

# 건강한 라이프 스타일과 건강

융합소프트웨어학부

60192344 박찬규

60192346 서장호

60221353 윤지원

# 목차

---

- 문제 정의

- 주제 변경 이유
- 주제 선정 및 목표

- 데이터 수집 및 설명

- 활용 데이터 소개
- 데이터 전처리

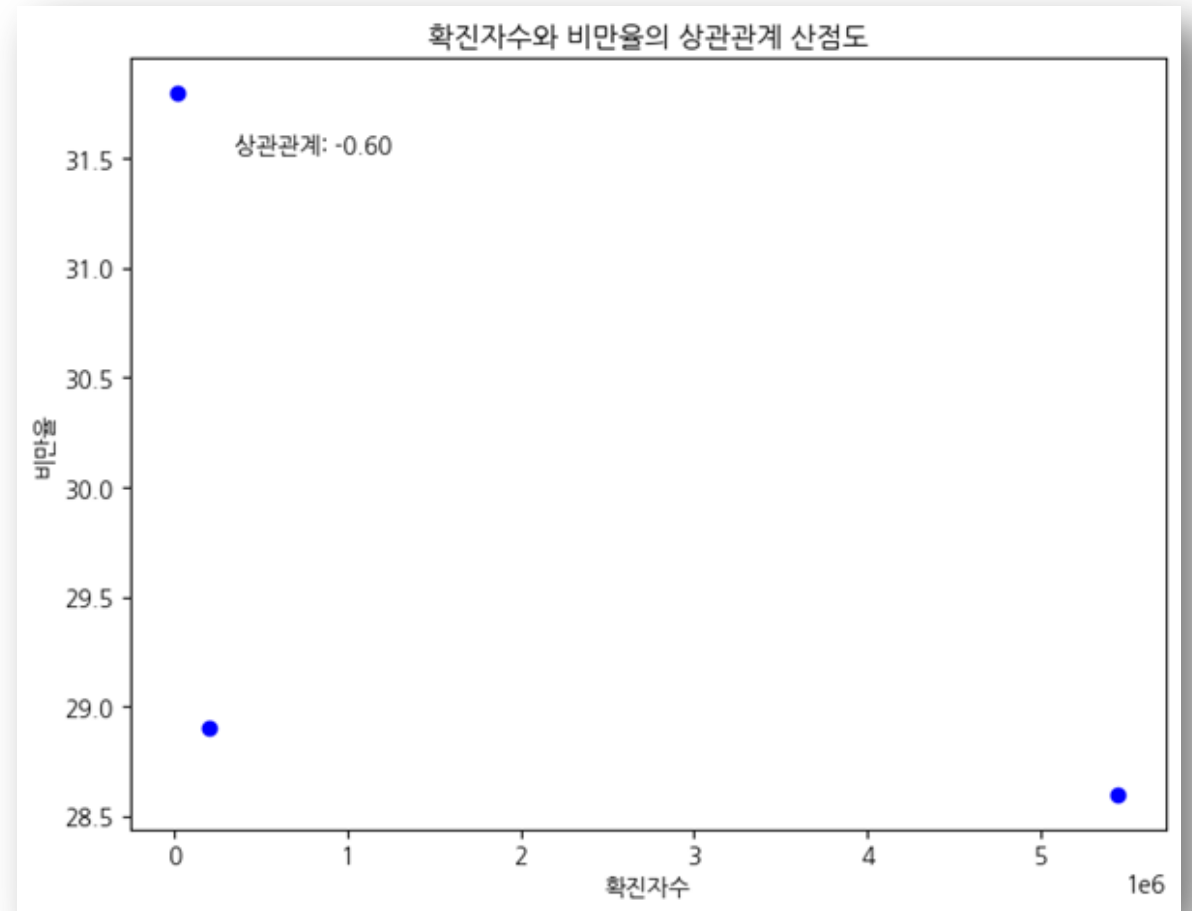
- 데이터 분석

1. Happiness levels에  
가장 큰 영향을 미치는 요인 분석
2. 비만율과 라이프 스타일 요소들의 관계
3. 비만율과 섭취하는 음식 종류의 관계
4. 영양결핍률과 섭취하는 음식 종류의 관계
5. Lifestyle과 영양결핍률의 관계
6. COVID-19 확진과 비만율의 관계
7. 영양결핍률과  
COVID-19 치명률, 사망률의 관계

- 데이터 분석 결과

# 문제 정의 – 주제 변경 이유

1. 데이터셋 표본의 크기가 너무 작아  
유의미한 검정 결과 도출 불가
2. 숫자형 데이터를 가진 Column의  
수 부족



# 문제 정의 – 주제 선정 및 목표

## ‘행복수준에 영향을 주는 요인’

- 문제 정의

: 어떤 요인들이 행복수준에 영향을 주는지 알아보고자 한다. 행복수준을 포함한 다른 수치들에 근거해 도시별 건강한 라이프스타일을 수치로 나열한 healthy\_lifestyle\_city\_2021의 데이터를 바탕으로 행복의 근거들을 예측하고자 한다. 또한 행복에서 우리가 가장 중요하다고 예측한 건강에서 식단, 질병, 영양상태 등의 요소들의 중요성과 다른 생활모습과의 연관성을 찾고자 한다.

# 데이터 수집 방법

: Kaggle을 통해서 두 가지 데이터셋을 수집

1. Healthy Lifestyle Cities Report 2021 - healthy\_lifestyle\_city\_2021.csv

<https://www.kaggle.com/datasets/prasertk/healthy-lifestyle-cities-report-2021>

2. COVID-19 Healthy Diet Dataset - Food\_Supply\_kcal\_Data.csv

<https://www.kaggle.com/datasets/mariaren/covid19-healthy-diet-dataset>

The Kaggle logo, consisting of the word "kaggle" in a lowercase, rounded, blue sans-serif font.

# 데이터 소개 - life

: 상위 44개 도시의 건강한 라이프스타일 지표를 나타내는 데이터셋

총 45개의 row, 12개의 column으로 구성

City	Rank	Sunshine hours(City)	Cost of a bottle of water(City)	Obesity levels(Country)	Life expectancy(years) (Country)	Pollution(Index score) (City)	Annual avg. hours worked	Happiness levels(Country)	Outdoor activities(City)	Number of take out places(City)	Cost of a monthly gym membership(City)
dam	1	1858	£1.92	20.40%	81.2	30.93	1434	7.44	422	1048	£34.90
lney	2	2636	£1.48	29.00%	82.1	26.86	1712	7.22	406	1103	£41.66
anna	3	1884	£1.94	20.10%	81.0	17.33	1501	7.29	132	1008	£25.74
olm	4	1821	£1.72	20.60%	81.8	19.63	1452	7.35	129	598	£37.31
ngen	5	1630	£2.19	19.70%	79.8	21.24	1380	7.64	154	523	£32.53
sinki	6	1662	£1.60	22.20%	80.4	13.08	1540	7.80	113	309	£35.23
joka	7	2769	£0.78	4.30%	83.2	-	1644	5.87	35	539	£55.87
erlin	8	1626	£1.55	22.30%	80.6	39.41	1386	7.07	254	1729	£26.11
lona	9	2591	£1.19	23.80%	82.2	65.19	1686	6.40	585	2344	£37.80
uver	10	1938	£1.08	29.40%	81.7	24.26	1670	7.23	218	788	£31.04
urne	11	2363	£1.57	29.00%	82.1	25.90	1712	7.22	243	813	£36.89
ijing	12	2671	£0.26	6.20%	75.4	85.43	-	5.12	223	261	£38.62
gkok	13	2624	£0.22	10.00%	74.1	76.64	-	5.99	377	1796	£50.03
lires	14	2525	£0.57	28.30%	75.9	52.64	-	5.97	246	1435	£22.45
onto	15	2066	£1.09	29.40%	81.7	37.83	1670	7.23	174	1656	£32.64
idrid	16	2769	£1.30	23.80%	82.2	52.68	1686	6.40	216	2491	£34.54
arta	17	2983	£0.21	6.90%	68.5	84.39	-	5.28	114	833	£29.94

- 1.City (도시): 각 행의 도시 이름.
- 2.Rank (순위): 도시의 순위(예: 가장 낮은 순위일수록 높은 등급)
- 3.Sunshine hours (도시의 일조 시간): 해당 도시의 연간 평균 일조 시간
- 4.Cost of a bottle of water (도시에서의 물 병 가격): 해당 도시에서 물 병의 가격
- 5.Obesity levels (Country) (비만 수준 - 국가): 해당 도시가 속한 국가의 비만 수준
- 6.Life expectancy (years) (Country) (평균 수명 - 국가): 해당 도시가 속한 국가의 평균 수명
- 7.Pollution (Index score) (City) (대기 오염 지수 - 도시): 해당 도시의 대기 오염 지수 점수
- 8.Annual avg. hours worked (연평균 근로 시간): 근로자들의 연간 평균 근로 시간
- 9.Happiness levels (Country) (행복 수준 - 국가): 해당 도시가 속한 국가의 행복 수준
- 10.Outdoor activities (도시에서의 야외 활동): 해당 도시에서 즐길 수 있는 야외 활동의 다양성 수준.
- 11.Number of take out places (도시의 테이크아웃 장소 수): 해당 도시 내의 테이크아웃 장소의 수
- 12.Cost of a monthly gym membership (도시에서의 월간 체육관 회원비): 해당 도시에서의 월간 체육관 회원비용.

Healthy Lifestyle Cities Report 2021 - healthy\_lifestyle\_city\_2021.csv

# 데이터 전처리 - life

- 열 이름 변경  
: 'City' 열을 'Country'로, 그리고 다른 열 이름들도 주어진 dictionary에 따라 변경
- 결측치 정제  
: 'Sunshine hours', 'Pollution(Index score)', 'Annual avg. hours worked' 열의 결측치를 각각 해당 열의 평균값으로 대체
- 단위 기호 제거  
: 'Obesity levels', 'Cost of a bottle of water', 'Cost of a monthly gym membership' 열의 값에서 단위 기호를 제거하고 숫자형으로 변환
- 도시를 국가로 변환  
: 'Country' 열의 값 중 일부 도시를 해당 국가로 mapping

# 데이터 소개 - food

1. Country (국가): 각 행의 국가 이름
2. Alcoholic Beverages (주류 소비): 주류 소비량
3. Animal Products (동물성 제품 소비): 동물성 식품의 총 소비량
4. Animal Fats (동물성 지방 소비): 동물성 지방의 소비량
5. Aquatic Products, Other (수산물 및 기타 수산물 소비): 수산물 및 기타 수산물의 소비량
6. Cereals - Excluding Beer (곡물 - 맥주 제외): 맥주를 제외한 곡물의 소비량
7. Eggs (계란 소비): 계란의 소비량
8. Fish, Seafood (어류, 해산물 소비): 어류 및 해산물의 소비량
9. Fruits - Excluding Wine (과일 - 와인 제외): 와인을 제외한 과일의 소비량
10. Meat (육류 소비): 총 육류 소비량
11. Vegetable Oils (식물성 오일 소비): 식물성 오일의 소비량
12. Vegetables (채소 소비): 채소의 소비량
13. Obesity (비만율): 국가별 비만율
14. Undernourished (국가 내 영양실조율): 국가 내 영양실조율
15. Confirmed (확인된 COVID-19 확진자 수): 국가별 확인된 COVID-19 확진자 수
16. Deaths (COVID-19 사망자 수): 국가별 COVID-19 사망자 수
17. Recovered (COVID-19 회복자 수): 국가별 COVID-19 회복자 수
18. Active (활성 COVID-19 환자 수): 현재 활성 상태인 COVID-19 환자 수
19. Population (인구): 국가별 인구 수
20. Unit (all except Population) .

	Country	Alcoholic Beverages	Animal Products	Animal fats	Aquatic Products, Other	Cereals - Excluding Beer	Eggs	Fish, Seafood	Fruits - Excluding Wine	Meat	...	Vegetable Oils	Vegetables	Obesity	Undernourished
0	Afghanistan	0.0000	4.7774	0.8504	0.0	37.1186	0.1501	0.0000	1.4757	1.2006	...	2.3012	0.7504	4.5	29.8
1	Albania	0.9120	16.0930	1.0591	0.0	16.2107	0.8091	0.1471	3.8982	3.8688	...	2.8244	2.7508	22.3	6.2
2	Algeria	0.0896	6.0326	0.1941	0.0	25.0112	0.4181	0.1195	3.1805	1.2543	...	5.7638	2.0457	26.6	3.9
3	Angola	1.9388	4.6927	0.2644	0.0	18.3521	0.0441	0.8372	2.3133	2.9302	...	4.2741	0.3525	6.8	25
4	Antigua and Barbuda	2.3041	15.3672	1.5429	0.0	13.7215	0.2057	1.7280	3.6824	7.0356	...	4.6904	1.2960	19.1	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
165	Venezuela (Bolivarian Republic of)	0.8454	7.2303	0.6007	0.0	21.3126	0.2892	0.4449	2.3804	3.1368	...	7.5417	0.6674	25.2	21.2
166	Vietnam	0.7150	10.9806	0.9363	0.0	26.9833	0.2894	1.0385	1.8046	7.8311	...	1.3279	1.9578	2.1	9.3
167	Yemen	0.0000	3.4667	0.3394	0.0	32.0727	0.1455	0.1697	1.1879	2.0121	...	3.9515	0.3636	14.1	38.9
168	Zambia	1.1925	3.3043	0.3230	0.0	31.5528	0.1988	0.5714	0.2236	1.5155	...	3.0062	0.4472	6.5	46.7
169	Zimbabwe	1.4269	3.9356	0.6904	0.0	29.8044	0.1381	0.1611	0.4373	1.7491	...	6.2601	0.2532	12.3	51.3

170 rows × 32 columns

## COVID-19 Healthy Diet Dataset - Food\_Supply\_kcal\_Data.csv

: 각 나라 사람들이 어떤 종류의 음식들로부터 어느 정도의 energy(kcal)을 섭취하는지에 대한 퍼센트를 나타낸 자료

총 12개의 row, 45개의 column으로 구성



# 데이터 전처리

## - food

# 1. 'Undernourished' 열의 값 중 '<2.5'를 '1.0'으로 변환하고, 모든 값을 float64 자료형으로 변환합니다.

```
food['Undernourished'] = food['Undernourished'].map(lambda x: '1.0' if x == '<2.5' else x).astype('float64')
```

# 2. 'Unit (all except Population)' 열을 데이터프레임에서 삭제합니다.

```
food = food.drop('Unit (all except Population)', axis=1)
```

# 3. 결측치가 있는 행을 삭제합니다.

```
food = food.dropna()
```

# 데이터 전처리

## - merge; food and life

```
import pandas as pd

# 'Country'를 기준으로 두 데이터셋 merge
food_and_life = pd.merge(food, life, on='Country', how='inner')

food_and_life
```

# 01

Happiness levels에 가장 큰 영향을 미치는 요인 분석

## 1. Happiness levels에 가장 큰 영향을 미치는 요인 분석

# Happiness levels과 라이프 스타일 요소 간 관계

```
# 각 열에 대해 분석 수행
for col in cols:
    # 데이터 추출
    a = life[dependent_variable]
    b = life[col]

    # 정규성 검정 - Shapiro-Wilk test
    stat_a, p_a = shapiro(a)
    stat_b, p_b = shapiro(b)

    print(f'정규성 검정 결과 ({dependent_variable}): Statistic={stat_a:.4f}, p-value={p_a:.20f}')
    print(f'정규성 검정 결과 ({col}): Statistic={stat_b:.4f}, p-value={p_b:.20f}')

    # 등분산성 검정 - Levene
    stat_levene, p_value_levene = levene(a, b)
    print(f'Levene 등분산성 검정 결과 ({col}): Statistic={stat_levene:.4f}, p-value={p_value_levene:.20f}')

    # 등분산성 검정 - Fligner
    stat_fligner, p_value_fligner = fligner(a, b)
    print(f'Fligner 등분산성 검정 결과 ({col}): Statistic={stat_fligner:.4f}, p-value={p_value_fligner:.20f}')

    # 스피어만 검정 또는 피어슨 검정 선택
    if p_a < 0.05 or p_b < 0.05 or stat_levene < 0.05 or stat_fligner < 0.05:
        # 정규성 또는 등분산성 가정이 깨질 경우 스피어만 검정 수행
        correlation, p_value = spearmanr(a, b)
        print(f'Spearman 검정 결과 ({col}): 상관계수={correlation:.4f}, p-value={p_value:.20f}')
    else:
        # 가정이 만족할 경우 피어슨 검정 수행
        correlation, p_value = pearsonr(a, b)
        print(f'Pearson 검정 결과 ({col}): 상관계수={correlation:.4f}, p-value={p_value:.20f}')

    # 절대값이 0.3보다 클 경우 스캐터플롯 출력
    if abs(correlation) > 0.5:
        sns.regplot(x=a, y=b, scatter_kws={'alpha': 0.5, 'color': 'red' if correlation < 0 else 'blue'})
        plt.xlabel(dependent_variable)
        plt.ylabel(col)
        plt.show()

print("\n")
```

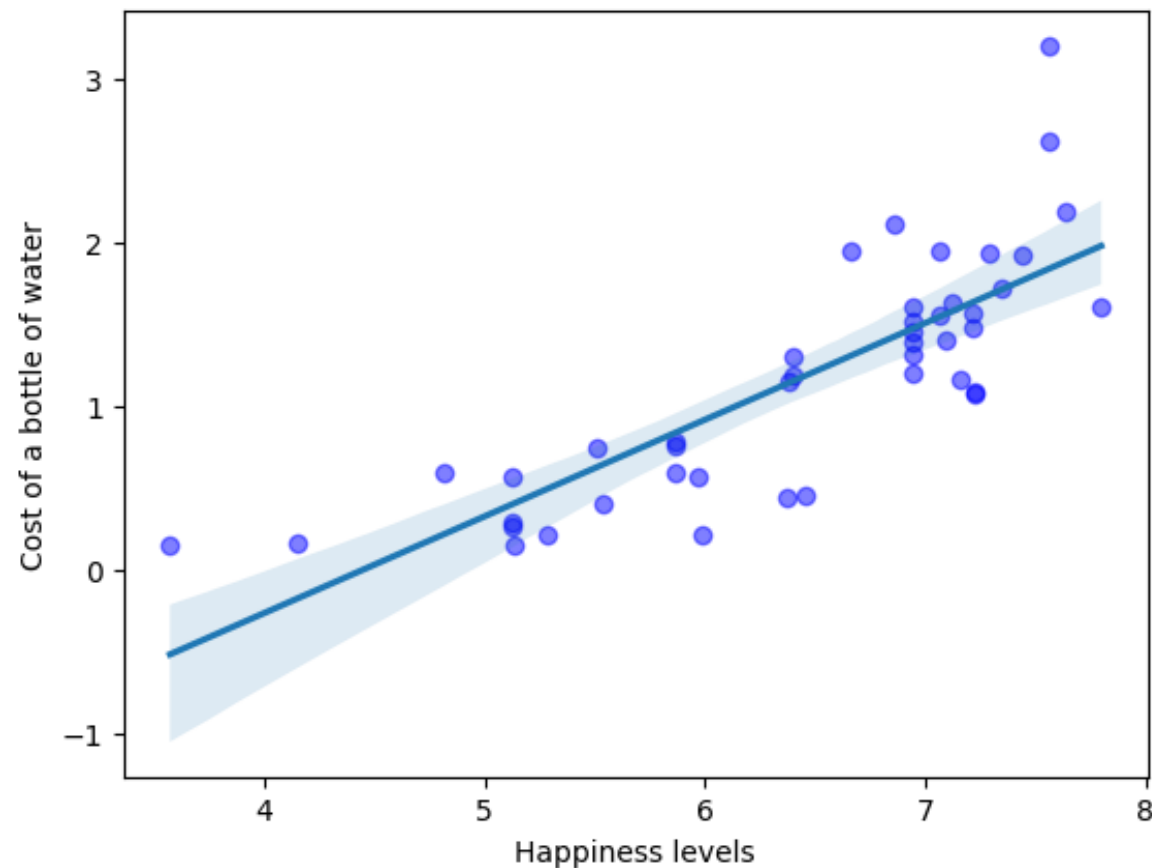
```
# 비교할 열들과 종속 변수 선택
```

```
cols = ['Sunshine hours', 'Cost of a bottle of water',
        'Obesity levels', 'Life expectancy(years)',
        'Pollution(Index score)', 'Annual avg. hours worked',
        'Outdoor activities', 'Number of take out places',
        'Cost of a monthly gym membership']
```

```
dependent_variable = 'Happiness levels'
```

