

## 잠재계층분석기법(Latent Class Analysis)을 활용한 영화 소비자 세분화에 관한 연구\*

구교령\*\* · 이장혁\*\*†

### Segmentation of Movie Consumption : An Application of Latent Class Analysis to Korean Film Industry

Kay Ryung Koo\*\* · Janghyuk Lee\*\*

#### ■ Abstract ■

As movie demands become more and more diversified, it is necessary for movie related firms to segment a whole heterogeneous market into a number of small homogeneous markets in order to identify the specific needs of consumer groups. Relevant market segmentation helps them to develop valuable offer to target segments through effective marketing planning. In this article, we introduce various segmentation methods and compare their advantages and disadvantages. In particular, we analyze "2009~2010 consumer survey data of Korean Film Industry" by using Latent Class Analysis(LCA), a statistical segmentation method which identifies exclusive set of latent classes based on consumers' responses to an observed categorical and numerical variables. It is applied PROC LCA, a new SAS procedure for conducting LCA and finally get the result of 11 distinctive clusters showing unique characteristics on their buying behaviors.

Keyword : Segmentation, Clustering Analysis, Mixture Model, Korean Movie Industry, Latent Class Analysis, K-Means, SOM

논문접수일 : 2011년 10월 01일 논문게재확정일 : 2011년 11월 24일

논문수정일(1차 : 2011년 11월 13일)

\* 이 논문은 2010년도 정부재원(교육과학기술부 인문사회연구역량강화사업비)으로 한국연구재단의 지원을 받아 연구되었음(NRF-2010-330-B00110).

\*\* 고려대학교 경영대학 경영학과

† 교신저자

## 1. 서 론

마케팅 활동과 관련하여 기업이 가장 먼저 주목해야 하는 점은 “나의 고객이 누구이며, 나의 고객이 어떠한 특성을 가지느냐”이다. Kotler and Armstrong[32]은 마케팅을 프로세스라고 정의하면서, 마케팅 전략을 실행하기 전 기업은 시장에 대한 이해와 고객이 필요로 하는 것을 밝혀내야 하며, 이에 기반한 마케팅 전략을 실행해야 한다고 설명한다. 이를 위해 기업이 우선적으로 착수해야 하는 것은 마케팅 조사에 기반한 시장 세분화이다. 1950년대에 시장세분화 개념이 처음 소개되면서 동일 상품 및 서비스 사용에 있어서의 사람들간의 이질적 욕구가 존재하는 현상에 대한 고찰이 시작되었다. 시장 세분화가 필요한 이유는 모든 잠재 소비자들이 동일하지 않은 니즈를 가지기 때문에 기업은 이질적 특징을 가지는 그룹 간에 차별화된 마케팅 프로그램을 실행해야 하기 때문이다[15]. 시장 세분화를 통해 분석된 각 군집은 이질적인 선호와 행동 패턴을 보이고 군집 내에 소속된 소비자들은 기업이 제공하는 제품이나 서비스로부터 얻고자 하는 욕구가 유사하며 기업이 실행하는 마케팅 활동에 대한 반응이 유사하게 나타난다[34]. 그러므로 기업은 해당 소비자들의 특성에 근거하여 시장을 세분화하고 기업에 적절한 타겟 세그먼트를 분별하여 마케팅 전략을 실행해야 한다. 따라서 시장 세분화 분석은 마케팅 전략을 수행하기 위한 가장 기초적인 단계이자 필수적 단계라 할 수 있겠다.

시장 세분화 분석 방법 중 가장 대표적인 방법인 K-means 방법은 그 활용도 측면이나 기술적인 측면에서의 용이성 때문에 다수의 학자들과 실제 기업들이 많이 사용하는 방법이다. K-means 방법의 경우는 등간 척도를 사용한 변수들을 기반으로 마케팅 조사를 진행 할 경우 유용하게 사용할 수 있다. 그 이유는 K-means는 유클리디안 거리(Euclidean distance)를 기반으로 유사성에 기초하여 소비자들을 그룹화하기 때문에 등간 척도로 측정된 변수들이 마케팅 조사에서 사용된 경우, 군집 내 거리는

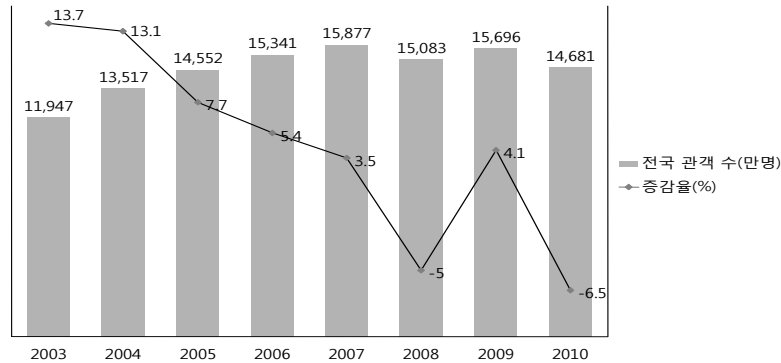
최소화하고 군집 간 거리는 최대화할 수 있는 거리를 측정하여 각 군집의 개수를 결정할 수 있기 때문이다.

그러나 범주형 변수와 수치형 변수가 동시에 사용되는 문항들로 구성된 설문지의 경우, K-means를 사용하여 분석하는 것은 그 한계를 가질 수 있다. 그 이유는 거리를 기반으로 소비자를 세분화하는 방식인 K-means의 경우, 임의로 숫자가 부여된 범주형 변수 간의 거리 계산이 불가능하기 때문이다. 따라서 K-means를 사용하여 숫자간의 차이를 거리로 지각하여 분석하는 것은 잘못된 세분화 결과를 도출 할 수 있다.

이러한 문제를 해결하기 위한 방법으로 본 연구에서는 1980년대부터 다양한 학자들에 의해 사용되고 있는 잠재계층분석 방법(Latent class analysis : 이하 LCA)을 소개하고자 한다. 해외에서는 다수의 학자들이 LCA를 사용하여 시장 세분화를 진행하고 있음에도 불구하고, 국내에서는 LCA를 활용한 시장 세분화에 대한 연구가 극히 제한적으로 진행되고 있으며[1], 특히나 LCA 분석 시, 범주형 변수 이외에 수치형 변수를 동시에 사용하여 분석한 사례는 찾아볼 수 없다. 따라서 본 연구는 기존 연구 방법을 확장[1], SAS LCA proc을 활용하여 범주 및 수치형 범주가 포함되어 있는 영화 산업 설문 자료를 토대로 시장 세분화를 하고 그 결과를 바탕으로 영화 소비자의 특성을 이해하고자 한다.

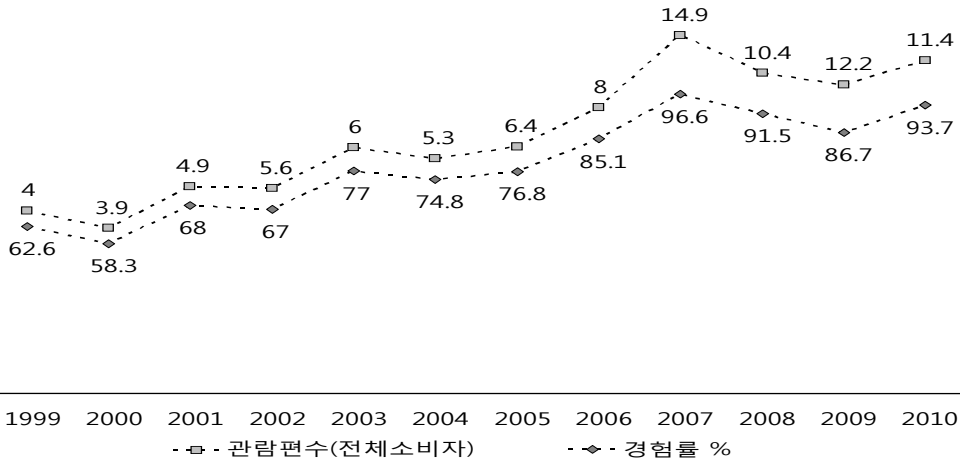
본 연구가 특히 영화 시장에 주목하는 이유는 다음과 같다. 국내 영화 시장은 1999년 이후부터 급격한 성장을 통해 엔터테인먼트 산업 중 최대 규모를 자랑하게 되었다. 그러나 2003년부터 산업 성장율은 더 이상 증가하는 형태가 아닌 감소하는 추세를 보이고 있으며 2010년에는 마이너스 성장율을 나타내고 있다[5].

또한 소비자들의 영화 구매 경험율이 평균 90% 수준을 유지하게 되면서[2, 7] 이미 영화 산업은 2007년 이후부터 성숙단계를 거쳐 쇠퇴 단계로 진입하고 있는 것으로 추론된다. 특히 2009년의 경우, 영화 산업의 총 이익이 -371억과 -13.12%의 투자수



출처 : [5]를 토대로 구성.

[그림 1] 산업 성장률



출처 : [2, 7]를 토대로 재구성.

[그림 2] 소비자 영화관람 경향

익률을 나타내고 있어 산업 위기론이 제기되고 있는 상황이다[4]. 이 같은 상황에서 영화 산업의 시장 세분화가 필요한 이유를 영화 상품만의 고유한 특성과 연관시켜 설명하면 다음과 같다.

첫째, 영화는 소비자들에게 유형의 상품을 제공하는 것이 아니라 “서비스 경험”을 소비하게 하기 때문에[3], 구매자가 직접 경험하기 전, 제품에 대한 특정 영화를 제대로 평가 내릴 수 없는 경험재의 특성을 가지고 있다[43, 46]. 따라서 소비자들은 매체를 통해 제시된 다양한 정보를 통해 영화 상품에 대한 인지와 간접적 평가를 진행하기 때문에 마케팅 활동에 대한 영향력을 상대적으로 크게 받을 수

있는 제품이라 하겠다. 이뿐만 아니라, 대부분의 영화가 개봉 첫째 주에 투자의 이윤을 회수하는 성격을 가지고 있어[48], 상대적으로 다른 제품에 비하여 짧은 제품수명주기를 가지고 있다. 그러므로 제품에 대한 이윤을 극대화하기 위해서는 적절한 고객 타겟을 대상으로 한 효율적 마케팅 전략을 진행하는 것이 필요하다 하겠다. 둘째, 영화 상품의 경우는 제품 구매 가격이 고정되어 있어 경쟁 상품간 가격 차별화 전략이 아닌 제품 차별화를 필요로 한다. 따라서 차별화된 제품 생산을 위해 기획 단계에서부터 적절한 마케팅 전략을 필요로 하는 상품이라 하겠다. 셋째, 해당 제품의 경우는 높은 투자를

요구하는 반면, 영화 소비의 정도가 개개의 영화마다 고정되어 있지 않아 수익의 정도를 미리 예측할 수 없다는 특징을 가진다[11]. 따라서 제품 생산 단계에서부터 마케팅 활동까지 고도화된 전략적 접근을 필요로 하는 상품이라 하겠다.

그럼에도 불구하고, 기존 영화 시스템은 20대 여성을 메인 타깃으로 한 영화 제작 및 홍보를 진행하고 있다. 이수범[8]에 따르면 총 7편 중 5편인 <공동경비구역 JSA(2000)>, <집으로(2002)>, <가문의 영광(2002)>, <올드 보이(2003)>, <태극기 휘날리며(2003)>가 20대 여성을 메인 타깃으로 홍보를 진행하였다고 설명한다. 즉, 영화 산업은 효율적인 마케팅 전략을 활용한 경우, 효과적인 수익 창출이 가능할 수 있는 상품임에도 불구하고, 시장을 세분화하여 적정한 타깃을 설정 한 후 마케팅 홍보를 진행하는 것이 아닌 대부분의 영화들이 헤비 유저(heavy user)를 대상으로 한 집중화된 시장 세분화 전략을 사용하고 있다.

결과적으로, 2009년의 경우, 총 118편의 영화가

한해 제작되었으나, 단 18편만이 손익분기점을 넘긴 암담한 결과를 나타내었다[4]. 이 같은 결과는 비단 2009년 한해만 해당하는 것이 아니라 매해 나타나는 결과로, 영화진흥위원회 보고에 따르면, 2005년은 전체의 32%, 2006년에는 전체의 14%, 2007년에는 전체의 11%, 2008년에는 전체의 7%만이 수익을 발생한 영화들이라고 설명하고 있다[4].

