

Report

나이와 행복은 관련이 있을까?



학과	수학과
학번	2018010705
이름	신영민

<목차>

I . 문제 정의

II . 데이터 준비

III. 구현 및 실험

IV. 결과 분석

V. 결론 도출

VI. 출처

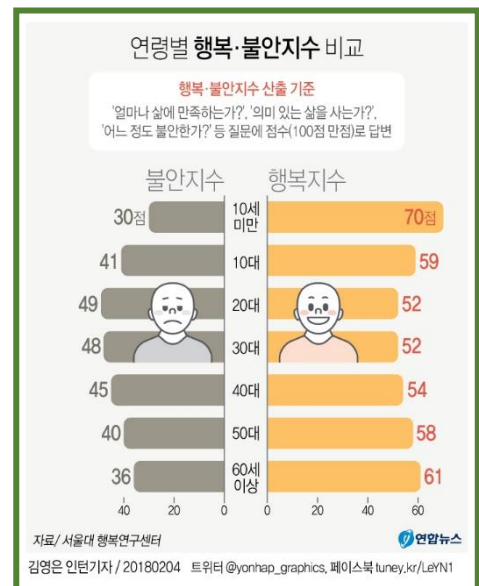
I. 문제정의

→ 행복과 나이의 관계를 선형회귀로 학습할 수 있을까?



- 위에 기사와 같이 노인 우울증을 앓는 경우가 증가하고 있다고 합니다. 반면에 어린이들을 생각하면 위에 사진과 같이 밝게 웃고 있는 모습이 연상됩니다. 단순히 생각하면 나이가 많아질수록 행복이 감소하는 건 아닐까 하고 생각해볼 수 있습니다. 그래서 행복지수를

검색해보았습니다. 예상외로 옆에 사진과 같이 행복지수가 나이가 증가할수록 증가하는 것을 알 수 있었습니다. 이를 통해 혹시 나이가 증가할 수록 행복이 증가하는 선형관계를 이루고 있지 않을까? 하는 의문을 가지게 되어 문제를 정의하고 이 주제를 정하게 되었습니다.



II. 데이터 준비

① 불안지수와 행복지수 데이터 수집.

-서울대 행복연구소에서 조사한 데이터를 기사로 쓴 연합뉴스
“연령별 행복 불안지수 비교” 기사에서 수집했습니다.

	행복지수	불안지수
10 대 미만	70 점	30 점
10 대	59 점	41 점
20 대	52 점	49 점
30 대	52 점	48 점
40 대	54 점	45 점
50 대	58 점	40 점
60 세 이상	61 점	36 점

② 코딩에 맞게 데이터 변형

- 종속 변수 : 10 대 미만, 10 대, ..., 60 세 이상으로 x 축을 설정할 수 없으므로 x 축의 값을 5, 15, 25, ...,55 로 바꿨고 60 대 이후는 65 로 정해서 진행했습니다.

- 독립변수 : y 는 점수 값만 사용했습니다.

	행복지수	불안지수
5	70	30
15	59	41
25	52	49
35	52	48
45	54	45
55	58	40
65	61	36

Ⅲ. 구현 및 실험

① 패키지 불러오기

```
7
8 import matplotlib.pyplot as plt
9 import numpy as np
10 from sklearn import linear_model as lm
11
```

② 변수설정

-나이 변수를 age 로 설정하고 정의했습니다.

- age=[[5],[15],[25],[35],[45],[55],[65]]

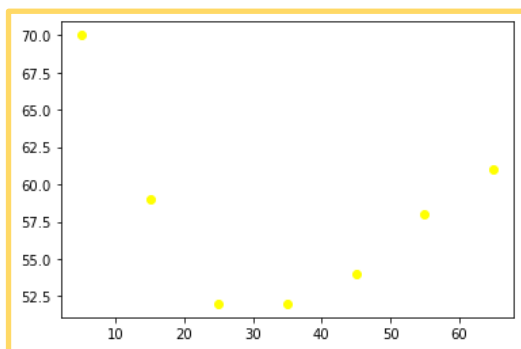
-행복지수와 불안지수를 각각 h 와 uh 로 설정하고 정의했습니다.