## 1. Happiness levels에 가장 큰 영향을 미치는 요인 분석

# Happiness levels VS Cost of bottle of water



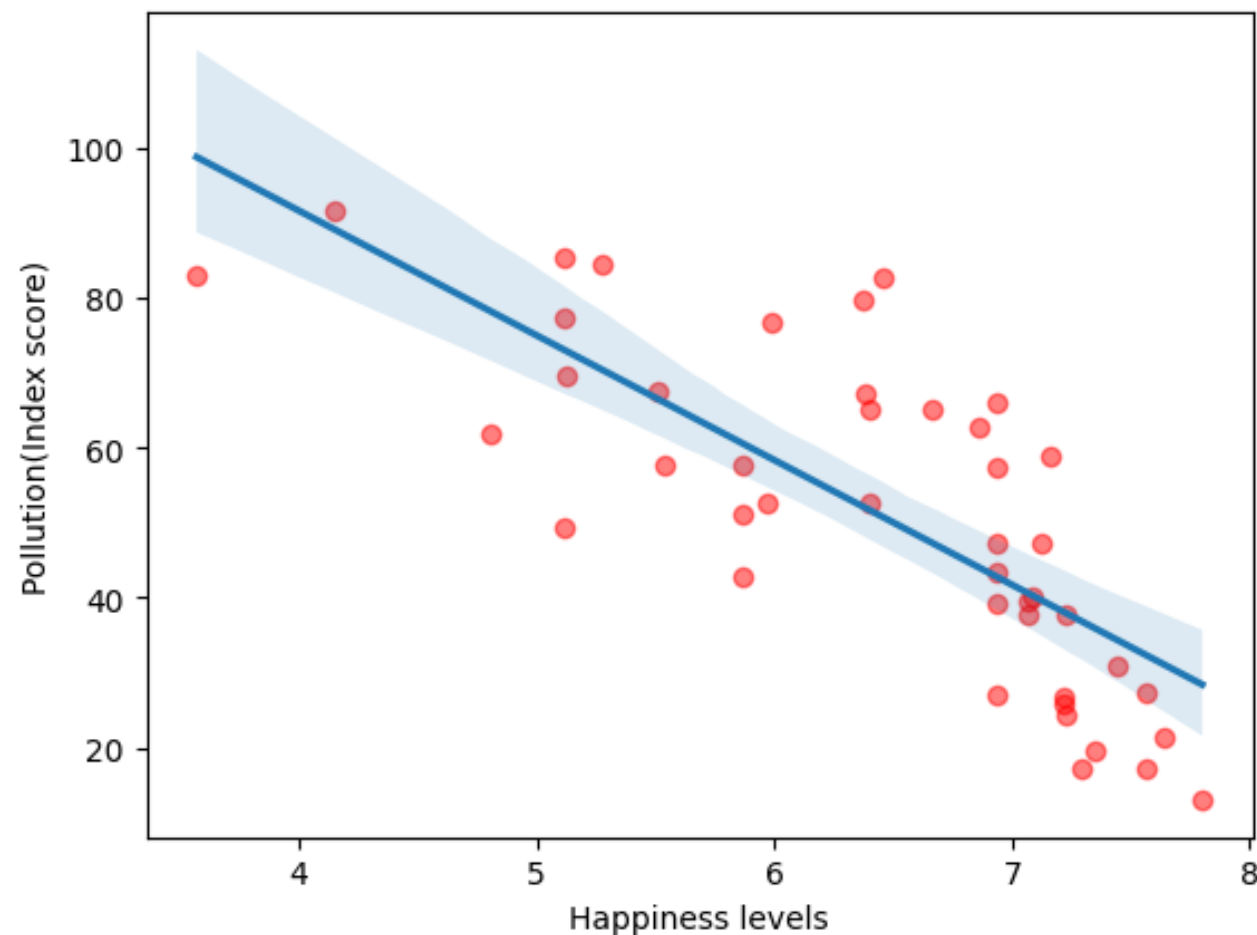
⇒ 물의 가격과 행복 수준 간의 **강한 양의 상관관계**

물의 가격이 상승할수록 행복 수준도 상승하는 경향이 있음

- 정규성 검정 결과 (Happiness levels): Statistic=0.9092, p-value=0.00211808481253683567
- 정규성 검정 결과 (Cost of a bottle of water): Statistic=0.9534, p-value=0.07345357537269592285
- Levene 등분산성 검정 결과 (Cost of a bottle of water): Statistic=2.2526, p-value=0.13704853205147365935
- Fligner 등분산성 검정 결과 (Cost of a bottle of water): Statistic=1.0598, p-value=0.30326166871322102203
- Spearman 검정 결과 (Cost of a bottle of water): 상관계수=0.8242, p-value=0.00000000000621643408

## 1. Happiness levels에 가장 큰 영향을 미치는 요인 분석

# Happiness level VS Pollution



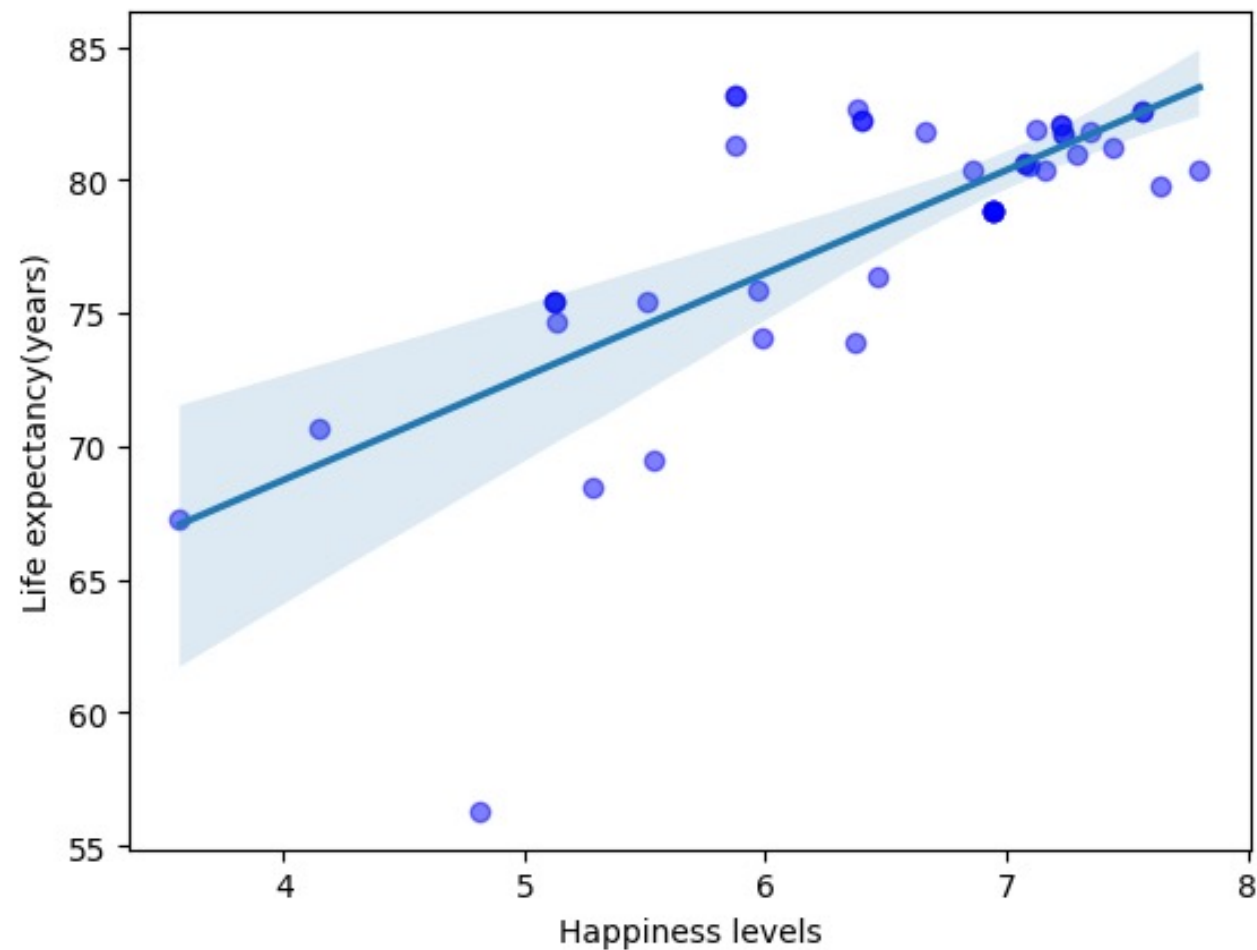
⇒ 오염정도와 행복 수준 간의 **강한 음의 상관관계**

오염도가 하락 할수록 행복 수준도 상승하는 경향이 있음

- 정규성 검정 결과 (Happiness levels): Statistic=0.9092, p-value=0.00211808481253683567
- 정규성 검정 결과 (Pollution(Index score)): Statistic=0.9667, p-value=0.23032510280609130859
- Levene 등분산성 검정 결과 (Pollution(Index score)): Statistic=99.8831, p-value=0.00000000000000046802
- Fligner 등분산성 검정 결과 (Pollution(Index score)): Statistic=51.0477, p-value=0.000000000000090146762
- Spearman 검정 결과 (Pollution(Index score)): 상관계수=-0.8232, p-value=0.00000000000692489846

## 1. Happiness levels에 가장 큰 영향을 미치는 요인 분석

# Happiness levels VS Life expectancy



⇒ 기대 수명과 행복 수준 간의 **강한 양의 상관관계**

기대 수명이 증가할수록 행복 수준도 증가한다는 관찰을 뒷받침하며, 질병 예방 및 의료 서비스의 향상이 행복에 긍정적인 영향을 미칠 수 있다는 가설을 뒷받침함

- 정규성 검정 결과 (Happiness levels): Statistic=0.9092, p-value=0.00211808481253683567
- 정규성 검정 결과 (Life expectancy(years)): Statistic=0.7850, p-value=0.00000140098029532965
- Levene 등분산성 검정 결과 (Life expectancy(years)): Statistic=16.0644, p-value=0.00012985288526244973
- Fligner 등분산성 검정 결과 (Life expectancy(years)): Statistic=21.6059, p-value=0.00000334828816850522
- Spearman 검정 결과 (Life expectancy(years)): 상관계수=0.5725, p-value=0.00004893824159814498

# 02

비만율과 라이프 스타일 요소들의 관계



# 비만율과 라이프 스타일 요소 간 관계

정규성 검정 결과 (Obesity levels): Statistic=0.9035, p-value=0.00140059425029903650

정규성 검정 결과 (Sunshine hours): Statistic=0.9439, p-value=0.03266043588519096375

Levene 등분산성 검정 결과 (Sunshine hours): Statistic=114.1480, p-value=0.00000000000000001893

Fligner 등분산성 검정 결과 (Sunshine hours): Statistic=55.5190, p-value=0.00000000000009255890

Spearman 검정 결과 (Sunshine hours): 상관계수=0.3346, p-value=0.02643802248436251737

정규성 검정 결과 (Obesity levels): Statistic=0.9035, p-value=0.00140059425029903650

정규성 검정 결과 (Cost of a bottle of water): Statistic=0.9534, p-value=0.0734535737269592285

Levene 등분산성 검정 결과 (Cost of a bottle of water): Statistic=57.7042, p-value=0.00000000003441344446

Fligner 등분산성 검정 결과 (Cost of a bottle of water): Statistic=37.3796, p-value=0.00000000097231435411

Spearman 검정 결과 (Cost of a bottle of water): 상관계수=0.1329, p-value=0.38991230829564726967

정규성 검정 결과 (Obesity levels): Statistic=0.9035, p-value=0.00140059425029903650

정규성 검정 결과 (Life expectancy(years)): Statistic=0.7850, p-value=0.00000140098029532965

Levene 등분산성 검정 결과 (Life expectancy(years)): Statistic=13.6262, p-value=0.00039067928112157302

Fligner 등분산성 검정 결과 (Life expectancy(years)): Statistic=12.3402, p-value=0.00044330354486013822

Spearman 검정 결과 (Life expectancy(years)): 상관계수=-0.0977, p-value=0.52810180226285252658

정규성 검정 결과 (Obesity levels): Statistic=0.9035, p-value=0.00140059425029903650

정규성 검정 결과 (Pollution(Index score)): Statistic=0.9667, p-value=0.23032510280609130859

Levene 등분산성 검정 결과 (Pollution(Index score)): Statistic=26.7047, p-value=0.00000151832257728926

Fligner 등분산성 검정 결과 (Pollution(Index score)): Statistic=18.1036, p-value=0.00002092107624157474

Spearman 검정 결과 (Pollution(Index score)): 상관계수=-0.1531, p-value=0.32098076912553197282

정규성 검정 결과 (Obesity levels): Statistic=0.9035, p-value=0.00140059425029903650

정규성 검정 결과 (Annual avg. hours worked): Statistic=0.9408, p-value=0.02517922781407833099

Levene 등분산성 검정 결과 (Annual avg. hours worked): Statistic=33.0273, p-value=0.00000013502312340704

Fligner 등분산성 검정 결과 (Annual avg. hours worked): Statistic=17.1226, p-value=0.00003504361941753612

Spearman 검정 결과 (Annual avg. hours worked): 상관계수=0.4358, p-value=0.00310950815783077317

정규성 검정 결과 (Obesity levels): Statistic=0.9035, p-value=0.00140059425029903650

정규성 검정 결과 (Happiness levels): Statistic=0.9092, p-value=0.00211808481253683567

Levene 등분산성 검정 결과 (Happiness levels): Statistic=54.0894, p-value=0.00000000010484684202

Fligner 등분산성 검정 결과 (Happiness levels): Statistic=34.7233, p-value=0.00000000380053059957

Spearman 검정 결과 (Happiness levels): 상관계수=0.2591, p-value=0.08949360359364867679

정규성 검정 결과 (Obesity levels): Statistic=0.9035, p-value=0.00140059425029903650

정규성 검정 결과 (Outdoor activities): Statistic=0.9444, p-value=0.03410947322845458984

Levene 등분산성 검정 결과 (Outdoor activities): Statistic=54.5332, p-value=0.00000000009129753103

Fligner 등분산성 검정 결과 (Outdoor activities): Statistic=47.4381, p-value=0.00000000000567676778

Spearman 검정 결과 (Outdoor activities): 상관계수=0.2476, p-value=0.10510926833665407143

정규성 검정 결과 (Obesity levels): Statistic=0.9035, p-value=0.00140059425029903650

정규성 검정 결과 (Number of take out places): Statistic=0.7449, p-value=0.00000022453011183643

Levene 등분산성 검정 결과 (Number of take out places): Statistic=23.6435, p-value=0.00000517979387231084

Fligner 등분산성 검정 결과 (Number of take out places): Statistic=52.6602, p-value=0.00000000000039655434

Spearman 검정 결과 (Number of take out places): 상관계수=0.1493, p-value=0.33354968548944485818

정규성 검정 결과 (Obesity levels): Statistic=0.9035, p-value=0.00140059425029903650

정규성 검정 결과 (Cost of a monthly gym membership): Statistic=0.9416, p-value=0.02694516256451606750

Levene 등분산성 검정 결과 (Cost of a monthly gym membership): Statistic=3.7645, p-value=0.05562676726684966794

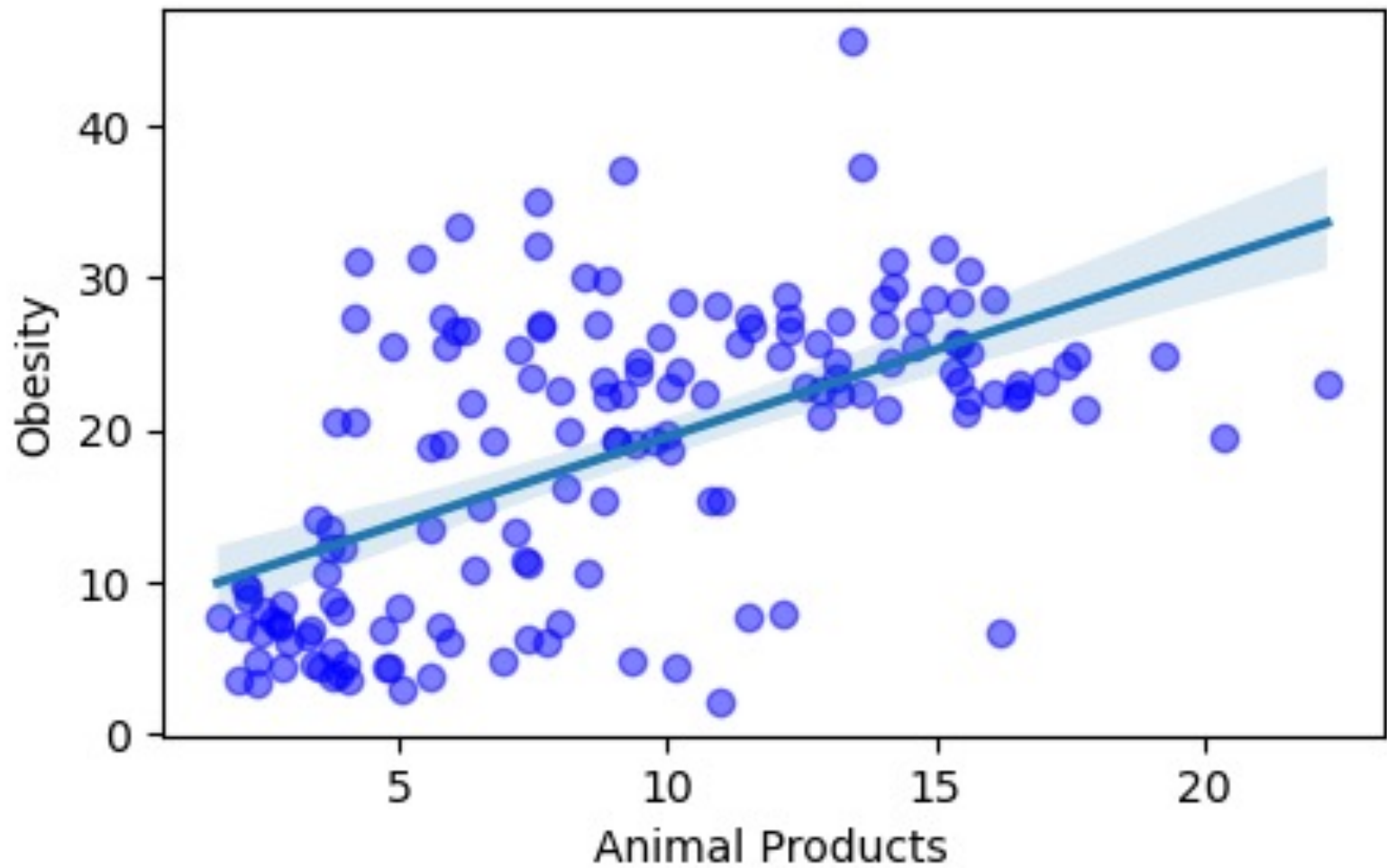
Fligner 등분산성 검정 결과 (Cost of a monthly gym membership): Statistic=3.5666, p-value=0.05895428064855148809

Spearman 검정 결과 (Cost of a monthly gym membership): 상관계수=-0.1020, p-value=0.50990577533540415445

# 03

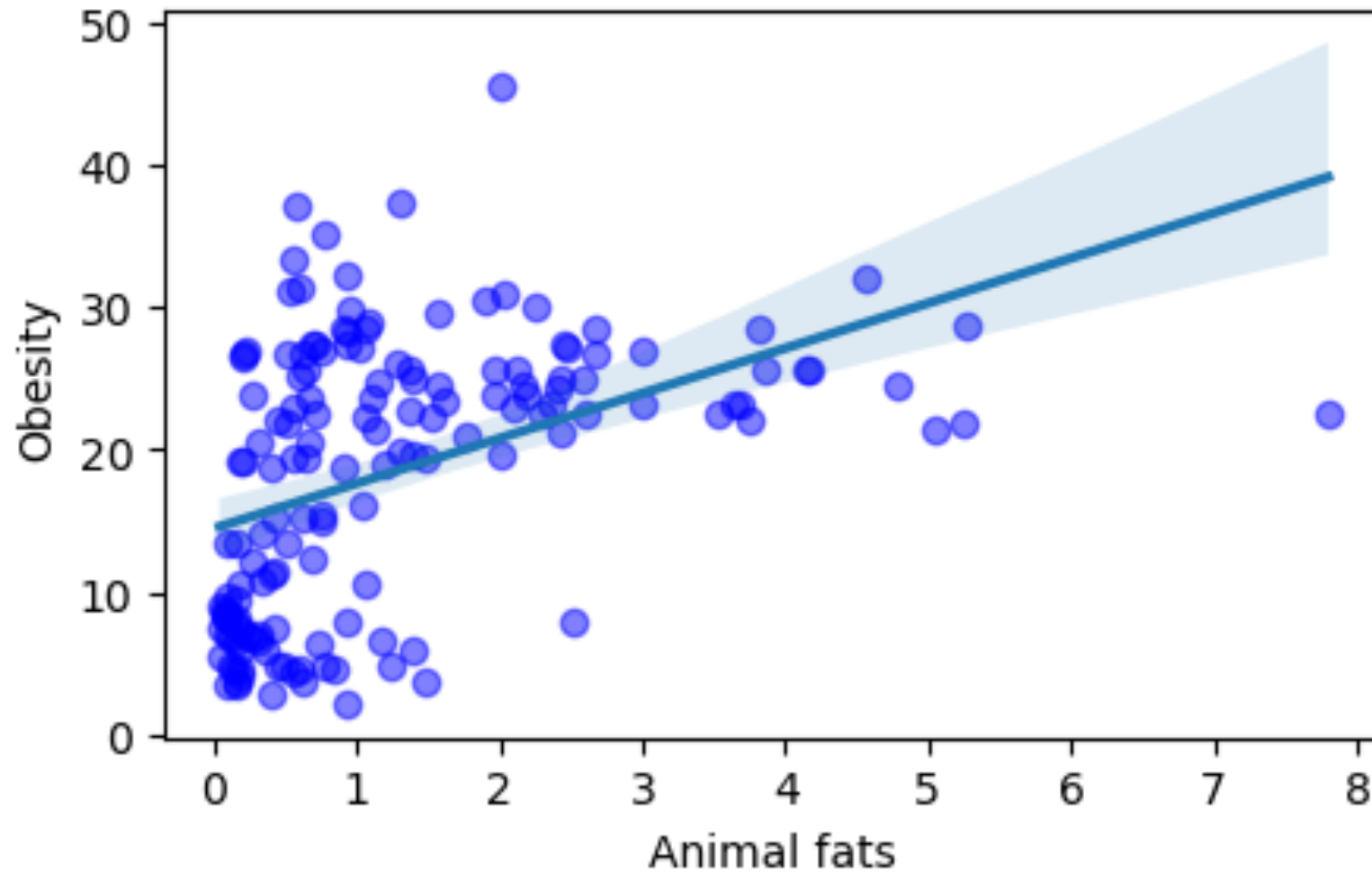
비만율과 섭취하는 음식 종류의 관계

3. 비만율과 섭취하는 음식 종류의 관계(Animal Products)



정규성 검정 결과 (Obesity): Obesity Statistic=0.9333, p-value=0.00000127858925225155  
정규성 검정 결과 (Animal Products): Animal Products Statistic=0.9575, p-value=0.00011729457037290558  
등분산성 검정 결과: Statistic=61.0055, p-value=0.00000000000000091630231081826332  
Spearman검정 (Animal Products vs Obesity) -> 상관계수: 0.5662 p-value: 0.0000000000000001960834

