

齒藥市場의 製品地位에 關한 集落分析*

An Application of Multivariate Analysis on Product Positioning with Emphasis on Cluster Analysis

成 三 慶**

目	次
I. 緒 言	③ Ward方法
II. 集落分析技法	2) 最適分離技法
1. 類似性和 거리의 測度	3. 最適 集落의 數
1) 類似性的의 測度	III. 齒藥製品에 對한 集落分析
2) 거리의 測度	1. 資 料
2. 分析技法의 種類	2. 分析過程과 結果
1) 系譜的 技法	3. 要因分析
① Centroid方法	IV. 結 論
② Average Linkage方法	

I. 緒 言

分類技法은 모든 학문의 발달에 큰 기여를 했다. 生物學과 動物學의 分類技法은 다윈에 의한 進化論의 提唱에 도움을 주었고 化學 元素 週期表는 原子構造를 이해하는데, 그리고 天文學에서는 크기에 따라 별들을 분류함으로써 星雲에 관한

* 本 研究는 文敎部 學術研究 助成費의 支援을 받아 쓰여졌으며 調查過程에서 助言을 하여준 력키금성그룹의 李 正載 常務계 감사드린다.
 ** 高麗大學校 經營大學 副敎授

理論을 밝히는데 도움을 주고 있다.

분류라고 하면 일정 대상의 개체들을 애초에 정확히 알지 못하는 집단으로 동질의 개체를 할당하는 과정을 말한다. 統計的 分類技法으로는 최근 20년내 개발되어온 集落分析(cluster analysis)이라는 기법이 있는데 多變量 資料중에서 자연스러운 集落을 찾아 자료에 관한 記述을 간단히 하고, 미래의 標本에 대한 假說을 세우며 또 集落에 기초를 두고 예측을 하는데 도움을 준다. 集落分析은 判別分析처럼 母集團내의 중요 차이점에 대한 사전적 가정이 없이 순수한 경험적 분류방법으로서 일종의 歸納的 기법이다. 어떤 학자들은 集落分析의 임의적 해법에 회의적 비판을 하지만 우선 분류하는 작업은 학문의 첫째 방법이자 최종의 방법이라는 말이 있듯이 이 기법에 대한 연구는 계속되고 있다.

집락분석기법은 경영학에도 이용되고 있으며 특히 마케팅연구분야에 대한 응용 사례는 Punj와 Stewart(1983)에서 찾아 볼 수 있다. 본 논문에서는 집락분석 기법에 대한 소개가 아직 넓지 못한 점에 비추어 기법에 대한 중요 개념을 간단히 밝히고 치약제품에 관한 標本을 갖고 集落分析을 통하여 제품의 地位에 대한 정보를 얻어 보았다. 본 연구의 한계와 기법의 적용상 제언은 結語에서 맺는다.

II. 集落分析技法

1. 類似性和 거리의 測度

集落分析技法은 어떤 標本の 個體(entity or individual)들 간의 동질이나 이질의 관계를 개체들 간의 거리의 개념으로 정보를 요약하려는 시도로 생각할 수 있는데 집락분석에서 이용되는 類似와 非類似를 측정하는 基準 尺度가 필요하다.

일반적으로 X_{ik} 를 i 번째 個體의 k 번째 變數에 대한 測定值라 하고 개체 i 의 p 개 변수에 대한 측정치의 집합을 i 번째 行벡터, $X_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ 로 표시할 때 n 개 개체 전체에 대한 $n \times p$ 資料行列, X 는 다음과 같이 표현하게 된다.

$$X_{n \times p} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & \dots \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & & & x_{np} \end{bmatrix}$$

여기서는 자주 사용되는 유사성과 비유사성의 측도 몇 가지만 소개한다.

1) 類似性的의 測度

유사성이라 함은 상대적 개념으로서 두 대상이 여러 측면에서 다른 것들에 비해 상대적 相似性的의 정도를 말한다. Gower(1971)는 定性和 定量자료에 두루 사용될 수 있는 類似性係數(similarity coefficient)를 다음과 같이 정의한다.

$$S_{ij} = \frac{\sum_{k=1}^p s_{ijk}}{\sum_{k=1}^p W_{ijk}}$$

여기서 W_{ijk} 는 (i, j, k) 에 대한 어떤 조건에 합당하는가의 여부에 따라 1과 0의 값을 갖는 加重值이며, S_{ij} 가 1에 가까울수록 두 개체(i, j)간의 유사성은 크게 된다. 定量 資料分析에 대한 유사점수 s_{ijk} 는

$$s_{ijk} = 1 - |x_{ik} - x_{jk}| / R_k$$

로 계산하고 이때 x_{ik} 는 k 번째 변수에 대한 i 번째 개체의 점수를 말하고 R_k 는 k 번째 변수가 갖는 범위를 나타낸다.

2) 거리의 測度(distance measures-metrics)

어떤 집합 E의 metric은 집합내의 두 점(i, j)들에 대해서 다음의 세 조건을 만족시키는 함수 d_{ij} 를 말한다. 즉,

(i) $d_{ij} \geq 0$; $d_{ij} = 0$ if $i = j$

(ii) $d_{ij} = d_{ji}$

(iii) $d_{ik} + d_{jk} \geq d_{ij}$

가장 흔히 쓰이는 거리의 測度로는 Euclidean metric을 들 수 있다. 즉 두 점 i 와 j 간의 거리 d_{ij} 는

$$d_{ij} = \left\{ \sum_{k=1}^p (x_{ik} - x_{jk})^2 \right\}^{1/2}$$

로 나타내는데 측정단위의 차이에 따른 영향을 배제하기 위하여 각 변수의 標準偏差로 측정치를 나누어 표준화된 자료를 사용하기도 한다.

Euclidean metric이 集落分析에서 가장 널리 사용되는 거리의 측도이나 경우에 따라서 Euclidean metric의 단점을 보완한 city block metric을 사용하며, Euclidean 거리를 共分散行列의 단위로 표준화시킨 Mahalanobis D^2 을 사용하기도 하고, Minkowski metric을 사용할 때도 있다.¹⁾

類似性的의 測度(S_{ij})와 거리의 測度(d_{ij})와의 가장 큰 차이는 유사성의 측도는 0과 1사이의 값을 갖는 반면 거리의 측도는 어떤 陽의 값을 갖는 점이다. 거리의 함수는 쉽게 유사성의 함수로 변환시킬 수 있으나 역방향의 변환은 그리 쉽지 않다. Euclidean metric이 모든 경우 항상 최적의 측도는 아니더라도 가장 널리 사용되고 있으며 꼭 유일한 유사성과 거리의 측도에 제한되지 않는 점이 集落分析技法의 장점의 하나이다.

