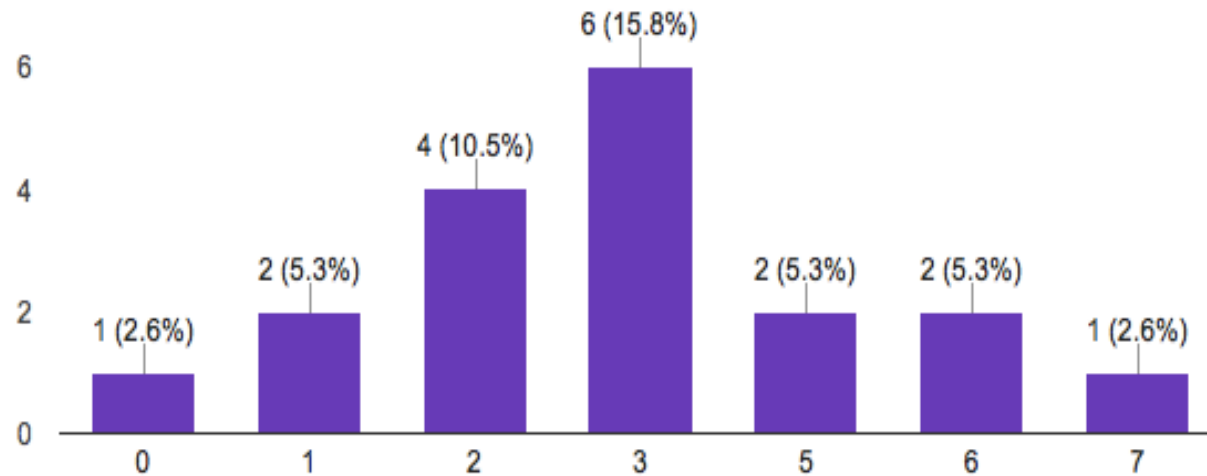


자료 가공 및 분석



## 문항별 기초자료 요약

3-1. 렌즈를 낀 지 몇 시간만에 안구건조를 경험하하셨습니까? (응답 18개)



Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	2.000	3.000	3.167	4.500	7.000

- 렌즈를 끼고 난 뒤에는  
대체로 세시간 안팎의 시간에서  
안구건조가 일어난 것을 알 수 있음



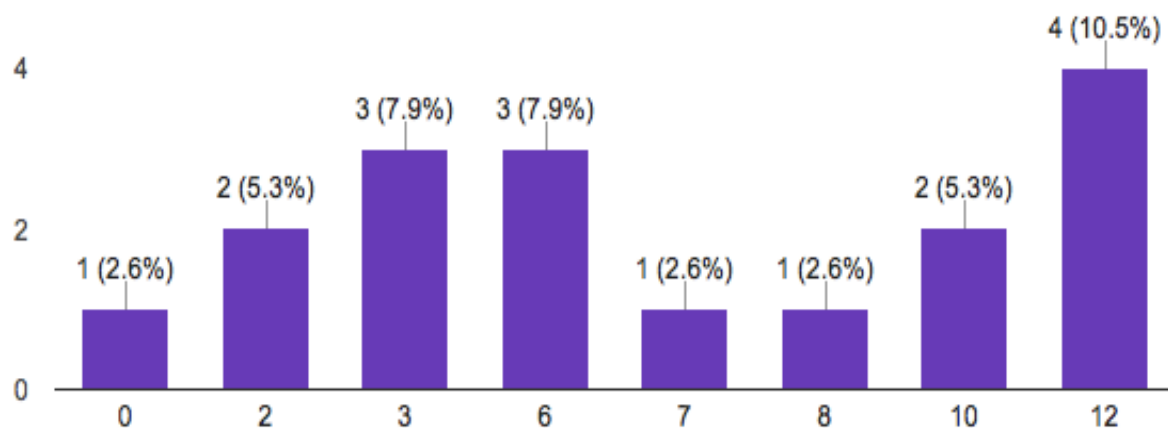
자료 가공 및 분석



## 문항별 기초자료 요약

3-2. 렌즈를 끼지 않았다면, 집에서 나온 지 몇 시간만에 안구건조를 경험하하셨습니까?

