

## 일반화된 제약다중대응분석을 이용한 범주형 마케팅 자료 분석

황 홍 선<sup>†</sup>

Claes Fornell International Group & Rice University

Hwang 과 Takane(2002)는 다양한 선형제약을 적용하여 다변량 범주형 변인들간의 상호 관계를 보다 다양한 측면에서 분석할 수 있는 탐색적 통계 방법으로 일반화된 제약다중대응분석을 개발하였다. 본 논문은 이 통계 기법의 실제 응용에 초점을 맞추고 있다. 특히, 마케팅 자료에 응용해봄으로써 일반화된 제약다중대응분석이 마케팅 조사에서 얻어지는 많은 범주형 변인들간의 관계를 보다 통찰력 있게 분석할 수 있는 유용한 통계 기법이라는 것을 보여주고자 한다.

주제어 : 다중대응분석, 선형제약, 범주형 마케팅 자료

### 서 론

다중대응분석(multiple correspondence analysis)은 두 가지 이상의 범주형 변인들간의 상호 관계를 알아보기 위한 유용한 통계 기법이다(Greenacre, 1984; Lebart, Morineau, & Warwick, 1984; Nishisato, 1980). 다중대응분석은 범주형 변인을 대상으로

하는 일종의 가중치를 부가한 주성분분석이라고 할 수 있다. 다중대응분석은 범주형 변인의 카테고리들(예를 들어, 성별이라는 범주형 변인에는 남자와 여자라는 두 가지 카테고리가 있다)에 최소제곱방식에 기초하여 구한 이상적인 대표값을 부여한다. 그렇게 얻어진 카테고리 대표값들을 다차원 공간상에 그림으로 표시한 후 각 대표값

<sup>†</sup> 교신저자 : 황 홍 선, hhwang87@hotmail.com

들의 상호 거리를 통해 서로의 관계를 알아본다. 예컨대, 만약 특정 카테고리 값들이 서로 가까이 위치하고 있다면 그 카테고리들은 보다 밀접하게 연관되어 있다고 해석한다.

다중대응분석에 대한 이해를 보다 돕기 위해 실제 마케팅 자료를 다중대응분석을 이용하여 분석한 예를 들어보고자 한다. 여기서 사용하는 자료는 미국의 한 대형 식품 제조 회사가 자사가 만들어낸 신상품에 대하여 소비자들이 얼마나 만족하고 있는가를 알아보기 위하여 2000년에 실시한 설문 조사 중의 일부분이다. 본 논문에서는 조사를 의뢰한 식품 제조 회사를 K회사라고 부르고 소비자 만족도를 알아보려고 했던 신상품을 상품A라고 부르기로 한다<sup>1)</sup>. K회사의 상품A는 누구나 쉽게 전자 렌지에 넣고 즉석에서 조리해서 먹을 수 있는 가공 식품이다.

실제 설문지에는 많은 범주형 변인들이 있었지만, 본 논문에서는 다음과 같은 세 가지 범주형 변인들간의 관계만을 알아보려고 한다: (1) 성별, (2) 가계 연간 총수입, (3) 상품A에 대한 소비자 만족도. 만약 특정 성별이나 특정 수입을 가진 가계들의 만족도가 다른 설문 응답자와는 다르게 나타난다면 이 후 상품A에 대한 마케팅 전략을 수립하는데 많은 도움이 될 수 있다고 보았기 때문에 이 세 가지 범주형 변인간의 관계를 분석하기로 결정하였다. 이 세가지 변인 중에서 성별은 남자와 여자의 두 가지 카테고리로 나뉘어졌다. 그리고 가계 연간 총수입은 다섯 가지 카테고리로 이루어졌다: 1 = \$20,000 미만, 2 = \$20,000 - \$34,999, 3 = \$35,000 - \$49,999, 4 = \$50,000 - \$74,999, 5 = \$75,000이상. 소비자 만족도는 원래 10점 척도의 연속 변인으로 측정하였지만, 본 연구를 위해서 다음과 같이 세 가지 카테고리를 갖는 비연

원래 연속 척도치 새롭게 구성된 비연속 카테고리

1 - 5	--->	1 = 낮은 만족도
6 - 8	--->	2 = 중간 정도의 만족도
9 - 10	--->	3 = 높은 만족도

속 변인으로 재구성하였다.