2. 分析技法의 種類

변수의 수, $p \leq 3$ 이라면 즉 2 혹은 3 차원에서는 평면이나 공간에 그려진 자료를 目測으로 어렵지 않게 集落을 가려 낼 수 있다. 실제로 변수의 수가 증가함에 따라 보다 합리적인 여러 수학적 기법이 개발되었는데 본 논문에서 직접 사용된 系譜的 技法(hierarchical technique)과 最適分離 技法(optimization-partitioning technique)만 살펴 본다.²⁾

1) 系譜的 技法

계보적 기법은 n 개의 개체를 연속적인 결합에 의해 여러개의 집락을 형성하여 결국 모든 대상을 포함하는 한개의 집락이 되어가는 併合的 技法과 n 개체를 전부 포함하는 하나의 집합에서 세분화된 集落으로 나누어 결국 오직 하나의 개체를 갖는 n 개의 집락으로 나누어 가는 分解的 技法이 있다. 물론 분석 과정중 최적 단계에서 가장 합리적인 집락을 찾게 된다. 거리의 측도를 사용한 여러가지

1) 類似性和 거리의 測度의 諸 方法에 관한 자세한 설명은 Everitt (1974)의 p. 49를 참조할 것

2) 系譜的 技法과 最適分離方法이외에 상대적으로 밀집된 개체들을 포함하는 영역을 찾아 집락을 구하는 稠密度 技法(density search technique)은 Everitt(1974)를 참조할 것. 가능하면 Everitt(1974)의 1980년도 版을 참조할 것. 집락끼리 서로의 중복을 허용하는, 즉, 한 개체가 한 개 이상의 집락의 구성원이 될 수 있는 重複技法(clumping or overlapping technique)는 Arabie, Carroll, De-Sarbo와 Wind(1981)을 참조할 것.

併合的 系譜的 技法 가운데 가장 유용하게 사용되는 centroid 방법, Ward 방법과 average linkage 방법만 알아 본다.³⁾

① Centroid 방법

이 방법은 원래 개체들의 집락보다 變數들의 集落을 찾기 위한 방법으로 제안된 것으로 집락간의 거리는 Euclidean 空間에서의 집락의 중심(centroid)간의 거리로 측정하여 centroid간의 거리가 가장 최소인 집락끼리 병합을 시작하여 自然的 集落을 형성하여 가는 방법이다.

예를 들어 5개의 개체($n=5$)가 2개의 변수($p=2$)에 측정되어 다음과 같은 資料行列을 갖는다고 가정하자.

	1	2
$X_1 =$	1.0	1.0
	2	2.0
	3	3.0
	4	2.0
	5	0.0

이때 각 개체간의 거리, d_{ij} 를 제공하여 얻은 거리행렬, D_1 은 다음과 같다.

	1	2	3	4	5
$D_2 =$	0.0	1.0	29.0	50.0	50.0
	2	0.0	26.0	49.0	53.0
	3	29.0	0.0	5.0	13.0
	4	50.0	49.0	0.0	4.0
	5	50.0	53.0	13.0	0.0

3) 본 논문의 세가지 방법 이외에 nearest neighbour (single linkage) 방법, furthest neighbour (complete linkage) 방법, median 방법, Lance & Williams Flexible 방법 등에 관해서는 Anderberg (1973)을 직접 참조할 것.

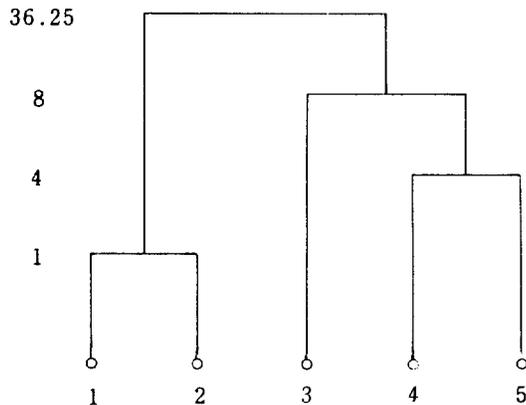
Jardine 과 Sibson (1971)은 single linkage 方法이 連續性을 유지하며 수 천의 개체(entities or individuals)를 가진 자료를 취급할 수 있는 계산상의 신속성과 융통성이 있으므로 가장 선용할 만한 방법으로 간주한다. 하지만 실제의 모든 상황에 대해 오직 한 방법만을 적용할 것이 아니라 둘 이상의 방법을 응용하여야 할 것이다. 둘 이상의 방법이 같은 형태의 集落群을 보여 준다면 그와 같은 自然的 集落이 존재한다는 確實性이 높아지기 때문이다. 반면에 Milligan (1980)은 Ward 方法과 average linkage 方法이 여러 연구 결과 가장 유용한 방법으로 지적된다고 말한다.

개체 1과 2의 거리 d_{12} 가 가장 짧으므로 1과 2가 併合되어 한 집락을 이루고 이 집락의 centroid는 이들 변수값의 平均이 되어 다음과 같은 축소된 새로운 資料行列 X_2 와 이에 따른 거리행렬 D_2 를 얻는다.

$$X_2 = \begin{matrix} & & 1 & 2 \\ (1,2) & \left[\begin{array}{cc} 1.0 & 1.5 \\ 3 & 6.0 & 3.0 \\ 4 & 8.0 & 2.0 \\ 5 & 8.0 & 0.0 \end{array} \right] \end{matrix}$$

$$D_2 = \begin{matrix} & (1,2) & 3 & 4 & 5 \\ (1,2) & \left[\begin{array}{cccc} 0.0 & 27.25 & 49.25 & 51.25 \\ 3 & 27.25 & 0.0 & 5.0 & 13.0 \\ 4 & 49.25 & 5.0 & 0.0 & 4.0 \\ 5 & 51.25 & 13.0 & 4.0 & 0.0 \end{array} \right] \end{matrix}$$

거리행렬 D_2 에서 가장 작은 원소는 개체 4와 5의 거리, d_{45} 이며 다시 한 집락을 이루게 되어 모든 개체가 하나의 집락으로 형성할 때까지 이와 같은 절차가 반복되어 결국 한 집락을 형성한다. 위의 예의 집락의 凝集過程을 dendrogram으로 표시하면 <그림 1>과 같다. dendrogram이란 단계적으로 집락들이 凝集 혹은 分解되는 과정을 2次元上の 도형으로 나타낸 것을 말한다.



<그림 1> Centroid방법 dendrogram

② Average linkage 방법

單一連結(single linkage) 방법에서는 집락간의 가장 가까운 관련을, complete linkage 방법에서는 집락간의 가장 먼 관련을 파악하여 집락들을 병합하여 나가지만 average linkage 방법은 집락내의 모든 가능한 원소들간의 모든 거리의 平均으로서 각 집락을 파악한다. 이 방법은 centroid 방법처럼 極限값을 갖는 개체에 대해 큰 영향을 받지 않으나 작은 分散을 가지는 집락을 형성하는 방향으로 偏倚를 갖게 된다.

③ Ward 방법

Ward는 개체를 집락화 시킴으로서 발생하는 情報의 손실은 각 개체가 속해 있는 집락의 平均으로부터 平方偏差의 총계를 각 단계에서 산출하여 봄으로써 측정할 수 있다고 생각한다. 각 단계마다 모든 가능한 쌍의 병합을 고려하여 平方誤差를 계산하고 最小平方誤差의 합을 갖는 쌍을 병합하여 나간다.