- h=[70,59,52,52,54,58,61],

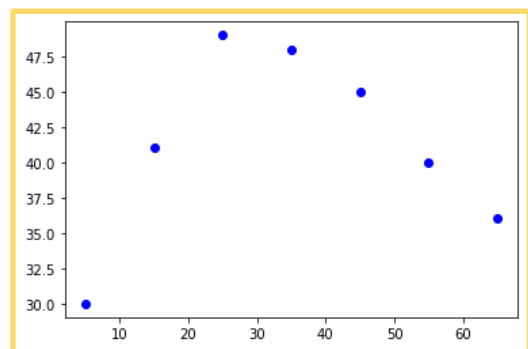
uh=[30,41,49,48,45,40,36]

```
15 age=[[5],[15],[25],[35],[45],[55],[65]]
16 age=np.array(age)
17 h=[70,59,52,52,54,58,61]
18 h=np.array(h)
19 uh=[30,41,49,48,45,40,36]
20 uh=np.array(uh)
```

③ 데이터 확인하기

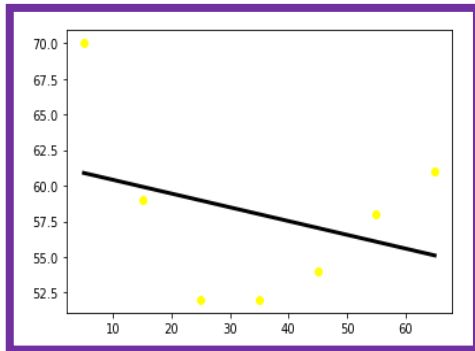


[행복지수]

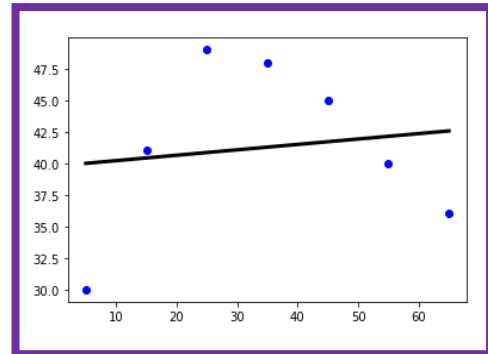


[불안지수]

④ 선형 회귀



[행복지수]



[불안지수]

```

22 h_reg.fit(age,h)
23 uh_reg.fit(age,u)
24
25 plt.scatter(age,h,color='yellow')
26
27 h_pred=h_reg.predict(age)
28
29 plt.plot(age,h_pred,color='black',linewidth=3)
30 plt.show()
31
32 plt.scatter(age,u,color='blue')
33 uh_pred=uh_reg.predict(age)
34
35 plt.plot(age,u_pred,color='black',linewidth=3)
36 plt.show()

```

⑤ 에러 확인해 보기

```

38 error_h=abs(h_pred-h)
39 error_uh=abs(uh_pred-u)

