

인구통계학적 특성과 질병유무에 따른 소비자 유형분류¹⁾

임세화*, 오민지**, 하지희*, 이경희***

*농촌진흥청 농산업경영과

**충북대학교 빅데이터학과

*** (주)힐링소프트

e-mail:shbreeze@korea.kr

Classification of Consumer Types by Demographic Characteristics and Disease Status

Se-Hwa Lim*, Min-Ji Oh**, Ji-Hee Ha*, Kyung-Hee Lee***

*Division of Farm&Agribusiness Management, Rural Development Administration

**Dept. of Bigdata, Chungbuk National University

***Healingsoft Inc.

요 약

본 연구는 농촌진흥청의 농식품 소비자패널 자료와 건강보험공단의 건강보험 데이터를 결합하여 인구통계학적 특성과 질병유무에 따라 소비자의 유형을 분류하고 유형의 특성을 분석하였다. 4개의 군집으로 분류할 때 타당성이 가장 높게 나타났다. 1번 군집은 40대 이상의 건강한 집단, 2번 군집은 50대 이상의 고혈압, 당뇨, 비만 질환이 있는 집단, 3번 군집은 50대 이상이며 비만한 그룹이지만 당뇨, 고혈압 진단은 받지 않은 집단, 4번 군집은 대상자 중 가장 젊은 집단으로 질환이 없는 집단으로 나뉘었다.

1. 서론

소비자는 소득, 나이, 지역 등에 따라 다른 소비패턴을 보인다. 서로 다른 특징을 가진 소비자에게 동일한 마케팅 전략 또는 정책을 제시한다면 기대와 전혀 다른 결과가 나타날 것이다. 따라서 특정한 기준에 따라 그룹을 나누고 소비자의 유형을 구분하여 유형별로 특징을 규정할 필요가 있다.

본 연구에서는 인구통계학적 특성과 질병유무에 따라 소비자 유형을 분류하고 유형별로 특징을 설명하였다. 소비자의 유형을 분류하는 방법은 시장세분화 등 마케팅 기법으로 다양하게 활용되고 있으나 인구통계학적 특성과 질병을 연계하여 소비자의 유형을 분류한 연구는 부족한 실정이다.

자료는 농촌진흥청의 농식품 소비자패널 자료와 건강보험공단의 건강보험 데이터를 활용하였다. 소비자패널 자료의 인구통계 정보와 건강보험 데이터의 건강정보를 결합하여 동일한 소비자의 인구통계 특성과 건강정보를 연결하고 인구

통계학적 특성과 질병유무에 따라 소비자 유형을 분류하였다. 두 데이터를 결합하기 위해 공공기관 생명윤리심의위원회(IRB) 승인을 받았다. 건강정보와 인구통계 관련 정보가 같이 있는 데이터는 국민건강영양조사가 대표적이나 국민건강영양조사는 건강에 대한 정보가 설문조사 방식이며, 본 연구에서 활용한 건강 관련 데이터는 병원진료 기록이므로 정확도가 더 높다고 할 수 있다.

연구대상은 농촌진흥청 소비자패널 중 두 데이터의 결합에 동의한 전국 총 463명의 소비자패널을 대상으로 분석하였다. 2015~2020년의 데이터를 연계하여 분석하였으며, 인구통계학적 특성은 지역, 나이, 가구원 수, 소득을 분석하였고, 질병은 당뇨병, 고혈압, 비만의 진단여부를 활용하였다.

본 연구는 소비자의 유형을 분류하고 특징을 설명하는데 그쳤으나 본 결과를 이용하여 소비자의 유형에 따른 농식품 소비 정책 등 다양한 분야에 활용할 수 있을 것으로 기대한다.

1) 본 성과물은(논문, 산업재산권, 품종보호권 등)은 농촌진흥청 연구사업(세부과제번호: PJ01533802)의 지원에 의해 이루어진 것임

2) 본 연구의 연구내용 및 수행에 대해 보건복지부 지정 공공기관 생명윤리심의위원회의 승인을 받았음(IRB 승인번호: P01-202012-21-101)

*** Corresponding Author : Kyung-Hee Lee(Healingsoft Inc.)

2. 분석방법

본 연구에서는 소비자의 유형을 나누기 위해 군집분석을 활용하였다. 군집분석은 모집단이 가지고 있는 다양한 속성의 유사도(Similarity)를 이용하여 동질적인 집단으로 묶어주는 방법으로서, 다양한 모집단을 몇 개의 동질적 집단으로 분류하여 집단 내에 있는 공통된 특성을 분석하고, 서로 다른 집단에 속한 개체들 간에는 서로 유사하지 않은 특성을 가지고 있다[1]. 범주형 변수로 군집분석을 실시하기 때문에 고위 거리(Gower Distance)[2, 3]를 사용한 계층적 군집분석을 수행하였다.

인구통계학적 특성에 대한 변수로 지역, 나이, 가구원 수, 소득, 질병유무는 건강보험 진료기록을 통해 당뇨병, 고혈압, 비만의 질환자 여부를 판단하였다.