### 3. 비만율과 섭취하는 음식 종류의 관계(Animal fats)



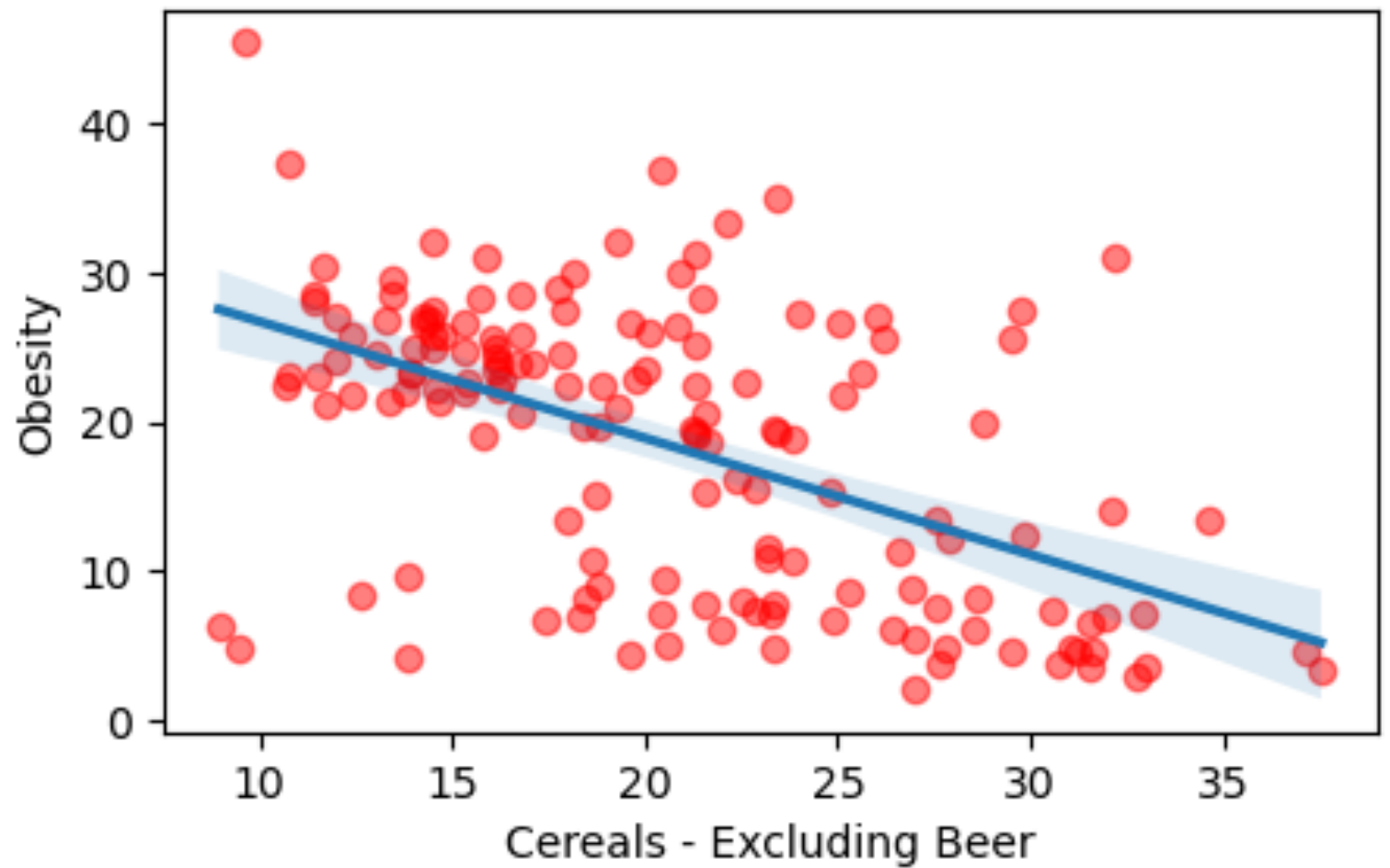
정규성 검정 결과 (Obesity): Obesity Statistic=0.9333, p-value=0.00000127858925225155

정규성 검정 결과 (Animal fats): Animal fats Statistic=0.8018, p-value=0.00000000000036137038

**등분산성 검정 결과:** Statistic=225.0912, p-value=0.000000000000000000000000000000

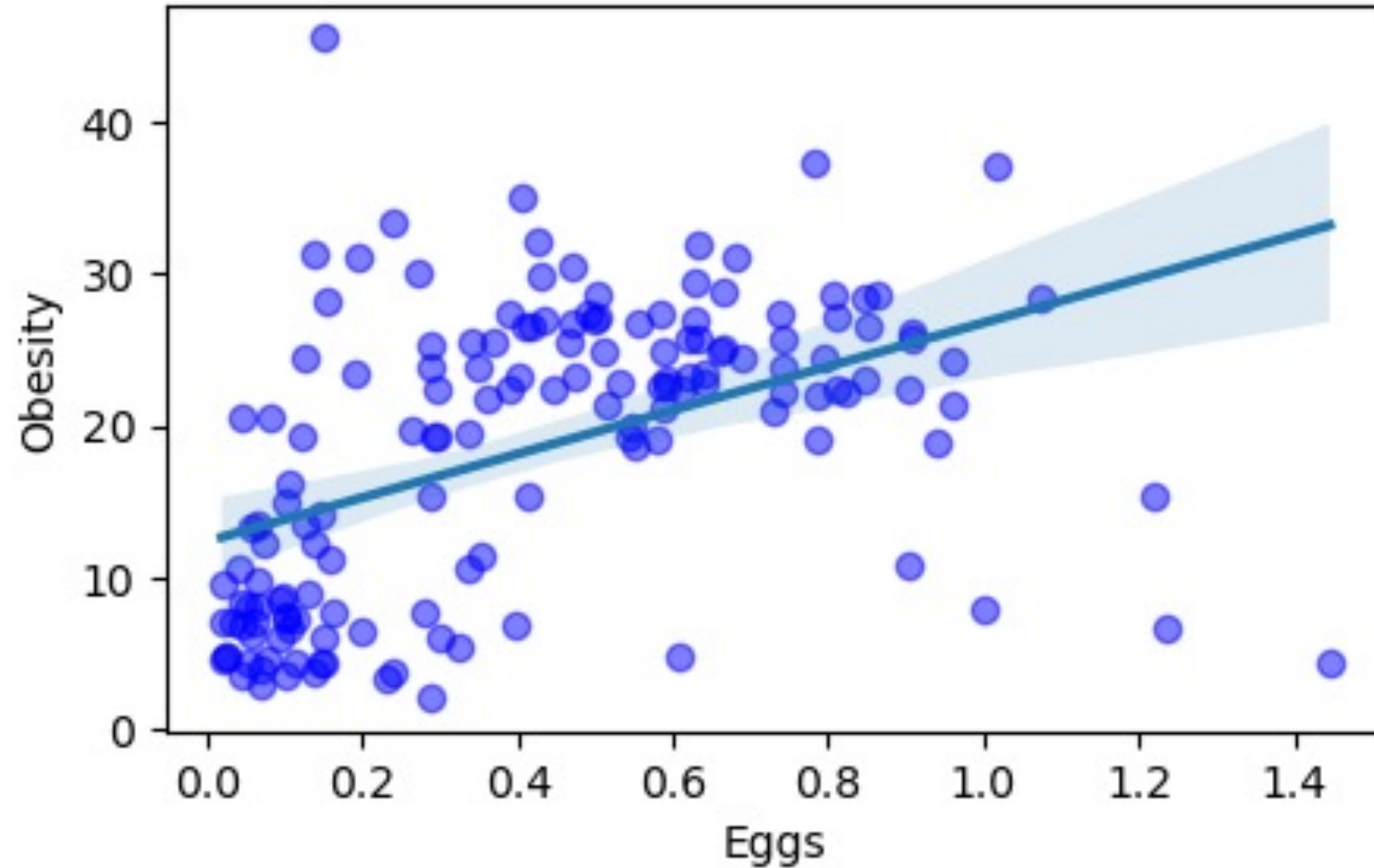
Spearman검정 (Animal fats vs Obesity) -> 상관계수: 0.5390 p-value: 0.000000000000055538151

3. 비만율과 섭취하는 음식 종류의 관계(Cereals - Excluding Beer)



정규성 검정 결과 (Obesity): Obesity Statistic=0.9333, p-value=0.00000127858925225155  
정규성 검정 결과 (Cereals - Excluding Beer): Cereals - Excluding Beer Statistic=0.9698,  
p-value=0.00183851050678640604  
등분산성 검정 결과: Statistic=23.3997, p-value=0.000002088834491180852211237055  
Spearman검정 (Cereals - Excluding Beer vs Obesity) -> 상관계수: -0.5047 p-value: 0.000000000002496821733

### 3. 비만율과 섭취하는 음식 종류의 관계(Eggs)



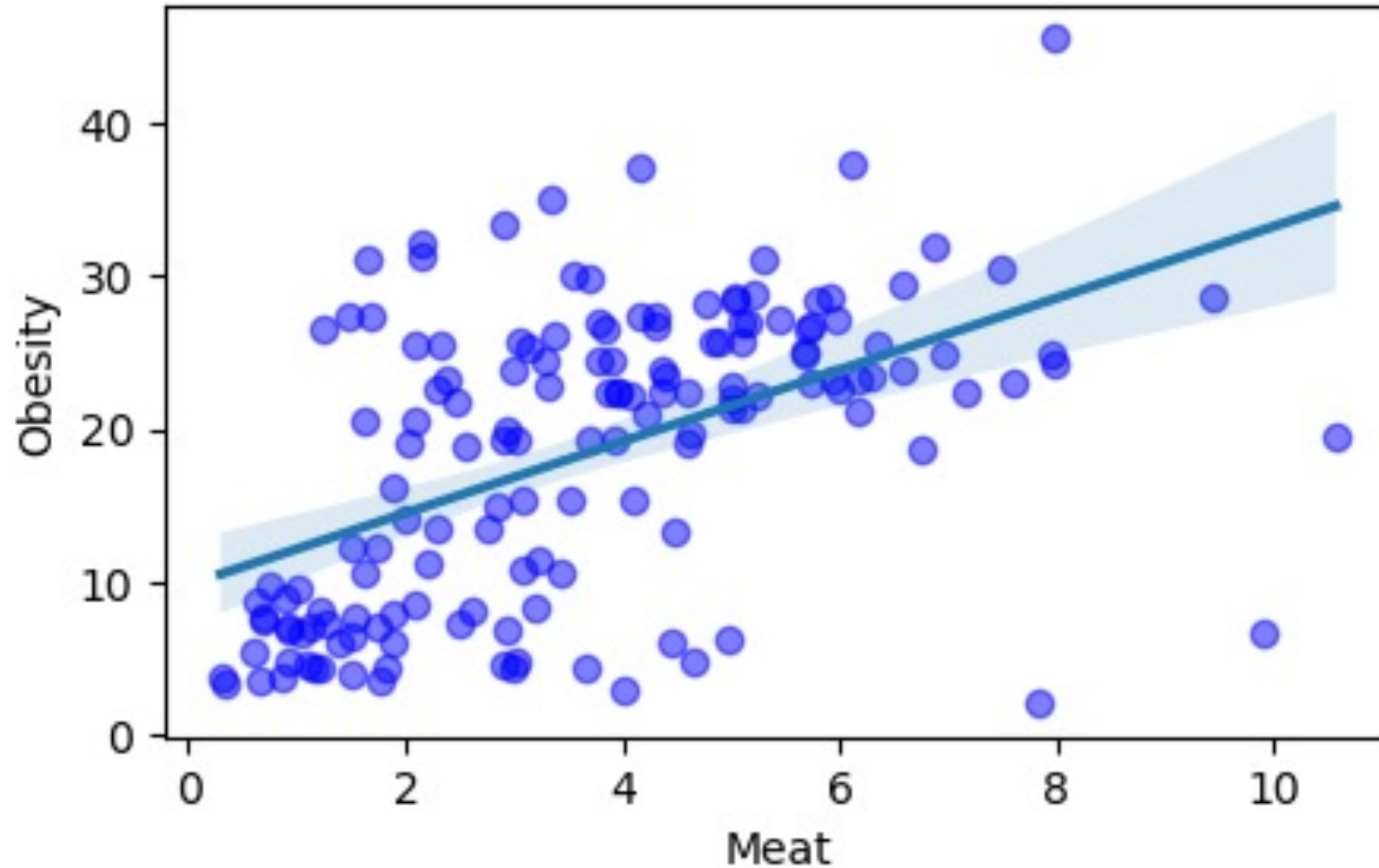
정규성 검정 결과 (Obesity): Obesity Statistic=0.9333, p-value=0.00000127858925225155

정규성 검정 결과 (Eggs): Eggs Statistic=0.9366, p-value=0.00000222735798161011

**등분산성 검정 결과:** Statistic=276.3927, p-value=0.00000000000000000000000000000000

Spearman검정 (Eggs vs Obesity) -> 상관계수: 0.5417 p-value: 0.000000000000040240672

### 3. 비만율과 섭취하는 음식 종류의 관계(Meat)



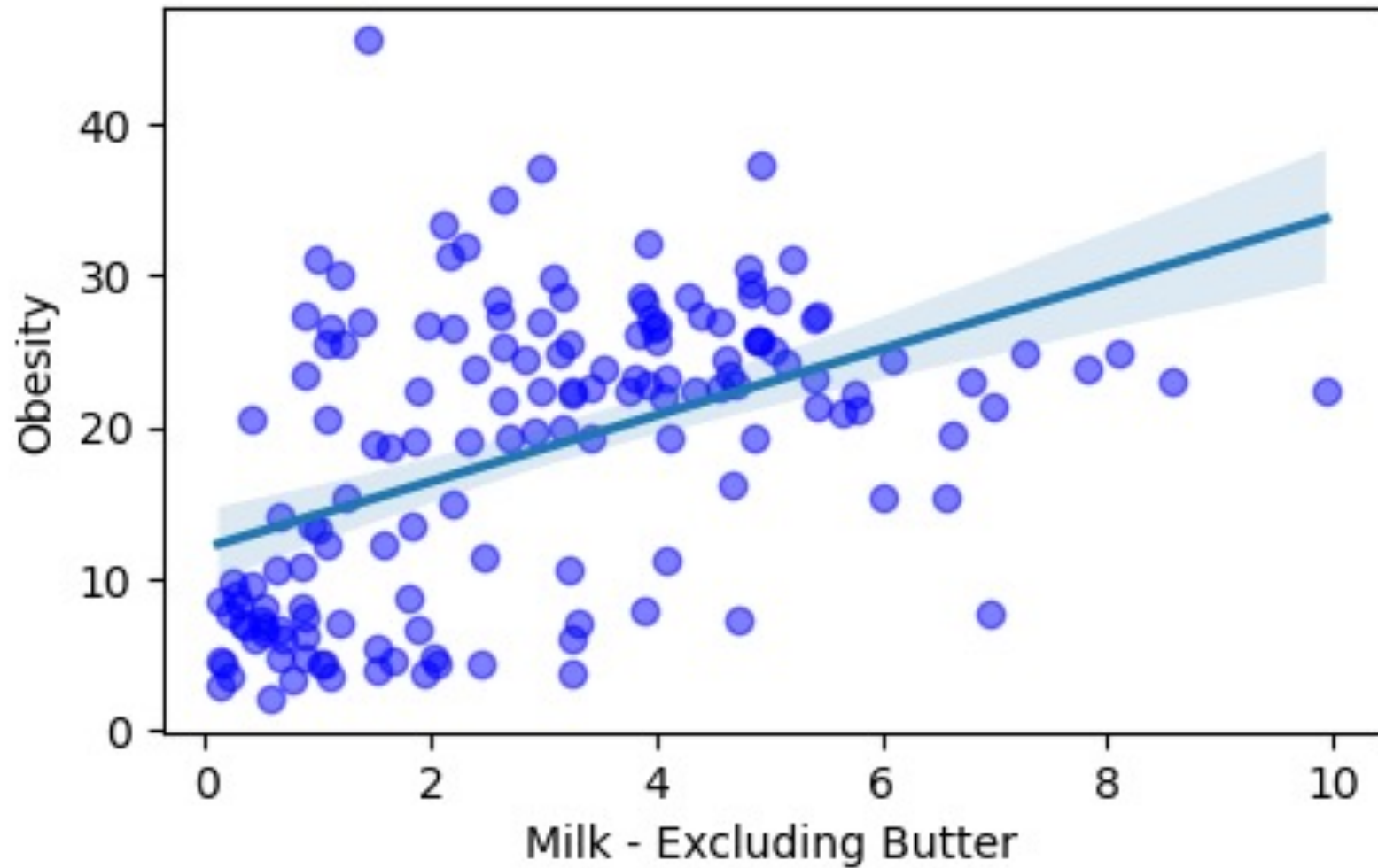
정규성 검정 결과 (Obesity): Obesity Statistic=0.9333, p-value=0.00000127858925225155

정규성 검정 결과 (Meat): Meat Statistic=0.9627, p-value=0.00035980451502837241

등분산성 검정 결과: Statistic=174.6879, p-value=0.0000000000000000000000000000000000

Spearman검정 (Meat vs Obesity) -> 상관계수: 0.5409 p-value: 0.000000000000044316741

### 3. 비만율과 섭취하는 음식 종류의 관계(Milk - Excluding Butter)



정규성 검정 결과 (Obesity): Obesity Statistic=0.9333, p-value=0.00000127858925225155

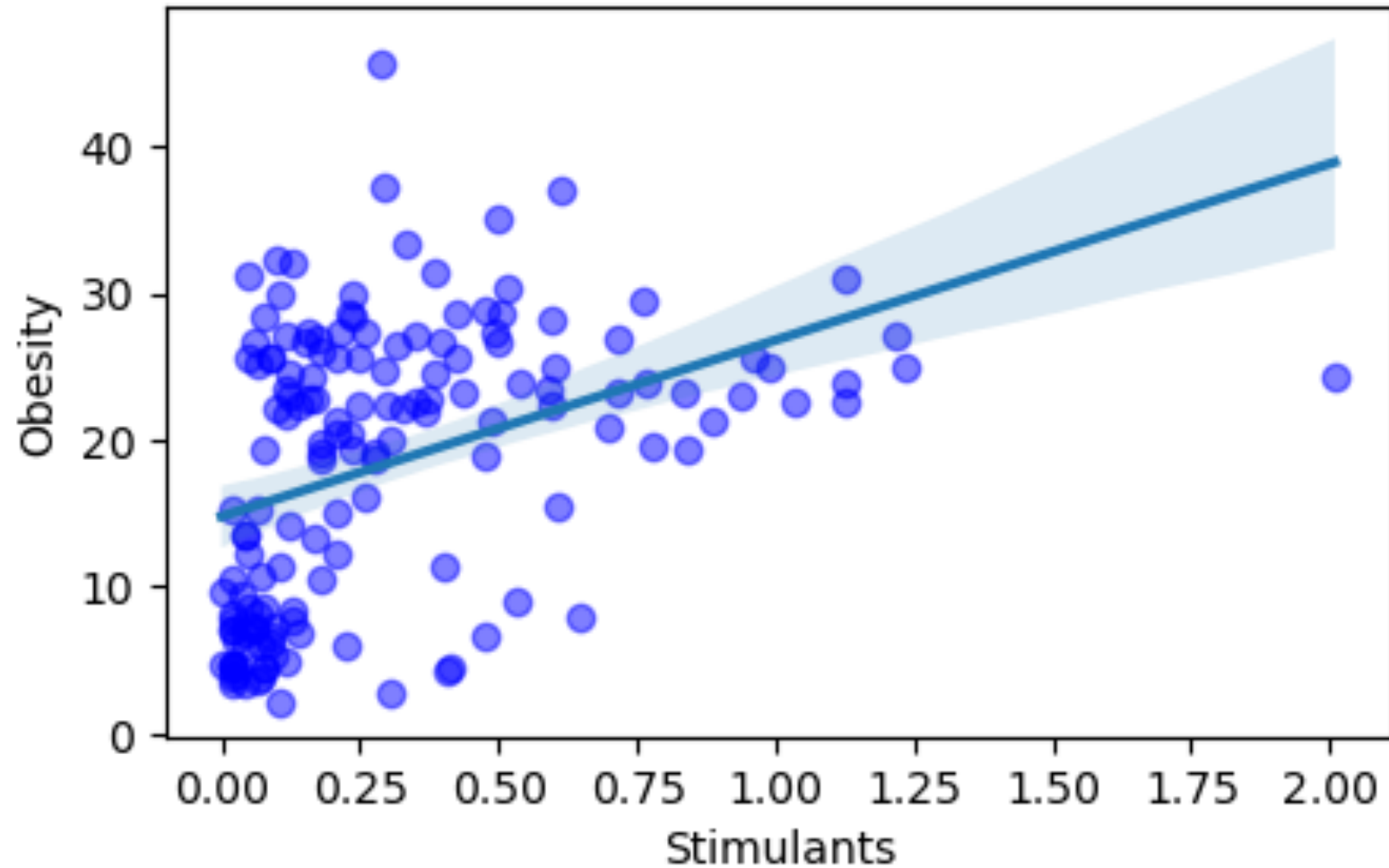
정규성 검정 결과 (Milk - Excluding Butter): Milk - Excluding Butter Statistic=0.9457, p-value=0.00001141097891377285

**등분산성 검정 결과:** Statistic=175.9500, p-value=0.00000000000000000000000000000000

Spearman검정 (Milk - Excluding Butter vs Obesity) -> 상관계수: 0.5025 p-value: 0.00000000003121375918