이 같은 현상이 나타난 근본적인 이유는 영화 상품 제작 시 사용되고 있는 비용이 효과적으로 지출되지 않고 있다는 것이다. 2009년 지출된 제작 비용을 살펴보면, 총 2829억 중 순제작비용이 1725억, 광고홍보비용이 490억, 기타비용이 613억이 사용되었다[4]. 전체 80% 정도의 비용을 광고비용과 제작비용이 차지하고 있음에도 불구하고, 차별화되지 않은 타깃 선정을 통한 제품의 제작과 마케팅 전략의 실행은 비용 대비 안정적인 수익을 보장하지 못하고 있다는 것을 알 수 있다.

따라서 마이너스 수익률을 기록하고 있는 문제점을 보완하기 위해서는 우선적으로 제작비용과 광고

〈표 1〉 영화 산업 현황

(단위 : 백만 원)

	2005	2006	2007	2008	2009
손익분기점이상 편수	26	15	12	8	18
전체제작편수	81	110	112	108	118
총 매출	336,786	388,058	266,927	200,839	245,821
총 비용	312,117	513,640	448,813	355,667	282,945
총 이익	24,669	-125,582	-181,886	-154,827	-37,123
투자수익률	7.90%	-24.46%	-40.53%	-43.53%	-13.12%

출처 : [4]를 토대로 구성.

〈표 2〉 영화 산업의 비용 지출 내역

	금액(백만 원)	총 비용대비					평균
		2009	2008	2007	2006	2005	
순제작비용 합계	172,579	60.99%	62.98%	62.96%	54.80%	60.53%	60%
광고홍보비	49,065	17.34%	19.51%	19.12%	21.75%	19.99%	20%
기타비용 합계	61,301	21.67%	17.51%	17.92%	23.45%	19.48%	20%
총 비용 합계	282,945	100%	100%	100%	100%	100%	100%

출처 : [4]를 토대로 구성.

비용의 효율적 사용이 필요하다. 그러나 기존의 영화 제작시스템은 차별화된 소재를 소비자들에게 제공한다고 보다는, 평균화된 기호나 욕구를 대상으로 대중화 및 정형화된 상품의 제작과 집중화된 마케팅 전략을 실행하였고, 결과적으로 소비자들의 영화에 대한 관심 저하 및 수익 경쟁력의 약화를 야기하였다. 그 결과, 잠재 성장력이 한계에 달한 영화 산업 시장에서의 차별화되지 않은 영화 상품들간의 무한 경쟁은 생산비용의 회수와 이윤 창출을 불가능하게 하였고, 결국 전체 관객 수 감소로 이어져, 한국 영화 산업 사양화의 발판을 제공하게 되었다. 그러나, 21세기 후반 들어, 관객들이 역동적으로 자신의 선호를 구축하고 변화시킴에 따라, 수익이 기대되지 않았던 <워낭소리(2008)>가 300만명에 육박하는 관객을 동원하며 190배에 달하는 수익을 창출 하였다. 이와 같은 예상외의 획기적 성공은 기존의 헤비 유저로 인식되고 있던, 20대 소비자 뿐 아니라, 중·장년층 관객들과의 공감대를 형성함으로써 가능하였다. 이는 흥행 시장의 관성에 굴복하지 않는 소비자들의 특성을 재발견한 사례이자, 잠재적 틈새 시장을 발견한 의미있는 사례로 평가된다[11].

이 성공 사례는 한국 영화 산업의 새로운 가능성을 보여주는 현상이라 할 수 있다. 즉, 보편화된 선호 이외에 ‘소비자 특성 별로 다양한 선호’가 존재한다는 것을 확인시켜주기 때문이다. 그러므로 다양한 취향을 가진 소비자들을 만족시킬 수 있는 새로운 전략의 필요성이 대두된다. 이러한 전략은 그동안 주목하지 않았던, 잠재적 영화 구매 집단의 극장 유인을 통해 한국 영화 산업의 전체 관객수 감소에 긍정적 변화를 유도할 수 있을 것이다. 따라서 영화 소비자를 대상으로 한 시장세분화 분석이 영화 시장에서 필수적으로 진행되어야 하며, 이를 통해 산업적 위기를 이겨 나갈 수 있을 것으로 예상된다.

상기 문제제기를 연구하고자, 본 논문은 다음과 같은 순서로 구성되었다. 제 2장에서는 시장세분화 분석 기법과 관련한 문헌연구들을 고찰하고, 이론

적 토대에 근거하여 본 연구에 가장 적합한 시장세분화 분석 기법을 고찰할 것이다. 제 3장에서는 실제 분석에 사용되는 데이터에 대한 설명과 연구 결과를 제시할 것이며, 마지막으로 제 4장에는 본 연구의 시사점과 향후 연구 과제에 대한 제언을 제시할 것이다.

## 2. 이론적 배경

### 2.1 시장 세분화 분석 방법

시장 세분화 분석 방법은 “어떻게 군집이 형성되는지”를 그 기준으로 하여 분류할 수 있다[50]. Hruschka[25]는 시장 세분화 분석 기법을 “nonoverlapping, overlapping, fuzzy”로 구분하였는데, nonoverlapping이란 용어 그대로 각각의 분석 대상이 하나의 군집에만 속할 수 있게 군집을 형성하는 것을 그 가정으로 한다. Overlapping과 fuzzy 기법은 분석의 대상이 하나 이상의 군집에 속할 수 있음을 가정하여 시장세분화를 진행한다. 즉 overlapping과 fuzzy 기법은 소비 상황에 따라 소비자가 군집에 속하는 경우가 상이해질 수 있다는 것을 고려한 분석 기법이라 할 수 있다[12]. 그러나 overlapping과 fuzzy가 그 차이를 나타내는 점은 overlapping의 경우 쌍으로 된 군집이 포개질 수 있다는 가정을 통해 하나의 분석 대상이 두 군집에 동시에 속할 수 있는 경우를 제시하고, 이에 따라 분석 대상은 두 개의 군집에만 속할 수 있다는 가정을 하는데 반해, fuzzy의 경우는 분석 대상이 각각의 군집에 속할 확률을 계산하기 때문에 세개 이상의 군집들에 분석 대상이 속할 확률을 제시한다는 차이점을 나타낸다.

구체적으로 살펴보면, nonoverlapping 방식에서 가장 대표적으로 사용되는 방법은 덴도그램을 활용한 위계적 군집 방식과 비위계적 군집 방식으로 불리는 K-means 방법이 있다. 덴도그램을 활용한 위계적 군집 방식의 경우는 하나의 분석 대상을 각각 군집으로 지정한 후, 하나의 군집으로 해당 대상들이 성공적으로 모두 묶일 때까지 반복적으로 개개

의 군집들을 응집화하는 과정을 거치는 병합적(agglomerative) 방식과 통합된 하나의 거대 군집을 개개의 군집으로 분류될 때까지 반복적으로 대상을 여러 개의 군집으로 나누는 분할적(divisive) 방식이 있다[21]. 이 같은 위계적 군집 방식을 사용할 경우, 연구자는 덴도그램을 도식화하여, 가장 이상적으로 추론되는 군집의 개수를 정하는데, 이때의 문제점은 소비자나 기업간에 위계적 관계 구조를 가지고 있음을 합리화할 수 있는 이론적 주장을 찾기 힘들다는 점이다[50]. 따라서 소비자 시장 분석 방법에서는 비위계적 방식이 위계적 방식보다 더 적합한 방식이라 하겠다[45]. 반면, 비위계적 방식인 K-mean는 연구자가 시장 분석을 하기 전 군집의 개수를 미리 정한 후, 대상을 무작위로 군집들에 분포시킨 후, 특정 기준이 적정한 수치를 나타낼 때까지 해당 대상들을 군집들에 지속적으로 재분배하는 과정을 진행한다. 이때 군집간의 차이와 유사성을 나타내는 기준으로 “유클리디안 거리”가 가장 많이 사용되고 있는데, 유클리디안 거리는 분석 대상이 시장 세분화 변수들에 응답한 정도를 거리로 해석하여 군집 내의 거리는 최소로 하고 군집 간 거리는 최대화하는 군집들로 분류하는 기준을 제시한다. K-means 방법의 가장 큰 문제는 미리 군집의 개수를 정해야 한다는 점에 있으나[41], 해당 방식은 다수의 학자들과 실업에서 통계 분석 시 사용하는 SPSS를 이용하여 계산할 수 있기 때문에 쉽게 분석할 수 있다는 이점을 가지고 있어 전통적으로 많이 사용되고 있는 분석 기법이라 하겠다.

Shepard and Arabie[47]에 의해 처음으로 소개된 overlapping 방식은 쌍으로 연결되는 유사성들을 기초로 하여 군집들을 서로 포개는 과정을 거치게 된다. 따라서 시장 세분화를 할 때, 모든 분석 샘플들을 대상으로 서로 쌍으로 연결이 될 수 있는 유사성이 분석되어야 하는데, 이 같은 방식은 샘플 사이즈가 커지게 되는 경우, 분석 상의 애로사항으로 인하여 실제로 시장 세분화 방식으로 자주 사용되지 않는 방법이다[50].

마지막으로 fuzzy 방식은 군집에 소속 대상이 들

어갈 정도를 계산하는 방법인 “fuzzy set”과[25, 51], 데이터가 여러 개의 혼합된 분포를 가지고 있다는 가정을 통해 각각의 군집에 속할 확률을 추정하는 “mixture model” 두 가지 방식으로 재분류될 수 있으며, 이 두 가지 방식은 모두 [0.1]까지의 값을 제시한다. Fuzzy set 방식 중, 가장 최근에 사용되는 방법으로, 데이터마이닝 분야에서 많이 사용되고 있는 분석 기법이 “Self-Organizing Map”(이하 SOM)이다. SOM의 경우, 군집 분석의 기능뿐 아니라 분석 결과를 시각적으로 표현할 수 있는 신경망 기법을 사용하고 있어, 해당 군집 결과에 대한 이해가 쉽고 해당 대상이 군집에 속할 가중치를 계산하기 때문에 기존의 전통적인 군집 방식보다는 논리적 설명을 제시한다 점에서 그 이점을 가지고 있다[9, 18, 39]. Kohonen[31]에 의해 처음 제시된 SOM은 군집화 기능을 통해, 데이터의 양을 축소시킨 후 이에 따라 저 차원 공간에 데이터간의 상관관계를 제시할 수 있어 K-means와 MDS(multi-dimensional scaling)의 특성을 모두 갖추고 있다는 특징을 가지고 있다[30]. 초창기, Kohonen[32]이 제안한 SOM/Kohonen 기법은 K-means와 유사하게 nonoverlapping 방식으로 분류될 수 있다[30, 49]. SOM의 알고리즘은 2단계 학습모형으로, 입력되는 분석 대상과 각 노드간의 유클리디안 거리를 계산하여 가장 가까운 거리의 노드와 대응시키고 유사 노드를 중심으로 해당 군집에 속할 가중치를 증가시키는 반복적 학습과정을 거치면서 하나의 군집을 형성한다[9]. 이때, 가우시안 함수를 사용하여, 뉴런간의 순위가 매겨지고, 이에 기반한 2차원 평면 공간이 시각적으로 도출되게 된다[9, 30]. SOM은 유클리디안 거리를 기준으로 분석 대상과 군집간의 관계를 정의한다는 점에서는 K-means 방식과 매우 유사한 특징을 보이고 있으나, K-means가 가장 가까운 거리에 있는 분석 대상만을 고려하는 것에 비해, SOM은 모든 분석 대상간의 거리를 고려한 후, 군집에 속할 가중치를 증가하는 방향으로 군집을 구성한다는 점에서 그 차이를 나타낸다[30]. 1980년대 후반에 이러한 SOM/Kohonen 방식이 데이터마이닝 분

