

# PCA에 의한 島嶼分類에 관한 研究(I)

## A Study on the Classification of Islands by PCA (I)

李 康 雨\*

Kang-Woo Lee

### 目 次

I. 序 論	3. 相關分析	
1. 研究目的	4. 主成分分析	
2. 研究方法	5. 主成分의 合成得點에 의한 島嶼分類의 節次	
3. 研究範圍	II. 結 論	
II. 本 論	1. 分析資料의 構造	參考文獻
1. 分析資料의 構造	2. 基本集計結果의 檢討	Summary

### I. 序 論

#### 1. 研究의 目的

韓國의 島嶼分類에 관한 研究는 다른 分野와는 달리 그 歷史가 매우 日淺하다. 지금까지의 韓國의 島嶼分類에 관한 現況<sup>1)</sup>을 살펴보면 그 分類의 基準이 人口規模, 面積, 交通手段의 數量, 電化實狀, 醫療施設 등으로 모두 單一尺度에 의한 市·道別 分類를 하고 있는 實情이다. 이와 같은 分類方法은 各 島嶼가 지니고 있는 諸特性 중에서 特定한 어느 하나의 特性만을 考慮하고 있기 때문에 眞正한 意味에서의 分類라고는 말할 수 없을 것이다. 여기서 먼저 分類란 무엇이며 왜 이와같은 分類를 하는가에 대해 살펴보기로 하자. 松原<sup>2)</sup>에 의하면 “分類란 客體의 觀測值에 의해 그 客體가 屬해 있어야 할 여러 種類의 카테고리 (혹은 部分母集團) 중에서 어느 하나에 이 客體를 가능한 한 正確하게 歸屬시키는 것이다.”라고 말하고 덧붙여서 “科學的 思考의 出發點은 分類에 있고, 이를 特徵 지을 수 있는 因子(factor)를 發見하는데 있으므로 分類는 훌륭한 科學基礎論的인 性格을 갖고 있다.”라고 지적하고 있다. 따라서 이와 같은 分類에 관한 研究는 分類對象인 客體의 諸特性值를 考慮하여야만 하기 때문에 多變量解析法(multivariate analysis)<sup>3)</sup>이라고 하는 數學的 手法의 導入이

\*釜山水產大學, 助教授.

1) 內務部, 島嶼·落島現況, 號1號, 1981. pp.136~145.

2) 松原 望, 意思決定의 基礎, 朝倉書店, p.155.

3) 多變量解析法이란 相互相關을 갖는 여러 種類의 特性值의 資料가 갖는 特徵을 要約하여 綜合的인 評價를 하는 手法이다.

불가피하다. 이와 같은 多變量解析의 手法<sup>4)</sup>은 今世紀 初부터 F. Galton, K. Pearson, R. A. Fisher, P. C. Mahalanobis, C. R. Rao, G. Wiegleb<sup>5)</sup> 등에 의하여 生物學分野에, C. Spearman, L. L. Thurstone, H. Hotelling에 의하여 計量心理學의 領域에 導入되었으며, 最近에는 컴퓨터利用의 急速한 發展에 의해서 工程解析, 市場調査, 官能檢査, 計量診斷法, 財務分析<sup>6)</sup> 등의 分野에 광범위하게 適用되기에 이르렀다.

本 論文에서는 各 島嶼가 지니고 있는 多次元的 特性值間의 相關關係를 究明하여 諸特性值에 適當한 荷重(weight)을 두어 合成得點 또는 綜合特性值를 구하는 多變量解析法의 基本的인 手法인 主成分分析法(principle component analysis)에 의하여 島嶼를 分類하는 것을 目的으로 한다.