이러한 비연속 변인으로서의 재구성은 실제 상당수의 응답자들이 상품A에 대하여 높은 만족도를 보이고 있다는 사실을 고려하여 이루어졌다. 전체 설문 응답자는 294명이었다.

이 세가지 변인들에 다중대응분석을 적용하였다. 그림 1은 다중대응분석을 통해 추정된 각 변인 카테고리 점수를 2차원 공간에 그려놓은 것이다. 이러한 2차원적 결과는 변인들간의 전체 변량의 약 99%를 설명하였다. 그러므로 2차원적 결과를 제시하는 것이 충분하다고 할 수 있다.

그림 1에서, 남자와 여자의 카테고리는 각각 MA와 FE로 표시하였다. 가계 연간 총수입의 다섯 가지 카테고리는 각각 i1, i2, i3, i4, i5로 표시하였다. 여기서 i의 숫자는 카테고리 숫자를 의미한다. 그러므로 이 숫자가 높을수록 고소득 가계를 나타낸다. 소비자 만족도의 세 가지 카테고리는 각각 LS = 낮은 만족도, MS = 중간 정도의 만족도, HS = 높은 만족도로 표시하였다. 그리고 그림 1에서 '+'이라는 기호는 2차원 공간상의 중심점을 나타낸다.

그림 1에서 중심점을 기준으로 오른쪽 중간 부분은 상품A에 대한 높은 만족도 (HS)와 관련된 것으로 보인다. HS는 가장 낮은 연간 수입을 갖는 가계 (즉, i1)와 가장 가깝게 위치하고 있다. 또한 남자 (MA)가 여자 (FE)보다 HS에 보다 가깝게 위치하고 있다. 이것을 통해 우리는 저소득층 가정과 남자들이 상품A에 대하여 상대적으로 보다 높은 만족도를 보이는 경향이 있다고 말할 수

1) 설문 조사 시기가 얼마 지나지 않은 이유로 식품 제조 회사와 상품명을 공개하는데 어려움이 있다.