### 3. 비만율과 섭취하는 음식 종류의 관계(Stimulants)



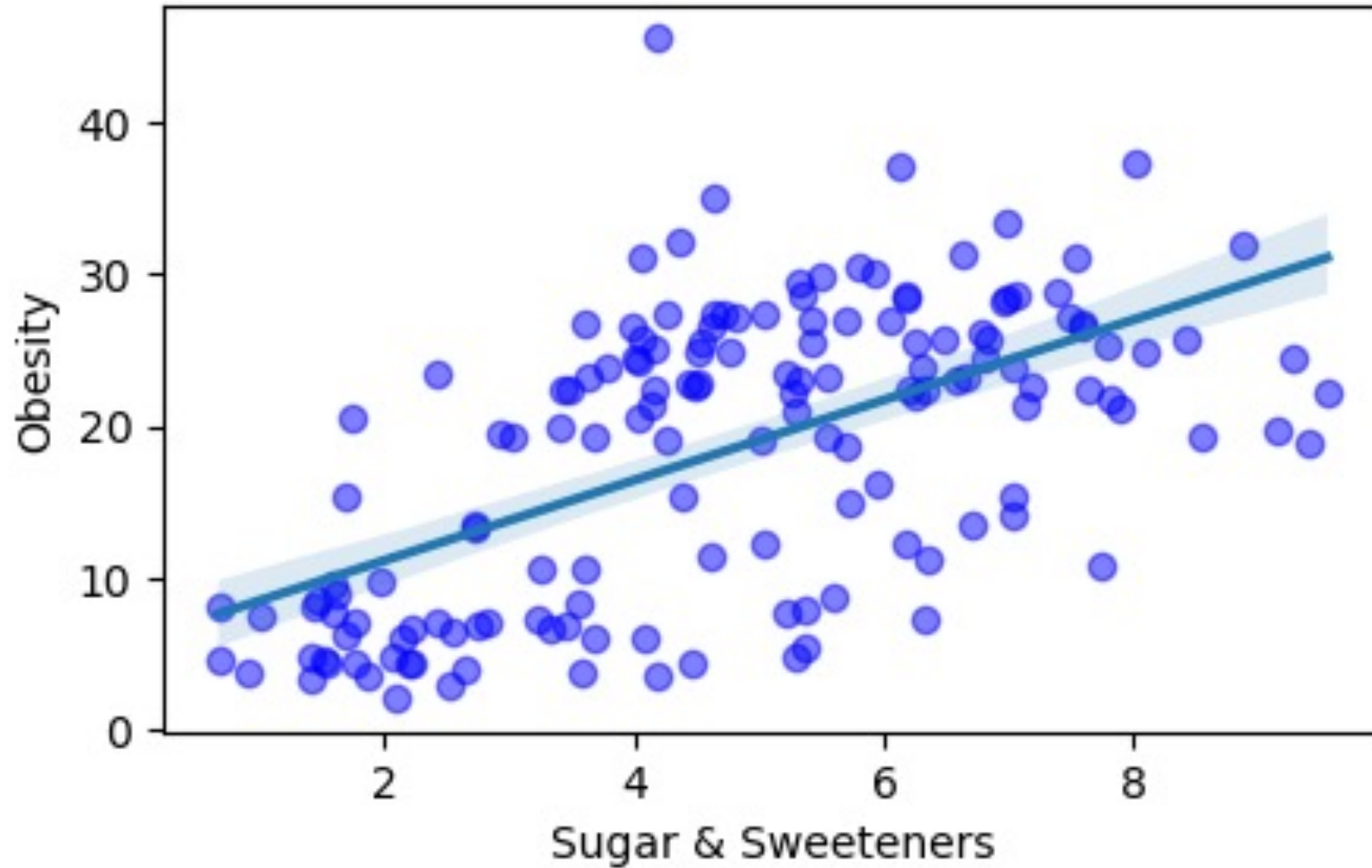
정규성 검정 결과 (Obesity): Obesity Statistic=0.9333, p-value=0.00000127858925225155

정규성 검정 결과 (Stimulants): Stimulants Statistic=0.8019, p-value=0.00000000000036255753

**등분산성 검정 결과:** Statistic=278.7930, p-value=0.00000000000000000000000000000000

Spearman검정 (Stimulants vs Obesity) -> 상관계수: 0.5163 p-value: 0.00000000000716440642

### 3. 비만율과 섭취하는 음식 종류의 관계(Sugar & Sweeteners)



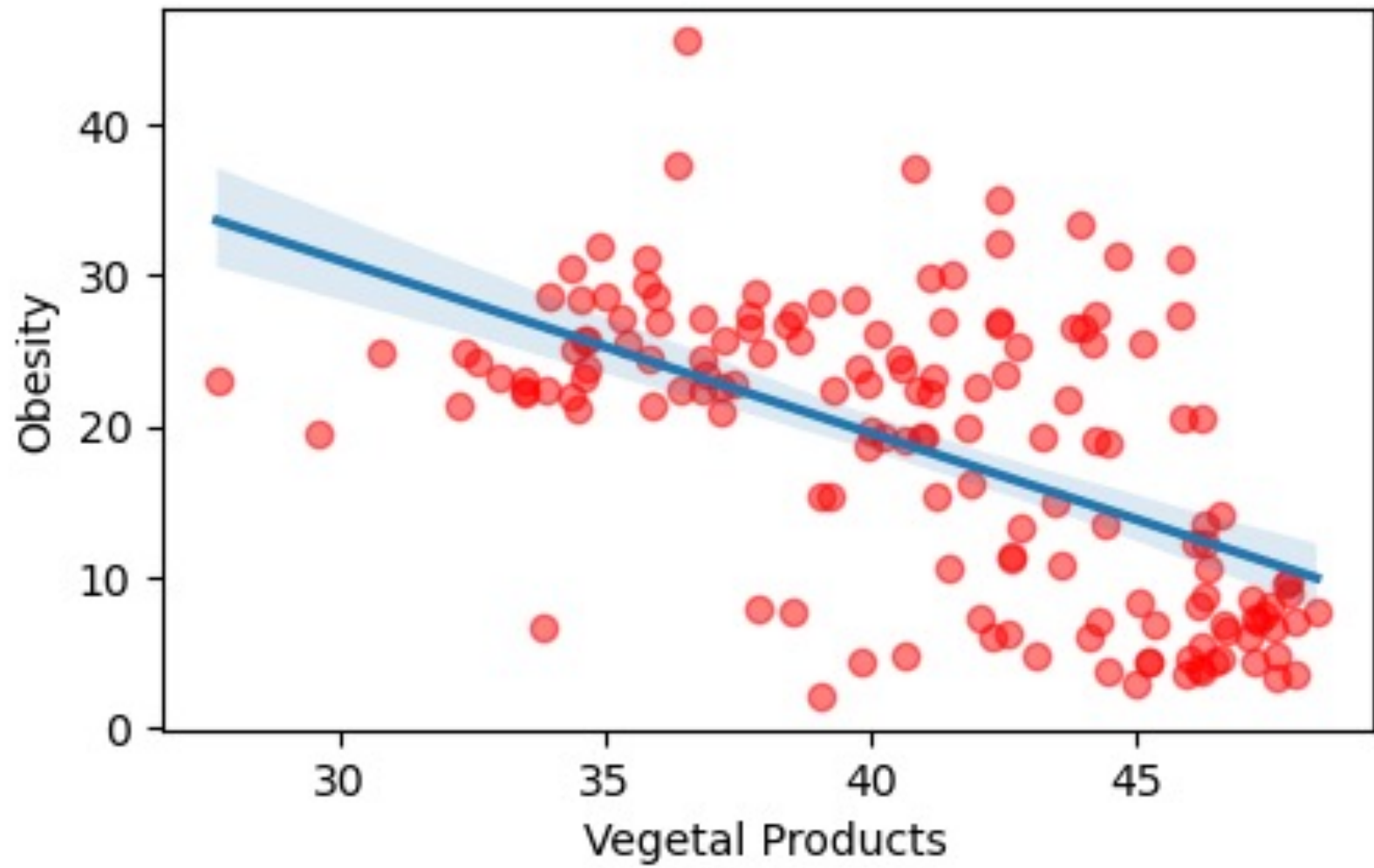
정규성 검정 결과 (Obesity): Obesity Statistic=0.9333, p-value=0.00000127858925225155

정규성 검정 결과 (Sugar & Sweeteners): Sugar & Sweeteners Statistic=0.9816, p-value=0.03807317465543746948

**등분산성 검정 결과:** Statistic=173.3899, p-value=0.0000000000000000000000000000000000

Spearman검정 (Sugar & Sweeteners vs Obesity) -> 상관계수: 0.5751 p-value: 0.0000000000000000619435

3. 비만율과 섭취하는 음식 종류의 관계(Vegetal Products)



정규성 검정 결과 (Obesity): Obesity Statistic=0.9333, p-value=0.00000127858925225155  
정규성 검정 결과 (Vegetal Products): Vegetal Products Statistic=0.9575, p-value=0.00011614019604166970  
등분산성 검정 결과: Statistic=61.0187, p-value=0.00000000000000091122346615395480  
Spearman검정 (Vegetal Products vs Obesity) -> 상관계수: -0.5660 p-value: 0.0000000000000002027253

# 비만율과 섭취하는 음식 종류의 관계

## 1. 결론

- 파란 그래프(비만율과 양적 상관관계): 동물성 음식(동물성제품, 동물성 지방, 달걀, 고기, 우유), 각성제, 설탕과 감미료
  - 빨간 그래프(비만율과 음적 상관관계): 식물성 음식(식물성 제품, 곡류)
- ➔ 비만율이 높은 이유는 동물성 음식, 설탕&감미료, 각성제로부터의 kcal 섭취량이 많고, 식물성 음식의 kcal 섭취량은 비교적 적기 때문이라고 추측할 수 있음

## 2. 시사점

### - 올바른 식습관 교육

- 인구에게 올바른 식습관 교육은 건강한 식단을 구성하는 데 중요하다.
- 동물성 음식과 설탕, 감미료, 각성제의 과도한 섭취를 줄이고, 식물성 음식을 적절히 섭취하는 교육이 필요하다.

### - 공공 보건 정책 수립

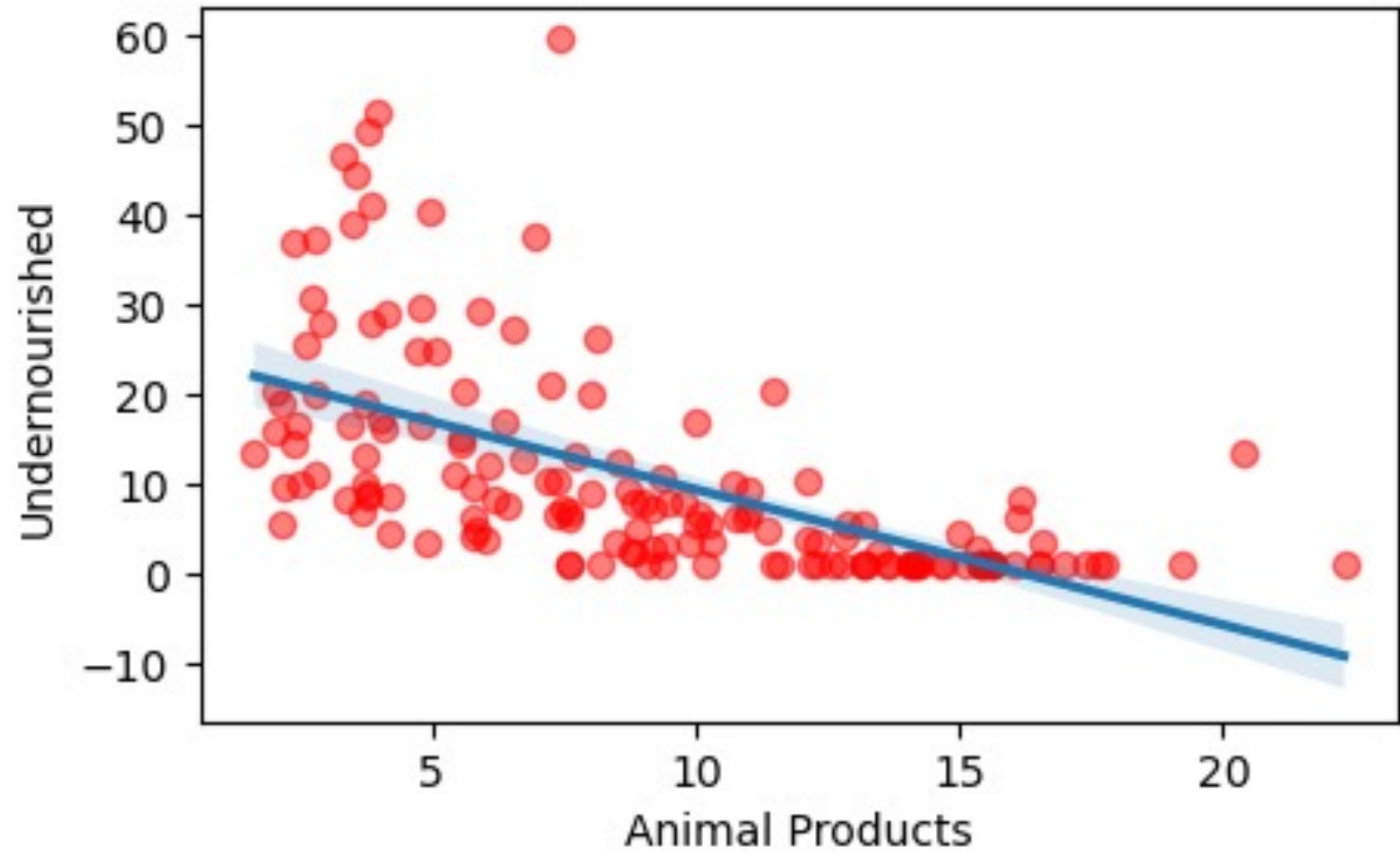
- 공공 보건 정책은 지역사회에서의 식품 소비에 영향을 미칠 수 있다.
- 비만율이 높은 지역에서는 동물성 음식과 설탕 등의 제한 정책을 강화하고, 식물성 음식의 접근성을 증가시키는 노력이 필요하다.

⇒ 이러한 결론과 시사점들을 고려하여 건강한 식습관을 촉진하고, 비만 관리에 효과적인 정책 및 교육이 이루어져야 한다.

# 04

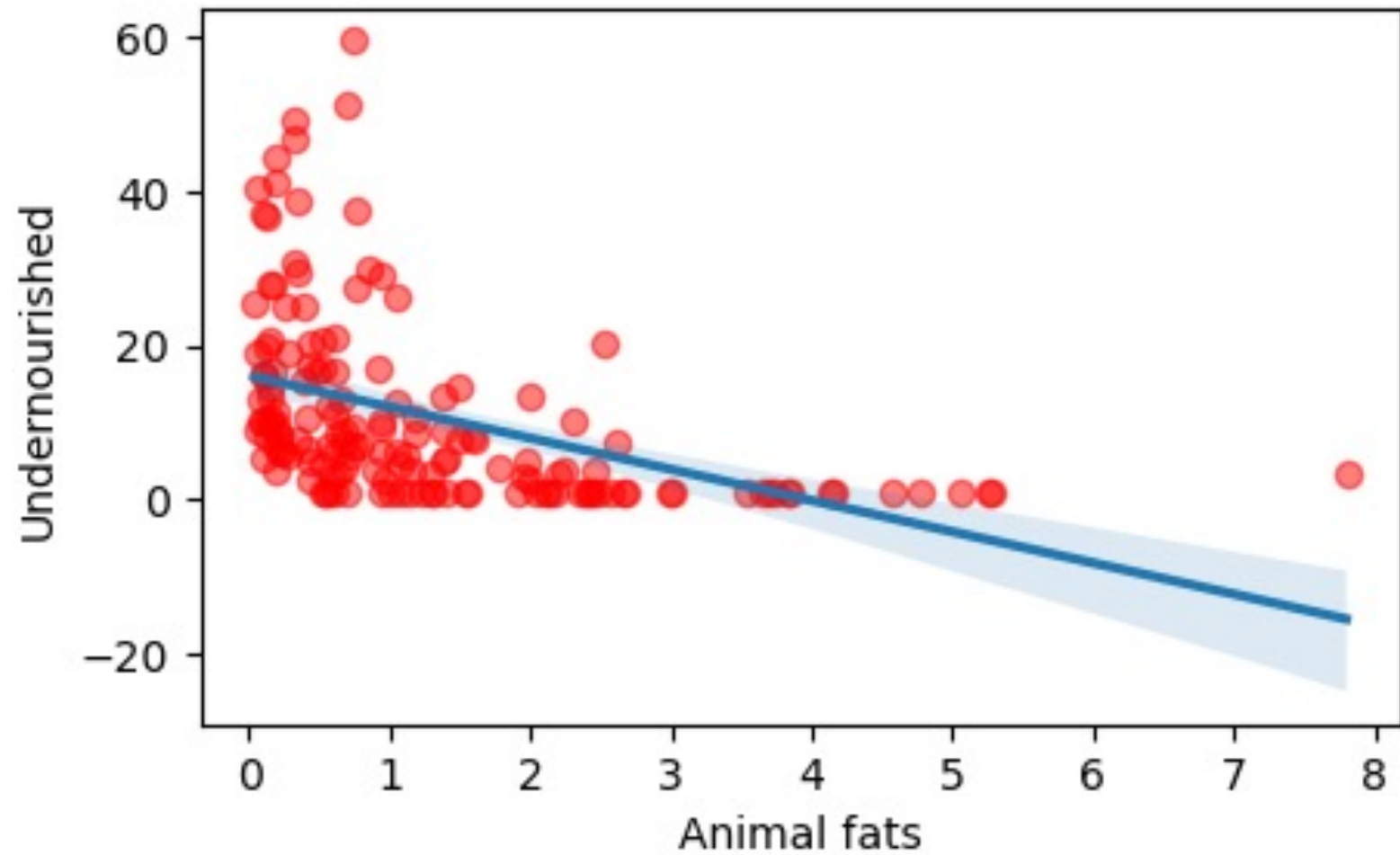
영양결핍률과 섭취하는 음식 종류의 관계

4. 영양결핍률과 섭취하는 음식 종류의 관계(Animal Products)



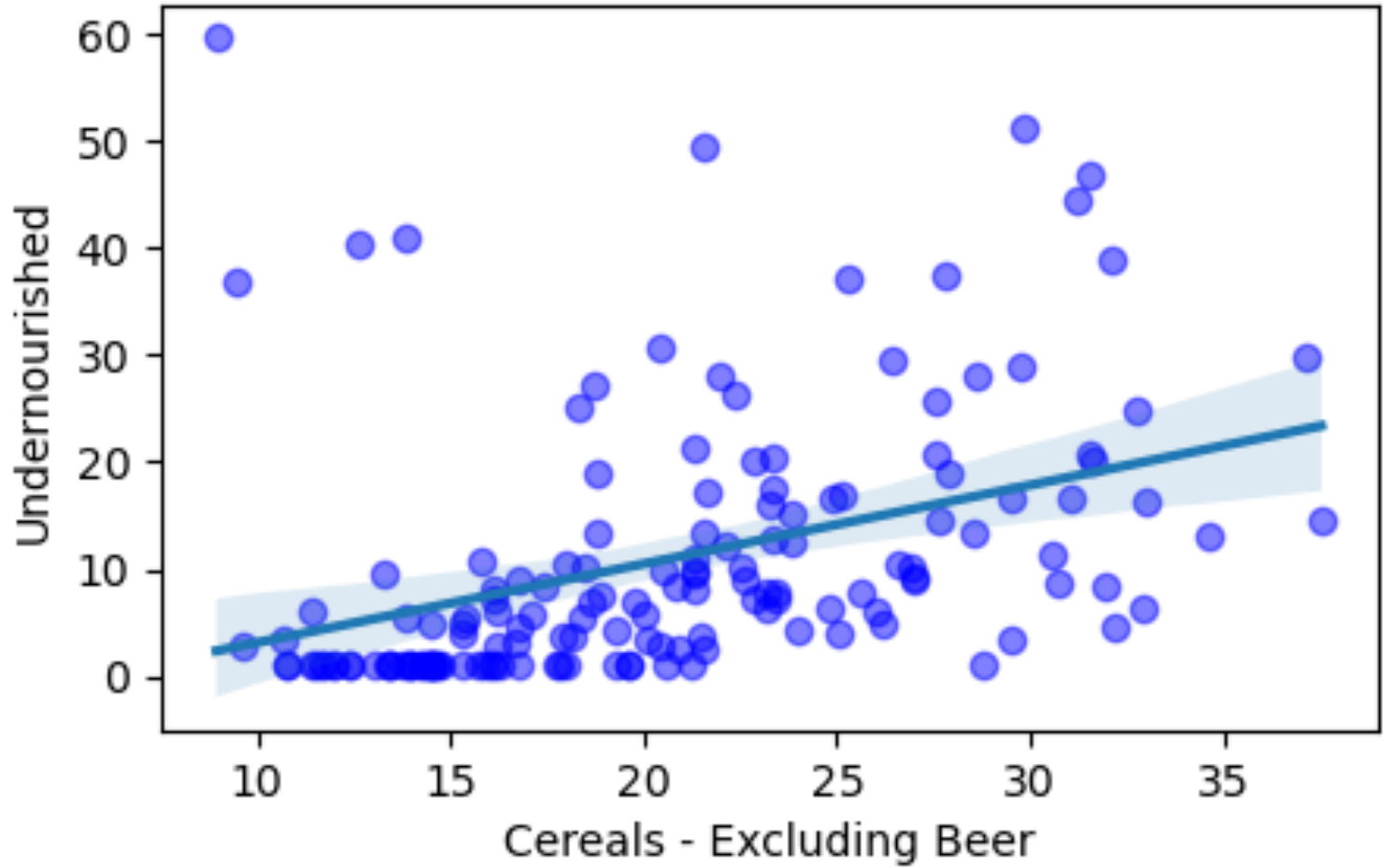
정규성 검정 결과 (Undernourished): Undernourished Statistic=0.7893, p-value=0.000000000000012983348  
정규성 검정 결과 (Animal Products): Animal Products Statistic=0.9575, p-value=0.00011729457037290558  
등분산성 검정 결과: Statistic=28.2978, p-value=0.000000201260870676462213233184  
Spearman검정 (Animal Products vs Undernourished) -> 상관계수: -0.7538 p-value: 0.000000000000000000000000

4. 영양결핍률과 섭취하는 음식 종류의 관계(Animal fats)