## 2. 研究方法

本 研究에서는 各 島嶼에 대한 12個의 特性值間의 相關行列(correlation matrix)을 算出하여 諸特性值間의 相關分析을 實施하여 諸特性值間의 一次的인 關係의 程度를 파악하고, 主成分分析에 의하여 各 島嶼에 대한 第1主成分의 得點과 第2主成分의 得點을 구한 다음 이를 利用하여 散布圖를 作成하였다. 이 散布圖를 바탕으로 相互 類似性을 갖고 있는 島嶼別로 分類를 實施하기로 한다.

여기서 島嶼分類를 實施하기 위하여 採用한 主成分分析의 概要<sup>7)</sup>를 살펴보면 다음과 같다.

$n$ 個의 個體(本 研究에서는 88個의 島嶼에 該當된다.)에 대해  $p$ 種類의 特性值 變量  $x_1, x_2, \dots, x_p$ 가 測定되어 <表 1>의  $p$ 個의 特性值와 같은 資料가 얻어졌을 때 次式으로 주어지는 綜合特性值<sup>8)</sup>를 抽出하는 手法을 主成分分析이라 한다. (단, 次式에서  $m \leq p$  이다.)

$$\left. \begin{aligned} z_1 &= l_{11}x_1 + l_{12}x_2 + \dots + l_{1p}x_p \\ z_2 &= l_{21}x_1 + l_{22}x_2 + \dots + l_{2p}x_p \\ &\dots\dots\dots \\ z_m &= l_{m1}x_1 + l_{m2}x_2 + \dots + l_{mp}x_p \end{aligned} \right\} \dots\dots\dots(1)$$

$$l_{k1}^2 + l_{k2}^2 + \dots + l_{kp}^2 = 1 \quad (k=1, 2, \dots, m) \quad \dots\dots\dots(2)$$

(1)式의 係數  $l_{ki}(i=1, 2, \dots, p)$ 는 主成分  $z_k(k=1, 2, \dots, m)$ 가 서로 無相關이란 條件下에서  $z_k$ 의 分散이 最大가 되도록 順次的으로 決定된다. 그 結果  $x_1, x_2, \dots, x_p$ 의 分散·共分散行列의 固有值<sup>9)</sup>(eigen value)를 크기가 큰 順으로

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_m \geq 0$$

4) 奥野忠一 外3人, 多變量解析法, 日科技連, p.6.  
 5) G. Wiegleb, Some applications of principal components analysis in vegetation; Ecological research of aquatic communities, Advances in vegetation science 2, 1980, pp.67~74.  
 6) 山田文道, 經營指標による企業評價のための新しい試み—多變量解析によるアプローチ, engineers, No. 262.  
 7) OR事典編集委員會編, OR事典, 日科技連, 1975, pp.167~168.  
 8) 이것을 主成分(principal component)이라 부른다.  
 9) 分散·共分散行列을  $V$ , 單位行列을  $I$ 라 하면  $|V-\lambda I|=0$ 의 行列方程式을  $\lambda$ 의 多項式으로 展開하면 다음과 같다.  
 $a_2 + a_1\lambda + a_2\lambda^2 + \dots + a_p\lambda^p = 0$   
 이를 만족하는  $\lambda$ 의 값은  $p$ 個가 됨을 알 수 있다. 또한 이  $p$ 個의 根은 모두 實數로서 非陰임이 알려져 있다. 이  $\{\lambda_k\}(k=1, 2, \dots, p)$ 를 行列  $V$ 의 固有值라고 한다.

