

# 지구온난화의 원인에 목축업/낙농업?

## 1. 서론

### A. 자료 조사

몇 년째, 심각해지고 있는 지구온난화. 과연 원인이 무엇인가?

기사: ['화석 연료, 사막화, 화학 비료'... 지구 온난화의 5 가지 원인] 지구 온난화의 핵심 원인 중 하나로 목축업/낙농업으로 인한 화학 비료 사용 및 메탄 가스 배출이라고 지적하고 있다.

출처: <https://news.joins.com/article/19454330>

### B. 참고 연구 방향

아래의 주소의 연구자의 지구온난화 연구 방법을 보면 알 수 있다시피, '지구 공전 궤도', '태양의 온도', '화산 활동', '미세먼지', '산림벌채', '오존' 등은 지구 온난화와 큰 상관관계를 가지 않는다. 그의 예로, 그래프가 같은 방향으로 진행하고 있지 않다. 출처:

출처: <https://brunch.co.kr/@thyoon/1>

하지만, '온실 가스'의 경우 그래프가 지구온난화의 지표와 동일한 방향으로 진행되고 있음을 알 수 있다. 이와 같이, 이번 발제에서도 유제품 생산량을 통해서 지구 온도변화와 얼마나 유사한 추이를 갖는지 알아보자!

## 2. 연구 계획

### A. 주제

지구온난화의 원인으로 주목되고 있는 목축업/낙농업이 얼마나 상관관계를 갖고 있는지 알아보자!

### B. 연구 방법

- i. 유제품 생산량, 온실가스 배출량, 지구온도 변화 데이터 수집
- ii. 각 데이터의 추이를 그래프로 표현
- iii. 서로 어떤 상관관계를 갖는지 분석. 얼마나 비슷한 추이를 보이는지 비교! Ols 회귀분석을 이용해보자! Coefficient 가 양수라면 서로 영향력을 갖는다는 것을 알 수 있겠다.

Ex) 유제품 생산량이 증가하면, 온실 가스 배출량도 증가하는지, 지구 온도는 올라갔는지

### C. 연구 데이터

- i. 유제품 생산량
- ii. 온실가스 배출량
- iii. 지구 평균 온도 추이

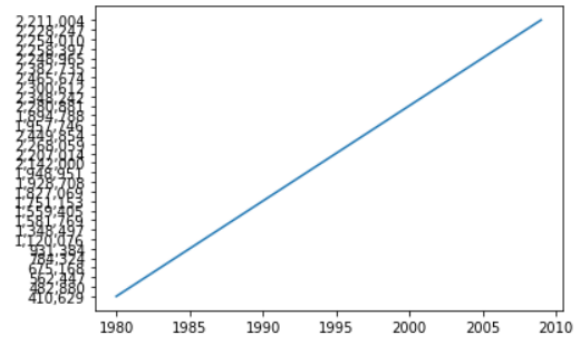
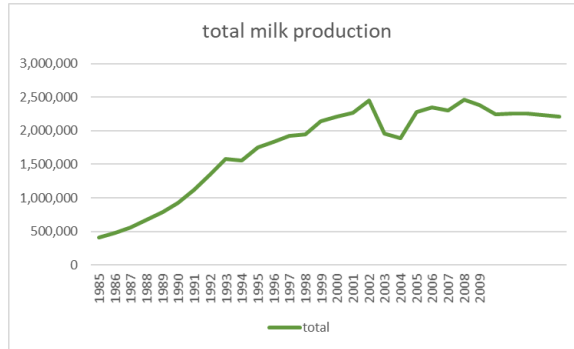
### D. 연구의 한계점

다른 요소를 제외하지 않고 회귀 분석을 할 경우 왜곡될 수 있지 않을까?  
통제변수?를 넣어야 하지 않을까 생각이 든다. 하지만, 아직 통제변수 다루는 법을 모르니... 일단 해보자! 결론적으로는, 목축업/낙농업이 지구온난화의 기여하는 부분이 있을 수 있지만, 그 정도를 왜곡할 수도 있다!