야에서 소개됨에 따라, SOM 방식은 nonoverlapping 만이 고려된 모델에서 fuzzy 모델로 발전하게 된다. Huntsberger and Ajijmarangsee[26]의 연구는 SOM/Kohonen 모델에 기반한 학습 모형을 발전시켜, Kohonen의 알고리즘에서 사용되는 학습 계수를 이용하여, 분석 대상이 뉴런의 군집에 할당될 멤버십을 “fuzzy c-means cluster 알고리즘(이하 FCM)”을 통해 추정하였다[49]. FCM이란 이미 정해진 군집의 수의 군집 중앙값을 계산하여 군집의 평균과 분석 대상간의 거리를 계산하고, 이 군집에 속할 멤버십의 정도를 추정하는 방법이다[50]. 이후에도 SOM은 초창기 nonoverlapping만을 추정하던 모형에서 분석 대상이 하나 이상의 군집에 속할 수 있고, 이 대상이 군집에 속할 확률은 [0, 1] 사이에 존재하며, 군집에 속할 전체 총합은 1이라는 기본적인 fuzzy 가정을 받아들이는 노력을 지속적으로 하고 있다[49]. SOM은 통계패키지인 SAS와 데이터마이닝 분야에서 많이 사용되고 있는 MATLAB을 통해 분석 가능하며, 특히 MATLAB을 사용할 경우 무료로 제공되고 있다(<http://www.cis.hut.fi/projects/somtoolbox>). 그러나 현재 경영학분야에서 해당 분석 기법을 사용하여 시장을 세분화한 경우는 극히 제한적으로 진행되고 있으며, 특히 마케팅 분야에서 해당 기법을 사용한 사례는 시장 구조 분석에 활용된 사례[40] 외에는 찾아보기 힘들다. 그러나 SOM이 K-means 방식보다 시각적으로 군집을 평면화해주어 더 쉽게 군집의 특성을 이해해줄 수 있다는 장점과 현실을 잘 반영한 fuzzy 분석 기법으로의 발전을 지속적으로 진행함에 따라, SOM이 현재보다 경영학 분야에서 자주 활용 될 수 있을 것으로 예상된다.

Mixture model에서 가장 대표적인 방법은 잠재계층방법(이하 LCA)이다. 근본적으로 SOM이 그 가정으로 하고 있는 fuzzy set 모형과 LCA 모형이 그 가정으로 하고 있는 mixture model간의 가장 큰 차이점은 시장 세분화의 군집 형성을 구체화하는데 있어서의 가정이 다르다는 점이다. Fuzzy set의 방식의 경우, 각각의 군집에 소비자가 부분적으로 속할 수 있는 가중치를 제시하기 때문에 실제로 분석

의 대상들이 두 개 이상의 군집에 속할 수 있다는 가정을 전제로 하는 반면, mixture 모델의 경우는 원칙적으로 소비자는 하나의 군집에만 속할 수 있다는 가정을 한다. 즉, mixture 모델은 데이터 상의 한계로 인해 데이터 자체가 소비자를 실제적으로 하나의 군집으로 제안할 수 없다는 문제점이 발생하므로, 이를 극복하고자 군집에 속할 수 있는 확률을 추정하고자 하는 목적을 가지고 있다는 점이다. 이 방법은 통계적 추론을 가능하게 하는 계량적 군집분석 방법으로, 군집의 개수가 정해지는 데에 대한 통계적 의미를 제시할 수 있고, 이에 따라 논리적인 시장 세분화가 가능하다는 장점을 가진다고 하겠다[26, 50]. Green et al.[22]에 의해 처음 시장 세분화에 적용된 LCA 방식은 관찰된 요인간의 관계를 통해 실제적으로 존재하지 않는 잠재적 군집간에 연결고리를 찾는 방식으로써, 해외에서는 이미 1980년대부터 다양한 경영학 학자들에 의해 활용되어진 방식이나[13, 16, 17, 23, 27-29, 34], 국내에서는 제한적으로 사용되고 있다[1]. 김영찬, 이두희[1]는 인터넷 이용자 중 인터넷에서 실질적으로 제품을 구매하는 소비자들의 특성을 분석하고자 인터넷 유저들을 대상으로 한 시장 세분화 분석을 LCA를 활용하여 진행하였다. 해당 연구가 LCA를 분석 기법으로 활용한 데에는 LCA가 다음과 같은 이점을 가지고 있기 때문이라 설명한다. 첫째, LCA가 모형학적 군집분석 방법이기 때문에 통계적 근거를 기반으로 군집의 개수를 설정할 수 있고, 둘째, 변수의 척도에 상대적으로 자유롭기 때문이며, 셋째, mixture 모델이기 때문에 LCA를 분석할 당시 포함되어 있지 않았던 다른 개체들이 데이터 업데이트를 통해 새롭게 분석의 대상으로 확장 될 때, 새로운 개체가 어떠한 군집에 포함될지를 예측할 수 있다는 장점을 가지고 있기 때문이라고 설명한다[1]. 구체적으로 살펴보자면 LCA는 모형학적 군집 방식이기 때문에 최대우도(maximum likelihood)를 통해 모수 추정이 이루어지고, 군집의 개수를 정하는 의사 결정 기준인 AIC, BIC 등이 사용되어 최적의 군집 개수가 결정되기 때문에 기존에 K-means의 가

장 큰 문제점인 군집의 개수를 정하는 것의 모호성을 논리적으로 해결할 수 있다는 특징을 가진다[24, 19]. 김영찬, 이두희[1]가 두 번째로 지적한 변수의 척도에 상대적으로 자유롭다는 의미는 앞서 살펴본 K-means나 SOM을 분석할 때 특정 척도로 설계된 설문지에서 사용할 경우 가장 정확한 시장 세분화 분석을 실행할 수 있는 것에 반해, LCA는 범주형 변수 혹은 수치형 변수를 모두 사용할 수 있다는 이점을 가지고 있다는 것을 뜻한다. 초창기 LCA 모형은 주로 범주형 변수를 기준으로 시장 세분화 분석을 실시할 때 많이 사용되었으나, 로지스틱 링크를 활용, 분석 대상이 잠재 군집에 속할 확률을 범주형 변수 이외에도 연속형 변수를 함께 고려하여 추정하는 방향으로 모형의 사용이 확장되고 있다[33]. 이를 “공변량이 함께 고려된 LCA 모형”이라 하며, 김영찬, 이두희[1]가 범주형 변수만 고려한 초창기 LCA 모형만을 사용하였던 것과는 다르게, 최근 연구는 범주형 변수와 연속형 변수를 모두 고려하는 LCA 모형을 사용하여 시장세분화를 진행하고 있다[16]. 마지막으로 새 데이터의 추가 부분에서 LCA의 장점은 fuzzy set의 경우는 이미 앞서 설명한 바와 같이 mixture model과 그 가정을 달리 하기 때문에 모형 자체가 분석 대상이 해당 군집에 속할 멤버십을 추정하고자 하는 모수인 데 반해(분석 대상은 둘 이상의 군집에 속할 수 있다는 가정을 반영), mixture model인 LCA 경우는 각 변수들 간의 관계를 통해 모델 모수가 추정되고 분석 대상이 사후적으로 해당 군집에 속할 확률이 계산되는 것이므로[24, 1], 새롭게 개체가 추가 된다고 할지라도, 기존의 추정된 모수는 유지되게 되므로, 추가된 새로운 분석 대상이 기존에 분류된 군집에 들어갈 확률을 예측할 수 있다는 특징을 가진다는 것이다. 그 외에도 LCA는 다양한 확률 분포에 대한 유연성을 가진다는 장점을 가지고 있어 시장 세분화를 할 때 적합한 분석 도구로 활용될 수 있다.

이 같은 장점에도 불구하고 LCA가 국내 경영학 분야에서 자주 접할 수 없었던 이유는 계산 상의 복잡성과 유료 통계 패키지를 사용해야 한다는 점에

서 그 한계를 나타냈기 때문이다. 그러나 최근 SAS에서 해당 프로그램을 무료로 배포하고 있어(<http://methodology.psu.edu>) 학계나 현업에서 현재보다 더 용이하게 해당 통계 분석 기법을 활용할 수 있을 것으로 예상된다.

## 2.2 시장 세분화 분석 시 고려사항

시장 세분화를 하기 위해서는 우선적으로 시장 세분화에 사용될 수 있는 변수가 무엇인지를 파악하여야 한다. 이러한 세분화 변수는 잠재적 소비자들을 동일한 그룹으로 분류하는 역할을 하게 되는데[50], 제품, 서비스 환경으로부터 독립적인 “일반적 요소”와 제품, 서비스, 특정 환경과 관련된 “제품 속성” 변수로 구분되며, 직접적으로 측정 가능한 “관찰 가능 변수”와 간접적으로 추론되어 측정되는 “관찰 불가능 변수”로 나눌 수 있다. 관찰 가능/일반적 요소의 예로는 사회 경제적 변수나 인구통계학적 변수 등이 존재하며, 관찰 가능/제품 속성 변수의 예로는 해당 제품의 사용 빈도, 혹은 로열티나 인터넷 사이트 방문 빈도 등이 있다. 반면, 관찰 불가능/일반적 요소로는 제품과 관련 없는 개인의 라이프 스타일, 개인의 가치관 등이 속할 수 있으며, 관찰 불가능/제품 속성인 변수로는 제품에 대한 속성 선호, 제품에 대한 가격 탄력성 등이 이에 속한다 하겠다[50]. 시장 세분화는 이러한 기초적인 변수들은 통해서 소비자들을 특정 그룹으로 구분하게 되는데, 시장 세분화 분석 기법을 선택 할 때에는 어떠한 변수를 선택해서 분석을 할 것인지를 결정하는 것 이외에도 해당 변수가 어떻게 측정되었느냐를 고려해야 한다.

앞서, K-means와 SOM이 데이터를 분석할 때, 유클리디안 거리를 기준으로 분석 대상을 군집에 할당한다고 설명하였다. 유클리디안 거리를 계산할 때 고려해야 할 가장 중요한 점은 해당 변수간의 차이가 수치적으로 그 의미를 가지고 있어야 한다는 점이다. 즉 변수가 어떻게 측정되었느냐에 따라서 분석 기법이 선택되는 것 또한 상이해진다는 의



미이다. 변수의 종류에는 명목척도, 서열척도, 등간척도, 비율척도가 있는데, 크게 범주형 변수와 수치형 변수로 다시 나뉘어 살펴볼 수 있다. 범주형 변수의 경우는, 해당 변수에 부여된 숫자가 임의적으로 부여된 숫자이므로, 숫자가 어떠한 의미론적 차이를 나타내지 않는다. 따라서 범주형 변수를 측정의 도구로 사용한 경우는 유클리디안 거리를 기준으로 유사성을 계산하는 K-means나 SOM과 같은 방식은 적합하지 않다는 것을 알 수 있다.

두 번째로 시장 세분화 분석 방법을 정하기 전 고려해야 하는 것은 데이터 상에 결측치가 얼마만큼 존재하느냐이다. 데이터의 결측치 양이 많은 경우, 예를 들어 특정 변수에서 데이터 결측치가 50% 가량 있는 경우, 나머지 50%를 가지고 데이터 분석할 경우 결과의 신뢰성에 문제가 나타날 수 있다. 그러나, 앞서 살펴본 K-means나 SOM은 모두 유클리디안 거리를 기준으로 분석 대상들을 군집화하기 때문에 결측치가 나타나게 되면, 거리 계산이 불

가능하기 때문에 결측치 양이 많은 경우는 해당 변수들을 제외하고 분석을 진행해야 하므로, 전체 수집된 데이터를 부분적으로만 활용할 수 있다는 약점을 가진다. 이뿐만 아니라, 이러한 유클리디안 거리를 기반으로 한 분석방법은 이상값(outlier)에 대해서도 큰 영향을 받기 때문에, 이상값의 여부에 따라서 시장 세분화 분석의 결과가 잘못 도출될 수 있다는 단점을 가진다.

### 2.3 시장 세분화 분석 기법에 따른 장단점

기존 문헌들을 고찰한 결과에 따라 각각의 시장 세분화 분석의 특징을 요약하면 <표 3>과 같다.

