

<https://doi.org/10.7236/IIBC.2018.18.1.203>

IIBC 2018-1-27

# 빅 데이터 분석을 활용한 회귀분석 기반 식품 첨가물과 암 발생 관계

## Relation between Regression-based Food Additives and Carcinogenesis using Big Data Analysis

이국형\*, 이준희\*\*, 김정준\*, 김영곤\*

Lee Kuk Hyung\*, Jun-Hee Lee\*\*, Jeong-Joon Kim\*, Young-Gon Kim\*

**요약** 최근 웰빙(Well-Bing) 열풍에 힘입어 화학적으로 가공되지 않은 천연 식품을 찾는 소비 형태가 많이 부각되고 있다. 하지만 현대 시대를 살아감에 있어 화학 식품 첨가물이 포함되지 않은 식품을 주변에서 찾기는 매우 힘들다. 또한 가공 조리된 식품에는 화학 식품 첨가물의 성분이 제대로 표기되어 있지 않거나 표기가 생략되어 나오는 제품이 상당수다. 본 논문은 한국의 연간 화학 식품 첨가물 소비량 데이터를 기반으로 첨가물 별 대두 되고 있는 암 발생을 빅 데이터 분석 기법 중 하나인 시계열 분석과 회귀 분석을 통해 검증 할 것이며, 이러한 식품 첨가물이 사회의 어떠한 요인 때문에 소비가 증가했는지에 대한 빅 데이터 분석을 실시해 검증한다.

**Abstract** Well-Bing Many consumer forms of natural foods that are not chemically processed due to heat are becoming popular. However, it is very difficult to find foods that do not contain chemical food additives in the modern. Also, there are many products that do not have proper marking of ingredients of chemical food additives or are omitted from marking. Based on the annual data on chemical food additive consumption in Korea, this paper will examine the occurrence of cancer, which is emerging by additives, through time series analysis and regression analysis, one of the big data analysis techniques. Of the total population.

**Key Words** : Big data, Cancer, Food additives, Regression analysis, Time series analysis

### 1. 서 론

오늘날 식생활이 간편해지고 식품 산업이 발달함에 따라 인스턴트식품과 가공식품의 이용이 많아지고 있어 이에 사용되는 식품첨가물의 종류와 사용량도 증가하고 있다<sup>[1]</sup>. 이에 각종 가공식품에서 다양한 목적으로 사용되고 있는 식품 첨가물의 섭취도 불가피하게 되어 소비자가 알지 못하는 사이에 매일 여러 종류의 식품첨가물을

섭취하고 있으며, 식품 첨가물의 안정성 평가와 섭취량에 대한 관심도 증가하고 있다<sup>[2]</sup>.

식품첨가물(food additives)이란, 일반적으로 그 자체를 식품으로서 섭취하지 않고, 영양적 가치에 상관없이 식품의 일반 성분으로서 사용되지 않는 물질을 의미하며, 식품의 '제조·가공·조리·처리·포장·보관' 시에 '기술적인 목적'을 달성하기 위해 식품에 첨가하여 효과를 나타내거나, 직접 또는 간접적으로 식품에 효과를 나타

\*정회원, 한국산업기술대학교 컴퓨터공학과

\*\*준회원, 한국산업기술대학교 컴퓨터공학과

접수일자: 2017년 2월 21일, 수정완료: 2017년 12월 30일

게재확정일자: 2018년 2월 9일

Received: 21 February, 2017 / Revised: 30 December, 2017

Accepted: 9 February, 2018

\*Corresponding Author: jkikim@kpu.ac.kr

Dept. of Computer Engineering, Korea Polytechnic University, Korea.

널 것으로 기대되거나, 그 부산물이 식품의 구성성분이 되거나, 식품의 특성에 영향을 미칠 수 있는 물질을 말한다<sup>[3]</sup>.