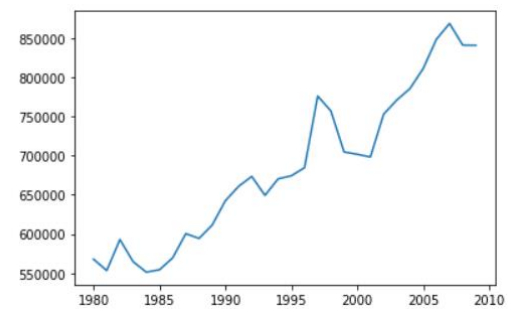
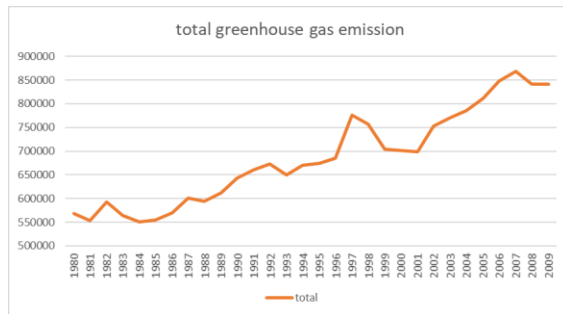
### 3. 본론

#### A. 데이터 그래프로 표현

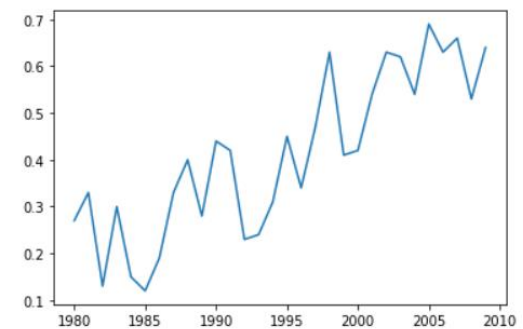
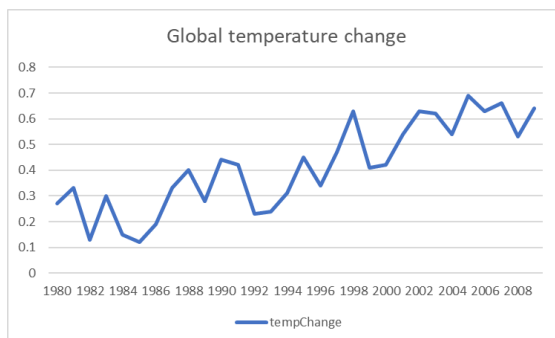
##### i. 유제품 생산량



##### ii. 온실가스 배출량



##### iii. 지구 평균 온도



## B. OLS 회귀분석

- i. 먼저, 각 데이터를 하나의 파일로 합친 후, 표로 나타낸다.

```
▶ import pandas as pd
from sklearn import datasets, linear_model
import matplotlib.pyplot as plt
import numpy as np
import statsmodels.api as sm
import statsmodels.formula.api as smf

totalData = pd.read_csv("csv/total.csv")
totalData
```

]:

	year	temperature	gas	milk
0	1980	0.27	567777.0807	410,629
1	1981	0.33	553345.5501	482,880
2	1982	0.13	592862.1000	562,447
3	1983	0.30	564652.2583	675,168
4	1984	0.15	551310.5385	784,324
5	1985	0.12	554318.2821	931,384
6	1986	0.19	569470.6118	1,120,076
7	1987	0.33	600408.9004	1,348,497
8	1988	0.40	594323.3854	1,581,769
9	1989	0.28	611381.4628	1,559,405
10	1990	0.44	642440.8789	1,751,153
11	1991	0.42	660724.9247	1,827,069
12	1992	0.23	673408.0558	1,928,708

```
In [12]: > model = smf.ols(formula = 'temperature ~ gas', data = totalData)
> result = model.fit()
> result.summary()
```

Out[12]: OLS Regression Results

Dep. Variable:	temperature	R-squared:	0.750			
Model:	OLS	Adj. R-squared:	0.741			
Method:	Least Squares	F-statistic:	84.06			
Date:	Tue, 14 Jan 2020	Prob (F-statistic):	6.31e-10			
Time:	17:00:49	Log-Likelihood:	31.625			
No. Observations:	30	AIC:	-59.25			
Df Residuals:	28	BIC:	-56.45			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	-0.6103	0.113	-5.422	0.000	-0.841	-0.380
gas	1.49e-06	1.62e-07	9.169	0.000	1.16e-06	1.82e-06
Omnibus:	4.236	Durbin-Watson:	1.878			
Prob(Omnibus):	0.120	Jarque-Bera (JB):	1.665			
Skew:	-0.113	Prob(JB):	0.435			
Kurtosis:	1.868	Cond. No.	4.89e+06			

Warnings:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 4.89e+06. This might indicate that there are strong multicollinearity or other numerical problems.

ii. OLS 회귀 분석을 해준다.