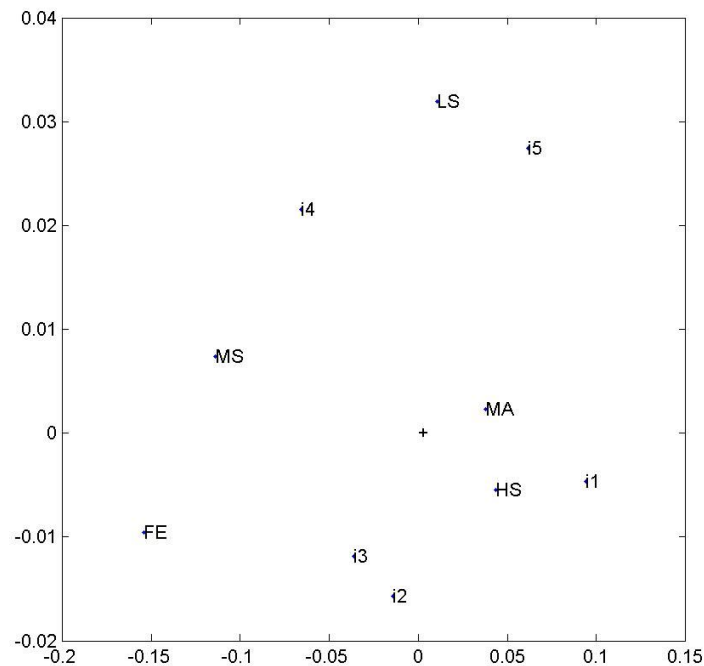


그림 1. 다중대응분석에서 얻어진 2차원적 변인 카테고리 점수

있다. 그러므로 이 두 집단이 상품 A의 주된 고객이 될 확률이 높다고 말할 수 있다. 반면에 그림 1의 상단 중앙 부분은 낮은 만족도(LS)와 관련되어 있는 것으로 보인다. LS는 고소득층 가계(i4와 i5)와 가장 가깝게 위치하고 있다. 그러므로 높은 연간 수입을 가지는 가정일수록 상품 A에 대해서 별로 만족하고 있지 못하다고 말할 수 있다. 마지막으로 그림 1의 왼쪽 중간 부분은 중간 정도의 만족도(MS)와 관련이 있는 듯 하다. 여자(FE)들은 남성들에 비해 상대적으로 MS에 가깝게 위치하고 있다. 그러므로 여자들이 남성들에 비해 상품 A에 대해 보다 낮은 만족도를 보이고 있다고 말할 수 있다. 그러나 중산층의 가계들(즉, i2, i3)은 HS와 MS 사이에 위치하고 있어서 이들 가계들이 어떤 확실한 만족도 유형을 가지고 있다고 해석하기가 힘들다.

위의 예에서 보았듯이, 그림으로 제시되는 다중대응분석의 결과들은 범주형 변인들간의 관계를 보다 손쉽게 컨설팅 의뢰자에게 전달할 수 있도록 도와준다. 실제 마케팅 조사 현장에서는, 연구자들이 주되게 분석하고자 하는 범주형 변인들(위의 예에서는 성별, 가계 연간 총수입, 상품 A에 대한 소비자 만족도) 외에도 여러 가지 부수적인 정보를 가지고 있는 경우가 많다. 예를 들어, 연구자는 흔히 설문 응답자의 여러가지 인구 통계적 정보(예들 들어, 나이, 교육 수준 등등)를 가지고 있는 경우가 많다. 또한 응답자들이 현재 사는 지역, 혹은 출신 학교등과 같은 특정 집단 소속 여부에 대한 정보를 가질 수도 있다. 더 나아가 이러한 응답자와 직접 관련되어 있는 정보 뿐만 아니라 범주형 변인들 자체간의 관계와 관련된 부가적인 가설이나 사전 지식들을 가질 수도

있다. 예를 들어, '변인들 간에는 상호 작용이 없다'라거나 '몇몇 변인들은 서로 동일하다'라고 하는 식의 가설을 사전에 가질 수 있다.

통계적으로는 이러한 부수적인 정보 혹은 가설들은 선형제약(linear constraints)의 형태로 실제 분석에 통합된다(Bockenholt & Bockenholt, 1990; Takane & Shibayama, 1991; Takane, Yanai, & Mayekawa, 1991; Yanai, 1986). 선형제약은 실제 통계 분석이 이루어지기 전에 자신이 가지고 있는 정보를 통해 변인들을 사전에 설명해주는 효과를 갖는다. 그러므로 연구자들은 선형제약에 의해 이미 구조화된 변인들을 분석하기 때문에 보다 간단한 결과를 얻을 수 있다. 또한 선형제약이 가지고 있는 정보에 기초하여 변인들간의 관계를 보다 다양한 측면에서 연구해 볼 수도 있다. 아울러 만약 연구자가 상정한 선형제약이 실제 자료의 성격과 일치한다면, 통계적으로 보다 신뢰도가 높은 모수 추정치를 얻어낼 수가 있다.

Hwang과 Takane(2002)는 다중대응분석에 선형제약들을 포괄적인 방식으로 통합시킬 수 있는 새로운 통계 기법을 개발하였다. 이른바 일반화된 제약다중대응분석(generalized constrained multiple correspondence analysis)라고 부르는 이 통계 기법은 모든 종류의 선형제약을 단일한 수학적 틀거리 속에서 통합시킬 수 있다. 그 결과 연구자가 가지고 있는 사전 지식이나 부가적인 정보들을 자유롭게 다중대응분석에 접목시켜 보다 다양한 측면에서 범주형 변인들간의 상호 관계를 알아볼 수 있다. 또한 일반화된 제약다중대응분석은 현재 개발되어 있는 모든 종류의 선형제약이 부가된 다중대응분석(예를 들어, Gifi, 1990; Nishisato, 1984; van Buuren & de Leeuw, 1992; Yanai & Maeda, 2000)들을 특수한 하부 분석으로 포함하고 있다.