정규성 검정 결과 (Undernourished): Undernourished Statistic=0.7893, p-value=0.000000000000012983348  
정규성 검정 결과 (Animal fats): Animal fats Statistic=0.8018, p-value=0.0000000000000036137038  
등분산성 검정 결과: Statistic=90.4982, p-value=0.00000000000000000000574578171530  
Spearman검정 (Animal fats vs Undernourished) -> 상관계수: -0.6544 p-value: 0.00000000000000000000003

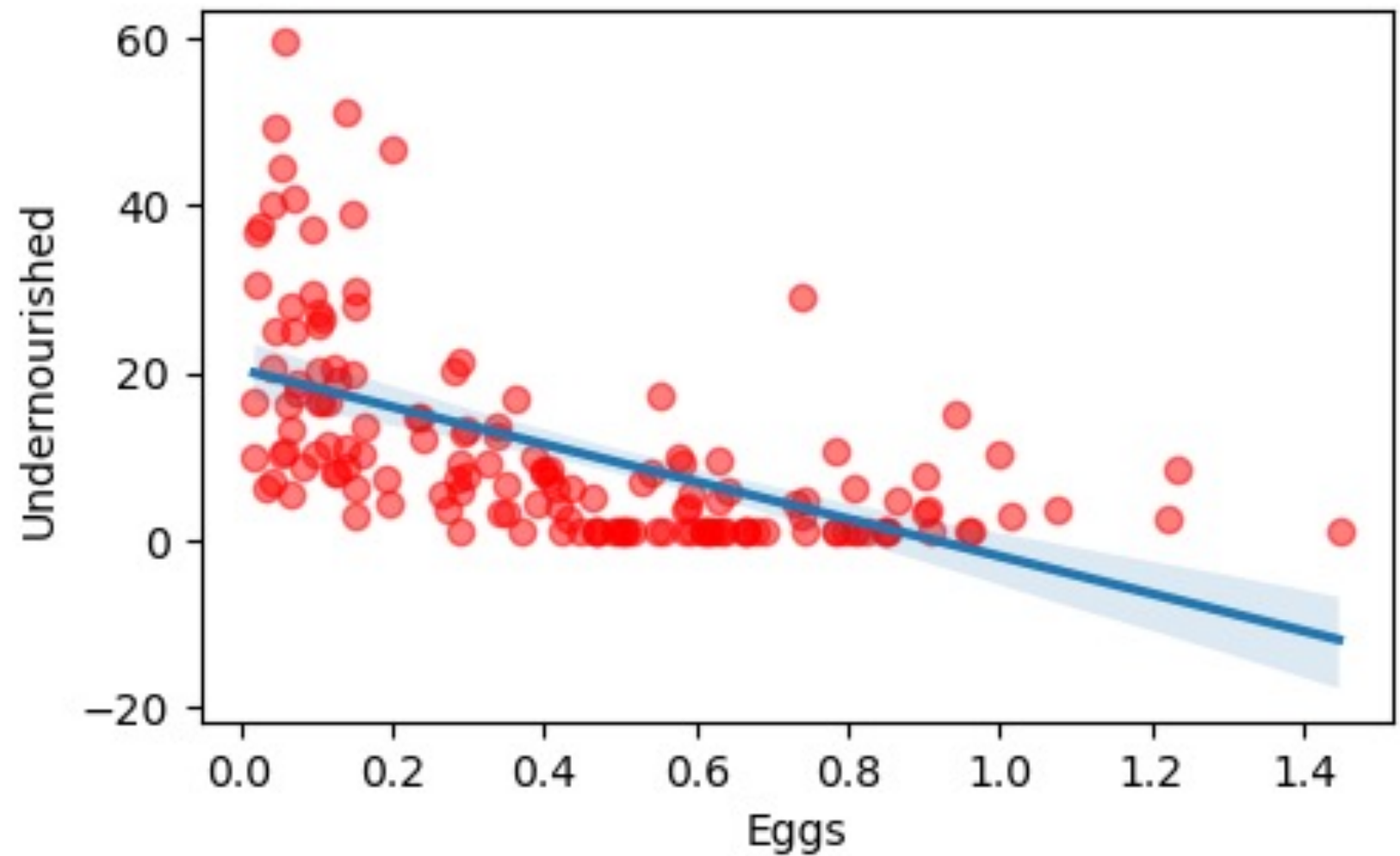
4. 영양결핍률과 섭취하는 음식 종류의 관계(Cereals - Excluding Beer)



정규성 검정 결과 (Undernourished): Undernourished Statistic=0.7893, p-value=0.000000000000012983348  
정규성 검정 결과 (Cereals - Excluding Beer): Cereals - Excluding Beer Statistic=0.9698, p-value=0.00183851050678640604  
등분산성 검정 결과: Statistic=12.6325, p-value=0.000438971915055249260750880325  
Spearman검정 (Cereals - Excluding Beer vs Undernourished) -> 상관계수: 0.5967 p-value: 0.00000000000000000031848

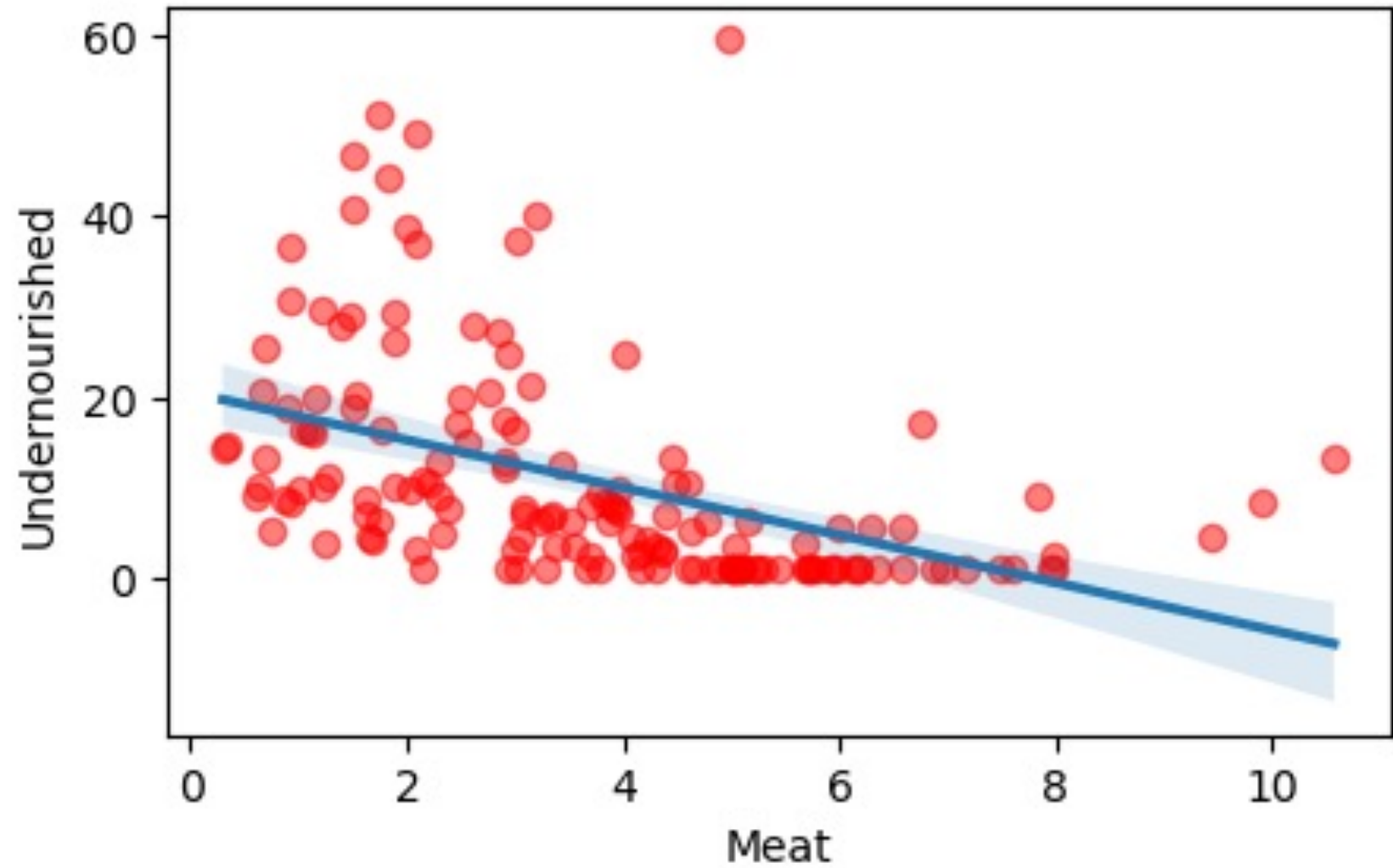


4. 영양결핍률과 섭취하는 음식 종류의 관계(Eggs)



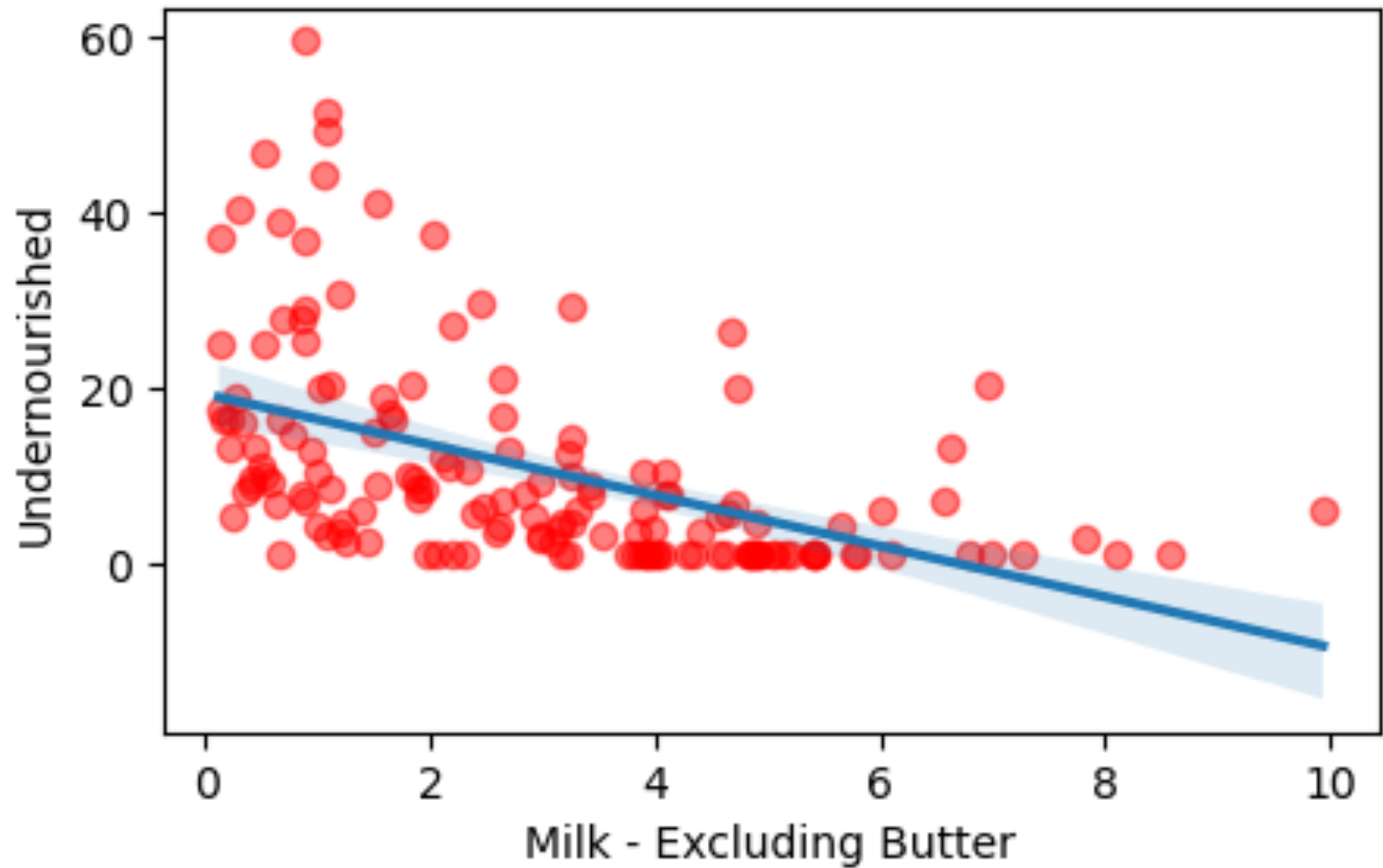
정규성 검정 결과 (Undernourished): Undernourished Statistic=0.7893, p-value=0.000000000000012983348  
정규성 검정 결과 (Eggs): Eggs Statistic=0.9366, p-value=0.00000222735798161011  
등분산성 검정 결과: Statistic=108.1912, p-value=0.000000000000000000000000676031017  
Spearman검정 (Eggs vs Undernourished) -> 상관계수: -0.7013 p-value: 0.000000000000000000000000

4. 영양결핍률과 섭취하는 음식 종류의 관계(Meat)



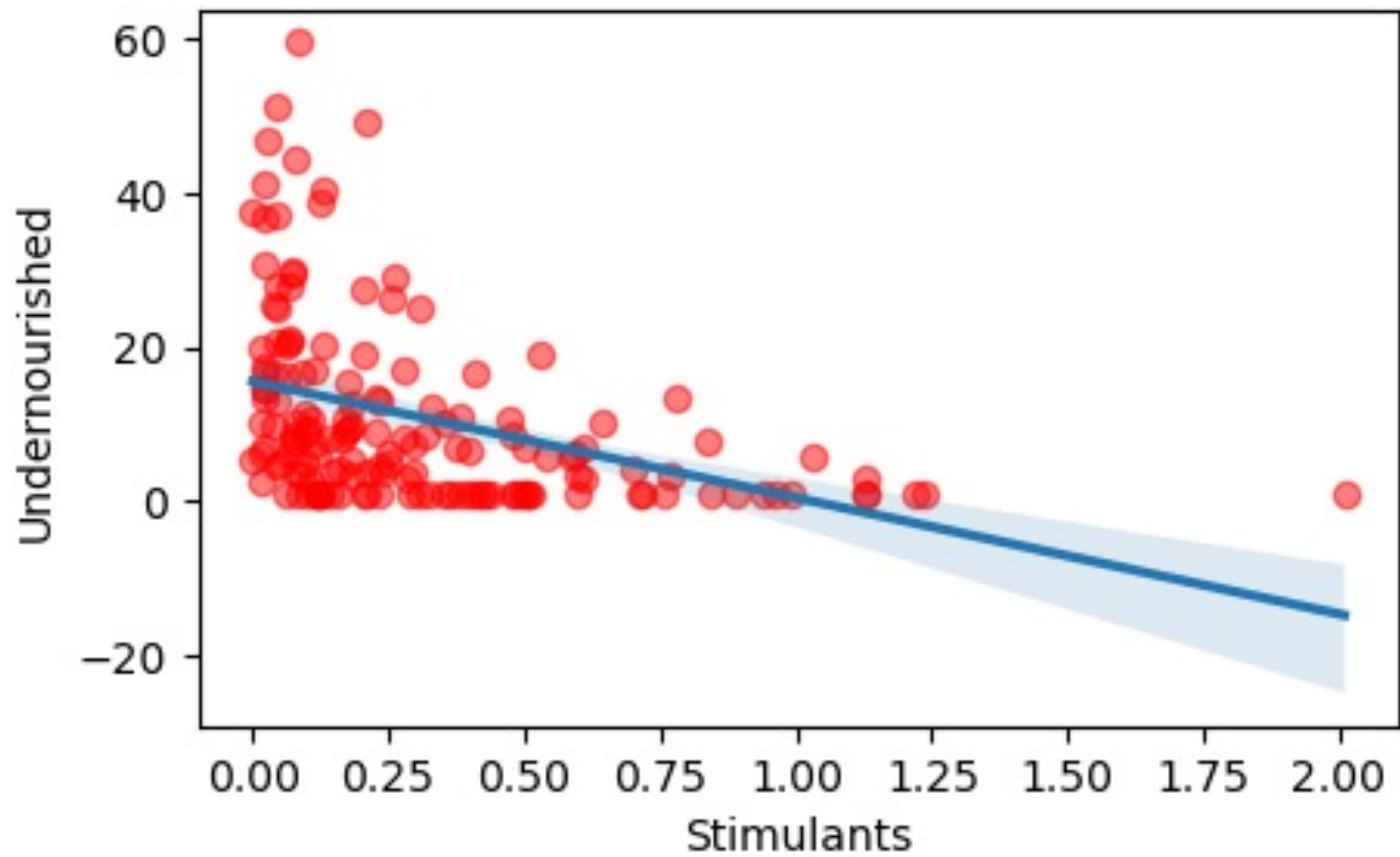
정규성 검정 결과 (Undernourished): Undernourished Statistic=0.7893, p-value=0.000000000000012983348  
정규성 검정 결과 (Meat): Meat Statistic=0.9627, p-value=0.00035980451502837241  
등분산성 검정 결과: Statistic=71.3593, p-value=0.0000000000000001220068585978986  
Spearman검정 (Meat vs Undernourished) -> 상관계수: -0.6527 p-value: 0.000000000000000000000005

4. 영양결핍률과 섭취하는 음식 종류의 관계(Milk - Excluding Butter)



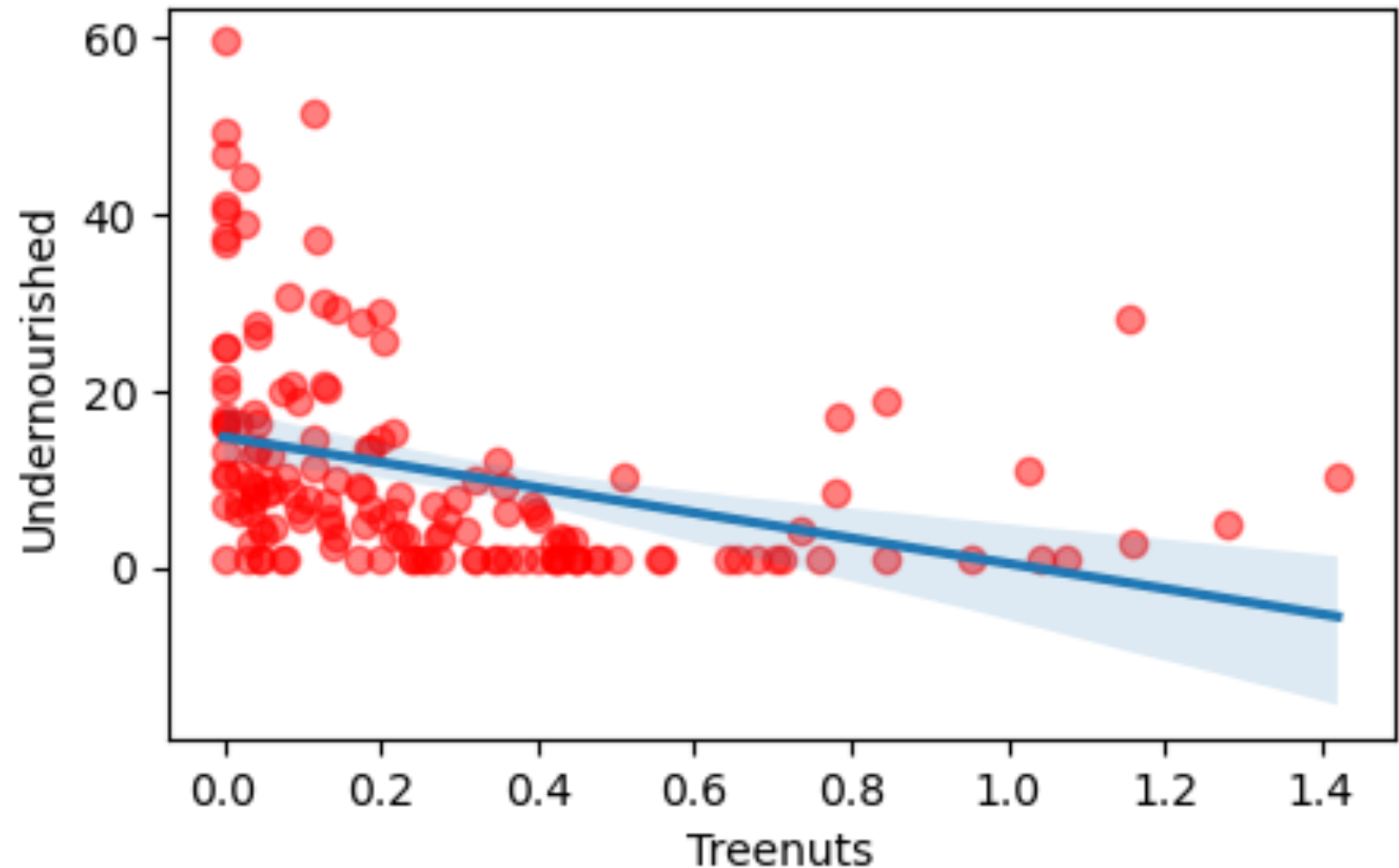
정규성 검정 결과 (Undernourished): Undernourished Statistic=0.7893, p-value=0.000000000000012983348  
정규성 검정 결과 (Milk - Excluding Butter): Milk - Excluding Butter Statistic=0.9457,  
p-value=0.00001141097891377285  
등분산성 검정 결과: Statistic=71.6792, p-value=0.00000000000000001069815745351464  
Spearman검정 (Milk - Excluding Butter vs Undernourished) -> 상관계수: -0.6237 p-value: 0.00000000000000000000563

4. 영양결핍률과 섭취하는 음식 종류의 관계(Stimulants)