기존 문헌에 따라 LCA의 장점은 다음과 같이 요약될 수 있다. 첫째, LCA는 관찰된 범주변인과 관찰되지 않은 잠재적 범주변인의 확률적 관계를 파악하고, 이에 근거하여 각 사례들이 잠재변인의 각 ‘잠재계층’에 속할 확률을 도출하는 데에 있어 우도

<표 3> 시장 세분화 분석 기법 비교

	K-Means	SOM	LCA
측정 방법	유클리디안 거리를 기준으로 해당 변수에 응답 유사성과 그 차이에 기초하여 군집을 형성함	유클리디안 거리를 기준으로 해당 변수에 응답 유사성과 그 차이에 기초하여 군집을 형성함	관찰된 범주변인과 잠재적 범주변인의 확률적 관계를 파악하고, 이에 근거하여 각 사례들이 잠재변인의 ‘잠재 계층’에 속할 확률을 도출함
특성	nonoverlapping	초기 모형은 nonoverlapping 방식을 차용하였으나 이후 fuzzy set 모형으로 발전하고 있음.	Fuzzy 모형 중, mixture 모형에 해당됨. LCA는 추정된 모델 파라미터와 관찰된 반응에 따라 사후적으로 군집에 속할 확률이 계산됨.
사용 가능 변수 종류	등간척도/비율척도	등간척도/비율척도	변수의 척도로부터 상대적으로 자유로움(명목, 등간, 비율 척도 모두 사용 가능함)
데이터 수집상의 문제	결측치와 outlier에 따라 결과치가 다르게 나타날 수 있음.	결측치와 outlier에 따라 결과치가 다르게 나타날 수 있음	결측치와 outlier에 따라 결과치가 상대적으로 영향받지 않음.
장/단점	사용하기가 편리하고 용이함. 분석자가 미리 군집 수를 정해야 하기 때문에 통계적 근거가 없음.	데이터마ining 분야에서 많이 사용되고 있으며, 시각적으로 군집의 분류를 이해할 수 있다는 점에서 가장 큰 특징을 가지며 또한 응답자가 해당 멤버십에 속할 가중치를 계산해 준다는 이점을 가짐 가중치를 계산해주기 때문에 통계적 근거의 일부를 제시해줌.	다양한 데이터 셋에 적합하고 최근 SAS 무료 proc으로 배포되고 있어 사용상의 용이성을 가짐. 우도를 최대화하는 군집을 계산해주므로 군집의 개수를 정할 때, 통계적 근거를 명확하게 제시해줌.

함수를 활용하여 이를 최대화하는 모수를 추정한다는 계량적 모델이라는 점, 둘째, 현실을 잘 반영한 fuzzy mixture model을 사용한다는 점, 셋째, 다양한 척도가 사용된 설문문의 경우 다른 기법에 비해 상대적으로 적용이 훨씬 더 쉽다는 점, 넷째, 우도 함수를 통해 모수를 계산하기 때문에 각각의 변수들을 독립적으로 처리할 수 있어 이상값과 결측치로부터 상대적으로 자유롭다는 점, 마지막으로 군집의 개수를 결정할 때 통계적 근거를 명확하게 제시해준다는 점이다.

따라서 본 연구는 해당 LCA 모형을 기본으로 영화 소비자의 특성을 파악하고, 해당 분석 기법을 사용할 때 주의사항과 해당 분석 결과를 해석하는 방법에 대한 논의를 진행하고자 한다. 김영찬, 이두희 [1]의 연구가 LCA를 활용시, 범주형 변수만을 사용하여 인터넷 구매자의 특성을 파악하였던 것을 확장하여, 본 연구는 “공변량이 함께 고려된 LCA” 모형을 사용하여, 범주형 외에도 연속형 변수들을 동시에 사용하여 분석을 진행하고자 한다.

### 3. 연구 방법 및 결과

#### 3.1 표본과 자료수집

본 논문에서는 영화진흥위원회의 “2009~2010년 영화 소비자 설문 조사” 자료를 사용하였고[6, 7], 총 4000명 응답자 중 최종 3606명을 대상으로 적정 고객군이 분석되었다. 사용 변수들은 크게 영화 속성 변수와 영화 관람 동기 변수로 분류될 수 있다. 영화 속성 변수의 경우는 영화의 경험제적 특성 때문에 발생하는 불확실성을 해소하기 위해서 소비자들이 접하는 다양한 속성 정보들이 측정에 고려되었다. 그 이유는 제품 속성이 소비자들의 영화 선택 위험을 축소시키는 정보를 제공함으로써, 영화 선택 기준을 제시하고, 영화의 흥행 여부를 결정하는 요인으로 활용되기 때문이다[20, 36, 37]. 또한 영화 관람 동기는 소비자들의 영화 수용 여부에 그 영향을 미치고 소비자가 선호하는 영화 속성에 그 영향

을 미치기 때문에[14, 42, 44] 이 두 가지 요인이 영화 시장 세분화 과정에서 소비자의 특성을 분석하는데 유용한 정보를 제공할 것으로 예상된다. 본 연구는 영화진흥위원회가 발간한 2차 자료를 활용하였기 때문에, 관람 동기를 설명하고자, ‘극장 관람 행태’ 관련 변수들을 사용하여 분석하였다. 구체적으로 사용된 변수들을 설명하면, 영화 선택시 고려되는 속성(출연배우, 감독, 내용 줄거리, 제작 국가, 흥행 성적, 영화 장르, 영화제출품 여부, 주위 평가)과 연간평균 영화 관람편수 및 극장 관람경험, 선호 장르 및 국적, 영화 정보 취득 경로, 관람 시간, 관람 동반자, 영화 선택 시기, 개봉일 기준 영화 구매 여부, 극장 선택 기준, 입장권 구입 방법이 사용되었으며 그 외의 인구통계학적 변수인 성별, 연령대, 소득수준, 결혼 여부가 사용되었다. 총 22개의 변수가 사용되었으며, 그 중 범주형 변수의 경우는 11개, 수치형 변수는 총 11개가 사용되었다.

#### 3.2 분석 모형

LCA의 가장 큰 장점은 범주형 변수와 수치형 변수를 동시에 사용할 수 있다는 점이다. 그러므로 범주형 변수와 수치형 변수를 동시에 사용하고 있는 영화진흥위원회의 설문을 분석하기에 적합한 방식이라고 할 수 있겠다. 본 연구에서 사용하고 있는 2차 자료인 영화진흥위원회의 소비자 조사는 범주형 변수와 수치형 변수를 모두 사용하여 설문을 진행하였다. 따라서 본 연구는 LCA를 활용하여 각 사례가 소속할 세그먼트의 반응 패턴의 확률을 제시하였고, 이를 기초로 합리적이고 객관적인 영화 소비자 세분화를 시도하였다.

##### 3.2.1. 잠재계층 모형(Latent Class Model)

$$P(Y_i = y | X_i = x) = \sum_{l=1}^{n_c} \gamma_l(x) \prod_{m=1}^M \prod_{k=1}^{\gamma_m} \rho_{mkl}^{I(y_m=k)} \quad (1-1)$$

$$\gamma_l(x) = P(L_i = l | X_i = x) = \exp\{\beta_{0l} + x\beta_{1l}\} / \sum_{j=1}^{n_c} \exp\{\beta_{0j} + x\beta_{1j}\} \quad (1-2)$$

$Y_i$ 는 M개의 범주형 아이টে에 반응하는 i의 반응을 의미하며,  $n_c$ 는 전체 세그먼트의 수, m은 변수의 개수, k는 범주의 개수로 정의된다.  $L_i$ 는 i의 잠재계층 멤버십을 의미한다. 전통적인 LCA 모델에서 추정되어야 할 모수는 잠재군집에 속할 확률인  $\gamma$ 과 라미터와 잠재 군집에 속할 조건하에 아이টে에 반응할 확률인  $\rho$ 과 라미터로 구성되어 있는데,  $\rho$ 과 라미터는 실제로 관측된 아이টে와 잠재 계층간의 관련성을 설명하는 모수라 하겠다. 즉, 범주형 변수만이 사용될 경우는 해당 두 모수만을 추정하게 되지만, 공변량을 통해 수치형 변수까지 동시에 고려하게 되면, 로지스틱 회귀 계수인  $\beta$ 가 추가적으로 추정된다.  $\beta$  모수가 포함되는 경우, 실질적으로는  $\rho$ 와  $\beta$  값만이 추정되며,  $\gamma$ 는  $\beta$ 와 공변량의 함수로써 계산되게 된다. 즉,  $\gamma$ 를 예측하기 위해 공변량 변수를 사용하여 로지스틱 회귀분석에서의  $\beta$ 값이 추정되게 된다[33].

따라서 본 모델에서 수치형 변수들은 공변량으로 처리하였으며,  $\rho$ 값과  $\beta$ 값만이 실질적으로 계산되었고, 각 분석대상이 멤버십에 속할 확률이 추정되었다. 마지막으로 LCA 가정에 따라 범주형 변수는 LCA 내에서 정규분포를 따른다고 가정하였고, 수치형 변수는 다항분포(multinomial distribution)을 따른다고 가정하였다[22]. 이외에 LCA를 분석할 때 고려되어야 할 사항은 LCA SAS 코드와 함께 부록에 제시하였다.

### 3.2.2 잠재계층 분석기준

잠재 세분시장의 개수를 정하는 기준은 BIC(Bayesian Information Criterion)와 AIC(Akaike Information Criterion)가 가장 대표적으로 사용되는 방식인데, AIC는 같은 자료를 두 개 이상의 모델로 비교할 때 우도가 가장 적게 감소되는 것을 선택하는 방식이다. 그러나 AIC의 경우 계산하는 방식에서 샘플 수를 직접적으로 고려하지 않기 때문에 샘플 수에 따른 점근적 일관성(asymptotic consistency)을 보여주는 데에 약점을 가진다[19]. 따라서 본 연구에서는 BIC 방법에 근거하여 세분 시장의

개수를 결정하고 세분 시장 별 모수를 추정하였다.

$$\text{BIC 수식} : -2\ln p(x|k) \approx \text{BIC} = -2\ln L + k\ln(n) \quad (2-1)$$

### 3.3 세분화 방법에 따른 결과 분석

본 연구에서는 영화진흥위원회의 “2009~2010년 영화 소비자 설문 조사” 자료를 토대로 총 4000명 응답자 중 최종 3606명을 대상으로 22개의 소비자의 행동 관련 변수를 활용하여 분석하였다. 분석 결과, 11개의 세그먼트가 BIC 값에서 가장 적은 숫자를 나타내어 11개의 세그먼트를 ‘최적군집숫자’라고 결정하였다.

<표 4> LCA 결과

군집숫자	로그-우도	BIC
7	-39116.520	79908.399
8	-38914.340	79749.474
9	-38775.910	79718.049
10	-38607.670	79627.003
11	<b>-38460.420</b>	<b>79577.938</b>
12	-38353.320	79609.271

이 같은 LCA의 결과가 K-means와 SOM의 결과보다 우월한지를 검증하기 위해, 본 연구는 추가적으로 동일한 표본을 대상으로 K-means와 SOM을 활용하여 군집 분석을 실행하였다.

<표 5>에 따르면 동일한 샘플을 대상으로 군집 분석을 실행함에도 불구하고, 각 군집 분석 방법별 군집 분류가 상이한 것을 알 수 있다. Magidson and Vermunt[38]의 연구의 경우, LCA와 K-means를 비교하기 위해, 시뮬레이션 기법을 활용하여, 시뮬레이션으로 형성된 세그먼트와 LCA와 K-means로 형성된 세그먼트의 정도를 직접적으로 비교하는 수단을 활용하여 LCA의 우월성을 검증하였으나, 본 연구는 실제 2차 자료를 대상으로 분석을 실행하였기 때문에 이 같은 직접 비교 수단이 아닌 간접 비교 수단을 사용하여 각 세그먼트 방법간의 효과 비교를 진행하였다. 이를 위해, 각 변수 별로 집단에 미치는 효과크기를 기준으로 각 분석 방법을 비교

〈표 5〉 군집 분석 방법 별 군집 크기 비교

군집 번호	LCA		SOM		K-Means	
	빈도	퍼센트	빈도	퍼센트	빈도	퍼센트
1	239	6.63	433	12.01	215	5.96
2	55	1.53	575	15.95	375	10.40
3	413	11.45	145	4.02	302	8.37
4	265	7.35	264	7.32	358	9.93
5	777	21.55	342	9.48	496	13.75
6	79	2.19	393	10.90	409	11.34
7	355	9.84	229	6.35	278	7.71
8	290	8.04	222	6.16	385	10.68
9	330	9.15	461	12.78	376	10.43
10	259	7.18	199	5.52	133	3.69
11	544	15.09	343	9.51	279	7.74

하고, 이에 따라 효과적인 세그먼트 방법을 검증하고자 한다.