이라 할 때 각각의 固有值에 對應하는 固有 Vector<sup>10)</sup>가 式(1)의 係數  $l_{1i}, l_{2i}, \dots, l_{mi}(i=1, 2, \dots, p)$ 가 된다.  $x_1, x_2, \dots, x_p$ 의 測定單位가 다를 경우는 各 變量을 標準化한 分散·共分散行列 즉 원개의 變量에 대한 相關行列  $R$ 로부터 固有值, 固有 Vector를 구한다. 어느 경우에도  $z_k$ 의 分散은  $v_k$ 와 一致하므로 第  $m$  主成分까지 취했을 때의 情報의 寄與率은  $\frac{\sum_{k=1}^m \lambda_k}{\sum_{k=1}^p \lambda_k}$ 로서 주어진다. 이때 分散은  $T_r(V)$ , 또는  $T_r(R)=p$ 가 된다. 이때  $m$ 의 작은 값에 대해 이 寄與率이 충분히 크면 원래의  $p$ 개의 變量이 갖는 모든 情報가  $m$ 개의 主成分에 의해 거의 要約된 것을 意味하므로 思考의 節約(parsimony of thinking)에 도움이 된다. 한편 主成分의 得點(score)  $\{z_{\alpha k}: \alpha=1, 2, \dots, n; k=1, 2, \dots, m\}$ 은 個體集團이 等質인 경우에는  $n$ 개의 個體의 細分化(segmentation)에 사용되며 個體集團이 異質인 경우는 異質性的 檢出에 사용된다. 따라서 本 論文에서는 島嶼라는 個體集團의 異質성을 檢出함으로써 島嶼의 分類를 實施하고자 한다. 결국 이와 같이 主成分分析法<sup>11)</sup>이란 주어진 1組의 變量群의 相關關係를 고려하여 많은 變量이 갖는 分散을 가능한 한 정확히 表現할 수 있는 少數의 合成得點을 구하는 方法이라 할 수 있다.

### 3. 研究範圍

本 論文에서 考慮하고 있는 分類個體인 島嶼는 慶尙南道內에 散在하고 있는 88개의 有人島嶼로 局限하고 各 島嶼에 關한 特性值는 島嶼카-드<sup>12)</sup>를 利用하였다. 한편 主成分分析에 사용한 各 島嶼의 特性值는 (1) 內陸과의 距離(km), (2) 總島嶼面積(ha), (3) 海岸線의 延長距離(km), (4) 道路의 길이(km), (5) 船着場의 延長길이(m), (6) 島嶼人口(名), (7) 戶當所得(千圓), (8) 住宅數棟, (9) 學生數(名), (10) TV台數(台), (11) 라디오台數(台), (12) 漁船隻數(隻)로서 모두 12개의 特性值를 利用하였다.

## II. 本 論

### 1. 分析資料의 構造

各 島嶼의 特性值에 대한 主成分分析을 實施하기 위한 資料構造는 <表 1>과 같다. 이 <表 1>에서 는 各 島嶼에 대한 特性值(變量)의 數로서  $p=12$ 이며,  $n$ 은 慶南道內에 位置한 有人島嶼의 數로서  $n=88$ 이다. 따라서 特性值  $x_{\alpha i}(\alpha=1, 2, \dots, 88; i=1, 2, \dots, 12)$ 는  $\alpha$ 라는 島嶼의  $i$ 번째의 特性值를 意味하게 된다. 여기서 各 特性值(變量)  $x_1, x_2, \dots, x_{12}$ 의 內容은 <表 2>에 表示되어 있다.

한편 <表 1>의 右上段에 表示되어 있는  $z_1, z_2, \dots, z_m$ 은 各 第1主成分, 第2主成分, ... , 第  $m$  主成分을 나타내며,  $z_{\alpha k}$ 는  $\alpha$ 라는 島嶼의 第  $k$  主成分의 綜合特性值(合成得

10) 固有 Vector는 記號  $L_k=(l_{k1}, l_{k2}, \dots, l_{kp})'$ 와 分散·共分散行列  $V$ 로 표시되는 行列方程式  $VL_k=\lambda_k L_k$ 의  $\lambda_k$ 에 固有值를 代入하고,  $\sum_{i=1}^p l_{ki}^2=1$ 의 조건하에서 求한 解  $L_k$ 를 固有值  $\lambda_k$ 에 對應하는 固有 Vector라 한다.

11) 高根芳雄, 柳井晴夫, 多變量解析法, 朝倉書店, 1978, p. 84.