일반화된 제약다중대응분석은 우선 각 범주형

변인 자료를 선형제약을 이용해 여러 가지 작은 부분들로 분해한다. 그리고 난 후 그 분해된 각각의 자료들을 다중대응분석으로 분석한다. 기술적으로 전자의 방식은 범주형 변인들의 직교사영(orthogonal projection)과 관련되어있고, 후자는 일반화된비정칙값분해(generalized singular value decomposition)와 상응한다고 할 수 있다. 실제 분해된 자료의 부분 중 어느 부분을 분석하느냐에 따라 그 다중대응분석의 종류가 달라진다. 예를 들어, 범주형 변인으로부터 특정 인구 통계적 변인의 효과를 선형제약의 형태로 제거하고 난 후에 남은 부분들을 다중대응분석으로 분석할 수 있는데, 이렇게 특정 변인이 설명하고 남은 부분을 분석하는 다중대응분석을 부분다중대응분석(partial multiple correspondence analysis; Yanai & Maeda, 2000)이라고 부른다. 또 다른 예로 몇몇 변인들의 카테고리들이 서로 같다는 가정을 선형제약 - 이런 경우에는, 상등선형제약(equality linear constraints)을 사용한다 - 을 통해 실제 분석에 포함시킬 수 있다. 이러한 변인들간의 관계에 대한 선형제약에 의해 구조화된 범주형 변인의 부분을 다중대응분석으로 다룬다면 이 경우를 제약다중대응분석(constrained multiple correspondence analysis; Gifi, 1990; Van Buuren & de Leeuw, 1992)이라고 부른다. 앞서 말했듯이 일반화된 제약다중대응분석은 이러한 여러 가지 특수한 형태의 다중대응분석들을 통합된 형태로 포함하고 있다. 일반화된 제약다중대응분석의 보다 자세한 기술적인 설명은 Hwang과 Takane(2002)를 참고하기 바란다.

본 논문의 주된 목표는 일반화된 제약다중대응분석을 설명하는 것이 아니라 그 통계 기법의 실제 응용에 있다. 특히 본 논문에서는 마케팅 설문 자료들을 일반화된 제약다중대응분석을 이용하여 분석하고자 한다. 다음 장에서는 앞서 제시한 소비자 만족도 자료를 일반화된 제약다중대

응분석을 이용하여 다중대응분석에서 다룰 수 없는 다양한 측면에서 분석하고자 한다. 이를 통해 일반화된 제약다중대응분석이 마케팅 조사에서 얻어지는 범주형 변인들간의 관계를 보다 통찰력 있게 분석하는데 유용한 통계 기법이라는 것을 보여주하고자 한다.

### 일반화된 제약다중대응분석의 응용

본 장에서는 일반화된 제약다중대응분석을 이용하여 앞서 예로 들었던 소비자 만족도 조사 자료를 보다 다양한 측면에서 분석해본다. 위에서 제시되었던 세 가지 범주형 변인 - 즉, 성별, 가계 연간 총수입, 상품A에 대한 소비자 만족도 -를 여전히 주된 분석 변인으로 사용한다. 본 논문에서는 지면상의 이유로 세 가지 분석의 예만을 제시한다.