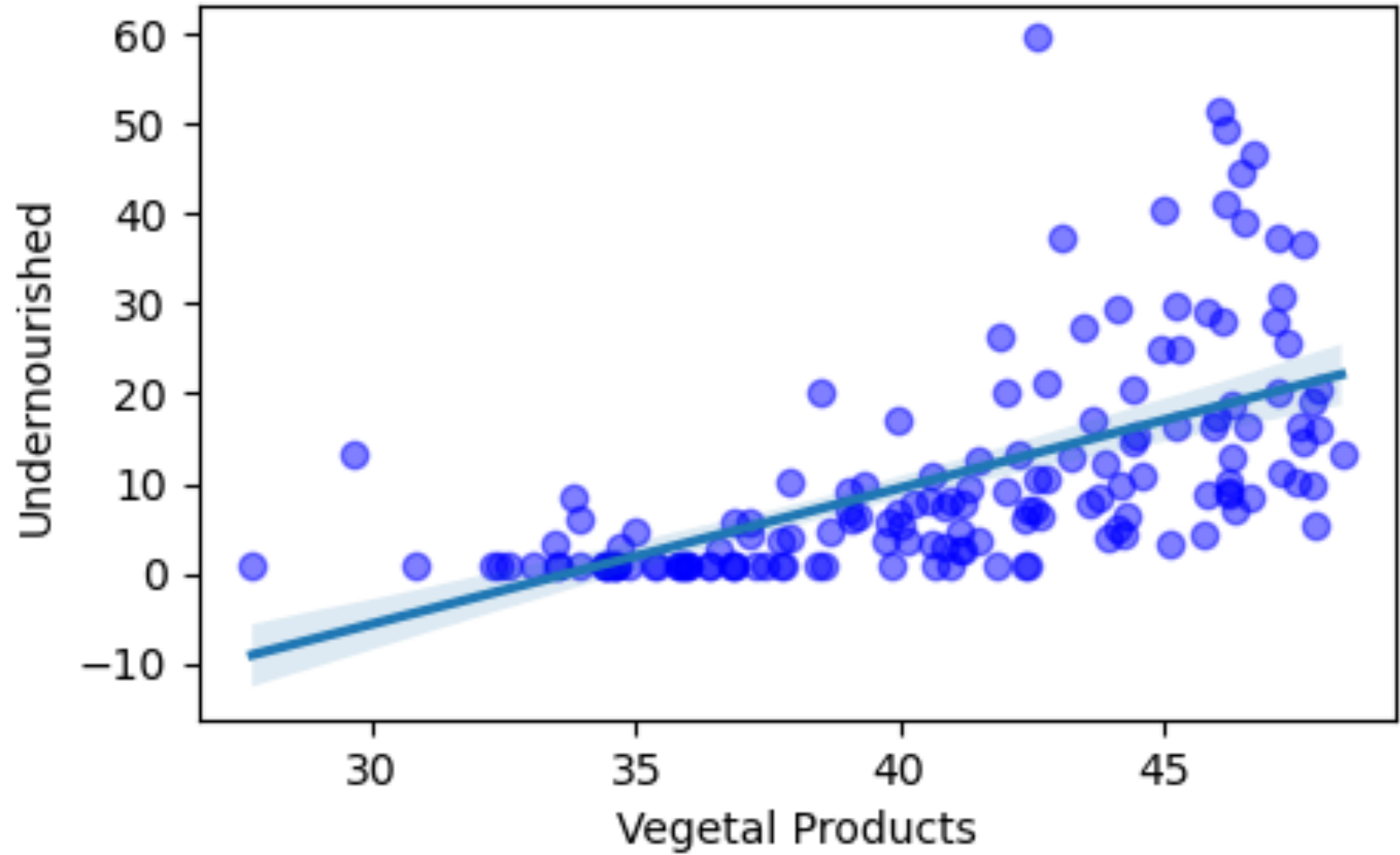
정규성 검정 결과 (Undernourished): Undernourished Statistic=0.7893, p-value=0.000000000000012983348  
정규성 검정 결과 (Stimulants): Stimulants Statistic=0.8019, p-value=0.0000000000000036255753  
등분산성 검정 결과: Statistic=109.1712, p-value=0.000000000000000000000000469388117  
Spearman검정 (Stimulants vs Undernourished) -> 상관계수: -0.5360 p-value: 0.0000000000000078277531

4. 영양결핍률과 섭취하는 음식 종류의 관계(Treenuts)



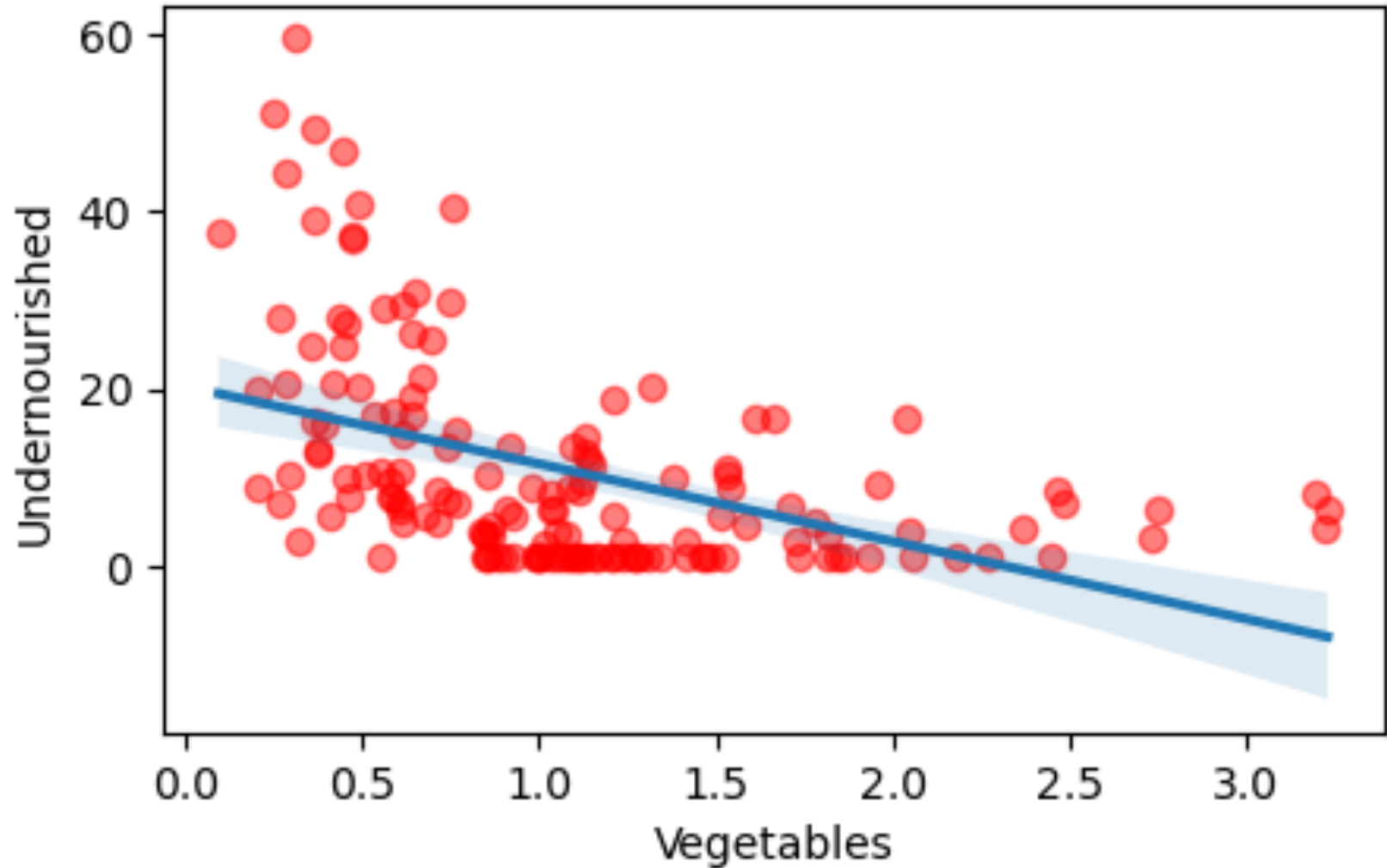
정규성 검정 결과 (Undernourished): Undernourished Statistic=0.7893, p-value=0.000000000000012983348  
정규성 검정 결과 (Treenuts): Treenuts Statistic=0.8157, p-value=0.000000000000118837253  
등분산성 검정 결과: Statistic=109.5890, p-value=0.000000000000000000000000401877064  
Spearman검정 (Treenuts vs Undernourished) -> 상관계수: -0.5552 p-value: 0.000000000000007899340

4. 영양결핍률과 섭취하는 음식 종류의 관계(Vegetal Products)



정규성 검정 결과 (Undernourished): Undernourished Statistic=0.7893, p-value=0.0000000000000012983348  
정규성 검정 결과 (Vegetal Products): Vegetal Products Statistic=0.9575, p-value=0.00011614019604166970  
등분산성 검정 결과: Statistic=28.3046, p-value=0.000000200618445949032710778353  
Spearman검정 (Vegetal Products vs Undernourished) -> 상관계수: 0.7533 p-value: 0.000000000000000000000000

4. 영양결핍률과 섭취하는 음식 종류의 관계(Vegetables)



정규성 검정 결과 (Undernourished): Undernourished Statistic=0.7893, p-value=0.000000000000012983348  
정규성 검정 결과 (Vegetables): Vegetables Statistic=0.9122, p-value=0.00000004967193945049  
등분산성 검정 결과: Statistic=102.1801, p-value=0.000000000000000000000006462438515  
Spearman검정 (Vegetables vs Undernourished) -> 상관계수: -0.5557 p-value: 0.00000000000000007390179

# 영양결핍률과 섭취하는 음식 종류의 관계

## 1. 결론

- 파란 그래프(영양결핍율과 양적 상관관계): 곡류, 식물성 제품
- 빨간 그래프(영양결핍율과 음적 상관관계): 동물성 음식(동물성 제품, 동물성 지방, 달걀, 고기, 우유, ), 각성제, 견과류, 야채
- 영양결핍율이 낮은 이유는 동물성 음식(동물성 제품, 동물성 지방, 달걀, 고기, 우유, ), 각성제, 견과류, 야채로부터의 kcal 섭취량이 적고 곡류, 식물성 제품의 섭취량은 많기 때문이라 추측할 수 있음

## 2. 시사점

- 다양한 식품군의 중요성 강조: 곡류와 식물성 제품 등의 과도한 섭취를 제한하고, 다양한 식품군의 균형 잡힌 섭취를 강조하는 정책 필요
- 지역적 특성 고려: 지역마다 음식 섭취 습관이 다르기 때문에 지역적 특성을 고려한 정책 및 교육 필요
- 영양 교육의 중요성 강조: 교육 기관과 정부는 영양 교육을 강화하여 국민이 올바른 식생활 습관을 형성할 수 있도록 지원해야 함

⇒ 이러한 결론과 시사점들을 고려하여 지역 사회의 영양 상태를 향상시키기 위한 종합적인 정책 및 교육이 이루어져야 함



# 05

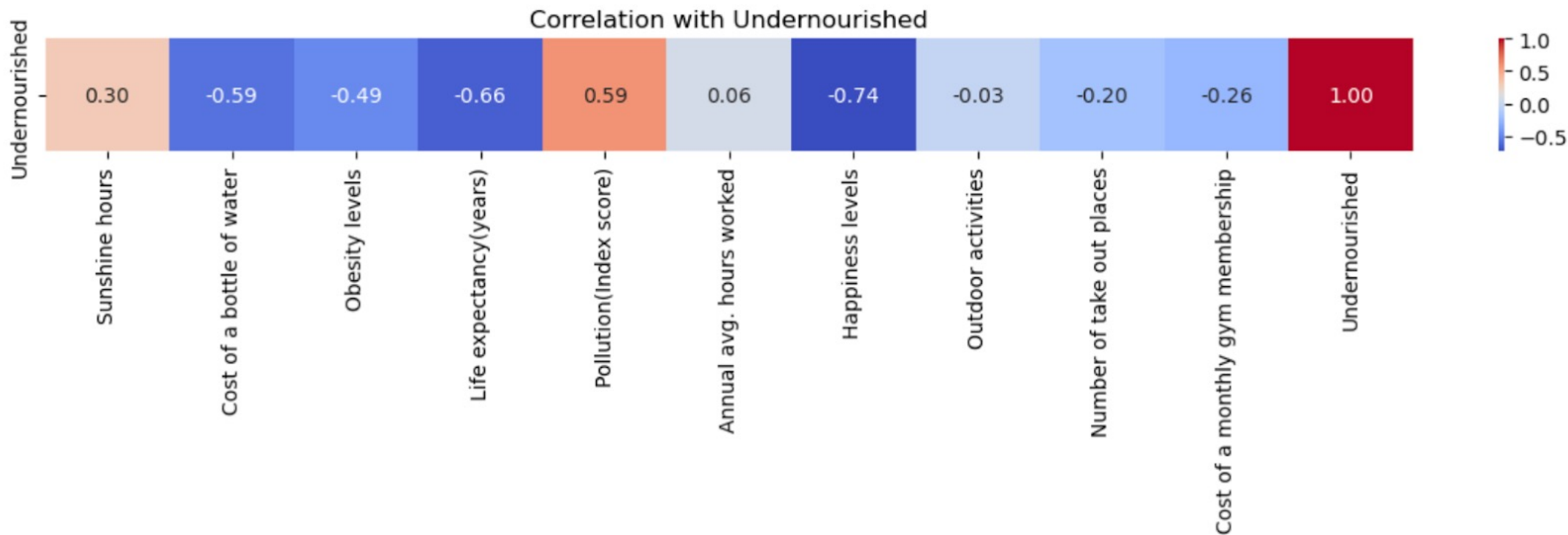
환경적 요인과 영양결핍률의 관계

5. 환경적 요인과 영양결핍률의 관계

① 환경적 요인과 영양결핍률의 상관관계 분석

```
# 영양결핍률의 관계가 더욱 돋보이도록 'Undernourished' 행만 선택
undernourished_corr = life_and_undernourished.corr()['Undernourished'].to_frame().T

plt.figure(figsize=(15, 1))
sns.heatmap(undernourished_corr, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation with Undernourished')
plt.show()
```



5. 환경적 요인과 영양결핍률의 관계

① 환경적 요인과 영양결핍률의 상관관계 분석

Significant Correlations:					
	Cost of a bottle of water	Life expectancy(years)	Pollution(Index score)	Happiness levels	Undernourished
Undernourished	-0.594109	-0.664252	0.58512	-0.744896	1.0

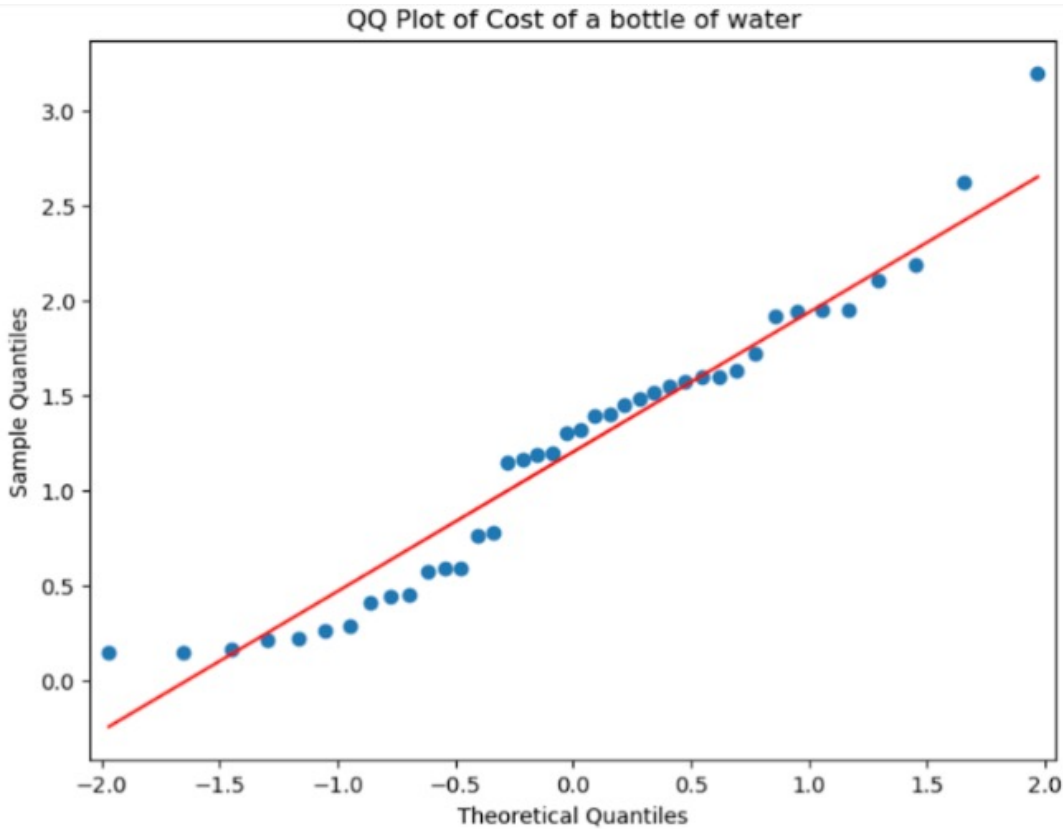
'Undernourished'와의 상관관계가 0.5보다 크거나 -0.5보다 작은 변수들  
즉, 상관관계가 높은 변수들만 선택

- ‘Cost of a bottle of water’
- ‘Life expectancy(years)’
- ‘Pollution(Index score)’
- ‘Happiness levels’

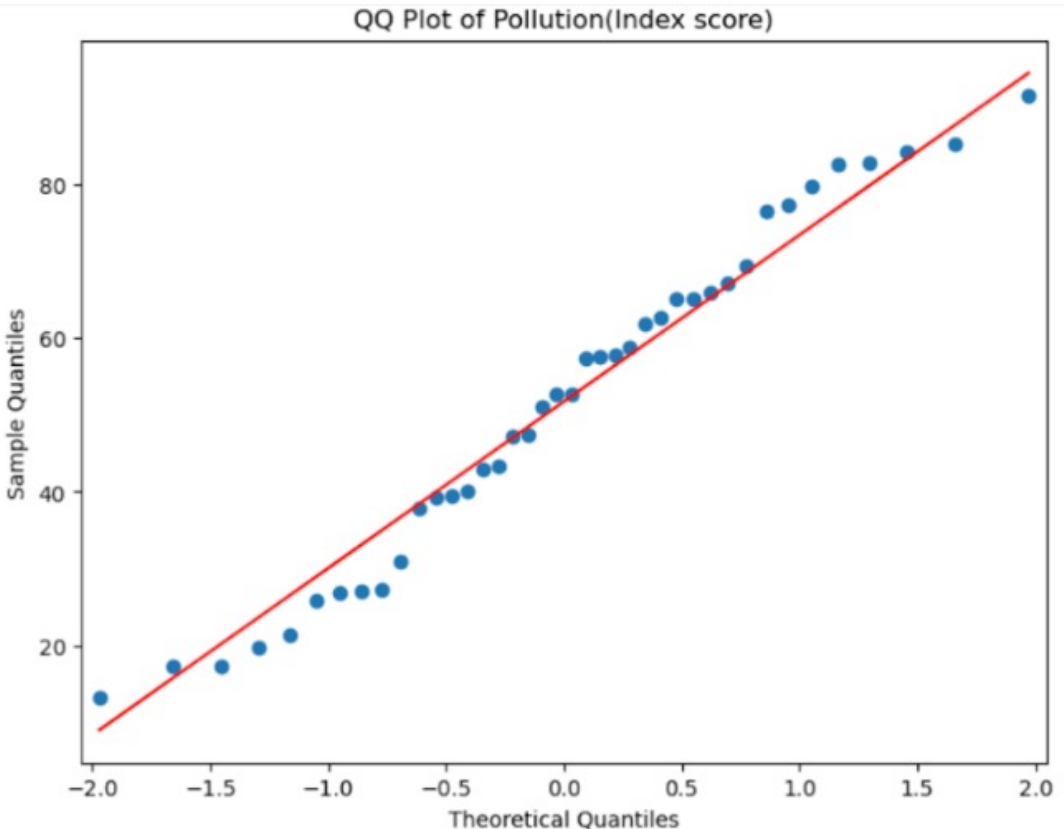
5. 환경적 요인과 영양결핍률의 관계

② 'Undernourished'와 상관관계가 높은 변수들 간의 오차항의 가정

- 정규성 검사(Q-Q plot)



Column: Cost of a bottle of water, Shapiro-Wilk 테스트 통계량: 0.9464963674545288, p-value: 0.057488057762384415



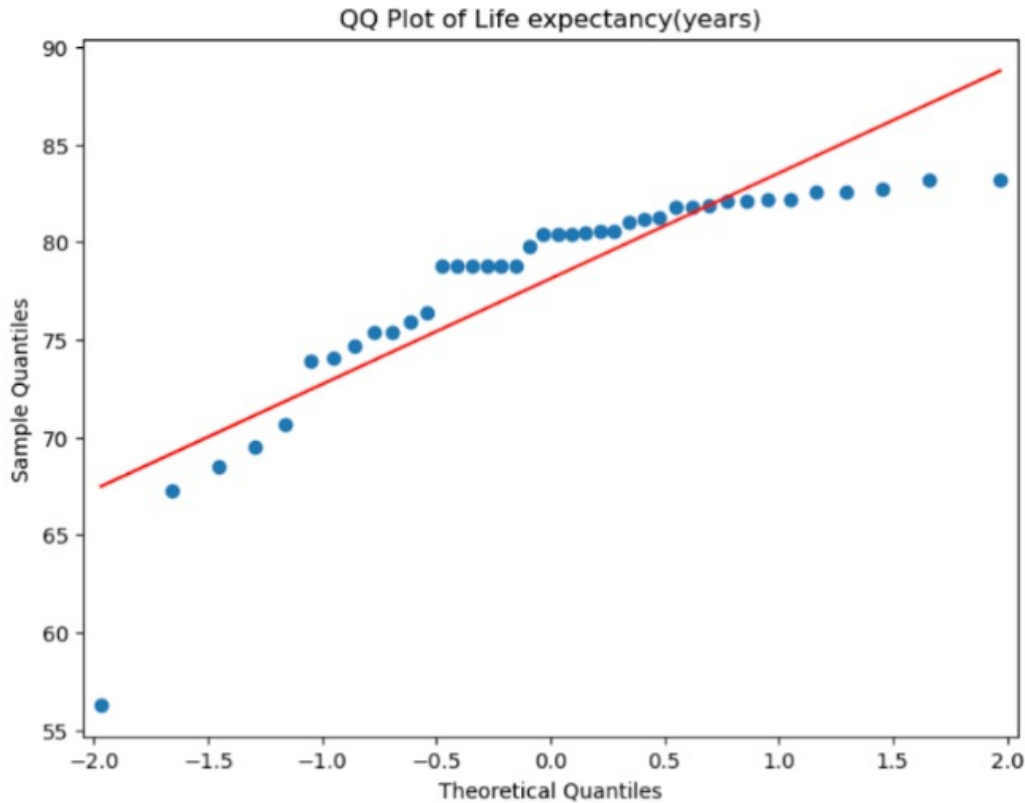
Column: Pollution(Index score), Shapiro-Wilk 테스트 통계량: 0.9643104672431946, p-value: 0.2344880849123001