현재 식품첨가물은 과학적인 근거에 의해 안정성이 검증되어야만 식품에 이용할 수 있도록 기준을 마련하고 허용 품목 및 식품별 사용량을 규제하고 지속적인 관리를 수행하고 있으나, 일반 소비자들은 식품첨가물에 대한 막연한 불안감을 가지고 있는 실정이다<sup>[4]</sup>. 하지만 화학 식품 첨가물에 대한 규제는 각 국의 나라마다 다를뿐더러, 최초 사용 시 인체에 무해하다고 판단되었지만 오랜 시간이 지나 다양한 연구를 통해 안전하다고 판단되어 사용되었던 화학 식품 첨가물이 인체에 해롭다고 밝혀지는 경우가 많다.

본 논문은 통계청에서 제시하는 조사 자료를 기반으로 빅 데이터 분석기법 중 하나인 시계열 분석과 회귀분석을 통해 첨가물 별 대두되고 있는 암 발생의 상관관계와, 식품 첨가물의 소비량이 급증한 원인에 대한 분석을 실시해 검증한다.

## II. 관련 연구

### 1. 시계열 분석<sup>[5]</sup>

본 논문에서 사용되는 통계자료는 특정 단위에 국한되지 않는 일반적인 비차원 수열에 적용될 수 있는 방법을 사용한다. 따라서 계산을 하기전에 시계열을 먼저 정규화 시켜 단위를 제거하는 과정이 필요하다. 우선 통계 자료를 통해  $\{v_0, v_1, \dots, v_{N-1}\}$ 라는 시계열을 얻었다고 하자. 이 시계열은 N개의 데이터 포인트로 구성되어 있으며 각각의 데이터 포인트는 년간 시간 간격을 두고 수집되었다. 이 시계의 평균  $\langle v \rangle$ 와 표준편차는 수식 1과 같이 정의할 수 있다.

$$\langle v \rangle = \frac{1}{N} \sum_{n=0}^{N-1} v_n \quad (1)$$

$$\sigma[v] = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} (v_n - \langle v \rangle)^2} \quad (2)$$

각 시계열의 각 요소  $v_n$ 을  $(v_n - \langle v \rangle) / \sigma[v]$ 로 대체함으로써 정규화 할 수 있다. 이후 사용되는 시계열

은 정규화 된 것이다.

이렇게 얻어진 정규화 된 시계열에 대해 먼저 자기 상관함수를 계산한다. 정규화 된 시계열에 대한 자기상관함수를 수식 3과 같이 계산할 수 있다.

$$\frac{1}{N-s} \sum_{n=0}^{N-s-1} v_n v_{n+s} \quad (3)$$

자기상관함수는 일반적으로 계의 기억효과를 추정하는데 사용된다. 자기 상관함수가 s가 증가함에 따라 빠르게 0으로 수렴할수록 시계열은 완전한 무작위 함수와 같다.

자기상관함수 다음에는 sample entropy(SampEn)를 계산한다. approximate Entropy(ApEn)의 경우 예측 가능성을 통해 시계열의 복잡성을 나타내는 척도이다. ApEn는 엄밀하게는 길이가 무한대인 시계열에 대해 정의되어 있고 시계열의 길이가 바뀌면 값이 달라지는 단점이 있다. 이러한 단점을 극복하기 위해 길이에 대한 의존성을 줄인 SampEn가 제안되었다.

예측 가능성을 측정하여 복잡성을 추정하기 위해 고안된 SampEn은 Bayes의 법칙으로부터 시작하여

$$P(B|A) = P(A|B) \frac{P(B)}{P(A)} \quad (4)$$

A를 {시계열 내에서 길이가 m인 시계열의 일부분이 같은 값을 가지며 반복되는 경우}, B를 {그 반복되는 시계열이 같은 데이터 포인트로 이어지는 경우}로 정의한다. 데이터가 같다는 의미는 비교하는 두 값의 차가 미리 정한 허용범위 r 보다 작은 경우를 의미한다.