### 3.3.1 시장 분석 방법 별 수치형 변수 결과

수치형 변수의 경우, 개념적으로 각 분석 방법간 사용상의 문제점이 발생하지 않으므로, 집단 분류에 미치는 영향 크기는 유사할 것으로 예상하였다. 이러한 예상을 검증하기 위해 다변량 분석을 실시하였다. 분석 결과, 집단 분류에 변수가 미치는 영향은 모두 통계적으로 유의하게 나타났다. 따라서 추가적인 검증을 위해 각 변수 별 군집 분류의 효과 크기를 추정하였다. 그 결과, 수치형 변수 중, 변수

간 가장 그 분산의 정도가 큰 연령 변수가 LCA를 비롯하여, K-means와 SOM이 집단을 분류하는데 미치는 효과가 가장 큰 것으로 나타났다. 즉, 각 시장 세분화 분석 방법별로 연령 변수가 해당 집단을 분류하는데 가장 큰 영향력을 미쳤다는 것을 알 수 있다.

그러나 SOM과 K-means의 경우 연령 변수를 제외하고는 집단 분류에 미치는 효과크기가 미비한 것에 반해 LCA의 경우, 분산이 큰 연령 변수의 효과크기의 가중치를 고려하고, 연령을 제외한 다른 변수들의 영향력은 균일하게 분포하는 경향을 나타내는 것을 알 수 있다.

### 3.3.2 시장 분석 방법별 범주형 변수 결과

LCA의 가장 큰 특징은 이미 이론적 배경에서도 밝혔듯이, 범주형 변수 분석에서의 상대적 우월성을 가지고 있다는 점이다. 수치형 변수에서의 결과만을 가지고는 LCA의 우월성을 검증하기에는 한계를 가지고 있으므로, 동일한 분석 방법을 기초로 각 집단 별 범주형 변수에 미치는 효과크기를 Cramer의 V 값으로 추정하였다.

〈표 7〉의 결과에서도 알 수 있듯이, 각 세분화 분석 방법 별로 모든 범주형 변수가 통계적으로 유의한 결과를 나타냈다. 그러나 범주형 변수 별 카이제곱 값의 경우, LCA의 값이 다른 군집 분석 방법 보

〈표 6〉 시장 분석 방법 별 수치형 변수 결과

	다변량 검정	LCA	SOM	K-means
		부분 에타 제곱	부분 에타 제곱	부분 에타 제곱
Q2_1	세부연령	0.73	0.98	0.99
Q19	영화 선택시 고려 요인-배우	0.12	0.12	0.06
Q20	영화 선택시 고려 요인-감독	0.11	0.15	0.10
Q21	영화 선택시 고려 요인-내용, 줄거리	0.10	0.02	0.03
Q22	영화 선택시 고려 요인-국가	0.11	0.07	0.04
Q23	영화 선택시 고려 요인-홍행성적 및 순위	0.12	0.23	0.13
Q24	영화 선택시 고려 요인-장르	0.10	0.04	0.06
Q25	영화 선택시 고려 요인-영화제 출품 및 수상	0.13	0.15	0.10
Q26	영화 선택시 고려 요인-주변인의 평가	0.10	0.14	0.07
Q103_2	월평균가구소득	0.12	0.08	0.08
Q104	1년간 영화관람 편수	0.21	0.08	0.08

〈표 7〉 시장분석 방법 별 범주형 변수 결과

	LCA			SOM			K-means		
	카이제곱	유의확률	Cramer의 V	카이제곱	유의확률	Cramer의 V	카이제곱	유의확률	Cramer의 V
성별	515.201	0.000	0.378	48.325	0.000	0.116	33.308	0.000	0.096
선호장르	2502.448	0.000	0.265	449.349	0.000	0.112	480.545	0.000	0.116
국적	1166.596	0.000	0.255	257.809	0.000	0.12	232.971	0.000	0.114
정보경로	1159.023	0.000	0.232	285.69	0.000	0.115	268.873	0.000	0.112
관람시기	626.673	0.000	0.295	91.269	0.000	0.112	97.936	0.000	0.117
관람동반	2058.092	0.000	0.419	1156.171	0.000	0.314	1116.031	0.000	0.309
선택시기	226.162	0.000	0.25	159.992	0.000	0.211	158.204	0.000	0.209
관람시기	1178.879	0.000	0.33	193.36	0.000	0.134	165.039	0.000	0.124
극장기준	1106.598	0.000	0.227	315.524	0.000	0.121	294.799	0.000	0.117
티켓구입	891.971	0.000	0.255	300.445	0.000	0.148	299.282	0.000	0.148
결혼여부	3540.339	0.000	0.991	2344.435	0.000	0.806	2291.125	0.000	0.797

다 훨씬 우월하다는 것을 알 수 있다. 이뿐만 아니라, 각 변수별 효과크기를 나타내는 Cramer의 V값 역시 SOM이나 K-means에 비교하여 LCA가 더 우월한 수치를 나타내는 것을 알 수 있다. 이 같은 결과가 의미하는 바는 범주형 변수를 사용하는데 있어서 LCA가 다른 분석 기법과 비교하였을 때, 각 군집을 구분하는데 있어 더 큰 영향력을 미쳤다는 것을 뜻한다. 즉, SOM이나 K-means의 경우, 범주형 변수가 통계적으로 유의한 결과를 나타내고는 있지만 실질적으로 변수간의 차이가 두드러지게 나타나지 않는다는 것을 알 수 있다. 따라서 범주형 변수를 사용할 때 해당 방법을 사용하는 것은 그 한계를 가지고 있다는 것을 간접적으로 제시하는 바이다.

### 3.4 LCA를 활용한 영화 소비자들의 세분군집 특성

시장 세분화 분석 방법 별 결과를 토대로 LCA가

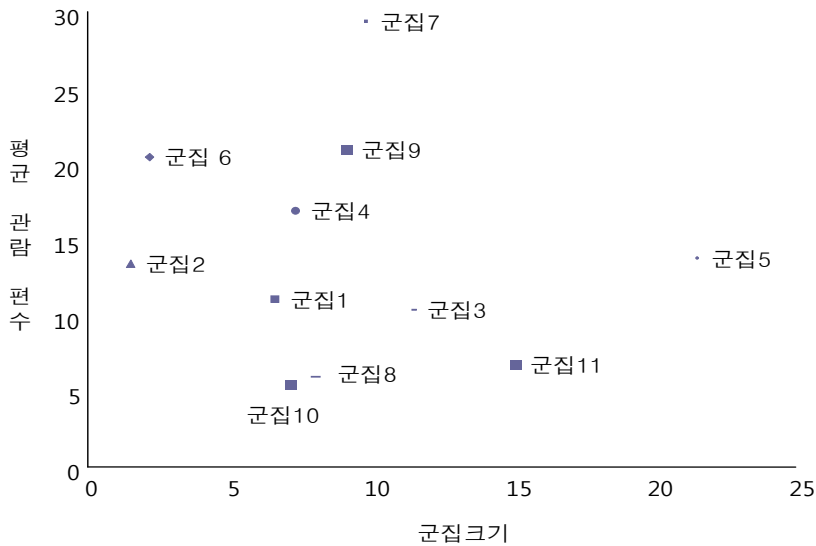
다른 방법에 비해 간접적으로 더 효과적이라는 것을 검증하였다. 따라서 본 항에서는 LCA에 기초하여 영화 소비자들의 세분 군집의 특성을 이해하고자 한다.

관측수 기준(세그먼트 크기)으로 고객군의 크기를 보면 군집 5가 21.5%로 가장 많은 고객수를 보유하고 있으며 군집 11(15.1%), 군집 3(11.5%)이 그 뒤를 잇고 있다. 평균 관람 편수 기준으로 보면 군집 7이 29.34편으로 가장 많은 관람 편수를 기록했고 군집 9(20.94편)와 군집 6(20.41편)이 2, 3위를 기록하고 있다.

고객군 당 투입되는 마케팅 비용이 일정하다는 가정 하에, 평균관람편수와 고객군 크기를 비교한 결과, 고객군의 사업적인 매력도를 비교해 보면, 관측수가 많아 고객군의 크기가 크고 평균 관람 편수가 높은 고객군일수록 투자 수익률이 높을 것으로 예상된다. 반면 향후 고객군별 전체 관람 편수를 증가시키기 위해서는 상대적으로 현재 관람 편수가 적

〈표 8〉 세그먼트 특성

군집	1	2	3	4	5	6	7	8	9	10	11	전체
관측수	239	55	413	265	777	79	355	290	330	259	544	3606
군집크기	6.6	1.5	11.5	7.3	21.5	2.2	9.8	8.0	9.2	7.2	15.1	100.0
평균관람편수	11.03	13.40	10.34	16.86	13.75	20.41	29.34	5.93	20.94	5.41	6.75	13.46
관람력지수	0.54	0.15	0.88	0.92	2.20	0.33	2.15	0.35	1.42	0.29	0.76	



[그림 3] 고객군 크기

은 하단에 위치한 고객군이 매력적으로 보일 수 있다. 단 문제는 해당 고객군의 낮은 평균 관람 편수를 높일 수 있는 적절한 마케팅 활동이 수반될 필요가 있다. 보조적인 지표로 고객군의 크기와 평균관람편수를 동시에 고려하여 ‘관람력 지수’로 계산해 보면 관람력 지수는 1을 기준으로 지수가 1보다 큰 경우 해당 고객군은 그 절대적 크기가 평균보다 크거나 해당 고객군에 속해있는 개인들이 전체 평균 보다 더 자주 영화를 관람한 것을 의미한다. 관람력 지수를 기준으로 보았을 때는, 군집 5가 2.20으로 가장 크고 군집 7(2.15), 군집 9(1.42)가 뒤를 따르고 있다.

### 3.5 고객군 특징 결과

범주형 변수의 경우, 변수와 군집간의 관계를 파악하고자 교차분석을 실시하여 카이제곱 값과 Cramer의 V 값을 분석하였다. 카이제곱 값의 통계적 유의성 이외에도 Cramer의 V가 변수간의 영향력을 설명하는 데에 사용된 이유는 변수의 유의수준을 설명하는 카이제곱의 경우 샘플 수에 민감하게 반응하기 때문에 샘플수가 많을수록, 모든 범주형 변수가 세그먼트를 잘 설명하고 있다고 분석된다. 그

러므로 카이제곱의 유의성만을 가지고는 각 변수별 세그먼트 분류에 미치는 영향력을 계산하기 어렵다는 단점을 가진다. 이를 보완하기 위해 본 연구에서는 샘플의 수가 모형에 고려된 Cramer의 V를 사용하여 세그먼트 분류에 가장 큰 영향력을 미친 변수들의 상위 랭크를 파악하였다. Cramer의 V 수치가 높은 변수가 군집간의 이질성을 대변하기 때문에, 군집의 특성을 이해하는 데에 유용한 지표를 제공하기 때문이다. <표 9>에 따르면, 각 변수가 군집 분류에 미치는 영향은 모두 통계적으로 유의한 것으로 나타나고 있으나, 각 변수가 집단에 미치는 영향력을 나타내는 Cramer의 V 값은 각 변수별로 상이한 것으로 나타나고 있다. 구체적으로 혼인 여부, 극장 관람 시 누구와 함께 관람하는지, 성별, 개봉일 기준 극장 관람 시기 등의 순서로 집단에 미치는 효과 크기가 큰 것으로 나타나고 있다. 이 같은 결과는 해당 변수들이 상대적으로 다른 변수들에 비해 군집 특성 파악에 더 큰 영향력을 미치는 것을 의미한다. 따라서 Cramer의 V 값에서 상위를 차지한 혼인여부, 극장 관람시 동반자, 성별, 개봉일 기준 극장 영화관람시기, 관람방법 별 관람시기 변수가 군집의 특성을 이해하는데 사용되었다.