(응답 17개)



Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	3.000	6.000	6.706	10.000	12.000

- 렌즈를 끼지 않은 날에는 렌즈를 끼지 않은 날과 비교하여 비교적 늦은 시간에 안구건조가 일어난 것을 확인할 수 있음
- 이는 중위수와 평균에서도 확연히 드러나는 차이임

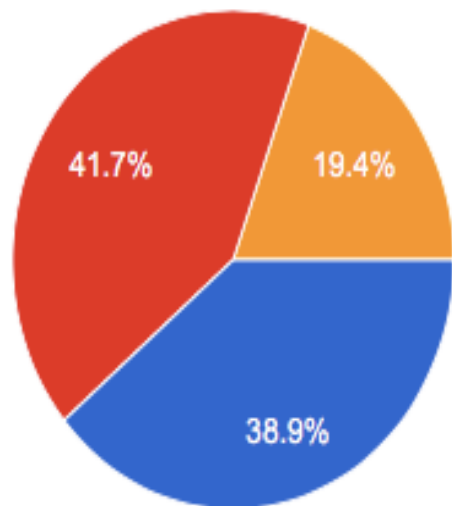


자료 가공 및 분석



## 문항별 기초자료 요약

4. 안구건조 증상을 극복하는 데 걸린 시간은 어느 정도입니까? (응답 36개)



- 눈 깜빡 몇 번만에 나옴
- 1분-3분
- 3분 이상

- 증상 극복 시간은 큰 차이를 보이지 않고 고루 분포됨
- 단독으로 유의미한 결과로 보기보다는 다른 문항들과 비교분석하였을 때 유의미한 데이터를 추출할 수 있을 듯



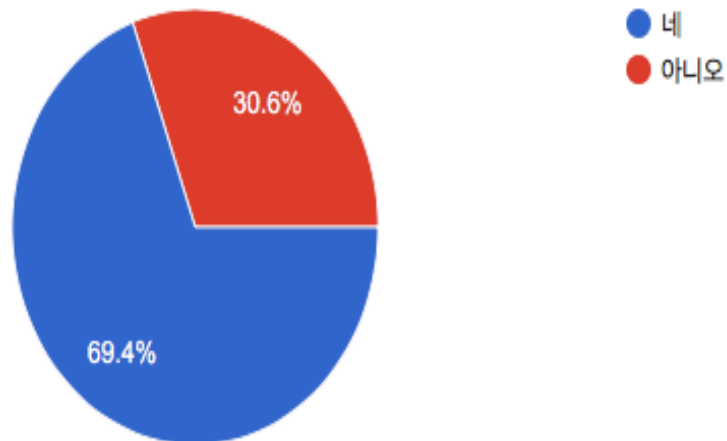
자료 가공 및 분석



## 문항별 기초자료 요약

5. 안구건조 증상 극복 이후 렌즈가 뿌애지거나 계속해서 눈이 시리는 등 일상생활에 불편이 있었습니까?

(응답 36개)



- 증상 극복 이후 계속해서 눈이 불편한 후유증이 남아 있는 정도를 조사하였는데, 전체 설문 의 흐름을 보았을 때 데이터 분석에는 크게 필요 없는 문항으로 보임

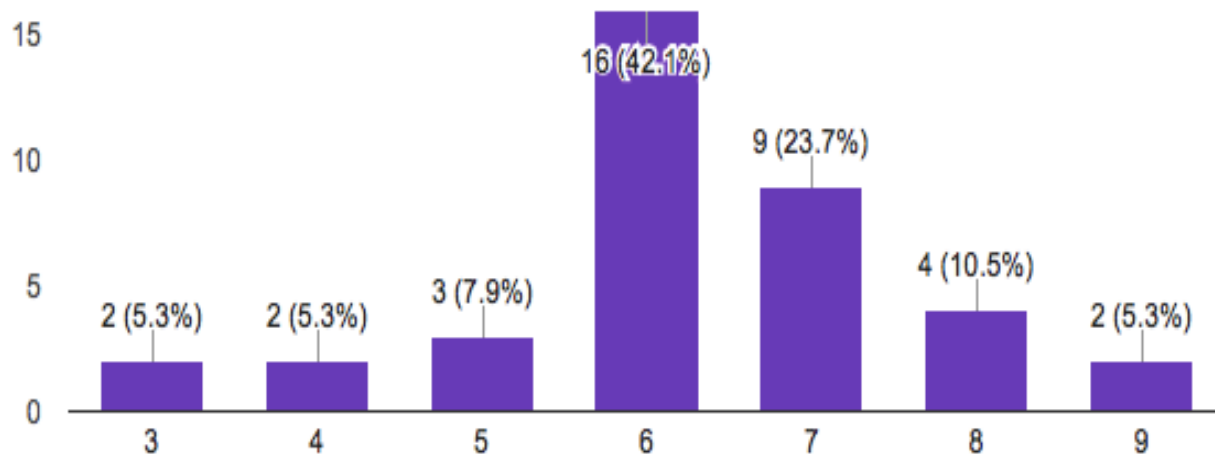


자료 가공 및 분석



## 문항별 기초자료 요약

6. 전날 잠은 몇 시간 잤습니까? (응답 38개)