### 2. 회귀 분석<sup>[6]</sup>

다중 회귀분석은 관련 변수들 간에 상호관련성을 알아보는 통계적 방법 중의 하나이다. 다중 회귀분석은 자료에 포함된 한 변수가 또 다른 하나이상의 여러 변수에 의해 어떻게 설명 또는 예측되는지를 적절한 함수식으로 표현하여 자료를 분석한다.

원래 데이터 분석은 일반적으로 많은 단계를, 때로는 몇 몇 단계를 거듭 반복하는 과정을 필요로 한다. 데이터의 각 변수 값의 분포 형태, 두 변수간의 플롯, 기타 자료의 특성을 알아보는 기술적(descriptive)인 단계를 거쳐서 통계적 모형을 세운 후 적합한(fitting)하는 단계에 들어

선다. 그리고 관심 있는 모수(parameter)에 대한 추론(추정과 감정)으로 흔히 데이터 분석을 끝맺음 한다. 그러나 데이터 분석이 과학적 과정(scientific process)이기 위해서는 모형 적합 이후 적합 모형에 대한 검토, 다른 말로 모형진단(model diagnostics)이 필요하다. 이러한 작업의 결과 적합 된 모형이 부적당하다고 판단되면 새로운 모형을 찾아보게 되고, 이후 모형적합이 이루어지며, 이에 대한 모형진단이 뒤따른다.

가능한 모든 후보 변수들을 독립변수로 사용하여 예측모형을 만드는 경우 일부 회귀 계수 추정치의 분산과 예측 값의 분산이 매우 커지게 되므로 이를 신뢰할 수 없기 때문에 불필요한 변수들이 들어있는 완전모형(full model)보다는 필요한 변수들만 들어 있는 축소 모형(reduced model)을 통해 모형을 보다 간결하게 하는 것이 바람직하다.

### III. 분석자료

#### 1. 한국 시도/성/연령/거주기간별 1인 가구

통계청에서 제공하는 한국 시도/성/연령/거주기간별 1인 가구 통계 자료는 각 조건에 따른 1인 가구 량이 5년 단위로 조사 되었다. 이 통계자료를 바탕으로 요약된 내용은 아래 표 1과 같다.

표 1. 한국 시도/성/연령/거주기간별 1인 가구  
 Table 1. Single-person households by provinces / provinces / age / residence period in Korea

year	persons	total
2000		2,224,433
2005		3,168,623
2010		4,142,165
2015		5,203,440

#### 2. 한국 연간 국민 다소비 식품 첨가물 순위

통계청에서 제공하는 한국 연간 국민 다소비 식품 첨가물 순위 통계 자료는 다소비 순위로 164개종의 국내에서 허가되어 사용되는 식품 첨가물에 대해 소비량이 톤(t) 단위로 조사 되어 있으며, 조사 기간은 1999-2011년 까지 매년 조사되었고 요약된 내용은 아래 표 2와 같다.

표 2. 한국 국민 다소비 식품첨가물 순위  
 Table 2. Ranking of Korean food additives

year	ton	Total	Average
1999		1,301,152	7,983
2000		1,407,650	8,636
2001		1,446,683	8,875
2002		1,583,229	9,713
2003		2,086,407	12,800
2004		1,580,056	9,694
2005		1,549,102	9,504
2006		2,666,075	16,356
2007		3,323,370	20,389
2008		4,416,392	20,959
2009		5,996,219	36,787
2010		4,654,753	28,557
2011		4,495,635	27,581

#### 3. 한국 61개 암종/성/연령 별 암 발생자수, 발생률

통계청에서 제공하는 한국 61개 암종/성/연령 별 암 발생자수, 발생률 통계자료는 국내에서 빈번히 발생하는 암 61개종을 대상으로 조사되었으며 단위는 명이다. 조사 기간은 1999-2013년까지 매년 조사되었고 요약된 내용은 아래 표 3과 같다.

