Journal of Korean
Data & Information Science Society
2003, Vol. 14, No.3 pp. 501~510

MDPREF and Perceptual Map via INDSCAL Method[1)]

S.Y. Hwang[2)] · J.H. Hahm[3)]

Abstract

Ordinary computer program in the context of MDS for providing separate graphs according to "individuals" is usually referred to as INDSCAL(individual difference scaling), which enables one to detect dissimilarities between "subjects". This article is concerned with applications of INDSCAL concept to MDPREF and perceptual map. Related algorithm is presented and is implemented through illustrative analysis of Korean credit card data.

Keywords : Credit card data, INDSCAL, MDPREF, Perceptual map

1. 서 론

응용성 높은 다변량 통계분석법인 다차원척도법(mutidimensional scaling ; MDS)은 개체(object)간 비유사성을 저차원의 공간에 기하학적으로 나타내는 다변량 그래픽 기법이다. 선호자료분석을 위한 다차원선호분석(multidimensional preference analysis ; MDPREF)은 소비자(subject)들의 제품(brand)에 대한 선호, 혹은 제품 속성에 대한 각 제품별 평가를 동시에 공통 공간상에 나타내 준다. MDS의 응용기법으로서 인지도(perceptual map) 역시, 제품의 좌표만을 그래프로 보여주는 것이 아니라 여기에 외부 정보가 들어와 벡터로서 이들을 함께 표시해 주는 기법이다. 앞서 서술한 기법들은 기본적으로 전체 표본에 대한 하나의 그래프를 작성한다. 그러나, 그룹화가 명확한 경우 전체를 대상으로 하나의 그래프를 얻는 위의 기법들은 큰 의미가 없다. 왜냐하면 유사성이나 선호도를 판단함에 있어서 다양한 개인별 또는 그룹별 차이가 고려되지 않기 때문이다. 따라서 본 논문에서는 MDS 분석에 있어서 개인별 차이를 고려하는 INDSCAL(individual difference scaling) 개념을 선호도 자료에 확장하여, 다차원선호분석과 인지도에서 이를 구현할 수 있는 방법론을 제시하고자 한다.

---

1) This work was supported by a research grant 2001 from Sookmyung Women's Univ.
2) Professor, Department of Statistics, Sookmyung Women's Univ., Seoul, 140-742, Korea.
E-mail : shwang@sookmyung.ac.kr
3) Graduate student, Department of Statistics, Sookmyung Women's Univ., Seoul, Korea.

2. INDSCAL의 개념

MDS는 비유사성 행렬 D의 측정 척도에 따라 계량형 다차원척도법(metric MDS)과 비계량형 다차원척도법(nonmetric MDS)으로 나눌 수 있으며 이들을 보통 이원 다차원척도법(two-way MDS)이라 한다. 이는 비유사성 행렬 D가 행과 열 즉, 이원으로 구성된 이원행렬이므로 이원행렬에 적용된 다차원척도법임을 뜻하는 것이다. 반면에 이원행렬 D를 주체(subject) 각각에 대해서 고려한 것을 삼원행렬이라 하며, 이에 적용한 다차원척도법을 삼원 다차원척도법(three-way MDS)이라 한다 삼원 다차원척도법을 제공하는 방법으로 가장 대표적인 방법이 INDSCAL 모형이다(최용석(1995)). 즉, INDSCAL 모형의 입력 자료는 삼원행렬 자료로서, 비유사성으로 이루어지는 이원 정방대칭행렬이 여러 주체에 따라 모아진 직육면체의 자료가 된다. 따라서 삼원 다차원척도법의 결과는 형상공간에서의 개체들의 좌표행렬, 즉 공통공간에서의 좌표와 더불어 주체공간에서의 좌표행렬이 구해진다. MDS에서의 INDSCAL 모형을 간략히 소개하면 다음과 같다. $d_{ij}^{(s)}$는 주체 s에 대해 i번째와 j번째 개체 사이의 비유사성을 나타내고 r은 축소된 형상공간의 차원수(보통 2로 선택)를 나타내며 x_{ik}는 공통공간에서의 i번째 개체의 k번째 좌표, w_{sk}는 차원 k에서의 주체 s의 가중치라 할 때 형상공간에서의 유도된 거리 $\delta_{ij}^{(s)}$ 는 다음과 같다.