⇒ 직선을 기준으로 데이터가 모여 있음

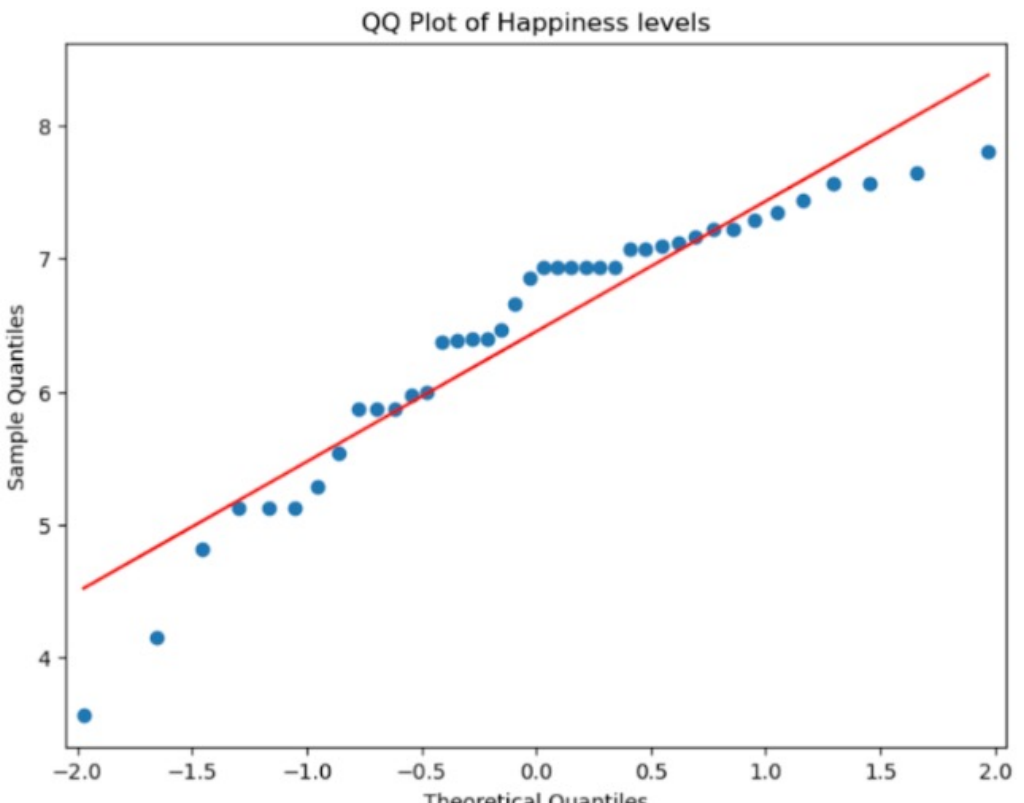
5. 환경적 요인과 영양결핍률의 관계

② 'Undernourished'와 상관관계가 높은 변수들 간의 오차항의 가정

- 정규성 검사(Q-Q plot)



Column: Life expectancy(years), Shapiro-Wilk 테스트 통계량: 0.7791827321052551, p-value: 2.564572696428513e-06



Column: Happiness levels, Shapiro-Wilk 테스트 통계량: 0.9068970084190369, p-value: 0.0030534756369888783

⇒ 직선을 기준으로 데이터가 약간 벗어나 있음

→ 정확한 판단을 위해 **Shapiro-Wilk** 테스트를 통해 **p-value** 값 도출

5. 환경적 요인과 영양결핍률의 관계

② 'Undernourished'와 상관관계가 높은 변수들 간의 오차항의 가정

- Shapiro-Wilk 테스트

Column: Cost of a bottle of water, p-value: 0.057488057762384415, Normality: 정규  
Column: Life expectancy(years), p-value: 2.564572696428513e-06, Normality: 비정규  
Column: Pollution(Index score), p-value: 0.2344880849123001, Normality: 정규  
Column: Happiness levels, p-value: 0.0030534756369888783, Normality: 비정규

‘Cost of a bottle of water’

‘Pollution(Index score)’

p-value가 0.05보다 큼 ⇒ 정규성을 따름

‘Life expectancy(years)’

‘Happiness levels’

p-value가 0.05보다 작음 ⇒ 정규성을 따르지 않음  
⇒ 비모수 테스트 진행

## 5. 환경적 요인과 영양결핍률의 관계

### ② 'Undernourished'와 상관관계가 높은 변수들 간의 오차항의 가정

- 비모수 테스트(Mann-Whitney U 검정)

'Undernourished'와 'Life expectancy(years)', 'Undernourished'와 'Happiness levels'

Mann-Whitney U 검정(Life expectancy(years)): 통계량=0.0, p-value=2.202472491963789e-15  
귀무가설을 기각하며, 'Undernourished'와 'Life expectancy(years)' 간에는 통계적으로 유의미한 차이가 있다.

Mann-Whitney U 검정(Happiness levels): 통계량=217.5, p-value=7.824605827737411e-09  
귀무가설을 기각하며, 'Undernourished'와 'Happiness levels' 간에는 통계적으로 유의미한 차이가 있다.

두 p-value 모두 0.05보다 작음

⇒ 정규성을 따르지는 않지만 통계적으로 유의미한 차이가 있다고 할 수 있음

## 5. 환경적 요인과 영양결핍률의 관계

### ② 'Undernourished'와 상관관계가 높은 변수들 간의 오차항의 가정

#### - 등분산성 검사

등분산성 검정(Cost of a bottle of water): Statistic=2.8793, p-value=0.09371036753678681219  
귀무가설을 기각하지 못하며, 'Undernourished'와 'Cost of a bottle of water' 간에는 등분산성이 있다고 할 수 있다.

등분산성 검정(Pollution(Index score)): Statistic=77.7544, p-value=0.00000000000024355064  
귀무가설을 기각하며, 'Undernourished'와 'Pollution(Index score)' 간에는 등분산성이 없다고 할 수 있다.

등분산성 검정(Life expectancy(years)): Statistic=5.8280, p-value=0.01811651131310790253  
귀무가설을 기각하며, 'Undernourished'와 'Life expectancy(years)' 간에는 등분산성이 없다고 할 수 있다.

등분산성 검정(Happiness levels): Statistic=1.8162, p-value=0.18166258330400680832  
귀무가설을 기각하지 못하며, 'Undernourished'와 'Happiness levels' 간에는 등분산성이 있다고 할 수 있다.

⇒ p-value가 0.05보다 큰 'Cost of a bottle of water'와 'Happiness levels'은 등분산성이 있다고 할 수 있음  
p-value가 0.05보다 작은 'Life expectancy(years)', 'Pollution(Index score)'은 등분산성이 없다고 할 수 있음

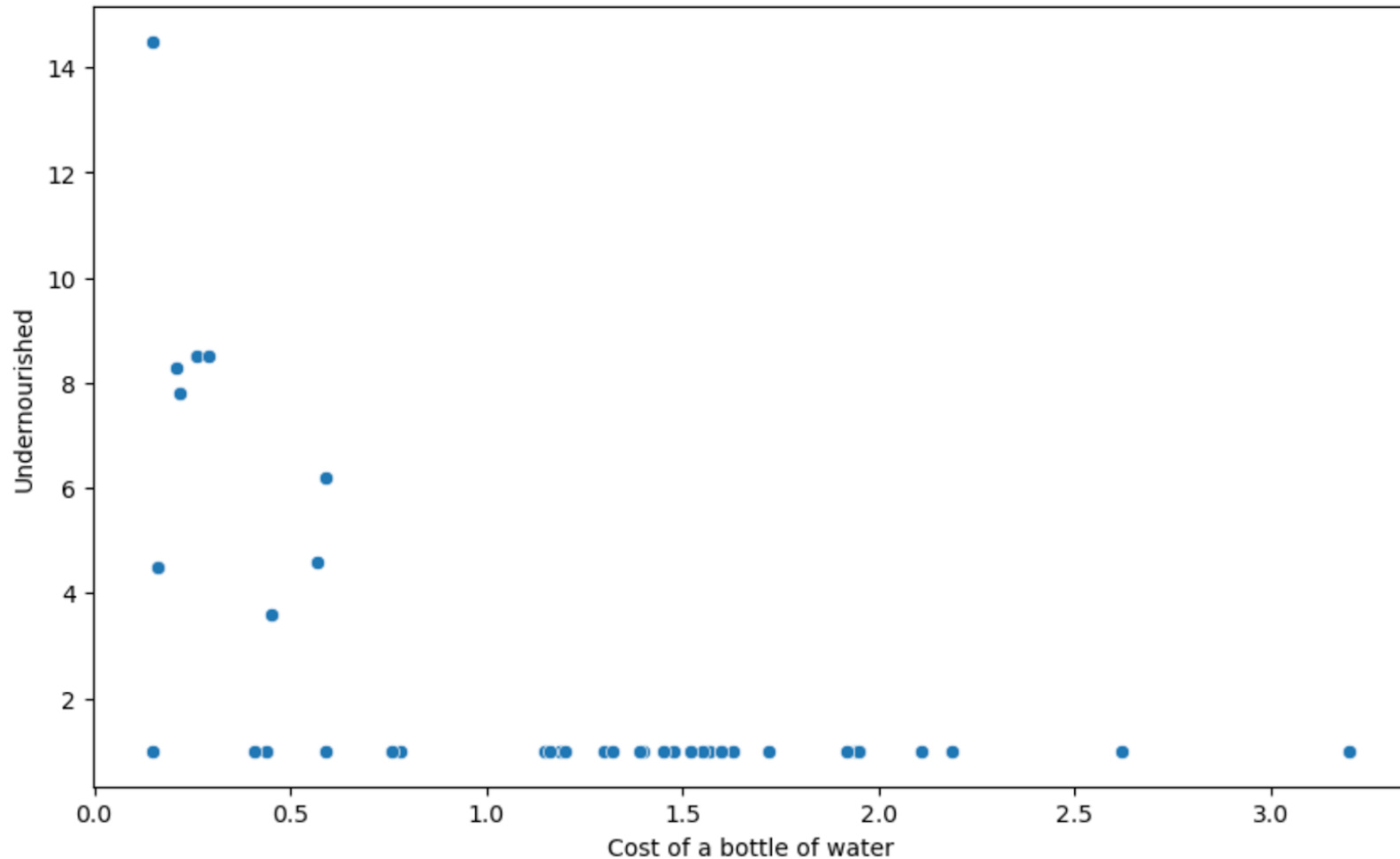


## 5. 환경적 요인과 영양결핍률의 관계

### ③ 'Undernourished'과 'Cost of a bottle of water'의 분석

#### 1) 산점도 분포 확인

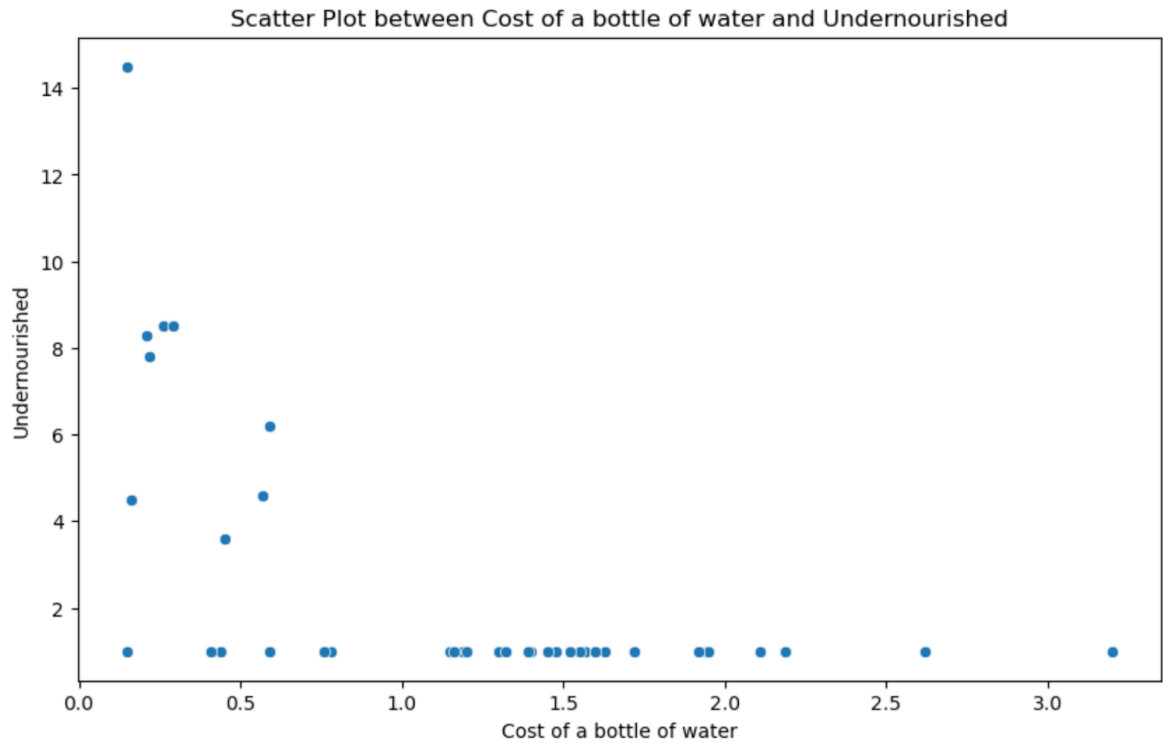
정규성과 등분산성을 모두 만족하는 변수 = Cost of a bottle of water



5. 환경적 요인과 영양결핍률의 관계

③ 'Undernourished'과 'Cost of a bottle of water'의 분석

1) 산점도 분포 확인



점들의 분포가 비교적 왼쪽 중반부에서 오른쪽 하단으로 하강하는 대각선 형태를 볼 수 있음  
Cost of a bottle of water을 나타내는 x축의 0.0과 1.0 사이에는 2 이상의 Undernourished의 점들이 밀집  
⇒ Scatter plot의 분포가 뚜렷하지 않아 **상관계수 검정** 진행

## 5. 환경적 요인과 영양결핍률의 관계

### ③ 'Undernourished'과 'Cost of a bottle of water'의 분석

#### 2) 피어슨 상관계수 검정

정규성과 등분산성을 모두 만족하는 변수 = Cost of a bottle of water

Pearson 상관계수:  $-0.5941$  p-value:  $0.00005307035307877404$



음의 상관관계



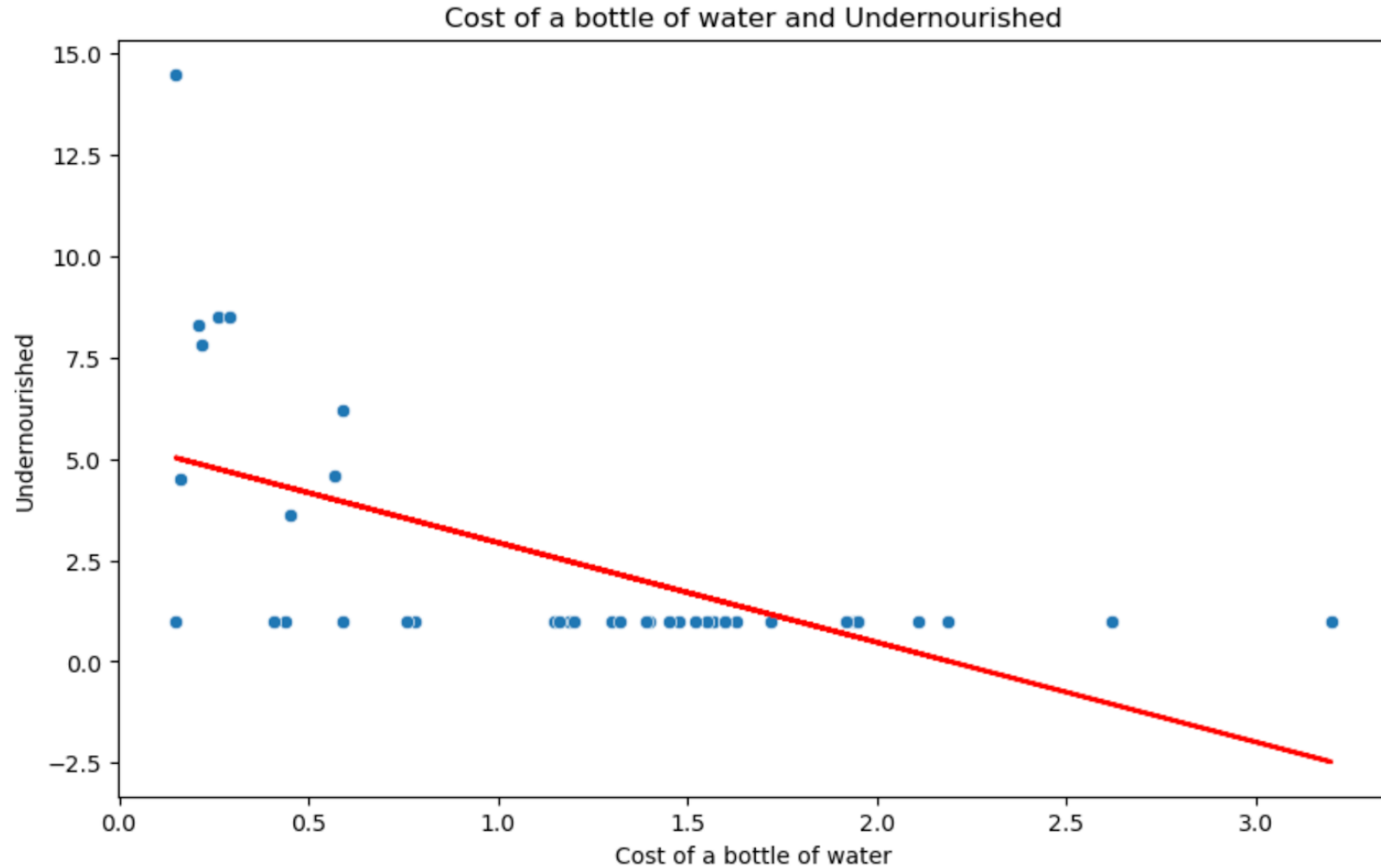
0.05보다 작으므로  
귀무가설 기각

⇒ 'Undernourished'와 'Cost of a bottle of water' 간에는 유의미한 선형 관계가 있음

## 5. 환경적 요인과 영양결핍률의 관계

### ③ 'Undernourished'과 'Cost of a bottle of water'의 분석

### 3) 선형회귀분석



# 환경적 요인과 영양결핍률의 관계

## 1. 결론 및 시사점

- 'Cost of a bottle of water'와 영양결핍률은 음의 상관관계를 보이며, 물의 가격이 높을수록 영양결핍률이 감소하는 경향이 나타남.
- 물의 가격 인하는 영양결핍률을 낮추는 데 중요한 역할을 할 수 있음.
- 추가적인 연구에서는 다양한 변수를 고려하여 모델을 개선하고 결과의 신뢰성을 높일 필요가 있음.

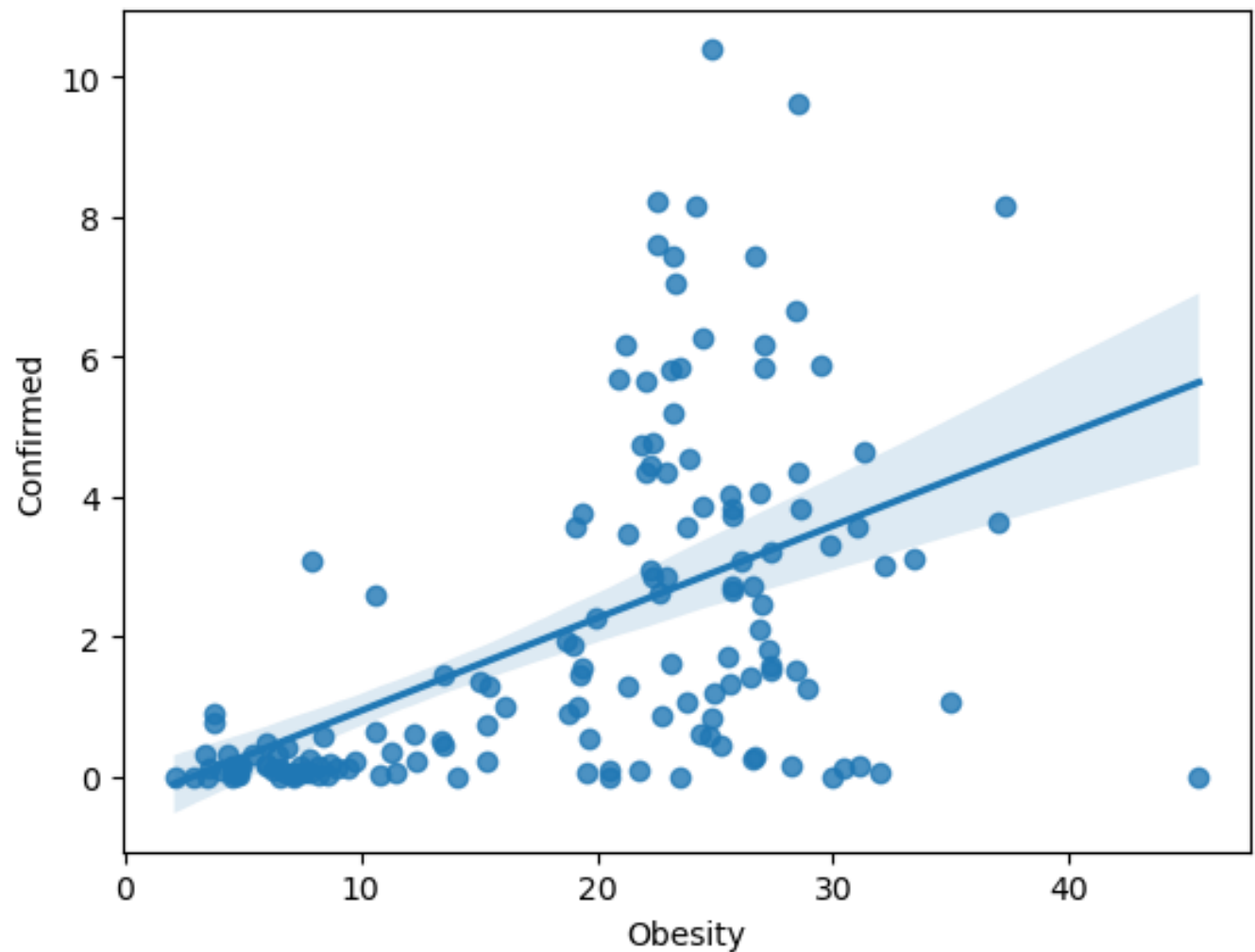
## 2. 한계 및 주의사항

- 데이터 한계: 결과는 관측된 데이터에 기반하며 인과관계를 단정하기에는 한계가 있음.
- 검정 결과의 한계: 정규성과 등분산성 검정 결과는 표본 크기 등에 영향을 받을 수 있음, 모델의 한계와 함께 정확한 해석이 요구됨.