표 3. 한국 61개 암종/성/연령 별 암 발생자 수, 발생률  
 Table 3. Number of cancer incidence and incidence by 61 cancer types / sex / age in Korea

year	person	Total	Male	Female	Average
1999		101,032	57,594	43,438	1,656
2000		101,772	58,016	43,756	1,668
2001		111,234	62,841	48,393	1,823
2002		117,089	65,531	51,558	1,919
2003		125,707	69,558	56,149	2,060

2004	134,476	73,730	60,746	2,204
2005	146,858	79,663	67,195	2,407
2006	154,583	82,800	71,783	2,534
2007	167,675	88,354	79,321	2,748
2008	182,129	94,682	87,447	2,985
2009	195,842	100,836	95,006	3,210
2010	207,085	105,511	101,574	3,394
2011	221,013	111,590	109,423	3,623
2012	226,216	113,305	112,911	3,708
2013	225,343	113,744	111,599	3,694

#### 4. 한국 시군구별 비만을

통계청에서 제공하는 한국 시군구별 비만율 통계자료는 국내 병원에서 측정된 비만 인구를 퍼센트 단위로 나타낸 것이다. 측정 인원에 따라 비만율이 조금씩 다르게 나타나지만, 전체적인 측정 인원의 변화가 많지 않기 때문에 신뢰할 수 있는 자료라 생각되어 사용하였다.

표 4. 한국 시군구별 비만율  
Table 4. Obesity Rate by Korean City Area

year person	Number of person	Obesity rate	Obese population
2008	21,586	20.6%	4,447
2009	23,152	21.4%	4,955
2010	22,636	21.7%	4,912
2011	22,822	22.7%	5,181
2012	22,611	23.7%	5,359
2013	22,827	23.2%	5,296
2014	22,660	23.9%	5,416
2015	22,677	24.7%	5,601

### IV. 본론

본 논문의 데이터의 분석처리는 R 언어를 이용하였고,

특히 R 언어를 쉽게 사용할 수 있게 해 주는 도구로써 R-Studio를 사용하였다. 분석 자료로는 통계청으로부터 한국 암 61개종 연간 발생 통계 자료(1999-2013), 한국 연간 화학 식품 첨가물 소비량 통계자료(1999-2011)를 사용하였다.

빅 데이터 분석 기법을 활용하기에 앞서 매년 별로 암 발생과 화학 식품 첨가물 소비량의 평균을 내었고 이 평균 자료를 이용해 막대그래프를 그려 암 발생과 화학 식품 첨가물 변화 추이를 표현했다. 그래프는 아래 그림 1, 그림 2와 같다.

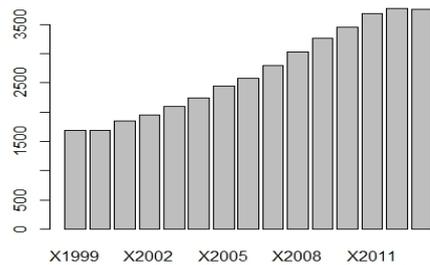


그림 1. 한국 암 61개종 연간 발생 평균  
Fig. 1. 61 Korean cancer types Annual occurrence rate

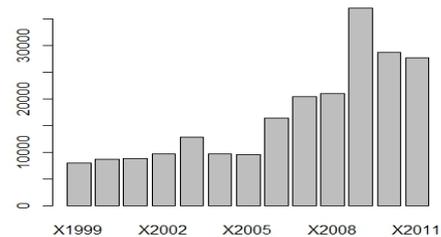


그림 2. 한국 화학 식품 첨가물 연간 소비량 평균  
Fig. 2. Korea Chemical Food Additives Annual Consumption Average

위 그래프를 확인해 보면 암 발생율은 1999년부터 매년 조금씩 증가하는 추세이고 화학 식품 첨가물 소비량은 2006년부터 급격하게 증가하였으나, 2010년부터 조금씩 줄어드는 추세이다. 이는 2005년부터 1인가구가 급증했기 때문에 2006년의 식품 첨가물 소비량이 자연스럽게 증가했음을 알 수 있다. 또한 2009년 이후 식품 첨가물에 대한 소비자의 관심이 증가함에 따라 식품 첨가물 소비량이 잠시 줄어들음을 알 수 있다. 1인 가구 변화량에 대한 그래프는 아래 그림 3과 같다.