$$\delta_{ij}^{(s)} = \Big[\sum_{k=1}^{r} w_{sk}(x_{ik} - x_{jk})^2 \Big]^{\frac{1}{2}} \tag{2.1}$$

식(2.1)은 INDSCAL 모형의 차원 축소된 형상공간의 비유사성이 유클리드 거리 대신에 가중 유클리드를 고려하고 있음을 보여준다. 그러므로 공통공간에서의 좌표에 각 주체별로 가중치를 고려한 좌표는

$$y_{ik}^{(s)} = \sqrt{w_{sk}}\, x_{ik} \tag{2.2}$$

에 의해서 제공된다. 식(2.2)의 각 주체에 대한 공통공간상의 좌표에 의해서 제공되는 공간을 개인지각공간이라 한다. 자세한 내용은 최용석(1995,1999)를 참조하라. 따라서 주체들의 비유사성은 각 주체 자신의 공통공간상의 좌표인 식(2.2)에 의해서 유클리드 거리

$$\delta_{ij}^{(s)} = \Big[\sum_{k=1}^{r} (y_{ik}^{(s)} - y_{jk}^{(s)})^2 \Big]^{\frac{1}{2}}$$

로 나타낼 수 있다. 요약하면, INDSCAL 모형은 표본 전체의 유사성 자료를 분석함에 있어서 모든 사람들에게 공통되는 공간이 있지만, 각 개인은 그 공간의 각 차원에 대해서 고유한 가중치를 갖는다고 가정하는 것이다. 따라서 표본 전체에 공통이 되는 공통공간과 그 그래프의 각 차원에 대한 개인별 가중치(individual weights)를 제공해

준다. 그러므로 각 주체별로 축에 대한 가중치를 반영하여 개체간의 비유사성이 어떻게 형상화 되는지를 그래프를 통해서 알 수 있게 된다. 이렇듯 INDSCAL 에서는 주체별 형상공간의 좌표를 얻을 수 있으므로 조사자가 개인간의 혹은 집단간의 차이에 관심이 있는 경우에 유용한 MDS 기법이라 하겠다.

3. MDPREF 기법에서의 INDSCAL

본 절에서는 유사성 자료에 있어서 개인별 차이를 분석하는 INDSCAL 개념을 선호도 자료에 확장하여, 다차원선호분석과 인지도에서 이를 구현할 수 있는 방법론을 제시하고자 한다. 경쟁제품사이의 비유사성을 기초로 한 MDS는 시장구조나 경쟁구조에 대한 탐색적인 분석을 가능하게 하지만 소비자의 제품에 대한 선호도가 나타나 있지 않으므로 신제품이나 제품수정의 새로운 포지셔닝, 마케팅믹스 전략수립 등을 위한 정보를 제공해 주지 못한다는 단점이 있다. 따라서 소비자의 인지, 선호, 선택의 일련의 과정들을 분석하기 위해서는 제품의 인지도 혹은 포지셔닝 맵에 개인의 선호도를 포함시킬 필요가 있으며 다차원선호분석(MDPREF)기법은 이러한 목적에 적합한 분석 기법이라 할 수 있으며 자료의 비정칙값 분해(singular value decomposition ; SVD)를 이용한 저차원 근사화가 핵심 내용을 이루고 있다. 자세한 내용은 황선영 et al. (2001)을 참고하기 바란다.

전체 자료가 몇 개의 그룹으로 구성되어 있는 경우를 생각해 보자. 특히 각 그룹별 차이가 뚜렷한 경우에는, 전체 자료에서 얻은 하나의 다차원선호분석의 결과보다는 각 그룹별로 세분화된 그래프가 그룹별 차이를 볼 수 있어 좀더 심층적인 해석이 가능하다. 전체 자료가 l개의 그룹으로 구성되어 있으며, m개의 상품에 대해서 각 그룹별 $n_j (j=1,\cdots, l)$명에게 얻은 선호도 자료를 생각해 보자.

$$X_1 = \underbrace{\boxed{\quad\cdots\quad}}_{m\times n_1} \quad X_2 = \underbrace{\boxed{\qquad\cdots\qquad}}_{m\times n_2} \quad \cdots \quad X_l = \underbrace{\boxed{\quad\cdots\quad}}_{m\times n_l}$$