〈표 9〉 범주형 변수 결과

	카이제곱	유의확률	Cramer의 V
혼인여부별	3540.339	0.000	0.991
극장 관람 시 동반자	2058.092	0.000	0.419
성별	515.201	0.000	0.378
개봉일 기준 극장 영화 관람 시기	1178.879	0.000	0.330
관람방법별 관람 시기_극장	626.673	0.000	0.295
관람방법별 자주 보는 영화 장르_극장	2502.448	0.000	0.265
선호하는 영화 국적	1166.596	0.000	0.255
극장 입장권 구입 방법	891.971	0.000	0.255
극장 관람 영화 선택시기	226.162	0.000	0.250
영화 선택 시 정보 취득 경로	1159.023	0.000	0.232
극장 선택 기준	1106.598	0.000	0.227

〈표 10〉 수치형 변수 결과

Covariate	Exclusion LL	Change in 2×LL	자유도	유의확률
세부연령	-39746.052	2571.271	10	0.000
영화 선택 시 고려 요인_배우	-38492.837	64.843	10	0.000
영화 선택 시 고려 요인_감독	-38558.757	196.683	10	0.000
영화 선택 시 고려 요인_내용, 줄거리	-38502.221	83.611	10	0.000
영화 선택 시 고려 요인_국가	-38472.603	24.374	10	0.007
영화 선택 시 고려 요인_흥행성적 및 순위	-38494.596	68.361	10	0.000
영화 선택 시 고려 요인_장르	-38481.860	42.887	10	0.000
영화 선택 시 고려 요인_영화제 출품 및 수상	-38482.434	44.037	10	0.000
영화 선택 시 고려 요인_주변인의 평가	-38474.137	27.442	10	0.002
월평균 가구소득	-38505.815	90.798	10	0.000
1년간 영화관람 편수	-38621.119	321.406	10	0.000

반면, 수치형 변수에서는 해당 분석시 고려된 인구통계학적 변수(연령대, 소득수준)와 연간평균 관람편수 및 영화 선택시 고려 기준 변수들이 집단 분류에 미치는 영향력을 로지스틱 회귀분석을 사용, 95% 신뢰수준을 기준으로 분석하였다. 그 결과, 분석에 포함된 각 변수들이 통계적으로 유의한 수치를 나타내는 것으로 분석되었다.

이 같은 결과를 토대로 각 군집의 특성을 요약하면 다음과 같다. 군집 1의 경우는 친구와 함께 영화 관람을 즐기는 30대 초반 남성으로 구성되어 있으며, 영화의 엔터테인먼트적 요소를 선호하는 경향을 나타냈다. 군집 2의 경우는 40대 초반의 남성 집단으로 다양한 영화 장르를 선호하고 상대적으로 예술영화를 선호하여, 주류 영화보다는 비주류 영화를 선호하는 경향을 나타냈다. 즉, 영화의 예술적

측면을 중요하게 생각하는 경향을 가지고 있으며, 자신만의 장르적 취향이 있고, 영화 소비를 통해 학습적 지식 탐구의 동기를 나타내고 있다. 군집 3의 경우, 30대 후반 남/여로 구성되어 있으며, 주로 배우자와 함께 영화를 관람하는 집단으로 영화를 여가 시간 활용의 수단으로 활용하는 경향을 보인다는 것을 알 수 있다. 반면 군집 4는 커플들로 구성되어 있는 10대 후반의 집단으로, 사회적 친목 교류 및 데이트의 일환으로 영화를 소비하는 경향을 보였으며 상황에 따라 적합한 영화 장르를 선택하는 경향을 나타내고 있다. 군집 5는 20대 초반의 미혼자로 구성되어 있고, 친구와 영화를 관람하려 하는 것을 즐기는 집단으로 나타났으며, 주로 커뮤니케이션 목적으로 영화를 소비하는 경향을 가지고 있다. 군집 6의 경우, 기혼자로 구성되어 있으나, 배

우자보다는 친구와 함께 영화를 관람하는 것을 즐기는 30대 후반의 집단으로, 혼자 영화를 관람하는 비율이 상대적으로 높고, 다른 군집에 비해 영화제출품 여부를 중요하게 생각하는 군집 특성을 보여, 영화 상품의 고관여적 성향의 특징을 가지고 있다고 해석된다. 20대 후반의 미혼여성으로 구성된 군집 7의 경우, 다른 집단에 비해 개봉 일주일 이내 극장을 찾는 경향이 가장 높게 나타났으며, 연간 29편의 영화를 관람하는 고관여적 특징을 나타내어 영화 제품의 얼리어답터 성향을 나타내고 있다. 군집 8의 경우는 배우자와 함께 극장을 찾는 40대로 구성되어 있으며, 배우자와의 여가 생활을 위해 영화를 관람하는 경향을 나타냈고, 다른 집단에 비해 연간 영화 관람율이 저조한 것으로 나타났다. 군집 9는 결혼한 30대 후반의 여성들이 주를 이루고 있었으며, 배우자 뿐만 아니라 친구와 함께 영화를 관람하는 것을 즐기는 집단으로, 가족과의 의사소통 목적뿐만 아니라 친구와의 사회적 교류 목적으로 영화를 관람하는 경향을 나타내고 있다. 군집 10은 40대의 여성들이 배우자나 친구와 함께 소통하는 수단으로 위해 영화를 구매하는 것으로 나타났으며, 주말뿐만 아니라 주중에도 영화를 관람하는 특징을 나타내고 있다. 마지막으로 군집 11은 30대 후반의 여성 집단으로, 다양한 영화 장르를 선호하고 특히 애니메이션과 같은 가족 영화를 선호하는 경향을 나타내고 있어 자녀와의 영화 관람이 높은 것으로 추론된다. 세부적인 분석 사항은 부록에 기입하였다.

## 4. 연구의 시사점 및 향후연구 제언

### 4.1 연구의 시사점

이 연구는 시장세분화 분석 기법 사용시 고려되어야 할 사항과 이에 따라 활용 가능한 시장세분화 분석 기법을 소개하고 영화 소비 경험에 따른 소비자 세분화 연구를 시도하였다. 이 연구는 소비자의 영화 선택 의사 결정 과정에 영화의 제품적 속성과 영화 관람 동기가 그 영향을 미침을 가정하고,

이 두 요인에 따라 영화 소비자 특성간의 분류가 가능함을 시사하였다.

본 연구는 매해 영화 진흥위원회에서 실시하는 관객 소비자 조사 데이터를 토대로 영화 소비자 특성을 규정하였고 이를 토대로 구체적인 영화 소비자의 특성 분석 방향을 제시하였다는 점에서 그 실용적 의의가 있다. 이 연구의 가장 큰 특징은 LCA를 활용하여 시장 세분화 분석에서의 문제점을 최소화하고 자의적 군집 수의 결정이 아닌 계량적 방법으로 최적군집숫자를 제시하였다는 것이다. 연구 결과 잠재계층 모형을 통해 분석된 영화 산업 시장의 최적군집숫자는 11개였으며, 이는 BIC값을 기준으로 결정되었다. 이 연구는 다양한 시장 세분화 기법을 소개하고 각 분석 기법의 장단점을 설명하여, 현업에서 시장 세분화 분석을 진행할 시 유용한 정보를 제공할 것으로 예상된다.

특히, 이 연구는 영화시장을 그 타깃으로 분석하였기 때문에 영화 관련 종사자들에게 유용한 정보를 제공할 것으로 예상된다. 첫째, 이 연구는 각 세그먼트별로 전략적 콘텐츠 개발을 가능하게 하는 방향성을 제시하였다는 점에서 그 의의를 가진다. 즉, 각 세그먼트 별로 보이는 특징을 토대로 영화 소재와 주제 선정의 방향성을 제시하고, 고객 맞춤형 콘텐츠 개발로 인한 투자 자본의 효율적 관리를 가능하게 할 것으로 예상된다. 또한 각 군집을 대상으로 한 콘텐츠를 개발할 경우, 해당 군집의 크기를 비교하여 매력도를 평가할 수 있으므로, 개봉 이전 해당 제품의 수요의 상당부분을 예측할 수 있고, 또한 투자자 입장에서 해당 영화에 대한 투자 대비 수익성을 미리 예측하여 제작을 진행할 수 있으므로, 투자 수익률을 증대시킬 수 있을 것이라고 예상된다. 두 번째로 마케팅 관점에서 이 연구가 제시하는 시사점은 영화 산업은 그 제품의 특징상 상품이 출시 되기 전, 제작 단계에서부터 다양한 경로로의 영화 정보 노출이 가능하기 때문에, 군집 별로 효과적인 영화 정보 노출을 가능하게 하여, 영화 홍보의 효율성을 확대시킬 수 있다는 것이다. 연구 결과에 따르면 각 세그먼트는 빈번하게 이용하는 영화 정



보의 경로가 상이한 것으로 나타난다. 이러한 결과는 타깃 세그먼트에 따라, 영화 정보의 집중적 노출의 차별화 전략 수립을 가능하게 한다 하겠다.

#### 4.2 연구의 한계점 및 향후 연구 제언

이 연구는 영화진흥위원회가 배포한 2차 자료를 분석함으로써 각 세그먼트의 특징을 살펴보았다. 이 자료는 다양한 인구통계학적 변수들을 토대로 표본들을 구성하고 있기 때문에 본 연구의 질을 향상시키는 데에 적합한 자료 선정이었지만, 동시에 영화진흥위원회의 설문 문항 사용으로 인한 측정의 한계를 가지고 있다. 첫째, 50세 이상의 영화 소비자 표본은 분석의 대상에서 제외되었기 때문에 해당 연령에 대한 분석의 한계를 가진다. 둘째, 영화진흥위원회가 배포한 2009~2010년 관객 조사 데이터의 경우, 선호하는 장르나 국적을 최대 2~3순위까지 복수 응답을 할 수 있도록 구성되어있었다는 점이다. 그러나 연구 분석에 있어서는 이 순위 데이터가 의미하는 '순위 간 유의미한 차이'를 제시할 수 없었다. 따라서 본 연구에서는 '선호하는 장르/국적 1순위' 데이터를 활용하여, 연구를 진행하였다. 향후 연구에서는 순위 데이터를 효과적으로 활용한 분석이 시도되어야 할 것이다. 마지막으로 본 연구는 2차 자료를 대상으로 분석을 실시하였기 때문에 시장세분화 분석 방법 간의 차이를 직접적으로 비교할 수 없었다는 한계를 가진다. 따라서 향후 연구에서는 이 연구가 제시한 3가지 시장세분화 기법을 시뮬레이션 기법을 활용하여 직접적으로 비교 분석하는 연구가 수행되어야 할 것이다.