3. 분석결과

3.1 인구통계학적 특성별 질병유무 분석

4가지 인구통계학적 특성 중 지역, 가구원 수는 질병 유무에 유의미한 영향을 주지 않았고, 나이(출생년도)와 소득에 따라 질병을 가진 사람 비율이 큰 차이를 보였다. 당뇨, 고혈압, 비만 모두 1960년대에 출생한 나이가 많은 소비자의 진단비율이 높았고, 소득이 가장 높은 16~20분위에서 진단비율이 높게 나타났다.

[표 1] 소비자패널의 출생년도 및 소득별 질병유무 비율

질병	질병 유무	출생년도				소득(분위)			
		1960s	1970s	1980s	1990s	0~5	6~10	11~15	16~20
당뇨	있음	9.3	5.6	0.4	0.2	3.9	2.2	4.3	5.0
	없음	28.6	42.0	12.3	1.5	18.4	17.1	18.2	29.0
고혈압	있음	13.6	8.7	0.6	0.0	5.0	4.5	4.5	8.9
	없음	24.2	39.0	12.1	1.7	17.3	14.7	18.0	25.1
비만	있음	21.6	14.9	2.6	0.4	10.8	6.1	7.6	14.3
	없음	16.2	32.7	10.2	1.3	11.5	13.2	14.9	19.7

3.2 인구통계학적 특성과 질병유무에 따른 군집 분석
63명의 패널에 대한 고위거리 기반 군집화하기 위해 군집 k를 설정하였다. 실루엣 방법(silhouette Method)를 활용하여 최적의 군집 개수를 4로 정하였다.

전체 소비자패널 중 1번 군집에 포함된 비율이 41.9%로 가장 높았으며 3번 군집이 다음으로 높았다.

1번 군집은 1970년대 이후 출생이며 건강한 집단, 2번 군집은 대사성질환 집단(당뇨, 고혈압, 비만 진단 비율 높음)으로 분류되었다. 3번 군집은 당뇨, 고혈압은 진단비율이 낮으나 비만 진단비율이 높으며 1960년대생 중심이고, 4번 군집은 1980년대생이 가장 많으며 질환자 비율도 낮은 젊고 건강한

집단으로 분류되었다.

모든 군집에서 당뇨와 고혈압의 진단비율이 비슷한 흐름을 보였으나 비만은 당뇨, 고혈압과 다른 방향의 군집으로 분류되기도 하였다.

[표 2] 인구통계학적 특성과 질병유무에 따른 군집 분석 결과

구 분		군집 1	군집 2	군집 3	군집 4
출생년도 비율	1960s	12.9	58.3	70.7	7.7
	1970s	79.9	38.3	19.1	23.1
	1980s	5.2	3.3	8.3	65.4
	1990s	2.1	0.0	1.9	3.8
소득 분위별 비율	0~5	18.0	41.7	21.6	21.6
	6~10	18.0	28.3	19.0	47.1
	11~15	37.0	15.0	45.8	15.7
	16~20	27.0	15.0	13.7	15.7
당뇨*	평균	0.05	0.67	0.12	0.06
고혈압*	평균	0.11	0.78	0.21	0.10
비만*	평균	0.12	0.87	0.68	0.04
군집크기(명, %)		194(41.9)	60(13.0)	157(33.9)	52(11.2)

4. 결론

본 연구는 농촌진흥청의 농식품 소비자패널 자료와 건강보험공단 건강보험 데이터를 결합하여 인구통계학적 특성과 질병유무에 따라 소비자의 유형을 분류하고 유형의 특성을 분석하였다. 그 결과를 요약하면 다음과 같다.

첫째, 4가지 인구통계학적 특성 중 지역, 가구원 수는 질병 유무에 유의미한 영향을 주지 않았고, 나이(출생년도)와 소득에 따라 질병을 가진 사람 비율이 큰 차이를 보였다.

둘째, 4개로 군집화한 결과, 1번 군집은 40대 이상의 건강한 집단, 2번 군집은 50대 이상의 고혈압, 당뇨, 비만 질환이 있는 집단, 3번 군집은 50대 이상이며 비만한 그룹이지만 당뇨, 고혈압 진단은 받지 않은 집단, 4번 군집은 대상자 중 가장 젊은 집단으로 질환이 없는 집단으로 나뉘었다.

향후 건강한 군집과 질환이 있는 군집의 농식품 소비 특성을 분석하여 질환 예방 또는 관리를 위한 건강한 농식품 소비를 제안할 수 있을 것이다.

참고문헌

- [1] 송민경 · 장훈, “군집분석을 이용한 소도권 도시의 유형화에 관한 연구”, 한국지형공간정보학회, 제 18권 1호, pp. 83-88, 2010년
- [2] Everitt, Brian S., et al. "Cluster analysis 5th ed." (2011).
- [3] Tuerhong, Gulnabai, and Seoung Bum Kim. "Gower distance-based multivariate control charts for a mixture of continuous and categorical variables." Expert systems with applications 41.4 (2014): 1701-1707.