이 자료행렬은 $\sum_{j=1}^{l} n_j$명의 사람에게서 얻은 선호도 자료이므로 각 열별로 표준화한다. 각 행렬의 표준화된 행렬을 Y_1, Y_2 ,⋯라 하면, 입력 자료는

$$Y = [\ Y_1 \vdots Y_2 \vdots \cdots \vdots Y_l\ ]$$

로서, 크기가 $m\times \sum_{j=1}^{l} n_j$ 인 표준화된 자료행렬이 된다. 행렬도를 그리기 위해 Y 를 다음과 같이 비정칙값 분해한다.

$$Y_{m\times(n_1+\cdots+n_l)} = P_{m\times r}\,\Delta_{r\times r}\,Q^T_{r\times(n_1+\cdots+n_l)}$$

여기서 r 은 행렬 Y 의 계수이며, 행렬 $P=(p_1, p_2, \cdots, p_r)$의 열벡터는 YY^T 의 정칙고유벡터(eigenvector)들이고, 행렬 Q의 열벡터는 Y^TY의 정칙고유벡터이다. 즉, $P^TP=Q^TQ=I_{r\times r}$ 이다. 그리고 행렬 Δ의 대각원소는 비정칙값(singular value) $\delta_1 \geq \delta_2 \geq \cdots \geq \delta_r > 0$ 로 구성되어있는 대각행렬이다. 이제 P, Q에서 처음 두 개의 열과 가장 큰 비정칙값 δ_1, δ_2 만을 이용하여 다음 근사식을 생각하자.

$$\begin{aligned} Y \simeq P_{(2)}\,\Delta_{(2)}\,Q^T_{(2)} &= P_{(2)}\,\Delta^c_{(2)}\,\Delta^{1-c}_{(2)}\,Q^T_{(2)} \\ &= G_{(2)}\,H^T_{(2)} \end{aligned}$$

이때 행 표시자와 열표시자는 각각 $G_{(2)}=P_{(2)}$과 $H_{(2)}=Q_{(2)}\Delta_{(2)}$으로 선택한다.

이제부터는 이렇게 얻어진 공통공간으로부터 그룹별(주체별)공간을 끌어 낼 수 있는 가중치를 찾는 방법을 제시하도록 하자. 공통공간에서 상품의 좌표를 행렬 $R_{m\times 2}$, 각 그룹별(주체별)로 축에 대한 가중치 행렬을 W_1 , $\cdots$, W_l 이라 하면, 개인지각공간은 RW_1 , $\cdots$, RW_l 로 나타낼 수 있게 된다. l개 그룹들의 가중치 행렬은

$$W_1=\begin{pmatrix}\sqrt{w_{11}} & 0 \\ 0 & \sqrt{w_{12}}\end{pmatrix}, \ \cdots, \ W_l=\begin{pmatrix}\sqrt{w_{l1}} & 0 \\ 0 & \sqrt{w_{l2}}\end{pmatrix} \tag{3.1}$$

로서 대각행렬이며, 다음 조건을 만족하도록 정하도록 한다.

$$\frac{1}{l}\sum_{j=1}^{l} W_j = I_{2\times 2}, \quad I_{2\times 2}=\begin{pmatrix}1 & 0 \\ 0 & 1\end{pmatrix} \tag{3.2}$$

식(3.2) 제약 조건의 의미는 개인지각공간에서의 좌표의 평균은 공통공간에서의 좌표의 평균과 같게 되는 것을 의미한다(Gower and Hand(1996, p. 219)). 몇 개의 그룹으로 이루어져 있는 전체 자료를 각 그룹별로 독립된 자료로 하여 각각에 대해서 다차원선호분석을 사용하는 경우에는 각 그룹별 결과물의 축이 같다는 보장이 없다. 즉 이것은 INDSCAL의 개념이 아니다. 이처럼 축의 의미가 같지 않은 상태에서는, 그룹별 비교가 불가능하기 때문에 전체 자료를 공통공간상에 나타내고 여기서 다시 축에 대한 가중치를 부여하여 개인지각공간을 이끌어 내야한다. 논의를 간단히 하기 위해 $l=2$ (즉, 두 그룹)라 가정하면서 다음과 같은 목적함수를 제시한다.

$$[\ |RW_1 - R_1T_1|^2 + |RW_2 - R_2T_2|^2\ ] \tag{3.3}$$

식 (3.2) 제약조건(물론 $l=2$)하에서 식(3.3)을 최소화 시키는 W_1과 W_2가 그룹별 축에 대한 가중치 행렬이 된다. 여기서,

$$\begin{aligned} R &\equiv \text{전체자료의 SVD에서 유도된 2차원상의 좌표} \\ R_1 &\equiv \text{첫 번째 그룹자료의 SVD에서 유도된 2차원상의 좌표} \qquad (3.4) \\ R_2 &\equiv \text{두 번째 그룹자료의 SVD에서 유도된 2차원상의 좌표} \end{aligned}$$