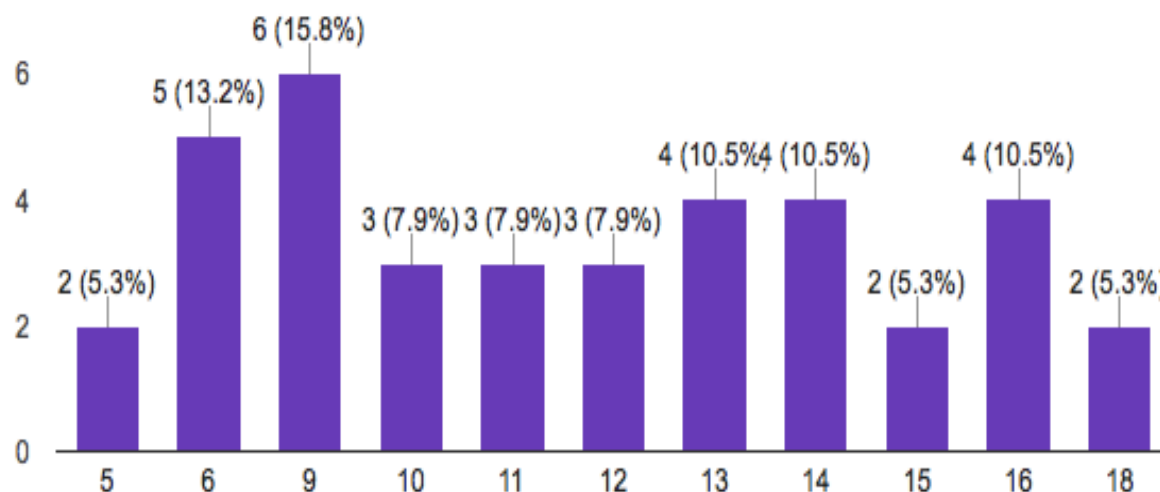
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
3.000	6.000	6.000	6.263	7.000	9.000

- 수면 시간은 대체로 6시간  
전후에서 이루어졌으며, 이 데이터  
또한 단독으로 분석하기보다는  
다른 데이터와 함께 분석하였을 때  
유의미한 데이터가 추출 될 것

## 문항별 기초자료 요약

7. 오늘 전자기기는 도합 대략 몇 시간 사용하였습니까? (노트북 및 핸드폰 화면 등)

(응답 38개)



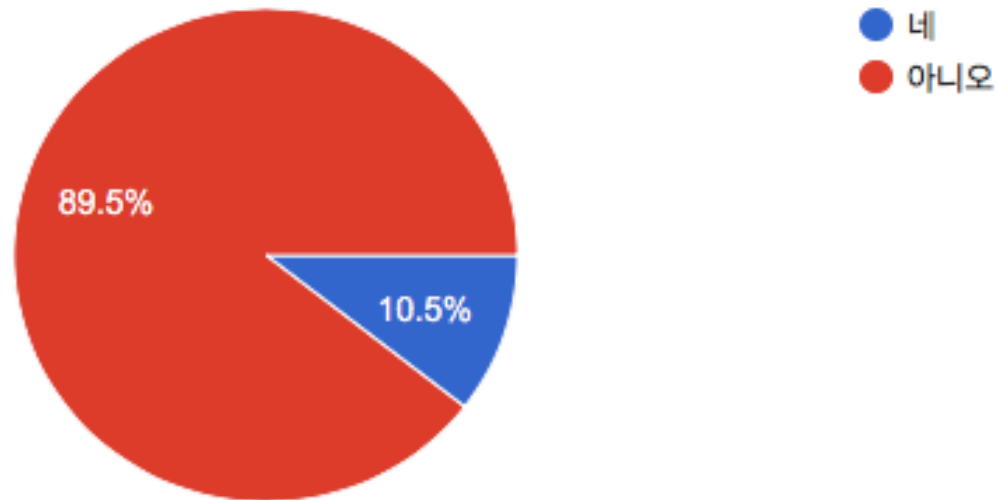
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
5.00	9.00	11.50	11.34	14.00	18.00