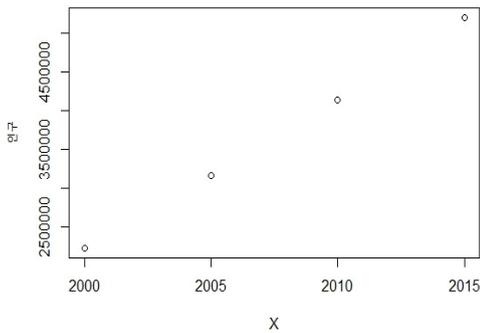


그림 3. 1인 가구 변화량  
 Fig. 3. one-person Household variation

위의 그림 1과 그림 2를 보면 식품 첨가물 소비량이 대체로 증가하는 추세에 있고 암 발생 율 또한 꾸준히 증가하는 추세를 볼 수 있다. 이를 토대로 식품 첨가물과 암 발생에 관한 상관성이 존재한다는 것을 알 수 있다. 하지만 그래프의 증가량만을 통해 연관성을 찾아내는 것은 신뢰성이 부족하다. 또한 암의 종류도 다양하고, 식품 첨가물의 종류도 다양하기 때문에 모든 식품 첨가물이 모든 암 발생과 연관 있다고 단정하기 어렵다. 이를 해결하기 위해 본 논문은 그동안 매스컴에서 보도된 특정 암 발생과 연관 있다 고려되는 식품 첨가물 몇 가지를 선택했다. 암 발생 종류와 의심되는 식품 첨가물의 종류는 아래 표 5와 같다.

표 5. 암과 의심되는 식품 첨가물들 (1)  
 Table 5. Cancer and Suspected Food Additives(1)

Cancer type	Food Additives
Liver cancer	Potassium sorbate Sodium benzoate Salicylic acid Sodium dehydroacetate Sodium nitrite Tar color
Uterine cancer, bladder cancer	Cyclamate Saccharin series
Male genital cancer	Bleaching powder Sodium hypochlorite
Colon cancer	Carboxymethylcellulose sodium Carboxymethylcellulose calcium
Thyroid cancer	Tar color
Other carcinogenicity	Butylhydroxyanisole Butylhydroxytoluene

이 중 한국 국민 다소비 식품 첨가물 순위 통계자료에 포함되어 있는 물질은 사카린나트륨계제(혼합계제류), 사카린나트륨/용성사카린, 카르복시메틸셀룰로오스나트륨, 카르복시메틸셀룰로오스칼슘, 타르색소가 존재한다. 이에 대한 소비량은 아래 표 6, 표 7과 같다.

표 6. 암과 의심되는 식품 첨가물들 (2)  
 Table 6. Cancer and Suspected Food Additives(2)

year ton	Saccharin sodium mixed preparation	Saccharin sodium water-soluble saccharin	Tar color
1999	878	489	41
2000	753	410	54
2001	823	363	45
2002	928	472	55
2003	656	258	65
2004	885	193	60
2005	812	201	75
2006	90	298	87
2007	92	233	85
2008	1,078	106	100
2009	877	140	140
2010	1,235	165	165
2011	1,152	348	215

표 7. 암과 의심되는 식품 첨가물들 (3)  
 Table 7. Cancer and Suspected Food Additives(3)

year ton	Carboxymethyl cellulose sodium	Carboxymethyl Cellulose calcium
1999	566	48
2000	663	40
2001	531	45
2002	637	55
2003	662	65
2004	507	60
2005	333	75
2006	405	86
2007	259	88
2008	101	100
2009	243	136
2010	285	165
2011	355	215

본 논문은 위의 표에서 언급한 발암 의심 식품 첨가물과 의심되는 암 별로 빅 데이터 분석을 실시했다. 먼저 사용된 기법은 시계열 분석이며, 시계열 분석을 사용한 이유는 암 발생 율 데이터와 식품 첨가물 데이터가 연속적인 시간별로 정리되어 있고, 2개 이상의 변수(특정 암, 특정 식품 첨가물 들)로 구성되어 있기 때문이다.