```

```

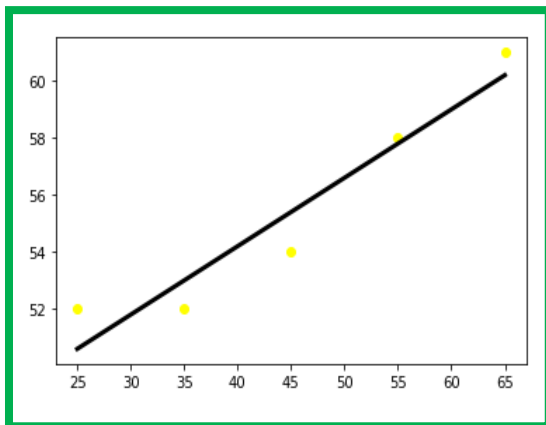
In [19]: error_h
Out[19]:
array([9.10714286, 0.92857143, 6.96428571, 6.
3.03571429,
1.92857143, 5.89285714])

In [20]: error_uh
Out[20]:
array([10.
0.57142857, 8.14285714,
6.71428571, 3.28571429,
2.14285714, 6.57142857])

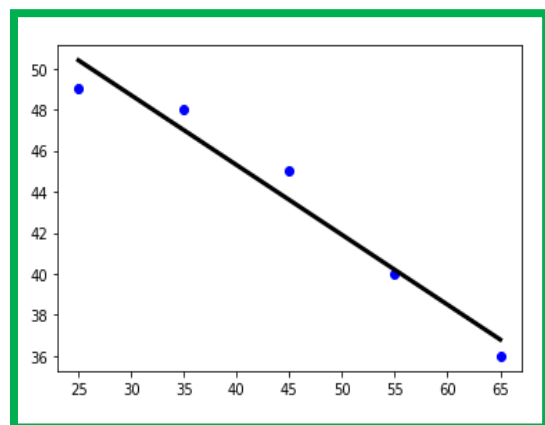
```

IV. 결과 분석

- 위의 그래프를 보면 선형회귀가 잘 이루어지지 않았음을 알았습니다. 그 이유는 값들이 나이에 따라서 증가만 하거나 감소만 하지 않고, 증가하다가 감소하고 감소하다가 증가하기 때문입니다. 행복지수와 불안 지수 모두 20 대를 기점으로 오직 감소 또는 증가만 하기 때문에 20 대부터 60 대 이후까지 선형회귀 시켜봤습니다. 아래 그래프를 통해 선형 회귀가 잘됐다는 것을 알 수 있었습니다.



[행복지수]



[불안지수]

에러도 구해보면

```
In [25]: error_h
Out[25]: array([1.4, 1. , 1.4, 0.2, 0.8])

In [26]: error_uh
Out[26]: array([1.4, 1. , 1.4, 0.2, 0.8])
```

범위를 나눠서 했을 때 에러가 더 작다는 것을 알 수 있습니다.

V. 결론 도출

앞에서 본 결과로부터 20 대 미만까지는 행복과 나이가 선형 관계가 아니지만 20 대 이후부터는 선형관계를 가진다는 결론을 얻을 수 있었습니다. 그리고 20 대 미만 지 포함을 시키고 싶다면 선형이 아닌 다른 비선형 함수로 피팅을 해야 잘 될 것 같습니다. 그리고 처음 생각과는 다르게 나이가 많을수록 행복지수가 높아지는데 이유가 무엇일지도 궁금해서 고민해본 결과 불안지수는 나이가 많아질수록 작아지는 것을 통해 나이가 증가함에 따라 삶의 안정감을 얻어 이러한 결과가 나오는 것을 알게 되었습니다. 20 대 미만에는 학교를 다니고 미래를 고민하는 시기인 만큼 안정감이 없어서 행복지수가 낮고 불안지수가 높지 않을까 하는 결론을 얻었습니다. 이를 통해 행복에는 안정감이 중요하고, 행복과 나이는 관계가 있다는 것을 알게 되었습니다. 그리고 더 나아가 나이 대 별이 아닌 나이별 데이터로 하면 더 정확한 결론을 얻을 수 있을 것 같아 다음에는 더 많은 데이터를 가지고 해보고 싶습니다.

VI. 출처

[1] [90 대 우울증 환자, 10 년새 4 배↑... 젊은층과 증상 달라]

http://health.chosun.com/site/data/html_dir/2020/10/05/2020100501966.html (2020.10.05)

[2] [연령대별 행복·불안지수 비교]

<https://www.yna.co.kr/view/GYH20180202001100044> (2018.02.04)