예를 들어 다음과 같이 5개의 개체가 하나의 변수에 의해 측정된 資料行列, X가 있다고 가정하자.

$$X = \begin{matrix} & 1 \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{bmatrix} 1 \\ 2 \\ 7 \\ 9 \\ 12 \end{bmatrix} \end{matrix}$$

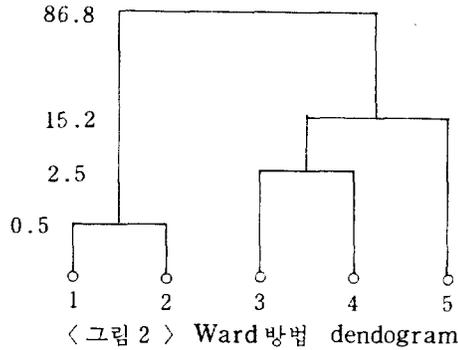
誤差平方合(error sum of squares; ESS)은 x_i 가 i 번째 개체의 값 일때 다음과 같이 계산한다.

$$ESS = \sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum x_i)^2$$

첫번째 단계에서 각 개체는 그것대로 하나의 집락으로 간주되므로 誤差平方合은 0이 된다. 다음으로 E.S.S.의 最小增分은 개체 1과 2를 병합함으로써 E.S.S.는 0.5가 된다. 다음 단계는 개체 3과 4를 병합하여 E.S.S.는 2가 증가되어 2.5가 되고 개체 5가 3과 4와 다시 병합하여 E.S.S.는 12.7이 증가된 15.2가 된다. 마지막으로 두 집락(1, 2)와 (3, 4, 5)가 병합하여 전체 E.S.S.는 71.6이 증가된 86.8이 된다. dendrogram으로 이와 같은 과정을 요약

하면 <그림 2>와 같다.

E.S.S.의 最小增分은 병합된 집락의 centroid간의 Euclidean 거리의 자승에 비례하는데 계산과정에서 centroid간의 거리의 가중치를 사용한다는 점이 centroid 방법과 다르다.



2) 最適分離 技法

이 방법은 미리 존재할 것같은 集落의 수에 맞추어 사전에 규정된 判定基準에 맞도록 集락은 最適化시키는 방법이다. 따라서 개체의 재할당을 통하여 미비한 初期 分割을 나중 단계에서 개선할 수 있게 된다.

p次元 空間에서 추정되는 k개의 集락의 중심(centroid)을 k개의 集落의 種子(cluster seed)로 시작하여 개체들을 가장 가까운 종자를 갖는 集락에 割當하면서 동시에 集락의 중심을 추적, 구성원의 위치까지 고려하여 다시 계산한다. 그후 단계의 할당과정에서 集落內平方sum이 특히 큰 集락을 再分離하여 集落內平方sum이 작은 集락과 병합할 수 있어 고정되었던 集락의 수 k 자체를 변화시킬 수도 있다.

개체를 재배치하는데 사용하는 判斷基準은 다음과 같은 等式에서 유도한다.

$$T = W + B$$

여기서 T는 全體散布行列(total scatter or dispersion matrix)이라 하고 그 元素 t_{ij} 는 변수 i와 j간의 誤差交積合(sum of squares and cross products)이다. W는 集落內의 散布行列(within-groups dispersion matrix)이고 W_i 가 i번째 集落內의 散布行列이므로 $W = \sum_{i=1}^k W_i$ 가 되고 B는 集落間 散布行列(between-groups dispersion matrix)이다.

행렬 T는 일정한 자료에 대해서 고정이므로 판단기준은 B와 W의 함수로 된다. 再配置하는데 사용되는 첫째 기준은 trace(W)가 최소가 되게 재배치하는 것으로 集落內 散布度의 sum이 최소가 되도록 유도한다. 결국 행렬 B의 trace의 최대화를 추구하는 것과 동일한 기준이다.

둘째로 행렬 W의 行列式(determinant)가 최소가 되도록 재배치하는 기준인데 이는 사실 Wilks의 λ 統計量, $|W| / |T|$ 의 최소화 과정과 같다.

마지막 기준으로는 $\text{trace}(BW^{-1})$ 을 최대화하는 것으로 λ_i 를 BW^{-1} 의 固有值(eigenvalue)라고 할 때

$$\text{tr}(BW^{-1}) = \sum_{i=1}^p \lambda_i$$

이므로 λ_i 의 합을 최대화하는 것과 같다.

지금까지 분석기법으로 系譜的 技法과 最適分離 技法을 살펴 보았는데 系譜的 技法의 한 단점은 分析處理의 처음 단계에서 잘못 분류된 개체를 재배치하여 바로 잡아 줄 절차가 없는 것이다. 生物學的 자료처럼 계보의 구조가 명확히 존재할 수 있는 자료에 系譜的 技法이 가장 적합하게 이용될 수 있다.

最適分離 技法의 단점은 분리기준을 따라 분석하는 과정에서 部分 極大化에 머물고 마는 경우가 흔하다는 것이다. 조그만 자료에 한해서는 動態計劃 技法(dynamic programming algorithm)이 개발되어 언제나 最適值를 찾을 수 있으나 전반적으로 이 기법은 전자계산기의 시간의 소모가 너무 큰 단점이 있다.

적용하는 기법에 따라 상호 다른 결과를 얻는 경우가 있기 때문에 기법을 선택하기 이전에 우선 그 기법의 배후 가정을 검토하여 適合性을 판단하여야 한다. 상식적으로는 여러 기법을 병행하여 적용한 후 다수의 기법이 일치된 결과를 보이는 集落分析을 취하고 혹은 자료를 無作爲로 둘로 나누어 분석하여 나누어진 두 자료에서 자료전체를 분석할 때와 거의 동일한 집락분석의 결과를 얻을 수 있는가 시험해 봄으로써 결과의 신빙성을 타진할 수도 있다.

3. 最適 集落의 數

集落分析에서 집락의 수를 결정하기는 要因分析에서 要因의 수를 결정하는 문제보다 더욱 어렵고 아직 만족할 만한 방법이 없다. 系譜的 技法은 집락의 수보다 자료가 어떠한 系譜的 構造를 갖는가 하는데 더 많은 관심을 갖는 수가 많다. 이러한 경우에는 dendogram을 검토하여 적절한 집락의 수를 결정한다. 예를 들면 Ward의 방법을 사용할 경우에 집락수에 대한 E.S.S.의 增分을 검토하여 급격한 증분을 갖는 위치에 대응하는 집락의 수로 결정한다. 부적합한 단계에서 병합을 시도할 때 E.S.S가 급증하기 때문이다.