시계열 분석을 실시한 결과 사카린 계열의 식품 첨가물의 시계열 그래프가 전체적으로 여성 자궁관련 암과 방광암에 대한 그래프와 비슷한 모양을 보였다. 먼저 사카린에 대한 그래프를 살펴보면 사카린의 경우 2004-2006년을 기준으로 소비량이 급락하는 것을 볼 수 있다. 이후 2010년을 기준으로 사카린 섭취량이 급증했다. 자궁 관련 암, 방광암, 사카린 섭취량에 대한 시계열 그래프는 아래 그림 4와 같다.

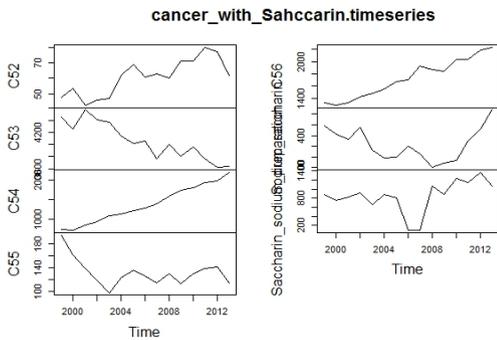


그림 4. 자궁 관련 암, 방광암, 사카린에 대한 시계열 그래프  
Fig. 4. A time series graph for uterine cancer, bladder cancer, saccharin

위의 표들을 보면 대부분의 자궁 관련 암과 방광암이 사카린의 섭취가 줄어든 시기에 따라 함께 어느 정도 감소했음을 볼 수 있다. 이를 통해 시계열 분석을 통한 사카린과 자궁 관련 암, 방광암에는 전체적인 연관성이 존재한다는 가정이 어느 정도 성립함을 알 수 있다. 하지만 시계열 분석은 시간에 따른 증감률을 비교하는 방식이기 때문에 완전한 관계성이 있다고 확정할 수 없다. 이를 해결하기 위해 본 논문은 다중 회귀분석을 사용했다.

우선 자궁암과 방광암의 관계성을 확인하기 위해 자궁암의 한 종류와 나머지 자궁암, 방광암에 대한 다중회귀 분석을 실시했다. 실시해 나온 결과의 p-value, 즉 오차범위는 0.0023 정도로 자궁암과 방광암 간의 증감에 대한 연관성이 존재함을 알 수 있다. 다중 회귀분석을 통해 분석한 자궁암과 방광암의 결과는 다음 그림 5와 같다.

```
> model<-lm(C52~., data = test_model)
> summary(model)

Call:
lm(formula = c52 ~ ., data = test_model)

Residuals:
    Min       1Q   Median       3Q      Max
-13.0778  -2.4289  -0.7617   2.8729   7.4800

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 205.15300   92.22099   2.225  0.0503 .
C53          -0.03499   0.01549  -2.258  0.0475 *
C54           0.01236   0.01609   0.768  0.4600
C55           0.09339   0.07982   1.170  0.2691
C56          -0.01832   0.02925  -0.626  0.5452
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.34 on 10 degrees of freedom
Multiple R-squared:  0.7828, Adjusted R-squared:  0.696
F-statistic: 9.012 on 4 and 10 DF, p-value: 0.002373
```

그림 5. 자궁암과 방광암 다중 회귀 분석  
Fig. 5. Multiple regression analysis of uterine cancer and bladder cancer