우선 그림 1에서 보여준 다중대응분석 결과에 기초하여, 우리는 가계 연간 총수입의 다섯 가지 카테고리를 단순화시키기 위한 두 가지 가설을 설정하였다. 우선 고소득층 가계를 나타내는 두 카테고리(즉, i4와 i5)를 하나로 묶기 위해 그 두 카테고리는 서로 같다고 가정하였다. 이렇게 해서 얻은 새로운 카테고리를 HI(High Income)라고 이름 붙였다. 또한 중산층 가계를 의미하는 두 카테고리(즉, i2와 i3) 역시 서로 같다고 가정하여 하나의 카테고리로 묶었다. 이렇게 만들어진 새로운 카테고리를 MI (Middle Income)라고 이름 붙였다. 우리는 이러한 가설을 상등선형제약의 형태로 일반화된선형제약다중대응분석에 적용하였다. 실제 이러한 방법은 제약다중대응분석과 동일하다.

그림 2는 이러한 제약다중대응분석을 통해 추정된 2차원적 카테고리 점수들을 보여주고 있다. 위에서 제시한 다중대응분석과 마찬가지로 2차원

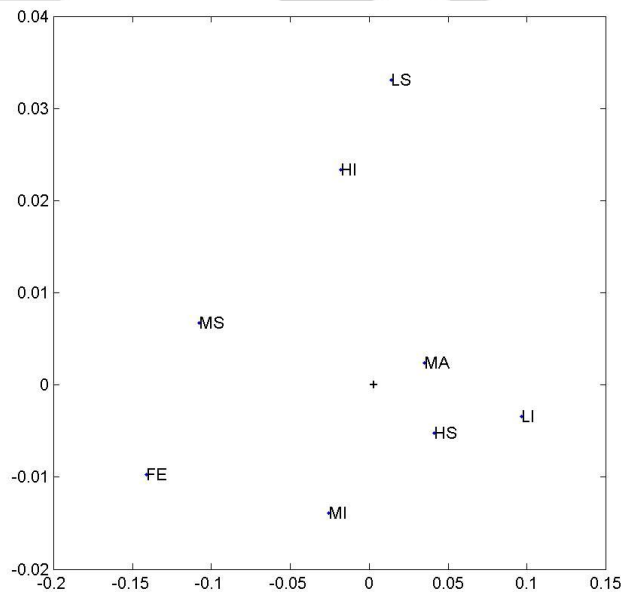


그림 2. 제약다중대응분석에서 얻어진 2차원적 변인 카테고리 점수

결과는 변인들의 전체 변량의 99% 정도를 설명하였다.

그림 2에서는 가계 연간 총수입과 관련하여 새롭게 만들어진 두 가지 카테고리(HI와 MI)와 비교가 쉽도록 저소득층 가계를 나타내는 II를 LI(Low Income)로 바꾸어 불렀다. 그림 2의 해석은 상등선형제약을 주었음에도 불구하고 기본적으로 그림 1의 해석과 동일하다고 할 수 있다. 즉, 상품A에 대하여 저소득층일수록 높은 만족도를, 고소득층일수록 낮은 만족도를 보여주고 있다. 그리고 여자보다는 남자가 보다 높은 만족도를 보이고 있다. 이러한 해석상의 동일함은 가계 연간 총수입의 몇 몇 카테고리들이 서로 같다는 가설이 분석 결과를 크게 왜곡시키지 않는다는 것을 의미하며, 그러므로 우리의 가설이 타당하다는 것을 경험적으로 증명해주고 있다. 게다가 이러한 가정들 덕분에 추정해야할 모수치의 수(여기

서는 이차원적 변인 카테고리 점수들의 수와 동일하다)가 줄어 이 전의 다중대응분석에 비해 보다 간단한 해석을 내릴 수 있었다.