를 나타낸다. T_1과 T_2는 회전각도 θ와 회전방향에 따라 사용될 수 있는 직교변환 행렬 중에서 시계바늘 반대방향의 회전을 선택한다. 목적함수를 이와 같이 세운 이유는 R과 R_1, 그리고 R_2는 비정칙치 분해를 통해 쉽게 얻을 수 있으며, 여기에 축의 의미를 조정하면 INDSCAL 개념과 비슷한 좌표를 유도할 수 있기 때문이다. 좌표행렬 R_1과 R_2는 INDSCAL 개념의 개인지각공간이 아님에 유의하라. R은 전체 자료의 행 좌표이고 R_1은 첫 번째 그룹의 행 좌표이며 R_2는 두 번째 그룹의 행 좌표이긴 하지만, 각각의 분석을 통해서 얻어진 좌표행렬들이므로 좌표들의 축의 의미가 서로 다르기 때문이다. 그러나 전체 자료에는 첫 번째 그룹과 두 번째 그룹이 포함되어 있다는 점에 착안하여 R_1과 R_2를 회전을 통해 R에서의 축의 의미와 가장 비슷해지도록 유도한다. 회전에 앞서, R에서 좌표들의 벡터길이를 기준으로 하여 R_1과 R_2에서의 좌표들의 벡터길이를 조정하는 과정을 먼저 거친다. 이러한 사전 작업 후에 θ_1과 θ_2를 1°씩 변화해가면서 R과 R_1T_1이 가장 유사할 때의 회전각도 θ_1을 선정한다. 같은 방법으로 R과 R_2T_2가 가장 유사할 때의 회전각도 θ_2를 선정하여 $TT' = T'T = I$를 만족하는 직교변환 행렬 T_1과 T_2를 결정한다. 본 논문에서는 R과 R_1T_1이 가장 유사할 때의 기준을, R에서의 m개의 상표점에 대응하는 R_1T_1에서의 m개의 상표점과의 거리의 합을 최소로 하는 각도로 삼았다. 회전각도를 변화시켜주면서 새로운 좌표를 생성한 뒤, 각각의 대응하는 상표점과의 거리를 계산하여 그 합을 최소로 제공해주는 θ_1과 θ_2를 선택하고 제약 조건에 의해서 $\sqrt{w_{11}}$과 $\sqrt{w_{12}}$를 순차적으로 변화시켜가면서 목적함수를 최소화시키는 $\sqrt{w_{11}}$과 $\sqrt{w_{12}}$를 선택하여 각 그룹별 가중치 행렬을 구하도록 한다. 이상의 내용을 다음과 같이 요약할 수 있다.

[1단계] 목적함수 수립 ; 식 (3.3)

[2단계] R, R_1, R_2 생성

[3단계] R_1, R_2 의 길이 조정

[4단계] 최소거리 합을 제공하는 회전각도 선정

[5단계] 목적함수를 최소화시키는 가중치 결정.

4. 예 제

현재 유통되고 있는 5가지 신용카드에 대한 선호도 자료를 가지고 3절에서 제안된 방법을 적용시켜 보자. 삼성카드, LG카드, 국민카드, BC카드, 현대카드를 조사대상으로 삼았으며, 신용카드 선택과 관련 있는 속성으로는 전반적인 서비스와, 광고이미지의 2개를 고려하였다. 소비자는 통계학과 교수 7명(그룹 1)과 통계학과 대학원생 20명(그룹 2)을 선택하여 총 27명으로 이루어져 있다. 조사된 자료는 5가지 상품 개체에 대한 소비자의 선호자료이며, 자료를 우선 열 표준화 시키도록 한다. 열 표준화된 자료를 비정칙값 분해하여 행 표시자를 구하고 이를 2차원으로 근사시켜 전체자료의 형상공간에서의 좌표행렬(R)을 먼저 구한다. 같은 방법으로 그룹1의 좌표행렬(R_1)과 그룹2의 좌표행렬(R_2)을 구한다. 이 세 개의 좌표행렬들은 그 축의 의미가 서로 같지 않으므로, R에서의 축의 의미와 가장 비슷해지도록 변환행렬 T_1, T_2를 먼저 정하도록 하자. 그 구체적인 방법으로는 R을 기준으로 R_1, R_2의 벡터길이를 조정한 후 각도를 변화시켜 가면서 대응하는 5개의 상품의 거리합을 구하여 그 값이 최소가 되는 각도를 정한 결과 $\theta_1 = 172^\circ$, $\theta_2 = 343^\circ$ 가 선택되었다. [표4.1]은 0° 에서 360° 까지의 회전 후 나온 거리합 중 일부를 수록한 것이다.