각 자궁암과 방광암의 관계성을 다중 회귀 분석을 통해 확인했기 때문에 다음으로 사카린 첨가물과 자궁, 방광암의 관계성을 확인하기 위해 암과 사카린 첨가물에 대한 다중회귀 분석을 실시했다. 실시한 결과는 다음 그림 6과 같다.

```
> model<-lm(C52~., data = test_model)
> summary(model)

Call:
lm(formula = c52 ~ ., data = test_model)

Residuals:
    Min       1Q   Median       3Q      Max
 -8.3823  -2.5489  -0.0943   2.2814   7.0751

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 212.126525   80.824909   2.625  0.0304 *
C53          -0.036681   0.014028  -2.615  0.0309 *
C54           0.008125   0.016149   0.503  0.6285
C55           0.118762   0.083184   1.428  0.1912
C56          -0.014764   0.024412  -0.605  0.5621
Sodium_saccharin
Saccharin_sodium_preparation  -0.024249   0.009628  -2.518  0.0359 *
                                0.004653   0.005876   0.792  0.4512
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.167 on 8 degrees of freedom
Multiple R-squared:  0.8846, Adjusted R-squared:  0.7981
F-statistic: 10.22 on 6 and 8 DF, p-value: 0.002192
```

그림 6. 자궁암과 사카린 다중 회귀 분석  
Fig. 6. Multiple regression analysis of uterine cancer and saccharin

각 자궁암들과 사카린 첨가물인 사카린나트륨 혼합제제, 사카린나트륨 용성제제에 대한 다중회귀 분석의 p-value는 0.0021정도로 오차범위가 0.05 미만임을 확인할 수 있다. 오차범위가 0.05 미만 일 경우 다중 회귀분석에서는 연관성이 있다고 생각하고, 하나가 증감 하게 되면 다른 요소들도 함께 증감한다고 볼 수 있다. 이를 통해 사카린 식품 첨가물들은 자궁, 방광암에 영향을 끼친다는 분석결과를 도출할 수 있었다.

그렇다면 과연 사카린 첨가물은 어떤 요인으로 증감했는지에 대한 의문점이 발생하게 된다. 사카린 첨가물은 설탕을 대체하기 위한 식품으로 1895년에 발명되었다. 사카린은 1g으로 설탕의 약 300배 이상의 단맛을 내며 칼로리가 한없이 0에 가까운 제품으로 오랜 시간동안 사용되어 왔다. 하지만 합성 식품 첨가물에 대한 위험성이 대두되면서 사카린에 대한 위험성 연구가 진행되었다. 이후 사카린은 흰 쥐를 대상으로 한 실험에서 방광암에 대한 위험성이 발견되었다. 이후 국내 매스컴을 통해 사카린에 대한 위험성이 보도 되고, 이와 함께 여러 식품 첨가물에 대한 위험성이 많이 다뤄졌다. 2005년부터 매스컴을 통한 화학 식품 첨가물에 대한 위험성이 연일 다뤄지면서 사카린의 소비가 많이 감소했다. 하지만 최근 사카린이 안전 할 지도 모른다는 연구 결과와 함께 다시금 사카린의 소비가 증가하고 있다.

사카린 첨가물을 소비하는 대상으로는 다이어트 식품으로써 소비자에게 가장 많이 소비되고, 당뇨병 환자의 치료 목적으로 많이 사용된다. 위에서 언급된 표 4와 같이 비만을 또한 점점 증가하고 있고, 암 발생 율, 사카린 소비량 또한 증가하고 있다. 이를 통해 비만에 대한 관심도가 증가함에 따라 사카린 소비량 또한 증가하고 있음을 알 수 있다.

## V. 결론

오래간 화학 식품 첨가물과 암 발생에 관한 연구가 많이 이루어 졌다. 암 전문 학자들은 식품 첨가물 보다 흡연과 음주, 대기 오염이 암 발생에 주요한 요인이라고 생각했다. 하지만 일반인들은 가장 많이 접하는 식품 첨가물에 암 발생 요인이 존재한다는 생각을 많이 했다. 국내에서 사용되는 합성 식품 첨가물의 경우 많은 테스트와 동물 실험을 통해 안정성이 확보된 첨가물은 정량을 정해 사용한다. 하지만 이러한 식품 첨가물들 중 몸에 쌓여 배출되지 않는 첨가물이 존재하고, 사회적 트렌드에 따라 소비량이 급증하는 식품 첨가물들도 존재한다.