우리는 더 나아가 세 가지 주된 변인들간의 관계를 다른 제 4의 범주형 변인과 연관시켜 새롭게 분석해보고자 하였다. 특히, 우리는 설문 응답자들의 현재 고용 상태가 이 세 가지 변인들간의 관계에 어떤 영향을 미치는가를 일반화된 제약다중대응분석을 이용하여 알아보려고 하였다. 현재 고용 상태를 나타내는 범주형 변인은 두 가지 카테고리를 가지고 있었다: 1 = 현재 직장에 고용되어 있음, 2 = 현재 직장에 고용되어 있지 않음. 우선 우리는 현재 고용 상태라는 변인이 세 가지 주된 연구 변인들에 대해 가지고 있을 효과를 선형제약의 형태로 모두 제거하였다. 그리고 나서 현재 고용 상태라는 변인을 제거하고 남은 세 변인들의 부분을 분석하였다(이것은 공변량

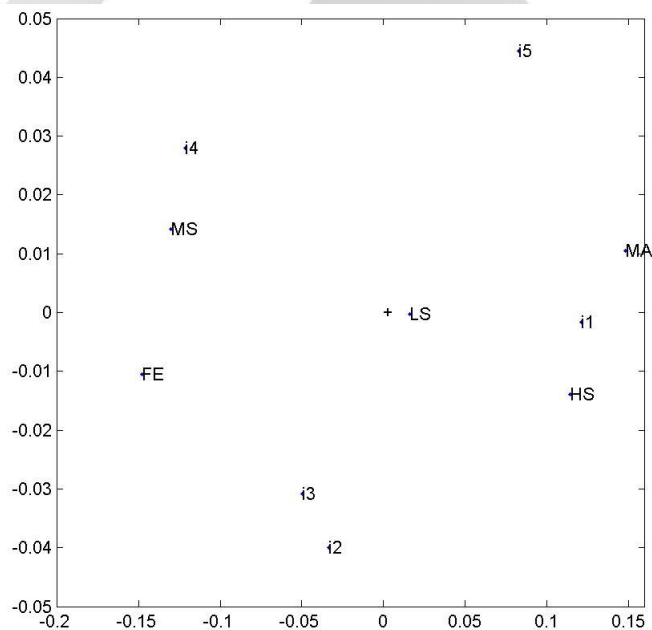


그림 3. 부분다중대응분석에서 얻어진 2차원적 변인 카테고리 점수

분석에서 '다'라는 변인의 효과를 '가'와 '나'라는 변인에서 동시에 제거한 후 남은 '가'와 '나'의 부분간의 관계를 알아보는 것과 같은 원리라고 할 수 있다.). 이러한 방법은 부분다중대응분석과 동일하다고 할 수 있다.

앞선 두 가지 분석들과 마찬가지로, 부분다중대응분석에서도 2차원적 결과가 전체 변인들간의 변량의 대부분을 설명하였다. 그림 3은 부분다중대응분석을 통해 추정된 2차원적 카테고리 점수들을 보여주고 있다.

그림 3에서 볼 수 있듯이, 현재 고용 상태라는 변인의 효과를 세 가지 범주형 변인들로부터 제거시킨 결과, 전체적으로 카테고리 점수들의 배열이 앞의 두 가지 분석들과는 다르게 나타났다. 그러나 그림에도 불구하고, 기본적으로 높은 만족도(HS)와 중간 정도의 만족도(MS)에 상응하는 부분들은 이 전의 분석들과 동일하게 해석할 수 있다. 반면에, 낮은 만족도 (LS)를 나타내는 점은

이 전의 분석 결과와는 매우 다르게 위치하고 있다. 즉, LS는 고소득층의 가계 (i4 와 i5)와는 멀리 떨어져 나와 2차원 공간의 중심점 쪽으로 이동해 있다. 이렇게 현재 고용 상태라는 변인의 효과를 제거한 후 LS만이 주되게 변화하였다는 것은 설문 응답자의 현재 고용 상태가 낮은 만족도라는 카테고리에만 주된 영향을 주고 있다는 것을 의미한다. 아마도 이것은 낮은 만족도를 보이는 응답자들이 그렇지 않은 응답자에 비해 훨씬 더 많이 비고용 상태에 있다는 사실과 관련이 있는 듯하다. 빈도 조사 결과 낮은 만족도를 나타낸 응답자의 약 63%가 비고용 상태에 있는 반면에 중간 정도의 만족도를 보이는 응답자 중에서는 약 21%가, 그리고 높은 만족도를 보이는 응답자 중에서는 약 40%만이 비고용 상태에 있었다. 그러므로, 상품A에 대한 낮은 만족도를 보이는 응답자에 대해서는 현재 고용 상태에 대하여 추가적인 분석을 해보는 것이 바람직하다고 할 수 있