합리적 집락의 수를 찾기 위한 여러가지 방법중, Calinski와 Harabasz는

다음과 같은 C비율을 고안했다.⁴⁾ 즉,

$$C = \frac{\text{trace}(B)}{g-1} \bigg/ \frac{\text{trace}(W)}{N-g}$$

여기서 C의 값이 집락의 수를 나타내는 g의 증가에 따라 單調增加를 보이면 이는 자료가 集落的 構造가 아님을 보이고 C의 값이 g에 대해 單調減少를 보이면 系譜的 集落構造를 갖는다고 한다. 따라서 g₀에서 C의 값이 최대값을 갖으면 g₀개의 집락이 존재한다고 판단하게 되지만 집락들이 거의 球形(spherical)集落的 모습을 갖는 경우에 한하여 유용하다.⁵⁾

본 논문에서는 SAS(1982)의 집락분석절차에서 사용되는 立體集落判斷基準(cubic clustering criterion, CCC)을 따른다. Arnold(1979)는 Monte Carlo 방법을 사용하여 集落內 散布行列의 determinant, 즉, |W|에 기초를 둔 집락수를 판정하는 基準値의 分布表를 작성하였다. 이와 달리 SAS의 분석은 행렬 W의 行列式보다 trace의 極小를 피하는 절차를 택하고 標本은 超四角形(hyperrectangle)상에서 一樣分布의 形態를 취하기 때문에 집락들은 거의 超立方體(hypercube)와 같은 형태를 갖는다고 가정한다. 이런 가정하에서 集落內 平方誤差合(within-cluster sum of squares)에 대한 期待値의 近似値를 구하여 집락의 수에 대한 判斷基準을 세운 것을 CCC라고 한다.

立體集落判斷基準(CCC)의 값을 2 내지 개체의 총수의 10분의1 정도의 집락의 수에 상응하도록 도표를 그려봄으로써 最適 集落的 數를 판단할 수 있는데 집락당 평균 개체의 수가 10보다 작을 때는 CCC는 큰 변화의 폭을 갖고 움직인다. CCC가 2나 3보다 큰 값에서 도표상에 갖는 尖峰에 대응하는 숫자가 적절한 집락의 수이고 0과 2 사이에도 가능한 집락수가 있을 수 있다. 系譜的 集落은 여러개의 尖峰을 갖는다. 이에 반해 非系譜的 集落은 尖峰에 앞서 급격히 상승했다가 완만히 하락한다. 만약 집락의 수가 증가함에 따라 CCC의

4) Everitt(1974), p. 60

5) Mojena와 Wishart(1980)는 집락의 수를 결정하기 위한 여러가지 heuristic methods를 제외하고, Binder(1981)는 Bayesian approach를 보여준다. 자료가 多變量 正規分布에서 얻어진 無作爲 標本으로 간주한 접근방법은 Wolfe(1978)와 Lee(1979)에서 찾아 볼 수 있고 Hartingan(1978)과 Arnald(1979)는 正規分布보다 一樣分布(uniform distribution)에서 오히려 더 적절한 歸無假說을 설정할 수 있다고 말한다.

값이 單調增加하면 자료가 너무 많이 반올림됐거나 몇자리 숫자가 안되는 자료 들일 경우에 일어나는 현상이다.

이상에서 살펴 본 집락의 수를 결정하기 위한 방법은 아직도 확고한 이론적 妥當性과 실제적 有用性에 關한 입증을 얻기 위한 연구가 進行중이다.

Ⅲ. 齒藥製品에 對한 集落分析

1. 資 料

우리나라 齒藥市場에 있어서의 製品地位(product positioning)와 아울러 市場分割(market segmentation)에 對한 본격적인 調査를 進行할 수 있을 정도 의 세심한 設問紙를 작성했지만 본 연구의 규모의 제약상 製品地位만을 파악하기 위한 손쉬운 標本抽出을 택했다. 購買되는 치약에 對한 標本은 高麗大學校 經營大學 2 學年 學生의 가정 250 세대 에 84 年 11 月 設問紙를 배포하여 完전한 標本의 가치 가 있도록 응답을 하여준 學生의 가정에서 구매되는 130개의 치약으

< 表 1 >

購買動機에 對한 문항

X ₁ .	단순한 치아세척과 그 밖의 구강위생을 위하여
X ₂ .	충치예방
X ₃ .	플라그(치석) 제거
X ₄ .	잇몸질환(치주염) 예방
X ₅ .	시린이 예방
X ₆ .	새하얀 치아를 원해서
X ₇ .	구취제거를 爲해서
X ₈ .	맛과 香기가 좋아서
X ₉ .	소독작용을 하는 소금끼가 있어서
X ₁₀ .	치약의 모양이 마음에 들어서
X ₁₁ .	값이 다른 치약보다 싸기 爲문에
X ₁₂ .	외국산이 품질이 더 나은 경우도 있어서

* 문항자체를 구매동기에 關한 변수로 표시

로 한했다.

購買動機에 관한 <表 1>의 12가지 문항에 대한 응답자의 생각과 느낌을 리커트型 7等級 尺度로(Likert-type 7-scale) 분류하여 해당란에 표시하도록 했다. 블렌닥스치약 29번, 클로즈업치약 16번, 하이얀치약 11번, 럭키치약 23번, 케디안치약 13번, 페리오치약 20번, 화이트치약 18번으로 총계 130으로 집계되었다.

각 제품에 대한 明細的 說明은 <附錄 1>을 참조할 수 있으며 <附錄 2>에서는 각 제품별로 소비자가 12가지 문항에 매긴 점수를 표준화하여 계산된 평균치를 볼 수 있다. <附錄 2>에서 變數 X_2 , 즉 충치예방을 위한 구매동기 문항에서 화이트치약에 대한 點數가 가장 높은 것은 화이트치약이 불소를 함유한 최초의 치약이라는 점과 일치하고 X_3 , 즉 치석제거를 위한 구매동기 문항에서 연마제를 다량 함유한 부광제약주식회사의 블렌닥스치약이 가장 點數가 높고, 일반 치아 위생용품인 비교적 저렴한 제품류에 속하는 럭키, 하이얀, 화이트치약등에 대해서도 消費者들의 정확한 가격의 高低에 대한 經濟性 認識을 變數 X_{11} 에서 볼 수 있는 등 소비자들의 제품에 대한 購買動機와 실제 제품의 용도별 명세가 거의 정확히 일치함을 보여준다.

2. 分析過程과 結果

系譜的技法의 Ward方法, average linkage方法, centroid方法 등 3개의 방법을 병행하여 12개의 變數로 표현된 130개의 개체를 SAS統計 프로그램중 集落分析편에서 'Cluster'절차를 사용하여 우선 몇개의 집락이 있는가, 즉 소비자들은 치약제품을 用度上 어떻게 地位를 나누어 인식하는가를 알기위한 시도를 해 보았다.

紙面관계상 생략한 Ward의 方法과 average linkage의 方法에 의한 CCC의 圖表에서, CCC의 값은 集落數 15까지 너무 상당한 陰數여서 예외적으로 극단적인 값의 개체들이 CCC의 檢定力을 약화시킨 경우처럼 해석이 된다. Centroid方法에 의한 CCC의 모든 값은 陰數이면서 계속적으로 감소하는 모습을 보여서 분포가 일양분포가 아니고 단봉분포임을 시사함으로 CCC에 의한 檢定의 假定에 어긋난다. 이같은 결과는 資料가 불필요한 變數를 많이 포함한 것이 원인일 수도 있다. 集落分析은 변수의 수가 너무 많은 資料에는 적합하지 않으므로 判別分析(discriminant analysis)으로 集落의 形成에 어떤 變

數들이 가장 큰 作用을 할 것인지 분석을 시도하여 보았다.