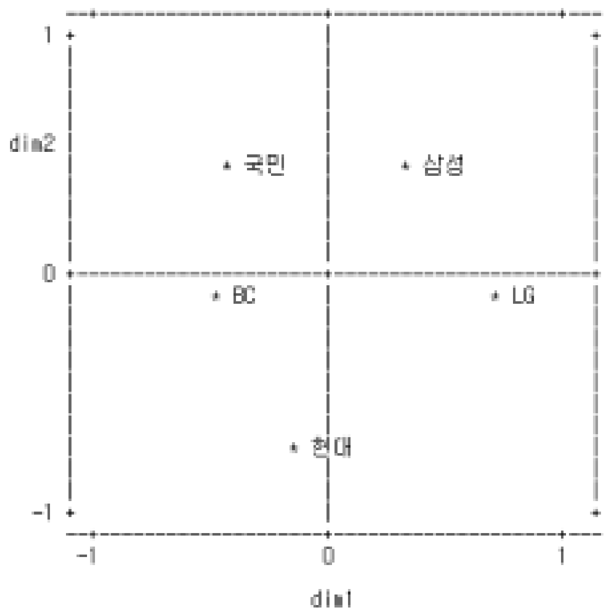
[표4.1] 각도변화에 따른 거리 합

회전각도	R_1 거리합	R_2 거리합	회전각도	R_1 거리합	R_2 거리합
30	4.82106	2.20405	210	3.41705	5.67319
60	4.84265	3.58861	240	3.74475	4.91225
90	4.53435	4.73725	270	3.81736	3.81668
120	3.91716	5.56318	300	3.62991	2.46111
150	3.41026	6.01013	343	4.13355	***1.03835***
172	***3.18164***	6.07833	360	4.46988	1.26523

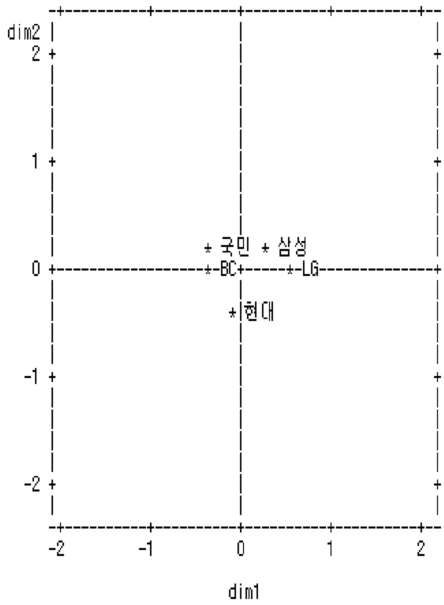
이어서 새로운 좌표행렬 $R_1 T_1$과 $R_2 T_2$를 이용하여 앞에서 제시한 목적함수를 최소화 시키는 W_1과 W_2를 찾아보자. 그런데 개인지각공간에서의 좌표의 평균이 공통공간에서의 좌표의 평균과 같도록 제약조건을 부여하였으므로 이를 만족시키는 가중치 행렬을 찾아보도록 한다. 즉, $\sqrt{w_{11}}$과 $\sqrt{w_{12}}$을 0부터 2까지 0.1씩 변화시켜가면서 식(3.5)를 최소화시키는 값을 선택하고 그 값에 따라 $\sqrt{w_{21}}$과 $\sqrt{w_{22}}$를 자동으로 결정한다. 이렇게 해서 선택된 값은 $\sqrt{w_{11}} = 0.7745967$, $\sqrt{w_{12}} = 0.4472136$이다. 선택된 가중치를 반영하여 공통공간의 그래프와 각 그룹별 그래프를 얻을 수 있다.