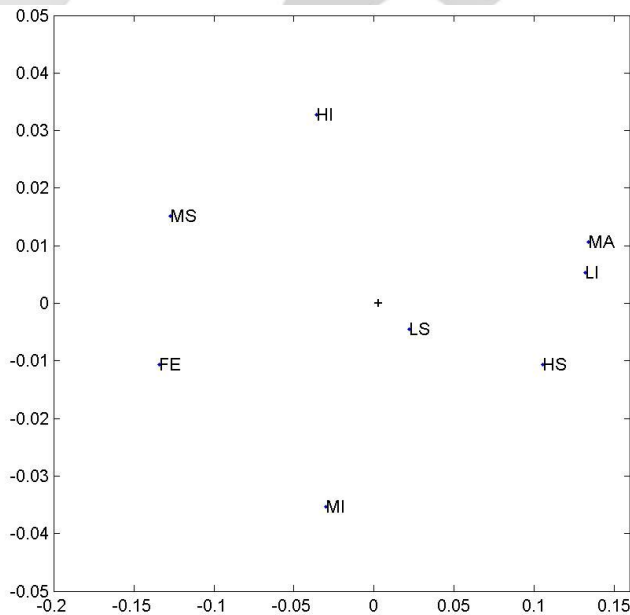


그림 4. 부분제약다중대응분석에서 얻어진 2차원적 변인 카테고리 점수

다. 특히 본 조사에서는 고소득층 응답자들이 상대적으로 낮은 만족도를 보이고 있는데, 고소득층 응답자 중 현재 직장을 가지고 있지 않는 응답자들이 보다 낮은 만족도를 보일 것이라고 유추해 볼 수 있다.

마지막으로 우리는 앞서 제시한 제약다중대응분석과 부분다중대응분석을 동시에 자료에 적용시켜 보기로 하였다. 이 경우는 부분제약다중대응분석 (partial constrained multiple correspondence analysis)이라고 부를 수 있다(Hwang & Takane, 2002). 앞에서 설명한 세 가지 분석 방법들이 이미 기존에 개별적으로 존재하고 있는 방법들이며 반하여 이 부분제약 다중대응분석은 전혀 새로운 종류의 다중대응분석이다. 이 방법은 전체 변인 중에서 제약 다중대응분석에서 사용한 상등선형 제약에 의해 설명되면서 동시에 현 고용 상태의 효과가 제거된 자료를 분석한 것이다.

이 부분제약다중대응분석의 결과는 그림 4와 나와 있다.

다른 분석들과 비슷하게 2차원적 결과가 변인들의 전체 변량의 대부분을 설명하고 있었다. 그림 4를 보면, 제약다중대응분석의 결과와 동일한 방식으로 가게 연간 총수입의 다섯 가지 카테고리가 세 가지 카테고리로 단순화 되어있다. 동시에 현 고용 상태의 효과를 제거한 결과, 부분다중대응분석과 유사하게 낮은 만족도를 나타내는 카테고리가 그림의 중앙으로 이동되어 있다. 전체적인 결과는 부분다중대응분석과 유사하다고 할 수 있으나, 상등제약조건 덕분에 보다 간단한 해석을 내릴 수 있다.