判別分析은 이미 각 個體가 각 集團에 關한 情報를 갖고 있어 이를 기초로 하여 判別函數를 설정하고 判別이 안된 個體들을 정해진 集團에 分類 혹은 그 소속을 예측하는데 목적이 있으므로 集團에 대해 전혀 事前情報가 없는 個體에 대하여 그 分類體系를 찾으려는데 근본목적이 있는 集落分析과는 다르다. 그러나 集落을 형성하는데 중요한 變數들을 찾아내는 非公式的 節次로서 判別分析을 이용할 수 있다.⁶⁾ SAS의 'Candisc' 節次를 이용하여 資料에 正準判別分析(canonical discriminant analysis)과 'Stepdisc' 節次를 이용하여 段階的 判別分析을 함으로써 <表 2>와 같은 結果를 얻었다.

< 表 2 >

正準判別分析의 結果

1. 유의적 判別함수를 위한 검정

判別函數	正準相關係數	F statistic	自 由 度	確 率 > F
1	.671	1.673	72, 615	.001
2	.408	.809	55, 527	.834

2. 標準 正準係數

判別函數	變數 X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
1	.113	-.317	.351	.400	.195	-.105
2	.344	.159	-.600	1.000	-.168	-.103

判別函數	變數 X ₇	X ₈	X ₉	X ₁₀	X ₁₁	X ₁₂
1	.041	.064	.163	.173	-1.081	.016
2	-.432	.344	-.000	.197	.191	.066

6) Everitt (1979), p. 171.

3. 製品別 判別點數의 平均

判別函數	B	C	H	L	M	P	W
1	.781	-.503	.012	-1.760	.551	.350	-.250
2	-.420	-.074	-.608	-.096	.166	.843	.180

* B는 블렌다스, C는 클로즈업, H는 하이얀, L은 럭키
M은 메디안, P는 페리오, 그리고 W는 화이트치약을 지칭

4. 段階的 判別分析에 의한 判別變數

단 계	導入變數	Partial R ²	F	Prob > F	Wilk's λ
1	X ₁₁	.307	9.087	.000	.693
2	X ₁	.172	4.236	.001	.573
3	X ₂	.108	2.447	.029	.511

< 表 2 - 1 >은 각각의 判別函數에 의한 그룹을 나타내는 變數集團과 判別變數集團에 각각의 線型結合된 變量간 최대의 單純相關關係를 나타내는 正準相關係數, 多變量資料에서 그룹간의 차이를 규명하는 통계량인 Wilks' Lambda로부터 변환시킨 F分布의 값과 自田度, 그리고 母集團에 集團間的 차이가 없다는 歸無假說하에서의 자료에 의해 계산된 F값을 F分布에 따르는 통계량이 초과할 확률을 나타낸다. 따라서 判別函數 1에 의해 나누어진 집단들의 차이가 없다는 귀무가설은 오차수준 .01에서 기각되며 최대 6개의 판별함수중의 미있는 판별함수는 둘로 결정된다.

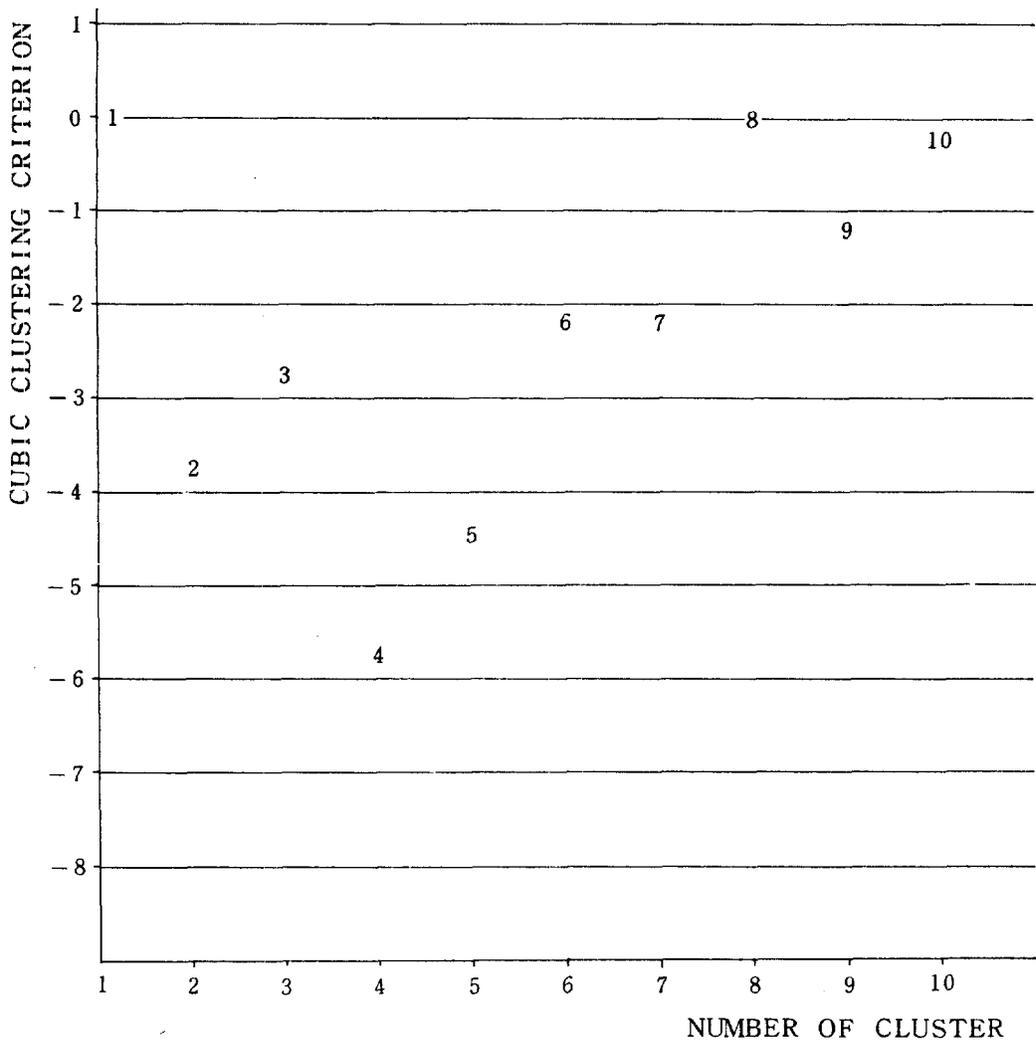
판별변수가 다른 변수에 비하여 판별점수에 미치는 상대적 공헌도를 말해주는 표준정준계수표의 표 <表 2-2>에서 판별함수 1의 표준정준계수를 보면 藥用性 구매동기 변수인 X₁, X₂의 공헌도가 크고 經濟性 變數 X₁₁은 역의 공헌도가 큼을 알 수 있다. 판별함수 1은 藥用性 구매동기와 經濟性이 일치한 판별기준, 다시 말하면 특수 약용성 제품은 일반 제품보다는 값비싼 제품이라는 판단으로 제품들을 분류하게 된다. 판별함수 2는 아직 명확하게 설명이 되지 않는다. < 表

2-3 >에서 藥用이며 값비싼 치약으로 블렌닥스, 메디안, 페리오치약이 꼽히고 저렴하며 약용치약이 아닌 일반 치아위생용 치약으로는 가장 전통있고 대중적인 렉키치약임을 알 수 있다.