- ➔ 회귀모형의 F 값이 84.06 유의확률( $Pr>F$ )은 0 으로 통계적으로 유의한 모형이라 할 수 있다.
- ➔ R-squared 가 0.75 즉, 75%의 설명력을 갖는다. 회귀모형의 유용성은 높은 편이라는 것을 알 수 있다.
- ➔ “[2] The condition number is large, 4.89e+06. This might indicate that there are strong multicollinearity or other numerical problems”를 보아하니, 다중공산성의 문제가 있는 것 같다. 다중공산성이란? 일부 독립변수(예측변수, 회귀자; regressor)가 다른 독립변수와 높은 상관관계를 가질 때 또는 그럴 때 발생하는 문제를 의미한다. 해결법?

[https://m.blog.naver.com/PostView.nhn?blogId=cto\\_hwangga&logNo=220969542325&proxyReferer=https%3A%2F%2Fwww.google.com%2F](https://m.blog.naver.com/PostView.nhn?blogId=cto_hwangga&logNo=220969542325&proxyReferer=https%3A%2F%2Fwww.google.com%2F) 여기를 참고해보자!

iii. OLS 회귀 분석 - milk

<b>Dep. Variable:</b>	temperature	<b>R-squared:</b>	0.409
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.065
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	1.188
<b>Date:</b>	Wed, 15 Jan 2020	<b>Prob (F-statistic):</b>	0.378
<b>Time:</b>	17:26:51	<b>Log-Likelihood:</b>	16.538
<b>No. Observations:</b>	20	<b>AIC:</b>	-17.08
<b>Df Residuals:</b>	12	<b>BIC:</b>	-9.111
<b>Df Model:</b>	7		
<b>Covariance Type:</b>	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
<b>Intercept</b>	0.4400	0.137	3.220	0.007	0.142	0.738
<b>milk[T.1,800,000]</b>	-0.0250	0.167	-0.149	0.884	-0.390	0.340
<b>milk[T.1,900,000]</b>	-0.0100	0.167	-0.060	0.953	-0.375	0.355
<b>milk[T.2,000,000]</b>	-0.2000	0.193	-1.035	0.321	-0.621	0.221
<b>milk[T.2,100,000]</b>	0.0300	0.167	0.179	0.861	-0.335	0.395
<b>milk[T.2,200,000]</b>	0.0350	0.153	0.229	0.823	-0.298	0.368
<b>milk[T.2,300,000]</b>	0.1620	0.150	1.082	0.300	-0.164	0.488
<b>milk[T.2,400,000]</b>	0.1000	0.158	0.634	0.538	-0.244	0.444

<b>Omnibus:</b>	0.617	<b>Durbin-Watson:</b>	1.731
<b>Prob(Omnibus):</b>	0.734	<b>Jarque-Bera (JB):</b>	0.337
<b>Skew:</b>	0.306	<b>Prob(JB):</b>	0.845
<b>Kurtosis:</b>	2.828	<b>Cond. No.</b>	14.0

➔ 이 표는 말썽쟁이 milk 의 데이터를 좀 수정해서, 1990~2009 년까지의 데이터로만 회귀분석을 돌려본 것인데... 보다시피 이상하게 나와요  $\pi\pi$  왜그런지는 저도 모르겠습다....

## 4. 결론

### A. 그래프를 그려보아서 비교해보자!

Milk graph 를 python 으로 그리는 것은 실패했지만! Greenhouse gas 와 temperature 은 그래프끼리 서로 비슷한 양상을 띄고 있음을 알 수 있다.

### B. 회귀식

$$Y_{temperature} = -0.6103 + 0.00000149 * greenhousegas$$

이게 얼마나 유의미한가...는 솔직히 잘 모르겠음 temperature 자체가 30 도 이런걸 나타나는게 아니라 작년 대비 얼마나 증가했는가 0.3 증가함! 이걸 나타내는 거라서 greenhousegas 의 베타값이 0.000000149 인게 과연 큰 값인가?!

### C. 원래는 milk 생산량에 대한 베타값을 구하고 싶었는데... jupyter notebook 을 보면 알다시피 milk 의 데이터만...이상하다!! 흑흑 $\pi\pi$ 이상해요 왜 애만 안되는지 모르겠네요 $\pi\pi$ 우리 모두 다같이! 문제점을... 찾아보길 바라요...