[그림 4.1]에서 보는바와 같이 제 1축(수평축)은 왼쪽의 은행카드(국민, BC)와 오른쪽의 비은행카드(삼성, LG)로 분할하고 있다고 해석할 수 있다. 그리고 제 2축(수직축)에 의해서 위쪽으로 국민, BC, 삼성, LG가 위치하고 아래쪽으로 현대가 위치하는 것으로 보아 축의 의미를

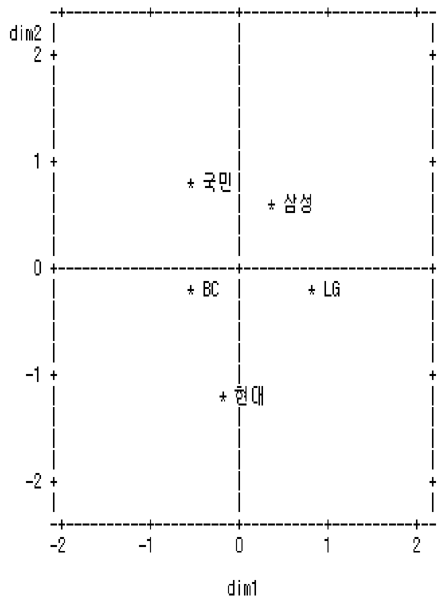
카드 보유 정도 혹은 카드의 대중성으로 해석할 수 있다. 사전 조사에서 대체적으로 위쪽에 위치해 있는 카드는 하나 이상은 보유하고 있는 반면에 현대카드를 가지고 있는 응답자는 거의 없었던 점으로 미루어 제 2축의 의미를 이렇게 정해도 무방하리라 생각된다. 또한 그룹별 그래프인 [그림 4.2, 4.3]을 살펴보면, 그룹1의 가중치는 그룹2에 비하여 제 1축이 상대적으로 크다. 또한 같은 제 1축에 대해서는 그룹2가 더 큰 값을 반영하고 있다. 이러한 사실로부터 그룹2가, 카드 선택에 있어, 은행카드인지 비 은행카드인지에 더 큰 의미를 두고 있음을 알 수 있다. 즉, 은행카드에 비해 비 은행카드들이 할인혜택이 다양한 점으로 미루어 학생그룹(그룹2)이 카드 선호에 있어 제 1축에 대한 가중치의 반영 정도가 그룹1과 차이를 보이는 것은 당연해 보인다. 이렇듯 INDSCAL 개념을 다차원선호분석에 확장하여 보다 심도 있는 해석이 가능해지는 이점을 얻을 수 있다.