<表 2-4 >는 'Stepdisc' 節次에서 선택 (Option)으로 'Step wise'를 골라 개별적으로 判別力이 크고 서로 情報를 적게 공유하는 3개의 변수를 보여준다. 첫 단계에서 R^2 가 .307이고 F값이 9.087인 변수 X_{11} 이 선택되고 이때 多變量資料에서, 특히 多變量分散分析 (MANOVA)에서 그룹내의 응집성과 그룹간의 차이를 동시에 고려하는 통계량인 Wilks' Lambda는 .693이다. 둘째 단계로 변수 X_{11} 를 共變量으로 하는 共分散分析의 結果로서 Partial R^2 가 .172이고 확률수준 .001에 대응하는 F값 4.236을 갖는 변수 X_4 가 모형에 들어오고 이와 관련된 λ 는 4.236이다. 다음 변수 X_3 가 도입되고 4번째 단계에서는 총치예방의 변수 X_2 의 F값 1.316이 'Stepdisc' 節次에서 사용되는 오차수준 .15를 훨씬 초과한 .255에 머물러서 들어오지 못한다. 따라서 X_{11} , X_4 , X_3 의 세개의 변수만이 선택된다.

이제 이 세 변수만을 가지고 'Cluster'의 절차를 이용하여 130개의 個體를 다시 세가지 系譜的技法으로 분석하여 CCC의 도표를 얻었다. Ward方法에 의

한 CCC의 도표는 單調增加의 양태를 보여 자료가 오직 홀자리 숫자만으로 작성된 이유때문인가 하는 추측이 가계 하고, centroid에 의한 CCC의 값은 모두 陰數의 값을 갖는 상태에 머물러 單峰分布에 근사한 분포임을 보인다. <그림 3>에서 볼 수 있듯이 average linkage에 의한 CCC도표에서는 集落의 數가 3인 경우와 8인 경우에 첨봉을 이룬다. 集落의 數가 3인 경우에는 CCC의 값이 陰數가 되고 8인 경우에는 CCC의 값이 0에 머무므로 8개의 集落이 존재한다는 결론을 얻게 된다.



<그림 3> Average linkage 方法에 의한 cubic clustering criterion 도표

다음으로 最適分離技法을 위한 'Fastclus' 節次를 이용하여 8개의 相互排斥的인 集落을 얻었다. 集落의 seed는 5번 반복해서 재계산 끝에 수렴되고 결과는 <表 3>과 같다.

<表 3>

1. 要 約

집 락	개 체 수	표본들간의 평균평방근	집락중심에서의 최대거리
1	12	.481	1.322
2	17	.399	1.018
3	14	.516	1.209
4	19	.441	1.525
5	27	.388	.958
6	21	.388	.990
7	13	.456	1.447
8	7	.584	1.337

2. 집락별 도수분포표

집 락	B	C	H	L	M	P	W	합 계
1	21 (6)	13 (2)	27 (3)	0	0	5 (1)	0	12
2	3 (1)	6 (1)	9 (1)	52 (12)	0	0	11 (2)	17
3	3 (1)	6 (1)	9 (1)	22 (5)	8 (1)	15 (3)	11 (2)	14
4	10 (3)	31 (5)	27 (3)	4 (1)	23 (3)	20 (4)	0	19
5	14 (4)	13 (2)	27 (3)	22 (5)	31 (4)	10 (2)	39 (7)	27
6	34 (10)	19 (3)	0	0	23 (3)	10 (2)	17 (3)	21
7	10 (3)	6 (1)	0	0	8 (1)	25 (5)	17 (3)	13
8	3 (1)	6 (1)	0	0	8 (1)	15 (3)	6 (1)	7
합 계	100 (29)	100(16)	100(11)	100(23)	100(13)	100 (20)	100(18)	(130)

* B는 블렌닥스, C는 클로즈업, H는 하이안, L은 럭키, M은 메디안

P는 페리오, 그리고 W는 화이트치약을 지칭

* 숫자는 제품의 집락별 백분율(괄호안의 숫자는 실제 빈도수

〈表3-1〉은 형성된 集落에 대한 구조를 요약하여 주고 〈表3-2〉는 집락별 빈도수를 나타내 준다. 실제로 제품은 7 가지이나 8 개의 集落으로 나타난 것은 標本의 소비자들은 제품에 대한 인식상 8 가지로 파악한다는 사실을 의미하는 것으로서 꼭 7 개의 集落이 되어야 할 이유는 없다.

〈表3-2〉의 집락별 度數分布表에서 하이얀과 럭키치약이 함께 값이 저렴하고 일반용인 제품의 特性을 갖는 것과 메디안과 페리오치약이 비슷한 價格에 들어 모두 치주염 예방의 特性을 살린 치약이라는 점을, 빈도수의 분포양태를 볼 때 소비자는 정확히 인식하고 있다는 것을 알 수 있다. 集落 1, 2와 3은 일반 치아위생용 치약에 관한 認識 集落으로, 集落 4, 5, 6과 7은 약용치약의 집락으로 구분하여 볼 수 있으나 集落 8에 대해서는 명확한 특성이 짙이지 않는다.

일반용 치약 集落群중에서 集落2는 저렴한 經濟性, 集落3은 전통적 一般性, 集落3은 개량된 一般性의 집락으로 해석하고 藥用齒藥 集落群에서 集落5는 치주염과 충치예방의 가장 대표적인 藥用集落, 集落6은 치석과 불순물제거용 치약의 集落, 集落7은 치주염예방 集落, 集落4는 일반적 藥用性의 집락으로 해석하여 본다. 〈附錄 1〉의 치약제품별 명세와 집락별 도수분포를 비교하여 보면 무리없는 집락에 대한 命名임을 알 수 있다. 만약 치약제품의 분류에 관한 일체의 정보가 없다고 가정하더라도 集落分析技法의 결과로 얻은 8개의 集落이 있다는 정보와 도수분포의 양태에 의한 하이얀과 럭키치약이 그리고 메디안과 페리오치약이 같은 用途로 인식된다는 정보는 무척 유용한 것이며 이 정보를 기점으로 더 많은 정보수집과 자료분석이 진행될 수 있다.

3. 要因分析 (Factor Analysis)

要因分析이란 多變量 資料의 여러 변수들간에는 相關關係를 생성시키는 숨은 要因들이 있어, 서로 독립이면서 어떤 개념상의 의미를 갖는 그러한 공통요인을 찾아 분석하는 통계적 방법이다.