#### 참 고 문 헌

- [1] 김영찬, 이두희, "인터넷 사용자가 모두 인터넷 구매자인가?", 『소비자학연구』, 제13권, 제1호(2002), pp.233-256.
- [2] 변현수, 『1999~2008년 한국 영화관객 성향변화 분석』, 영화진흥위원회, 2011.
- [3] Squire, J., 강인형 옮김, 『할리우드의 영화 산업-기획부터 상영까지』, 길벗, 1997.
- [4] 영화진흥위원회, 『2009년 한국 영화산업 실태 조사와 한국영화 투자 수익성 분석』, 영화진흥위원회, 2011.
- [5] 영화진흥위원회, 『2010년 한국 영화산업 결산』, 영화진흥위원회, 2011.
- [6] 영화진흥위원회, 『2009 영화 소비자 조사』, 영화진흥위원회, 2010.
- [7] 영화진흥위원회, 『2010 영화 소비자 조사』, 영화진흥위원회, 2011.
- [8] 이수범, 『영화마케팅 PR론』, 나남신서, 2005.
- [9] 이준배, 손찬수, 정택구, 이현경, "SKC 모형을 이용한 양돈 농가의 경쟁적 벤치마킹 전략 분석", 『농업경제연구』, 제50권, 제1호(2009), pp. 1-32.
- [10] Jowett G. and Linton, 김훈순 옮김, 『영화 커뮤니케이션(Movies as Mass Communication)』, 나남신서, 1994.
- [11] 최광희, "Special 상반기 한국영화 동향과 전망", 『이슈페이퍼, CINNO』, 제3호(2009).
- [12] Arabie, P., "Clustering Representations of Group Overlap," *Journal of Mathematical Sociology*, Vol.5, No.1(1977), pp.113-128.
- [13] Auerbach, K.J. and L.M. Collins, "A multi-dimensional developmental model of alcohol use during emerging adulthood," *Journal of Studies on Alcohol*, Vol.67, No.6(2006), pp. 917-925.
- [14] Austin, B.A., *Immediate seating : A look at movie audiences*, Wadsworth Publishing Company, Belmont, 1989.
- [15] Batra, R.J.M. and D. Aaker, *Advertising Management 5th edition*, Prentice Hall, NJ, 1996.
- [16] Chung, H., B.P. Flaherty, and J.L. Schafer, "Latent-class logistic regression : Application to marijuana use and attitudes among high-school seniors," *Journal of the Royal Statis-*

- tical Society : Series A*, Vol.169, No.4(2006), pp.723-743.
- [17] Chung, H., Y. Park, and S.T. Lanza, "Latent transition analysis with covariates : Pubertal timing and substance use behaviors in adolescent females," *Statistics in Medicine*, Vol. 24, No.18(2005), pp.2895-2910.
- [18] Collica, R.S., *CRM Segmentation and Clustering : Using SAS Enterprise Miner*, SAS Press Series, NC, 2007.
- [19] Dayton, C.M., *Latent Class Scaling Analysis*, Sage Publications Inc., CA, 1998.
- [20] Elberse, A. and J. Eliashberg, "Demand and Supply Dynamics for Sequentially Released Products in International markets : The Case of Motion Pictures," *Marketing Science*, Vol. 229, No.3(2003), pp.329-354.
- [21] Everitt, B.S. *Cluster Analysis*, Edward Arnold, London, 1992.
- [22] Green, P.E., F.J. Carmone and Y. Wind, "Subjective Evaluation Models and Conjoint Measurement," *Behavioral Science*, Vol.17, No.13 (1972), pp.288-299.
- [23] Grover, R. and V. Srinivasan, "A Simultaneous Approach to Market Segmentation and Market Structuring," *Journal of Marketing Research*, Vol.24, No.2(1987), pp.3-19.
- [24] Hagenaars, J.A. and A.L. McCutcheon, *Applied Latent Class Analysis*, Cambridge University Press, NY, 2002.
- [25] Hruschka, H., "Market Definition and Segmentation Using Fuzzy Clustering Methods," *International Journal of Research in Marketing*, Vol.3, No.2(1986), pp.117-134.
- [26] Huntsberger, T.L. and P. Ajjimarangsee, "Parallel self-organizing feature maps for unsupervised pattern recognition," *International Journal of General Systems*, Vol.16, No.4(1990), pp.357-372.
- [27] Jackson, K.M., J.J. Sher, H.J. Gotham, and P. K. Wood, "Transitioning into and out of large-effect drinking in young adulthood," *Journal of Abnormal Psychology*, Vol.110, No.3 (2001), pp.378-391.
- [28] Kamakura, W.A. and J.A. Mazzon, "Value Segmentation : A Model for the Measurement of Values and Value Systems," *Journal of Consumer Research*, Vol.18, No.2(1991), pp.208-218.
- [29] Kamakura, W.A. and T.P. Novak, "Value-System Segmentation : Exploring the Value of LOV," *Journal of Consumer Research*, Vol.19, No.1(1992), pp.119-132.
- [30] Kaski, S., "Data exploration using self-organizing maps," *Mathematics, Computing and Management in Engineering Series*, No.82, (1997), pp.1-57.
- [31] Kohonen, T., "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, Vol.43, No.1(1982), pp.59-69.
- [32] Kotler, P. and G. Armstrong, *Principles of Marketing 13th edition*, Prentice Hall, NJ, 2009.
- [33] Lanza, S.T., L.M. Collins, D.R. Lemmon, and J.L. Schafer, "PROC LCA : A SAS Procedure for Latent Class Analysis," *Structural Equation Modeling*, Vol.14, No.4(2007), pp.671-694.
- [34] Lehman, D.R., W.L. Moore, and T. Elrod, "The Development of Distinct Choice Processes Over Time : A Stochastic Modeling Approach," *Journal of Marketing*, Vol.46, No.2 (1982), pp.48-59.
- [35] Lilien, G. and A. Rangaswamy, *Marketing Engineering : computer-assisted marketing analysis and planning 2nd edition*, Pearson, 2002.

- [36] Linton J.M. and J.A. Petrovich, "The Application of the Consumer Information Acquisition Approach to Movie Selection : An exploratory Study," in Austin, B.A.(ed), *Current Research in Film : Audiences, Economics and Law*, Ablex Publishing, NJ, Vol.4(1988).
- [37] Litman, B.R. and L.S. Kohl, "Predicting financial success of motion pictures," B.R. Litmaned, *The Motion Picture Mega-Industry*, Allyn and Bacon, MA, 1989.
- [38] Magidson, J. and J.K. Vermunt, "Latent class models for clustering : A comparison with K-means," *Canadian Journal of Marketing Research*, Vol.20(2002), pp.37-44.
- [39] Markey, M.K., J.Y. Lo, G.D. Tourassi, and C. E. Floyd Jr., "Self-organizing map for cluster analysis of a breast cancer database," *Artificial Intelligence in Medicine*, Vol.27, No.2 (2003), pp.113-127.
- [40] Mazanes, J.A., "Neural market structure analysis : Novel topology-sensitive methodology," *European Journal of Marketing*, Vol.35, No. 7/8(2001), pp.894-916.
- [41] Milligan, G.W. and M.C. Cooper, "An Examination of Procedures for Determining the Number of Clusters in a Dataset," *Psychometrika*, Vol.50, No.2(1985), pp.159-179.
- [42] Moller, K.E.K. and P. Karppinen, "Role of Motives and Attributes in consumer motion picture choice," *Journal of Economic Psychology*, Vol.4, No.3(1983), pp.239-262.
- [43] Neelamegham, R. and D. Jain, "Consumer Choice Process for Experience Goods : An Econometric Model and Analysis," *Journal of Marketing Research*, Vol.36, No.3(1999), pp. 373-386.
- [44] Palmgreen, P. and P.A. Lawrence, "Avoidances, Gratifications and Consumption of Theatrical Films : the rest of the story," in Austin, B.A.(ed), *Current Research in Film : Audiences, Economics and Law*, Ablex Publishing, NJ, Vol.5(1991).
- [45] Punj, G. and D.W. Stewart, "Cluster Analysis in Marketing Research : Review and Suggestions for Application," *Journal of Marketing Research*, Vol.20, No.2(1983), pp.134-148.
- [46] Sawhney, M.S. and J. Elisashberg, "A Parsimonious Model for Forecasting Gross Box Office Revenues of Motion Picture," *Marketing Science*, Vol.15, No.2(1996), pp.113-131.
- [47] Shepard, R.N. and P. Arabie, "Additive Clustering : Representation of Similarities as Combinations of Discrete Overlapping Properties," *Psychological Review*, Vol.86, No.2 (1979), pp.87-123.
- [48] Shugan, S., "Recent Research on the Motion Picture Industry," *Proceedings of the Inaugural Business and Economics Scholars Workshop in Motion Picture Industry Studies*, (2000), pp.65-86.
- [49] Vuorimaa, P., "Fuzzy self-organizing map," *Fuzzy Sets and Systems*, Vol.66, No.2(1994), pp.224-231.
- [50] Wedel, K. and W. Kamakura, *Market Segmentation : Conceptual and Methodological Foundations 2nd edition*, Springer, 2000.
- [51] Zadeh, L.A., "Fuzzy sets," *Information and Control*, Vol.8, No.3(1965), pp.338-353.

## 〈부 록〉

### 1. SAS Proc 예시

〈표 11〉 SAS Proc 예시

	코드 예시
파라미터 추정	<pre> procclca data = sasoutest = est4 outparam = test4; nclass 4; items q1 q2 q3 q4 q5 q6 q7 q8 q9 q10 q11; categories 2 13 6 7 3 5 2 4 7 6 2; covariates a1 a2 a3 a4 a5 a6 a7 a8 a9 a10 a11; id no; seed 518165; run; </pre>
멤버십 부여	<pre> data seg4_mem; set datal; segm = max(of postlc1-postlc8); if segm = postlc1 then seg = 1; if segm = postlc2 then seg = 2; if segm = postlc3 then seg = 3; if segm = postlc4 then seg = 4; run; </pre>

SAS Proc은 파라미터 추정 용 코드와 멤버십 부여를 위한 코드 두 가지로 크게 분류될 수 있다. 파라미터 추정 코드의 첫줄을 보면, 해당 데이터를 불러올 때(data = sas) 필수적으로 입력해야하는 코드와 각각의 모수를 각각의 데이터로 불러드리기 위해 사용된 선택적 코드가 기입되어 있다(outest, outparam). 둘째 줄에 기입된 nclass는 세그먼트의 숫자를 기입하는 코드로 필수적으로 기입되어야 하며, items에 표시된 11개의 변수는 범주형 변수를 의미한다. categories라고 기입된 코드는 각각의 범주형 변수들의 카테고리 개수가 기입되어야 하는 코드로, 이 역시 필수적으로 기입되어야하는 사항임을 명심해야한다. 이때, 각 변수들은 데이터 상에 기입된 순서대로 코드를 기입해야한다는 것을 유의한다. covariates의 경우는 수치형 변수를 기입하는 코드로써, 해당 설문지에 수치형 변수를 사용하지 않은 경우는 분석 코드에서 제외된다. 그 밖에도 id, seed는 선택적으로 코드에 기입할 수 있다. 마지막으로 멤버십 부여 코드의 경우는 군집의 개수별로 postlc의 개수를 지정함을 유의해야한다.

### 2. 군집 별 분석 결과

#### 군집 1(크기 : 6.6%, 평균 관람 편수 : 11.03편, 관람력 지수 : 0.54)

- 모두 미혼이며, 군집의 78%가 남자로 구성됨.
- 친구와 함께 영화 관람을 즐기는 성향이 높고(87%) 주말/휴일에 극장을 찾는 경향이 높으며(90%), 주로 개봉 한 달 이내에 영화를 관람한다고 설명함(54%).
- 영화를 선택할 시, 영화의 내용과 즐거리를 가장 많이 고려하며(4.2), 장르(3.9), 주변인평가(3.9) 순으로 고려하는 특징을 나타냄.
- 평균 32세로 구성된 집단으로써, 월 270만 원의 소득 수준을 나타내고 평균 11편의 영화를 관람하는 것

으로 나타났음.

- 기타 : 액션 장르를 선호하며, 미국 영화를 상대적으로 선호하고, 인터넷을 통해 영화 관련 정보를 취득하는 경향이 높고, 당일 극장에서 티켓을 구매하는 경향을 나타내고 극장에 가기 전에 영화를 선택하며, 영화관의 위치가 극장을 선택하는 주요 요인으로 나타났음.

#### 군집 2(크기 : 1.5%, 평균 관람 편수 : 13.4편, 관람력 지수 : 0.15)

- 군집의 86%가 기혼자이며 대부분이 남성으로 이루어졌음(76.4%)
- 92%가 배우자와 함께 영화를 관람하는 집단으로, 요일 상관없이 영화 관람하는 비율이 73%를 차지함.
- 개봉일 무관하게 영화를 선택한다고 대답한 비율이 군집 전체의 49%를 차지하고, 나머지 42%가 개봉 일주일 이내에 영화를 관람하는 경향을 보인다고 설명하였음.
- 영화의 내용과 즐거리가 영화 선택에 가장 영향을 미치며(4.5), 그 외에도 영화의 감독(4.3), 영화의 장르(3.8)가 영화 선택에 가장 큰 영향 요인이라고 설명함.
- 연령은 41세로 구성되어 있고, 월평균 총 가구소득 수준은 280만 원 정도의 평균 분포를 보이고 연간 13편의 영화를 극장에서 관람.
- 기타 : 범죄/스릴러 장르를 가장 선호하나 실험/예술 영화에 대한 선호도가 상대적으로 높음. 국적에 상관없이 영화를 선택하고, 주로 인터넷을 통해 영화 정보를 습득하며, 극장에 가기 전 영화를 선택하고, 관람하고 싶은 영화를 상영하는 영화관을 선택하는 경향이 높음. 주로 인터넷 예매 사이트를 통해 극장 티켓을 구매함.