# 06

COVID-19 확진과 비만율의 관계

6. COVID-19 확진과 비만율의 관계



정규성 검정 결과 (Confirmed): Statistic=0.8204, p-value=0.000000000000179627016

정규성 검정 결과 (Obesity): Statistic=0.9333, p-value=0.00000127858925225155

등분산성 검정 결과: Statistic=160.8276, p-value=0.000000000000000000000000

Spearman검정 -> 상관계수: 0.5723 p-value: 0.00000000000000000897039

# COVID-19 확진과 비만율의 관계

## 1. 통계적 유의성

- 상관계수와 P-value: 분석에서 사용된 상관계수 및 P-value 모두 유의수준에서 통계적으로 의미 있음.

## 2. 결론 및 시사점

비만율과 코로나 확진율 간에 양의 상관관계가 있음을 확인할 수 있다. 이는 비만과 코로나 간의 상호작용이 있을 수 있으며, 비만율이 높은 지역에서 코로나 확진율도 높을 수 있다는 가능성을 시사함.

이를 통해 정부는 비만율이 높은 지역에서의 예방 및 건강 교육을 강화해야 할 것임.

건강한 생활습관과 비만 관리는 코로나 및 기타 감염병의 예방에 중요한 역할을 할 수 있을 것으로 예측됨.

## 3. 한계 및 주의사항

- 데이터 한계: 결과는 관측된 데이터에 기반하며 다양한 요인들이 결과에 영향을 미칠 수 있음.
- 다양한 요인의 영향: 다양한 요인들이 결과에 영향을 미치므로, 향후 연구에서 이를 고려한 추가적인 분석이 필요.

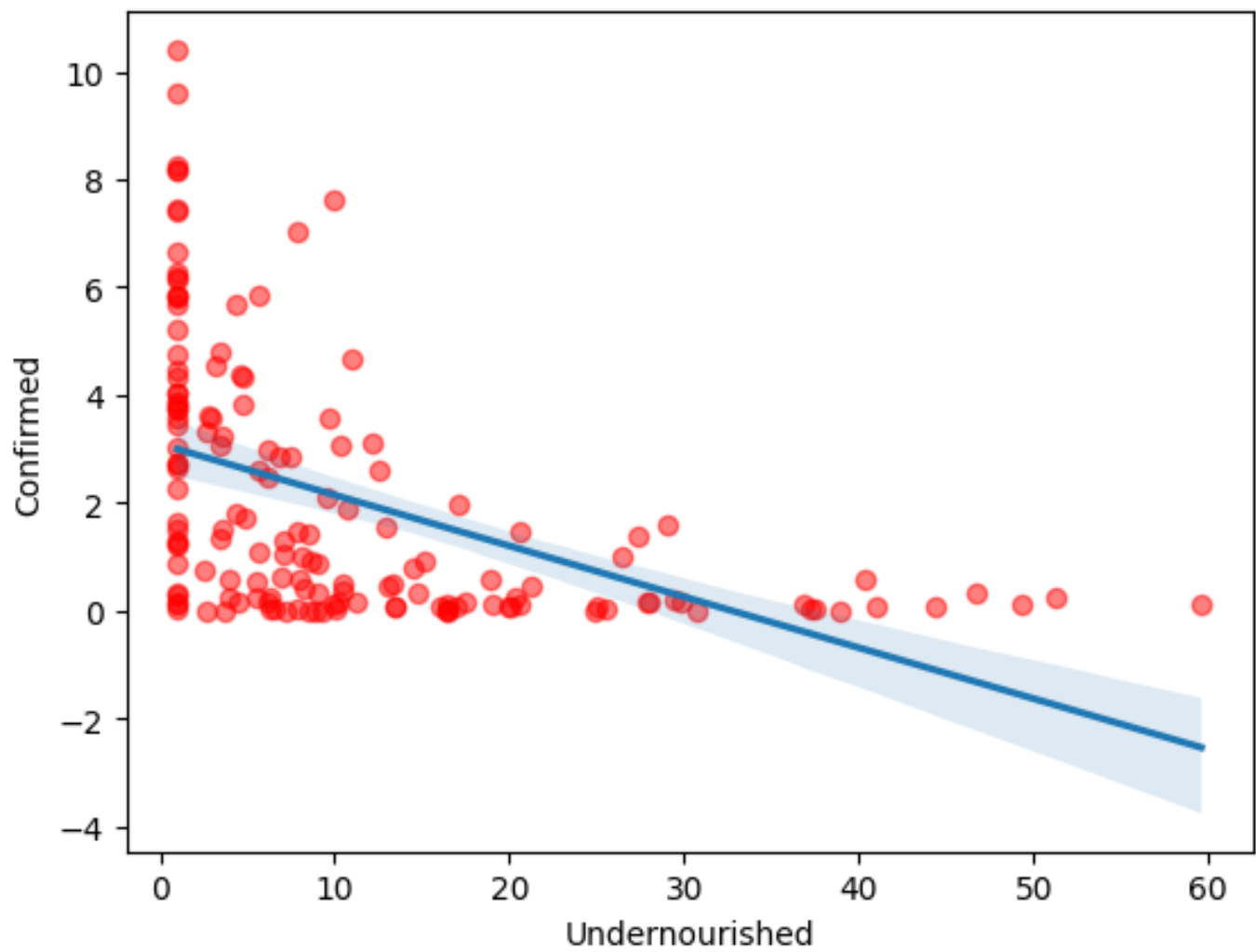


# 07

영양결핍률과 COVID-19 치명률, 사망률의 관계

7. 영양결핍률과 COVID-19 치명률, 사망률의 관계

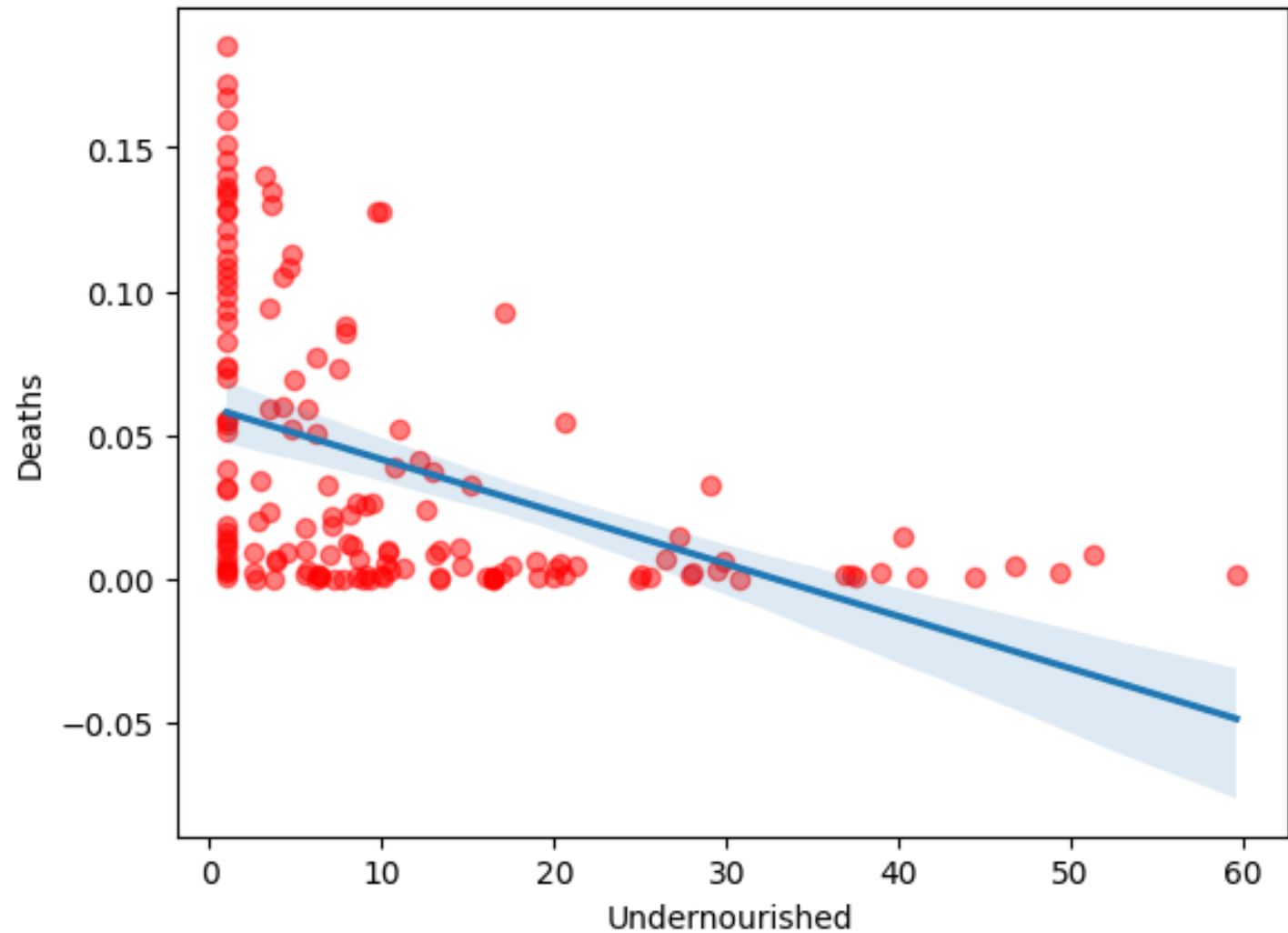
영양결핍률과  
COVID-19 확진율의 관계



- 정규성 검정 결과 (Undernourished): Statistic=0.7893, p-value=0.00000000000012983348
- 정규성 검정 결과 (Confirmed): Statistic=0.8204, p-value=0.00000000000179627016
- Levene 등분산성 검정 결과 (Confirmed): Statistic=68.0083, p-value=0.000000000000000486757
- Fligner 등분산성 검정 결과 (Confirmed): Statistic=95.6167, p-value=0.000000000000000000000
- Spearman 검정 결과 (Confirmed): 상관계수=-0.6184, p-value=0.000000000000000001275

7. 영양결핍률과 COVID-19 치명률, 사망률의 관계

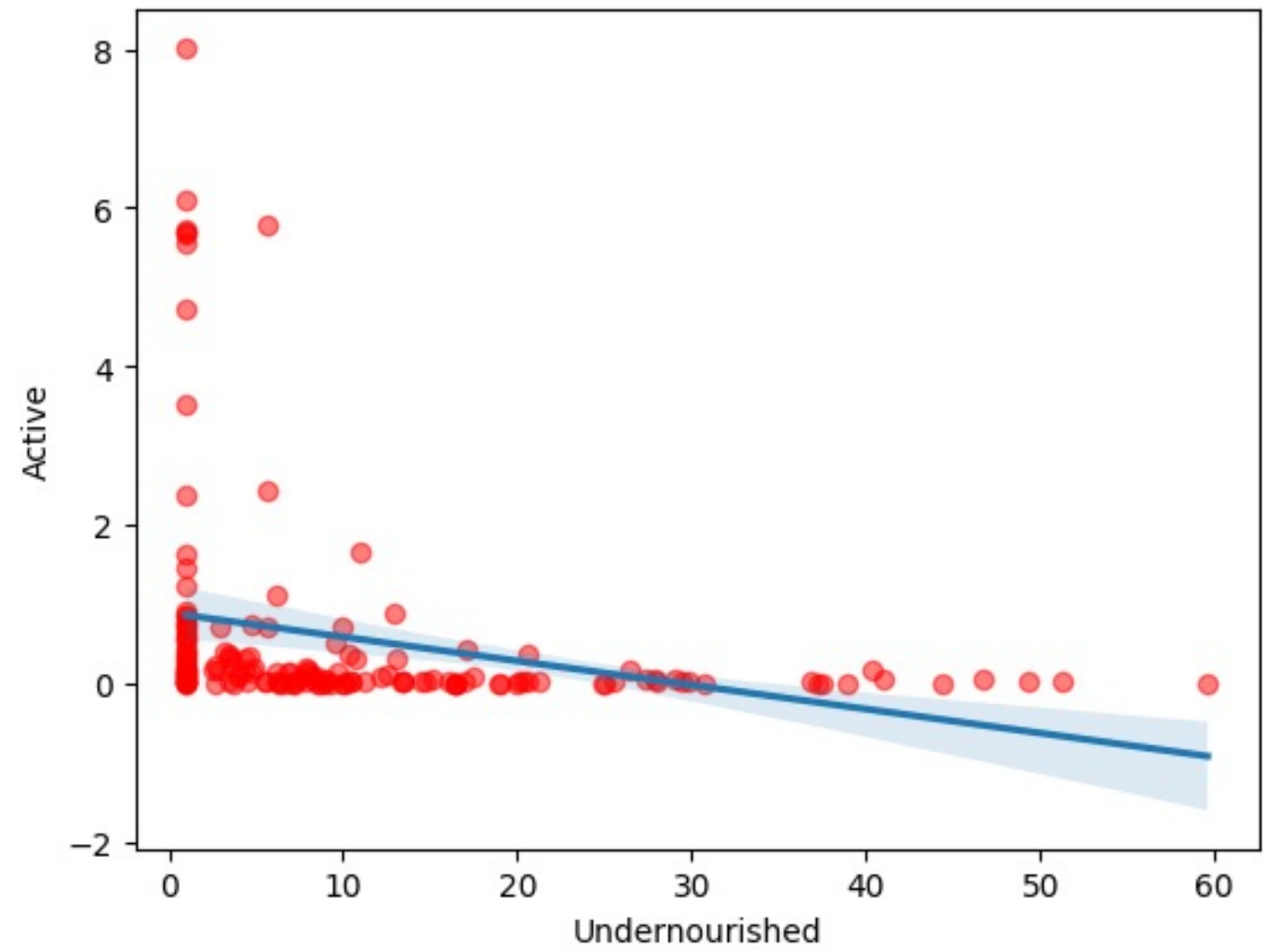
영양결핍률과  
COVID-19 사망률의 관계



- 정규성 검정 결과 (Undernourished): Statistic=0.7893, p-value=0.00000000000012983348
- 정규성 검정 결과 (Deaths): Statistic=0.7873, p-value=0.00000000000011030987
- Levene 등분산성 검정 결과 (Deaths): Statistic=114.3234, p-value=0.00000000000000000000
- Fligner 등분산성 검정 결과 (Deaths): Statistic=189.4837, p-value=0.00000000000000000000
- Spearman 검정 결과 (Deaths): 상관계수=-0.5689, p-value=0.000000000000001384265

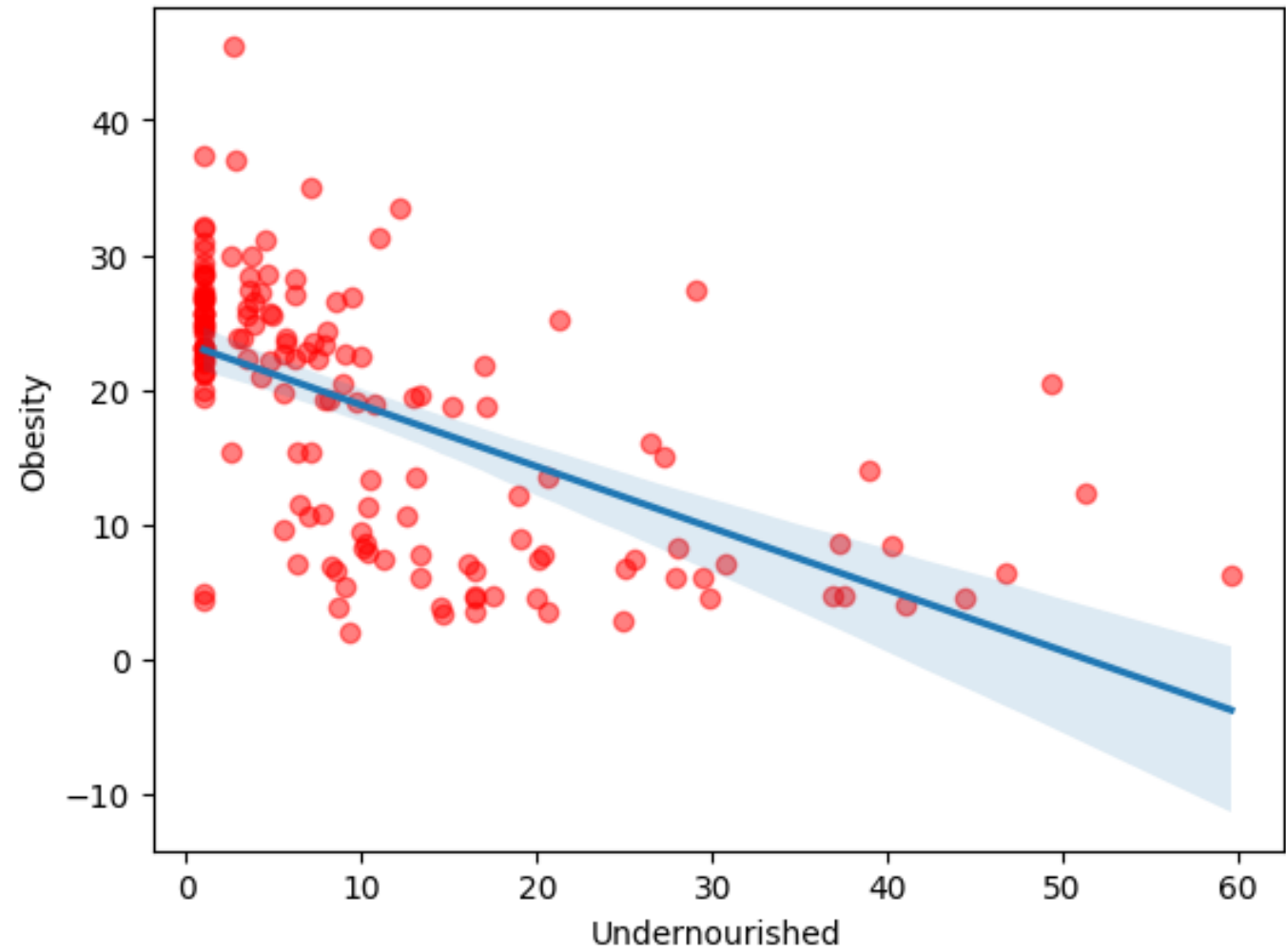
7. 영양결핍률과 COVID-19 치명률, 사망률의 관계

영양결핍률과  
COVID-19 치명률의 관계



- 정규성 검정 결과 (Undernourished): Statistic=0.7893, p-value=0.000000000000012983348
- 정규성 검정 결과 (Active): Statistic=0.4380, p-value=0.000000000000000000000
- Levene 등분산성 검정 결과 (Active): Statistic=99.0013, p-value=0.000000000000000000002
- Fligner 등분산성 검정 결과 (Active): Statistic=159.4200, p-value=0.000000000000000000000
- Spearman 검정 결과 (Active): 상관계수=-0.5488, p-value=0.000000000000017284293

# 영양결핍률과 비만율의 관계



- 정규성 검정 결과 (Undernourished): Statistic=0.7893, p-value=0.000000000000012983348
- 정규성 검정 결과 (Obesity): Statistic=0.9333, p-value=0.00000127858925225155
- Levene 등분산성 검정 결과 (Obesity): Statistic=0.0834, p-value=0.77287661809139818203
- Fligner 등분산성 검정 결과 (Obesity): Statistic=0.2724, p-value=0.60171574891311285782
- Spearman 검정 결과 (Obesity): 상관계수=-0.6572, p-value=0.0000000000000000000002

# 데이터 분석 결과

## 1. 결론 및 시사점

영양결핍률과 비만율을 식사와 환경적 요인으로 해결하고, 질병에 적극 대응한다면 건강한 환경이라는 인식을 줄 수 있고, 이는 행복수준을 향상시키는데 영향을 끼칠 수 있을 것이다.

## 2. 한계

- 데이터 한계: 결과는 관측된 데이터에 기반하며 다양한 요인들이 결과에 영향을 미칠 수 있음.
- 다양한 요인의 영향: 다양한 요인들이 결과에 영향을 미치므로, 향후 연구에서 이를 고려한 추가적인 분석이 필요.

# 참고문헌

1. <https://www.kaggle.com/code/ydalat/work-life-balance-survey-eda>
2. <https://www.kaggle.com/code/yadhua/happiness-life-prediction-and-clustering>
3. <https://www.kaggle.com/code/formeforu/team-4-eda-for-covid-19-health>