集落分析에 의해 분류된 각 집락에 대해 要因分析을 다시 함으로써 더 세부적인 정보를 얻을 수 있고 變數가 지나치게 많은 자료는 집락분석을 직접 하기전에 요인분석을 먼저 실시하여 얻은 각 個體에 대한 요인점수들 (factor scores)을 개체에 대한 원래의 수치대신 사용할 수도 있다. 이처럼 要因分析은 집락분석을 하는 과정에서 크게 도움을 줄 수 있으나 본 연구의 작은資料의 量과 크지않는 범위를 참작하여 集落分析의 일환이라기 보다는 본 자료를

< 表 4 >

要因分析의 結果

1. 要因別 固有值

요 인	1	2	3	4	5	6
고 유 치	3.093	1.987	1.331	.694	.365	.155
비 율	.519	.334	.223	.117	.061	.026

요 인	7	8	9	10	11	12
고 유 치	.011	-.153	-.281	-.305	-.425	-.516
비 율	.002	-.026	-.047	-.051	-.071	-.087

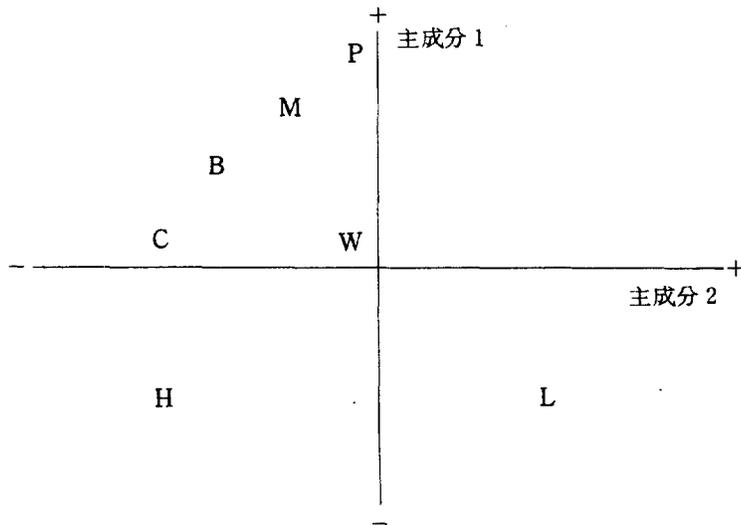
2. 要因負荷

	요 인 1	요 인 2	요 인 3
X ₁	.97 *	-.01	-.02
X ₅	.54 *	.28	-.13
X ₂	.47 *	-.03	.21
X ₃	.46 *	.24	.26
X ₉	.02	.85 *	.26
X ₁₂	.10	.49 *	-.10 *
X ₁₀	-.07	.46 *	.37 *
X ₈	.12	.37 *	.32
X ₁₁	-.07	.33 *	-.26 *
X ₇	.13	-.08	.61 *
X ₆	.15	.01	.45
X ₁	.16	-.15	.27

가지고 독립적인 요인분석을 하여 집락분석의 결과에 대한 추가적인 정보를 얻으려고 실시하여 보았다.

초기 共通要因을 추출하기 위하여 SAS의 'Factor' 절차를 사용하였다. 最尤法(maximum likelihood method)을 따라, 즉 요인뿐 아니라 모든 변수들이 정규분포를 따른다고 가정하고 표본에서 계산된 상관행렬을 갖는 尤度(likelihood)가 최대가 되도록 대응하는 母數, 즉 要因負荷(factor loading)를 추정했다. 초기 요인이 얻어진 후에도 좀더 구조적으로 단순하고 개념상 해석하기 쉬운 상태로 共通要因을 변경시키기 위해 각 변수의 共通性和 요인의 수를 고정시킨 상태에서 요인을 회전하였다. 直交回轉의 베리맥스(varimax)방법과 요인간의 相關性을 허용하는 프로맥스(promax)방법으로 회전하여 보았으나 해석상 도움이 되도록 크게 진전되지 않으므로 最尤法에 의한 초기의 결과만을 <表 4>에 요약했다.

要因分析에 더하여 SAS의 'Princomp' 절차에 의한 主成分分析을 하였을 때 要因分析과 크게 틀리지 않는 결과를 얻게 되므로 主成分 1과 主成分 2를 축으로 한 도표상에 분포된 개체들의 제품별 위치를 <그림 4>에 축약하여 보았다.



<그림 4> 주성분 1과 주성분 2를 축으로 한 제품별 위치

* B는 블렌더스, C는 클로즈업, H는 하이안
L은 럭 키, M은 메 디 안, P는 페리오
그리고 W는 화이트 치약을 지칭.

〈表 4 - 1〉에서 고유치가 1보다 큰 값에 대응하는 요인의 수를 확보하는 판정기준을 따라 3개의 의미있는 요인을 택한다. 〈表 4 - 2〉에서 다시 요인별 의미는 각 요인에 대한 변수들의 負荷를 관찰하여 찾는다.

要因 1에서 큰 負荷를 갖는 변수 X_4 , X_5 , X_2 와 X_3 는 각각, 잇몸질환, 시린이예방, 충치예방 그리고 플라그제거의 동기를 묻는 변수들이므로 藥用 動機의 요인이라 볼 수 있고, 변수 X_9 , X_{12} , X_8 와 X_{11} 는 각각 소독용 소금끼, 예상되는 외국산의 고급품질, 제품의 디자인, 맛과 향기, 그리고 저렴한 값을 구하는 변수들이므로 전통적 관념의 일반 위생 동기의 요인이라 간주되며, 변수 X_7 , X_6 , X_{10} 과 X_8 은 각각 구취제거, 새하얀 치아, 제품의 디자인과, 맛과 향기의 동기가 강조되는 요인이므로 美容 혹은 화장동기로 해석된다. 표본조사에서 사용한 12가지 문항이 소비자의 중요구매동기를 망라했다고 가정하면 결국 치약 소비자들은 藥用, 衛生, 그리고 美容의 세가지 중요구매동기를 갖는다고 말할 수 있다.

〈그림 4〉에서 약용의 主成分 1과 전통적 위생의 主成分 2를 축으로 한 평면에 대략적으로 잡아 본 각 치약제품의 위치를 보면 럭키치약과 하이얀치약은 약용이 아닌 점에서 같으나 럭키치약은 전통적 위생의 동기가 강한 반면 하이얀치약은 약하다. 지면관계상 생략한 것이나 위생의 主成分 2와 美容의 主成分 3을 축으로 한 공간에서 하이얀치약은 미용의 동기가 오히려 강하다. 페리오치약과 메디안치약은 전통적 위생동기에 反하고 약용성이 높은 위치를 차지하고 있다. 變數의 측면에서 集落關係를 파악한 〈그림 4〉에 대한 해석은 관찰치의 집락화 측면에서 분석한 결과의 〈表 3 - 2〉의 해석과 거의 동일한 것을 알 수 있다.

IV. 結 語

集落分析技法의 중요 개념과 논리를 소개하고 소비자가 인식하는 각 치약제품의 製品地位를 기법을 통하여 분석하여 보았다. 집락분석의 연구뿐 아니라 모든 연구조사에 두루 적용되는 네가지 문제점을 본 연구 결과를 위하여 다시 검토하여 본다.

첫째, 자료처리에 관한 문제로 본연구에서는 유사성의 측도 대신 거리의 측도를 사용했는데 많은 과거의 연구실적에 의하면 어느 것을 사용하던지 결과

에 별로 영향을 주지 않는 것으로 알려져 있다. 극단의 값을 갖는 개체가 많을 경우에는 자료를 표준화하여 주어야 하나 본 연구의 자료는 표준화할 필요성이 크지 않았다.