- 전자기기 사용시간은 대체로 11시간 전후에서 이루어졌으나 이 전후로 고르게 분포되어 중위수보다는 평균이 큰 의미를 가질 것으로 보임
- 이 데이터 또한 단독 분석보다는 다른 데이터와 함께 분석하였을 때 유의미한 데이터가 추출 될 것



## 문항별 기초자료 요약

8. 오늘 안약(인공눈물 포함)을 점안하하셨습니까? (응답 38개)



- 안약 점안 여부를 통해  
안구건조의 완화를 확인하고자  
했으나 안약 점안이 잘 이루어지지  
않은 관계로 큰 의미 없는  
데이터 수집이 되었음



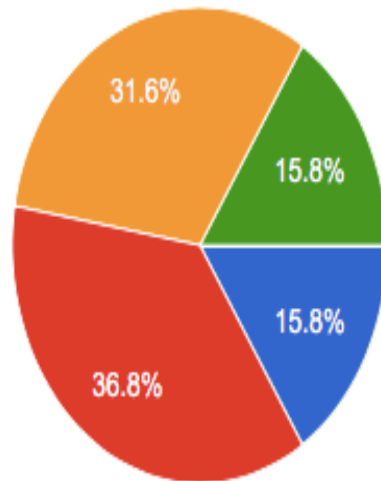
자료 가공 및 분석



## 문항별 기초자료 요약

### 9. 오늘의 미세먼지 농도는 어느 정도였습니까? (아시아의 대기오염 홈페이지 기준)

(응답 38개)



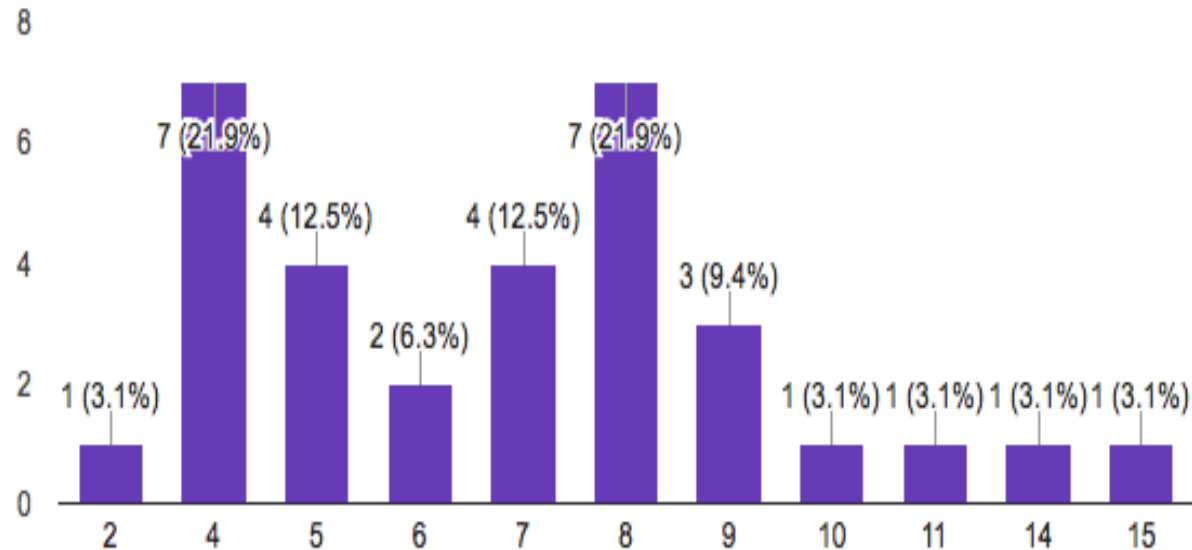
- 0-50 (좋음: 대기오염 관련 질환자군에서도 영향이 유발되지 않을 수준)
- 51-100 (보통: 환자군에게 만성 노출시 경미한 영향이 유발될 수 있는 수준)
- 101-150 (민감군영향: 환자군 및 민감군에게 유해한 영향이 유발될 수 있...)
- 151-200 (나쁨: 환자군 및 민감군(어린이, 노약자 등)에게 유해한 영향 유발, 일반인...)
- 201-300 (매우 나쁨: 환자군 및 민감군...)
- 300+ (위험: 환자군 및 민감군에게 응급...)