#### 군집 3(크기 : 11.5%, 평균 관람 편수 : 10.3편, 관람력 지수 : 0.88)

- 모두 기혼자로 구성된 군집으로, 남성이 53%, 여성이 47%로 구성됨.
- 개봉 한 달 이내에 영화를 관람한다고 설명한 비율이 가장 높았고(44%) 개봉 일주일 이내에 영화 관람 비율이 그 뒤를 이음(31%). 주로 주말/휴일에 극장 관람하는 경향을 나타냄(71%).
- 89%가 배우자와 함께 영화를 관람하는 것을 즐김.
- 영화의 내용 및 즐거리(4.5)와 영화의 장르(4.1)에 민감하며, 영화 출연배우(3.8), 주변인 평가(3.8) 등을 중요하게 생각함.
- 평균 연령이 37세이며, 월평균 총 가구소득 수준은 350만 원 선이고 연간 관람 편수는 10편임.
- 기타 : SF/판타지/무협에 대한 장르 선호도가 높고, 액션 장르를 선호함. 미국 영화에 대한 선호도가 높으며, 인터넷을 통해 영화 정보를 습득하고, 극장에 가기 전 영화를 선택하며, 영화관의 위치가 극장 선택의 주요 요인임. 주로 인터넷 극장 사이트를 통해 티켓을 예매함.

#### 군집 4(크기 : 7.3%, 평균 관람 편수 : 16.9편, 관람력 지수 : 0.92)

- 모두 미혼자이며, 남성이 59% 여성이 42%로 구성됨
- 친구와 함께 영화를 관람하는 비율이 가장 높고(76.3%), 상대적으로 다른 군집에 비하여 가족(부모/형제/자매/친척)과 함께 극장을 찾는 비율이 높음(19.3%). 주말/휴일에 영화를 극장에서 관람하며(82.6%) 개봉 한 달 이내에 영화를 관람하는 경향을 나타냄(52.8%).
- 영화의 장르가 가장 선택 시 큰 영향을 미치며(4.6), 그 뒤를 내용 즐거리(4.3), 흥행성적(3.8) 순이 영화 선택시 고려하는 사항이라고 응답함.

- 평균 19세, 월평균 총 가구소득 수준 290만 원 연간 극장관람 편수 17편임.
- 기타 : 액션 장르를 선호하고 SF/판타지/무협 장르를 선호함. 미국 영화에 대한 선호도가 상대적으로 높음. 인터넷을 통해 영화 정보를 습득하며, 극장에 가기 전 영화를 선택하고, 영화관의 위치가 극장 선택의 주요 요인임. 당일 극장에서 티켓을 구매하는 경향을 나타냄.

#### 군집 5(크기 : 21.5%, 평균 관람 편수 : 13.8편, 관람력 지수 : 2.20)

- 모두 미혼자로 구성되어 있고 친구와의 영화 관람이 74%이며 연인과의 관람이 16%으로 구성.
- 주로 주말/휴일에 영화를 극장에서 관람하며 (62.7%), 개봉 한 달 이내에 극장에서 영화를 관람하는 경향이 높다고 응답함(40.5%).
- 영화 선택 시 영화 내용/줄거리를 가장 중요하게 고려하며(4.4), 주변인 평가(3.9), 영화 장르(3.8)가 그 뒤를 이음.
- 평균 연령 약 23세이며, 월 평균 소득 수준이 290만원, 연평균 관람편수는 14편이라고 응답함.
- 기타 : 로맨틱/코미디 장르와 드라마를 가장 선호하나 그 외에도 액션, 코미디 멜로/로맨스 등의 다양한 장르에 대한 선호가 존재함. 한국 영화를 선호하는 집단임. 주로 인터넷을 통해 영화 정보를 취득하는 경향이 높으나 상대적으로 주변인을 통해 영화 정보를 습득하는 경향을 나타냄. 주로 극장에 가기 전에 영화를 선택하고, 영화관의 위치가 극장 선택의 주요 요인임. 당일 극장에서 티켓을 구매하는 경향을 나타냄.

#### 군집 6(크기 : 2.2%, 평균 관람 편수 : 20.41편, 관람력 지수 : 0.33)

- 86%가 기혼자이며 상대적으로 여성의 비율이 높은 집단(69.6%).
- 군집의 54%가 친구와 함께 영화를 보러 간다고 대답하였고, 배우자와는 약 30%가 함께 영화를 관람한다고 대답함. 또한 혼자 보는 비율이 다른 집단에 비해 가장 높게 나타났음(8.8%).
- 영화 선택 시 영화감독이 가장 중요하며(3.4), 출연배우(3.4), 주변인 평가(3.3), 내용/줄거리(3.1) 순으로 영화를 선택 평가함. 상대적으로 다른 군집에 비해 영화제 출품 여부를 중요하게 생각하는 군집 특성을 보임(3.15)
- 개봉 일주일 이내, 개봉 한 달 이내 관람하는 비율이 전체의 68%를 차지함. 주로 주말/휴일에 극장을 찾음(73.4%).
- 평균 연령 39세, 월평균 총 가구소득 수준 280만 원, 연간 극장 관람 편수 20편 수준임.
- 기타 : 액션 장르에 대한 선호가 높으며, 미국 영화를 상대적으로 선호함. 주로 TV를 통해 영화 정보를 취득하는 경향을 나타냄. 주로 극장에 가기 전에 영화를 선택하고, 영화관의 접근성이 극장 선택의 주요 요인임. 당일 극장에서 티켓을 구매하는 경향을 나타냄.

#### 군집 7(크기 : 9.8%, 평균 관람 편수 : 29.34편, 관람력 지수 : 2.15)

- 모두 미혼자로 구성되어 있고, 주로 여성으로 구성됨(61%)
- 대부분이 친구와 영화를 관람하고(63.7%), 연인이 그 뒤를 이음(20.1%).
- 주로 주말이나 휴일에 영화를 관람하는 집단이나(45.6%), 주중 혹은 요일에 상관없이 극장을 찾는다고 응답한 사람의 비율이 비슷하게 구성되어 있음(주중 관람 비율 : 26.5%, 요일 상관없음 비율 : 27.9%). 개봉 일주일 이내 극장을 찾는 경향이 가장 높게 나타났음(46%).

- 내용/줄거리 (4.6), 출연 배우(4.2), 영화 장르(4.1) 순으로 영화 선택 시 고려한다고 응답하였음.
- 연령은 29세이고, 월평균 총 가구소득 수준은 320만 원이며, 연간 29편의 영화관 관람을 하고 있음.
- 기타 : 장르에 대한 선호도가 균일하게 분포하는 특징을 나타냄. 한국영화를 가장 선호하나 미국 영화 역시 높은 비율로 선호하는 경향을 나타냄. 주로 인터넷을 통해 영화 정보를 취득하며, 주로 극장에 가기 전에 영화를 선택하고, 영화관의 위치가 극장 선택의 주요 요인임. 인터넷 극장 사이트를 통해 티켓을 구매하는 경향을 나타냄.

#### 군집 8(크기 : 8.0%, 평균 관람 편수 : 5.93편, 관람력 지수 : 0.35)

- 군집 8은 기혼자로 구성되어 있고, 82%가 남성으로 구성되어 있음.
- 93%가 배우자와 함께 극장을 찾는다고 응답하였으며, 주로 주말/휴일에 극장 관람하는 경향을 나타냈음(90%). 개봉 한 달 이내에 극장을 찾는 경향이 높게 나타났음(64.8%).
- 영화 선택 시 내용/줄거리(4.2)가 그 영향을 가장 많이 미치며, 영화 장르(4.1), 주변인 평가(3.91)이 영화 선택 시 고려되는 요인들임.
- 평균 연령은 43세이며, 월평균 총 가구소득 수준은 390만 원, 연간 영화관에서 5.9편의 영화를 관람함.
- 기타 : 액션 장르를 선호하며, 미국 영화를 선호하고, 주로 인터넷과 TV를 통해 영화 정보를 취득함. 주로 극장에 가기 전에 영화를 선택하고, 영화관의 위치가 극장 선택의 주요 요인임. 당일 극장에서 티켓을 구매하는 경향을 나타냄.

#### 군집 9(크기 : 9.2%, 평균 관람 편수 : 20.94편, 관람력 지수 : 1.42)

- 군집 9는 기혼자로 구성되어 있고, 62%가 여성으로 구성되어 있음.
- 대부분이 배우자와 함께 영화를 보는 것을 즐기고(57%), 27%가 친구와 함께 보러 간다고 응답하였음. 주말/휴일에 영화를 관람하는 비율이 높음(51.8%). 또한 개봉 일주일 이내에 영화를 관람하는 경향이 높음(56%).
- 영화 선택 시 내용/줄거리를 중요하게 생각하며(4.4), 영화 장르(4.3), 영화 출연배우(4.1) 등에 대한 고려도가 높음. 영화의 제작 국가가 상대적으로 높음 편임(3.6).
- 평균 연령 39.1세, 월평균 총 가구소득 수준 353만 원, 연간 20.9편의 영화를 극장에서 관람함.
- 기타 : 액션 장르를 선호하나, 멜로/로맨스 및 범죄/스릴러 장르에 대한 선호가 높으며 다른 군집에 비해 다큐멘터리 선호가 상대적으로 높음. 한국 영화를 선호하며, 인터넷을 통해 영화 정보를 취득함. 주로 극장에 가기 전에 영화를 선택하고, 영화관의 위치가 극장 선택의 주요 요인임. 당일 극장에서 티켓을 구매하는 경향을 나타냄.

#### 군집 10(크기 : 7.2%, 평균 관람 편수 : 5.41편, 관람력 지수 : 0.29)

- 모두 기혼자이며, 67.6%가 여성으로 구성
- 51.4%가 배우자와 함께 영화를 보는 것을 즐기며, 37.6%가 친구와 함께 영화관을 찾음.
- 요일에 무관하게 극장을 찾으러 온다고 응답하였으며(47.5%), 개봉 일에 무관하게 영화를 관람한다고 설명하였음(58%).
- 영화의 내용/줄거리에 따라 영화를 선택하는 비율이 가장 높고(4.4), 그 뒤를 장르(3.9), 주변인 평가(3.7)가 차지함.

- 평균 연령 43.8세이고, 월평균 총 가구소득 수준 311만 원, 연간 5.4편의 영화를 극장에서 관람함.
- 기타 : 액션 장르를 선호하는 경향을 나타내나 상대적으로 다양한 장르에 대한 선호를 가지고 있음. 한국 영화를 상대적으로 선호하고, 인터넷과 TV를 통해 영화 정보를 취득함. 주로 극장에 가기 전에 영화를 선택하고, 영화관의 위치가 극장 선택의 주요 요인임. 당일 극장에서 티켓을 구매하는 경향을 나타냄.

**군집 11(크기 : 15.1%, 평균 관람 편수 : 6.75편, 관람력 지수 : 0.76)**

- 모두 기혼자이며, 81.3%가 여성으로 구성
- 72%가 배우자와 함께 영화를 보는 것을 즐기며, 25%가 친구와 함께 영화관을 찾음.
- 주로 주말/휴일에 영화를 관람하러 극장에 간다고 응답하였으며(71.8%), 개봉일 한달 이내에 영화를 관람한다고 설명하였음(67.1%).
- 내용/줄거리에 따라 영화를 선택하는 비율이 가장 높고(4.3), 그 뒤를 주변인 평가(4.2), 흥행성적(3.9)이 차지함.
- 평균 연령 37.7세이고, 월평균 총 가구소득 수준 300만 원, 연간 6.8편의 영화를 극장에서 관람함.
- 기타 : 로맨틱/코미디 장르와 드라마 장르에 대한 선호가 높고, 상대적으로 다른 집단에 비해 애니메이션에 대한 장르 선호가 높음. 한국 영화에 대한 선호가 높으며, 주로, TV와 주변인을 통해 영화 정보를 취득하는 경향을 나타냄. 주로 극장에 가기 전에 영화를 선택하고, 영화관의 위치가 극장 선택의 주요 요인임. 당일 극장에서 티켓을 구매하는 경향을 나타냄.