둘째는 사용된 알고리즘(algorithm)에 관한 문제다. 가장 효율적이라고 추천되는 2段階 集落節次라는 것이 있는데 이것은 첫단계에서 계보적 기법의 Ward方法과 average linkage方法으로 임시 집락의 수와 각집락의 중심등 예비적 결과를 얻어 이것을 출발점으로 다시 집락의 수와 중심을 수정하는 反復分離方法(iterative partitioning method)을 통하여 집락분석을 하는 것이다. SAS는 반복분리 알고리즘이 없으나 본 논문에서 'Cluster' 절차와 'Fastclus' 절차를 이어서 사용한 점으로 일종의 2단계 분석이었다.

셋째로 변수의 선택에 관한 문제인데 본 연구에서는 'Stepdisc' 절차에 의해 적절한 세계의 변수를 택했으나 이 변수들이 개체를 언제나 최대로 분리시키는 변수는 아닌 것이고 엄격한 의미에서 소비자의 구매동기가 총망라 되었다고 확신할 수는 없다.

네째는 타당성의 문제이다. 전혀 다른 표본이나 규모를 확대한 새로운 표본과 분석 비교하여 봄으로서 결과에 대한 외부적 妥當性을 알아 볼 수 있다. 큰 표본을 사용할 경우는 본 연구에 사용된 7가지 이외의 다른 치약제품들이 분석에 포함될 수 있겠다.

이상의 관점에서 본 연구를 더욱 향상시킬 여지도 없지는 않으나 현단계에서는 集落의 數를 밝히고 집락별 도수분포에서 의미있는 해석을 찾을 수 있는 것으로 그친다. 集落分析은 정보가 충분하지 못한 자료를 분류하고 분석할 때 큰 도움을 줄 수 있고 그외에 要因分析, 判別分析, 分散分析과 연계하여 사용할 수 있는 기법이므로 이 기법의 개발과 진전에 각별한 주의를 줄 만한 것이다.

〈 附錄 1 〉 齒藥製品明細書

블렌다스치약 : 치아에 붙은 불순물을 닦아 내기 위한 연마제가 다량 포함된 제품, 그램당 10.16 원.

부광제약주식회사 제품

클로즈업치약 : 불소가 약간 함유된 제품으로 투명하게 보이는 치약, 그램당 6.27 원. 태평양화학주식회사 제품

- 하이안 치약 : 일반 위생용 치약, 그람당 3.25원.
태평양화학주식회사 제품
- 럭키치약 : 일반 위생용 치약, 그람당 2.53원.
럭키화학주식회사 제품
- 메디안치약 : 치주염등 치주질환 발생 억제 효과가 있는 알란토인(allantoin) 이 함유된 치약, 그람당 6.25원.
태평양화학주식회사 제품
- 페리오치약 : 치주염등 치주 질환 발생 억제 효과가 있는 알란토인이 함유된 치약, 그람당 5.87원.
럭키화학주식회사 제품.
- 화이트치약 : 충치예방을 강화하는 목적으로 국내 제품으로는 처음 불소를 함유한 치약, 그람당 3.00원 럭키화학주식회사 제품

* 그람당 제품가격은 1985년 6월말 강남구 압구정동 한양쇼핑센터에서의 소매가로 120g에서 150g짜리 제품에 기초함

< 附錄 2 > 標準化된 제품별 變수의 平均點數

	K1	K2	K3	K4	K5	K6
블렌더스	-0.0516	0.0039	0.0102	0.0995	0.0279	0.1034
클로즈업	0.0281	-0.0379	0.0265	0.0313	0.0373	0.0715
하이안	-0.1566	-0.0196	-0.0294	-0.0794	-0.0213	0.1009
럭키	-0.1097	-0.0179	-0.1099	-0.0796	-0.0221	-0.1159
메디안	-0.1539	-0.0754	0.0779	0.0298	-0.1406	0.1102
페리오	0.0259	0.0071	0.1029	0.0249	0.0467	-0.1049
화이트	0.1099	0.0917	-0.1116	0.0913	-0.0999	-0.0311
	K7	K8	K9	K10	K11	K12
블렌더스	0.1397	-0.0264	0.0483	0.1250	-0.0594	0.1234
클로즈업	0.0413	0.0295	-0.1396	-0.1223	-0.0259	-0.1077
하이안	0.0652	0.0425	-0.0213	0.1259	-0.0529	-0.0871
럭키	-0.0312	-0.0594	0.0217	-0.0961	0.0399	-0.1099
메디안	-0.1112	0.0694	0.0728	-0.0793	-0.0002	0.0412
페리오	-0.0363	0.0253	0.0876	0.0205	0.0917	0.1005
화이트	-0.0217	-0.0979	-0.0956	-0.0443	0.0641	-0.0259

參 考 文 獻

- 株式會社럭키(1983), 한국인의 구강보건실태와 구강위생의식에 관한 조사보고서, 서울, 주식회사 럭키
- Anderberg, M.R.(1983), *Cluster Analysis for Applications*, New York: Academic press.
- Arabie, P., Carroll, J.D., DeSarbo, W., and Wind, Jerry (1981)
"Overlapping Clustering: A New Method for Product Positioning," *Journal of Marketing Research*, Vol. XVII (August) 310-319.
- Arnold, S.J. (1979), "A Test for Clusters," *Journal of Marketing Research*, 16, 545-551.
- Binder, D.A.(1981), "Approximations to Bayesian Clustering Rules," *Biometrika*, 68, 275-285.
- Cormack, R.M. (1971), "A Review of Classification," *Journal of Royal Statistical Society A*, 134, 321-67.
- Everitt, B.S. (1974), *Cluster Analysis*, John wiley & Sons, New York
- Everitt, B.S. (1979), "Unresolved Problems in Cluster Analysis," *Biometrics*, 35, March, 169-181.
- Friedman, H.P. and Rubin, J. (1967), "On some invariant criteria for grouping data," *Journal of the American Statistical Association*, 62, 1159-1178.
- Gower J.C.(1971), "A general coefficient of similarity and some of its properties," *Biometrics*, 27, pp.857-872.
- Hartigan, J.A.(1978), "Asymptotic Distributions for Clustering Criteria," *Annals of Statistics*, 6, 117-131.
- Jardine N. and Sibson R. (1971), *Mathematical Taxonomy*, London, Wiley.
- Lee, K.L. (1979), "Multivariate Tests for Clusters," *Journal of the American Statistical Association*, 74, 708-714.

- McRae, D.J. (1971), "MICKA, a Fortran IV Iterative K-means Cluster Analysis Programs," *Behavioral Science*, 16. 423-424.
- Milligan, G.W. (1980), "An Examination of the Effect of Six Types of Error Perturbation on Fifteen Clustering Algorithms," *Psychometrika*, 45, 325-342.
- Mojena, R. And Wishart, D. (1980), "Stopping Rules for Ward's Clustering Method," *Compstat: Proceedings in Computational Statistics*, 4th Symposium, Wien: Physica-Verlag.
- Punj, Girish and Stewart David W.(1983), "Cluster Analysis in Marketing Research: Review and Suggestions for Application," *Journal of Marketing Research*, Vol.XX(May), 134-48.
- SAS Institute(1982), *SAS User's Guide: Statistics*, SAS Institute INC., Cary, North Carolina,
- Wolfe, J.H. (1978), "Comparative Cluster Analysis of patterns of Vacational Interest," *Multivariate Behavioral Research*, 13,33-44.