본 논문은 이러한 일반인들의 생각을 기준으로 식품 첨가물 소비량과 암 발생의 상관관계에 대해 빅 데이터 분석을 통해 접근했다. 빅 데이터 분석 결과 가장 연관성이 있는 식품 첨가물은 자궁암, 방광암과 함께 사카린 첨가물이 영향이 있었고, 대장암에 카르복시메틸셀룰로오

스나트륨, 카르복시메틸셀룰로오스칼슘이 있었다. 하지만 다중회귀 분석을 통해 분석한 대장암의 경우는 오차율이 0.5에 근접한 결과가 나왔고, 자궁암의 경우는 0.002 정도의 아주 낮은 오차율이 발생했다. 이 결과를 통해 사카린 나트륨(혼합제제)와 사카린 나트륨(용성제제)가 자궁암, 방광암에 영향을 끼친다는 분석 결과가 나왔다.

본 논문에서 나온 사카린의 유해성에 대한 결과에 따라 사카린의 섭취를 줄여야 하고, 다이어트 식품의 소비도 줄여나가 건강한 식생활 습관을 들여야 할 것이다.

## References

- [1] Park. E. J., "Development of risk communication materials (book, animation) on food additives for the parents of elementary", Journal of the Korean Room Education Society Vol. 23, No. 2, pp. 295-317, 2012.
- [2] Park. S. K., "Estimation of Dietary Intake of Food Red Colors - Food Red No. 2, No. 3 and No. 40 -", Korean food nutrition and academic journals Vol. 34, No. 1, pp. 75-80, 2005.
- [3] Kim. D. S., "Food Safety Policy through the Food Additives Norm System of the EU", International economic law research Vol. 11, No. 3, pp. 25-66, 2013.
- [4] So. Y. J., "A Survey on the Perceptions of Consumer Organizations to Promote Risk Communication for Food Additives", Korea Food Cooking Science Journal Vol. 29, No. 2, pp. 105-113, 2013.
- [5] Kim. I. D., "Time Series Analysis of Engine Test Data", Proceedings of the Conference on Promotion Engineering in Korea, pp. 241-245, 2011.
- [6] Kim. J. B., "The Relationship Between Odor Unit and Odorous Compounds in Control Areas Using Multiple Regression Analysis", Journal of the Institute of Environmental Health Vol. 35, No. 3, pp. 191-200, 2009.
- [7] Han. H. C., "Decision-making system for the

resource forecasting and risk management using regression algorithms”, The journal of the institute of Internet, broadcasting and communication, Vol. 15, No. 6, pp. 311-319, 2015. DOI: <http://dx.doi.org/10.7236/JIIBC.2015.15.6.311>

#### 저자 소개

##### 이 국 형(정회원)



• Lee Kuk Hyung received his BS in Computer Engineering at Korea Polytechnic University in 2013 and MS in Tourism Business Management at Kyunggi University in 2016, respectively. He currently holds a doctor's degree from Korea Polytechnic University. His research interests include Software Engineering, information communication system, object-oriented analysis and design, etc.

##### 이 준 희(준회원)



• Lee Jun-hee earned a bachelor's degree from Korea University in 2016. He is currently enrolled in the Master's program at Korea Polytechnic University in 2016. His major fields are image processing, embedded, and network, etc.

##### 김 정 준(정회원)



• Jeong Joon Kim received his BS and MS in Computer Science at Konkuk University in 2003 and 2005, respectively. In 2010, he received his PhD in at Konkuk University. He is currently a professor at the department of Computer Science at Korea Polytechnic University. His research interests include Database Systems, BigData, Semantic Web, Geographic Information Systems (GIS) and Ubiquitous Sensor Network (USN), etc.

##### 김 영 곤(정회원)



• Young Gon Kim received his BS in Electronic Engineering at Kyungpook University in 1983 and MS in Electronic Engineering at Yonsei University in 1985, respectively. In 2000, he received his PhD in at KAIST. He is currently a professor at the department of Computer Science at Korea Polytechnic University. His research interests include Software Engineering, information communication system, object-oriented analysis and design, etc.