[그림4.1] 공통공간에서의 좌표

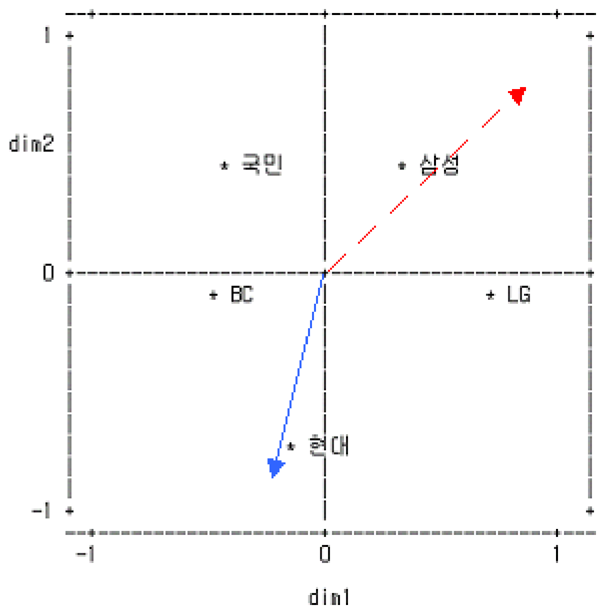


[그림4.2] 그룹1의 좌표



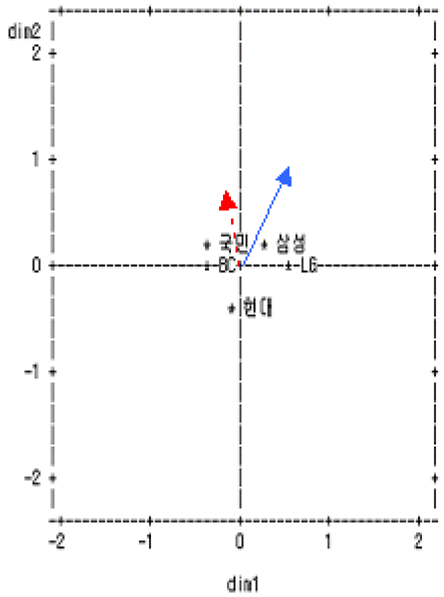
[그림4.3] 그룹2의 좌표

다차원선호분석을 통해 얻어진 그래프에 속성을 추가하여 그룹별 인지도를 그려 보도록하자. 먼저 식(3.4)에서 정의된 좌표행렬 R을 독립변수행렬로 하고 각각의 속성을 반응변수로 하여 회귀분석을 한 후 추정된 회귀계수를 그 속성의 좌표값으로 하여 벡터로서 그래프에 추가하여 전체 자료의 인지도를 그릴 수 있다. 자세한 내용은 황선영 et al.(2002)을 참고하기 바란다. 그룹1의 인지도는 R_1을 독립변수행렬로 하고 그룹1에서의 속성을 반응변수로 하여 회귀분석을 시행한다. 같은 방법으로 그룹2의 인지도를 얻을 수 있다. 속성은 서비스와 광고 이미지로서 서비스는 점선으로, 광고 이미지는 실선으로 표시하였다.

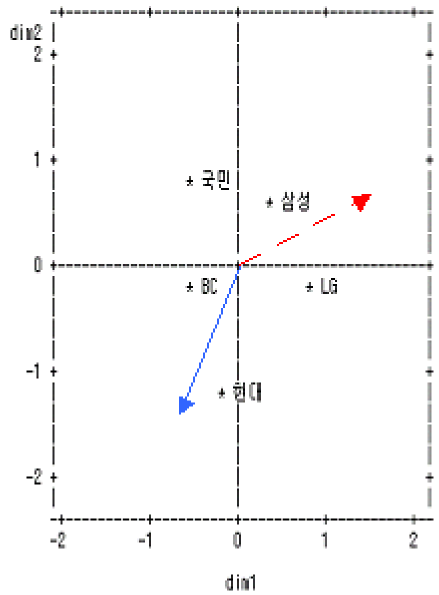


[그림4.4] 전체 자료의 인지도

카드를 속성 벡터에 직교 투사시킨 거리가 추정된 속성의 만족도 값으로 해석된다. 세 개의 인지도를 통해 그룹간에 속성의 만족도가 상반됨을 알 수 있다. 또한 전체 자료의 인지도 [그림 4.4]와 그룹2의 인지도 [그림 4.6]은 비슷한 경향을 보이는데, 그룹1은 상당히 다른 방향으로 속성 벡터를 나타내고 있다. 광고 이미지를 나타내는 속성에 있어, 그룹2는 현대카드가 이미지가 좋은 것으로 나타나 있는 반면에 그룹1에서는 현대카드의 광고 이미지가 가장 좋지 않다. 대신에 삼성카드나 LG카드의 광고 이미지가 긍정적으로 나타나 있음을 알 수 있다. 서비스 속성에 대해서는, 그룹2는 삼성카드와 LG카드가 서비스가 좋은 것으로 인식하고 있으며 그룹1은 [그림 4.5]에서 보듯이 BC카드나 국민카드가 서비스가 좋다고 응답하고 있음을 알 수 있다. 이러한 해석은 INDSCAL 개념의 도움 없이는 쉽지 않은 결과물이며 본 논문을 통해 그룹간의 숨어있는 차이를 알아볼 수 있는 방법론을 제시한 점에 의미를 부여하고 싶다.



[그림4.5] 그룹1의 인지도



[그림4.6] 그룹2의 인지도

감사의 글

논문을 심사해 주신 두 분의 심사위원님께 감사를 드립니다. 본 연구는 숙명여자 대학교 교내 연구비(2001)지원에 의해 수행되었음.

참고문헌

1. 최용석 (1995), *SAS다차원척도법*, 자유아카데미.
2. 최용석 (1999). *행렬도의 이해와 응용*. 부산대학교출판부.
3. 황선영, 정수진, 김영원(2001). 다차원선호분석의 최적척도화 및 부분수량화. *응용통계연구,* 14권, 305-320.
4. 황선영, 정수진, 김영원(2002). Perceptual map using weighted regression. *한국통계학회논문집*, 8권, 651-659.
5. Gower, J.C. and Hand, D.J (1996), *Biplots,* Chapman & Hall, London.

[2003년 5월 접수, 2003년 7월 채택]