## 결 론

본 논문에서는 일반화된 제약다중대응분석이

라는 새로운 통계 기법을 이용하여 범주형 마케팅 자료간의 관계를 분석해보았다. 앞의 예들에서 보았듯이 일반화된 제약다중대응분석을 이용하여 연구자는 자신이 가지고 있는 사전 지식(가설(예를 들어, 몇 몇 변인들의 카테고리가 같다)이나 부수적인 정보(예를 들어, 현 고용 상태)를 손쉽게 다중대응분석에 접목시킬 수 있다. 그 결과 단순히 다중대응분석만을 적용했을 때보다 훨씬 다양한 측면에서 범주형 마케팅 자료간의 상호 관계를 분석할 수 있다. 이러한 다양한 측면의 분석은 보다 세밀한 마케팅 전략을 수립하는데 도움이 될 것이다. 기존의 마케팅 조사 연구는 범주형 변인들을 그저 응답자들을 소집단으로 나누는 지표 (예컨대, 남자 집단과 여자 집단) 정도로 사용하거나 분석을 하더라도 빈도 분석 정도에서 그치고 마는 경우가 많다. 그러나 본 논문에서 볼 수 있듯이 일반화된 제약다중대응분석을 이용하여 범주형 변인 자체들간의 관계에 대해 보다 흥미롭고 다양한 분석을 할 수 있다.

본 논문에서는 일반화된 제약다중대응분석방법 중에서 몇 가지 경우만을 예로 들었다. 그러나 일반화된 제약다중대응분석의 큰 장점 중의 하나가 연구자의 흥미에 따라 얼마든지 다양하고 새로운 선형제약에 기초한 다중대응분석방법을 만들 수 있다는 것이다. 그러므로 마케팅 조사 현장에서 일반화된 제약다중대응분석의 보다 폭넓은 응용을 기대해본다<sup>2)</sup>.

2) 현재 일반화된 제약다중대응분석을 위한 매트랩(Matlab) 프로그램이 만들어져 있다. 프로그램이 필요하신 분은 본 저자에게 연락하시면 된다.



### 참고문헌

- Bockenholt, U., & Bockenholt, I. (1990). Canonical analysis of contingency tables with linear constraints. *Psychometrika*, 55, 633-639.
- Gifi, A. (1990). *Nonlinear Multivariate Analysis*, Chichester: Wiley.
- Greenacre, M. J. (1984). *Theory and applications of correspondence analysis*, London: Academic Press.
- Hwang, H., & Takane, Y. (2002). Generalized constrained multiple correspondence analysis, *Psychometrika*, 67, 215-228.
- Lebart, L., Morineau, A., & Warwick, K. M. (1984). *Multivariate descriptive statistical analysis*. New York: Wiley.
- Nishisato, S. (1980). *Analysis of categorical data: Dual scaling and its applications*. Toronto: University of Toronto press.
- Takane, Y., & Shibayama, T. (1991). Principal component analysis with external information on both subjects and variables. *Psychometrika*, 56, 97-120.
- Takane, Y., Yanai, H., & Mayekawa, S. (1991). Relationships among several methods of linearly constrained correspondence analysis. *Psychometrika*, 56, 667-684.
- van Buuren, S., & de Leeuw, J. (1992). Equality constraints in multiple correspondence analysis. *Multivariate Behavioral Research*, 27, 567-583.
- Yanai, H. (1986). Some generalizations of correspondence analysis in terms of projection operators. In E. Diday, Y. Escoufier, L. Lebart, J. P. Pages, Y. Schektman, & R. Thomassone (Eds.), *Data Analysis and Informatics IV*. (pp. 193-207). Amsterdam: North Holland.
- Yanai, H., & Maeda, T. (2000). Partial multiple correspondence analysis. *Proceedings of the International Conference on Measurement and Multivariate Analysis*, 110-113.

## Analysis of categorical marketing data by generalized constrained multiple correspondence analysis

Heungsun Hwang

Claes Fornell International Group & Rice University

Hwang and Takane (2002) proposed generalized constrained multiple correspondence analysis that allows to explore relationships among many categorical variables from more diverse perspectives by incorporating a variety of linear constraints into multiple correspondence analysis. The focus of the present paper is on the application of the method. In particular, this paper applies the method to marketing data, and aims to demonstrate that the method is useful to obtain more insightful information on relationships among categorical variables collected in marketing surveys.

*keywords* : Multiple correspondence analysis, linear constraints, categorical marketing data.