- 미세먼지는 안구건조 데이터와 함께 분석하였을 때 유의미한 데이터가 추출 될 것



## 문항별 기초자료 요약

10. (7번 데이터부터 추가) 컴퓨터 모니터는 몇 시간 보았습니까? (응답 32개)



Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2.000	4.750	7.000	6.969	8.000	15.000

- 모니터 응시 시간은 분포가 일정하지 않아 중위수나 평균의 의미를 찾기가 어려움
- 이 데이터 또한 단독 분석보다는 다른 데이터와 함께 분석하였을 때 유의미한 데이터가 추출 될 것



## 서로 다른 문항들을 이용한 분석: 개요

```
> str(data)
'data.frame': 38 obs. of 7 variables:
 $ Date          : Factor w/ 38 levels "2016.10.1","2016.10.10",...: 1 11 20 21
22 23 24 25 26 2 ...
 $ Happened      : Factor w/ 2 levels "N","Y": 2 2 2 2 2 2 2 1 2 2 ...
 $ Lens          : Factor w/ 2 levels "N","Y": 2 1 2 1 1 1 2 1 1 1 ...
 $ Sleep         : int  6 6 8 6 7 6 7 6 8 7 ...
 $ ElectricDevices: int  12 9 5 9 10 14 14 16 16 16 ...
 $ Fine.Dust      : Factor w/ 4 levels "0-50","101-150",...: 4 4 1 4 4 4 4 4 1
1 ...
 $ CompMonitor   : int  NA NA NA NA NA NA NA 4 8 14 7 ...
```

- 기존 데이터수집의 목적에 걸맞게 안구건조 여부와 안구건조에 영향을 주는 요인들로 데이터를 추출하였고, 어느 요인들이 안구건조에 영향을 주는지 분석하여 알아보고자 함



## 서로 다른 문항들을 이용한 분석: Decision Tree

```
> library(tree)
> set.seed(1234)
> ind <- sample(2, nrow(data), replace=TRUE, prob=c(0.7,0.3))
> trainData <- data[ind==1,]
> testData <- data[ind==2,]
> data_tree <- tree(Happened ~ Sleep+ElectricDevices+Fine.Dust+CompMonitor,
data=data)
> print(data_tree)
```

```
node), split, n, deviance, yval, (yprob)
* denotes terminal node
```

```
1) root 32 14.960 Y ( 0.0625 0.9375 )
```

```
2) Fine.Dust: 0-50,51-100 14 11.480 Y ( 0.1429 0.8571 )
```

```
4) Sleep < 6.5 7 8.376 Y ( 0.2857 0.7143 ) *
```

```
5) Sleep > 6.5 7 0.000 Y ( 0.0000 1.0000 ) *
```

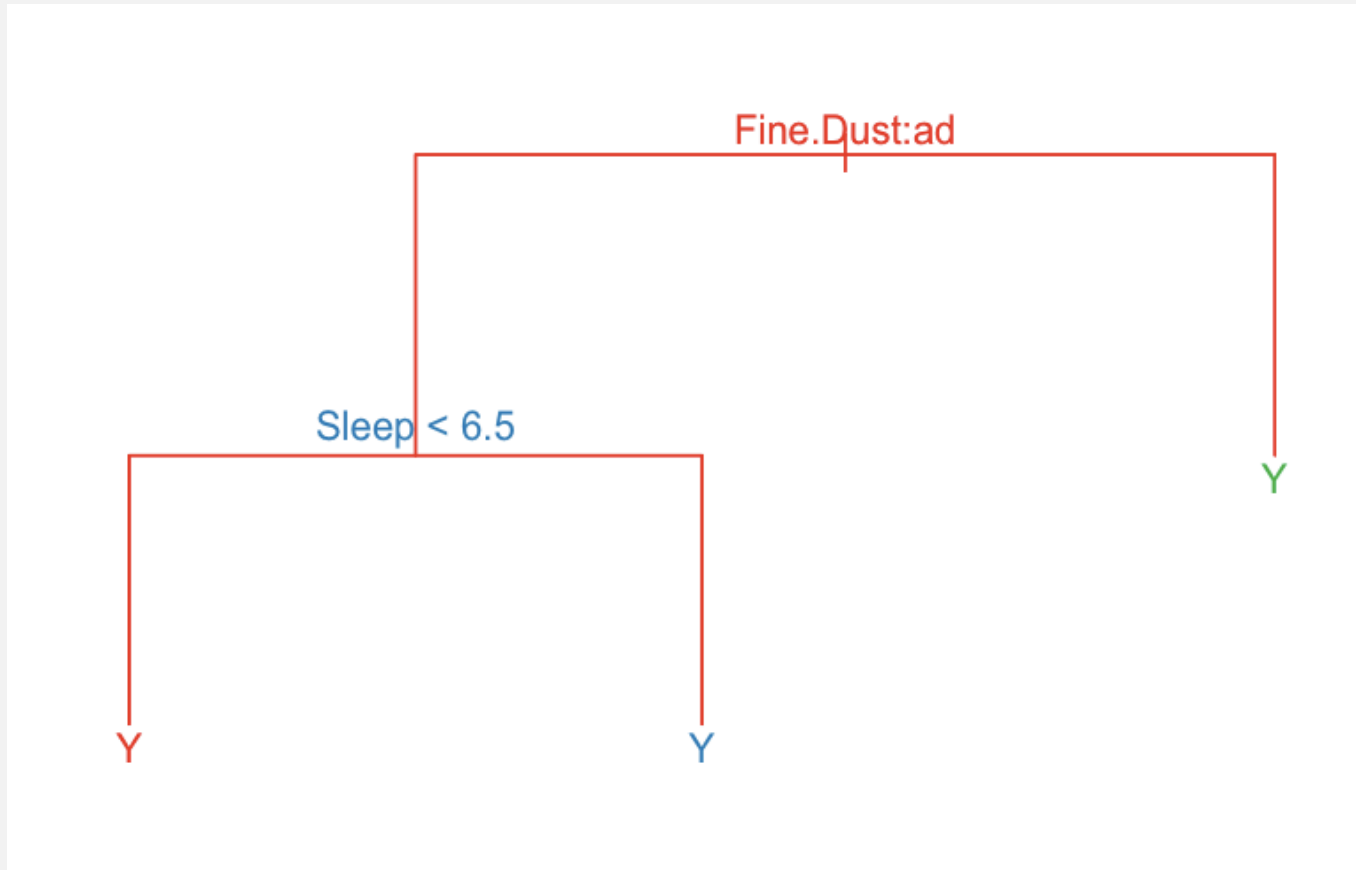
```
3) Fine.Dust: 101-150,151-200 18 0.000 Y ( 0.0000 1.0000 ) *
```

```
> plot(data_tree, col=pal)
```

```
> text(data_tree, col=pal)
```

- 미세먼지 농도가 101 이상일 때 무조건 안구건조 발생
- 미세먼지 농도가 100 이하일 때는 수면시간에 따라 안구건조가 발생하는데, 잠을 6.5시간 이상 충분히 잤을 때 안구건조가 100퍼센트 발생한다고 분석된다. 이는 상식에 맞지 않는 결과이기에 안구건조는 눈을 뜨고 잤는지의 여부로 판단하는 것이 더 좋을 것으로 보임.

## 서로 다른 문항들을 이용한 분석: Decision Tree



- 테스트 데이터가 적기 때문인지 Decision Tree 하면 생각나는 회황찬란하지 않은 나뭇가지 모형이 그려지지 않아 아쉬움
- 트리의 시각화보다 앞선 데이터를 확인하는 것이 이 데이터 분석에서는 더 나은 선택으로 보임



## 분석을 통해 밝혀진 사실 요약

### 추출

안구건조에 어떤 요인들이  
영향을 주는지 알아보기 위해

- 수면시간
- 전자기기 사용 시간
- 미세먼지 농도
- 모니터 응시 시간

네 가지의 데이터 추출

### 분석

Decision Tree를  
이용하여 어떤 요인이  
안구건조에 영향을  
미치는지 순서대로  
알아보고자 함

### 결과

미세먼지 농도가 가장 큰 영향:  
농도가 101 이상일 때 100%  
다음은 수면시간이나,  
잠을 잘 자면 안구건조가 준다는  
통념과 다르게 수면시간이 길 수록  
안구건조 확률이 100%이므로  
안구건조에는 '눈을 뜨고 잔 시간'이  
지표로서 더 기능할 것으로 보임





감사합니다