12) 島嶼카-드란 內務部에서 定期的으로 實施하는 全國 有人島嶼의 特性值에 關한 諸資料가 記入된 카-드를 말한다.

수 산 경 영 른 집

<表 1> p 變量 資料와 그 要約

島 嶼 No.	p 個의 特性值				主 成 分	
	$x_1$	$x_2 \dots x_i \dots x_p$	$z_1$	$z_2 \dots z_m$		
1	$x_{11}$	$x_{12} \dots x_{1i} \dots x_{1p}$	$z_{11}$	$z_{12} \dots z_{1m}$		
2	$x_{21}$	$x_{22} \dots x_{2i} \dots x_{2p}$	$z_{21}$	$z_{22} \dots z_{2m}$		
⋮	⋮	⋮	⋮	⋮		
$\alpha$	$x_{\alpha 1}$	$x_{\alpha 2} \dots x_{\alpha i} \dots x_{\alpha p}$	$z_{\alpha 1}$	$z_{\alpha 2} \dots z_{\alpha m}$		
⋮	⋮	⋮	⋮	⋮		
n	$x_{n1}$	$x_{n2} \dots x_{ni} \dots x_{np}$	$z_{n1}$	$z_{n2} \dots z_{nm}$		
平 均	$\bar{x}_1$	$\bar{x}_2 \dots \bar{x}_i \dots \bar{x}_p$	0	0 ..... 0		
標 準 偏 差	$s_1$	$s_2 \dots s_i \dots s_p$	$\sqrt{\lambda_1}$	$\sqrt{\lambda_2} \dots \sqrt{\lambda_m}$		
相 關 係 數		$r_{12} \dots r_{2i} \dots r_{ip}$		$r=0$		

<表 2> 資 料 의 要 約

特 性 值	平 均	標 準 偏 差	C. V.
1. 內陸과의 距離 (km); $x_1$	8.02	10.32	1.29
2. 總 面 積 (ha); $x_2$	259.79	752.94	2.90
3. 海 岸 線 延 長 距 離 (km); $x_3$	6.41	7.96	1.24
4. 道 路 길 이 (km); $x_4$	1.44	7.35	5.10
5. 船 着 場 길 이 (km); $x_5$	85.75	142.14	1.66
6. 人 口 (名); $x_6$	565.94	1,118.98	1.98
7. 戶 當 所 得 (千 圓); $x_7$	2,223.45	426.17	0.19
8. 住 宅 數 (棟); $x_8$	123.60	257.74	2.09
9. 學 生 數 (名); $x_9$	147.34	340.67	2.31
10. T V 台 數 (台); $x_{10}$	82.72	165.20	2.00
11. 라 디 오 台 數 (台); $x_{11}$	107.23	230.97	2.15
12. 漁 船 數 (隻); $x_{12}$	31.92	47.19	1.48

註) C. V.는 變動係數(coefficient of variation) 즉 標準偏差/平均値를 意味한다.

點)를 나타낸다.

2. 基本集計結果의 檢討

<表 2>에  $p=12$ 個에 대한 特性值의 平均·標準偏差·變動係數(C. V.)가 要約되어 있다. <表 2>에서 標準偏差가 특히 큰 特性值는 人口(名), 總面積(ha)·戶當所得(千圓)·學生數(名)·住宅數(棟)이고 標準偏差가 작은 特性值는 道路길이(km)·內陸과의 距離(km)·海岸線의 延長距離(km)이다. 여기서 各 特性值의 單位가 相異하므로 特性值間의 標準偏差의 크기에 의한 相互比較는 無意味함에 有意해야 한다. 特性值別로 標準偏差가 큰 特性值를 考察하여 보면  $x_2, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}$ 로서

PCA에 의한 島嶼分類에 關한 研究

戶當所得을 除外한 나머지의 特性值들은 島嶼의 크기(size)와 關係가 있는 特性值들임을 알 수 있다. 한편으로 散布程度의 相對的인 크기를 表示해 주는 變動係數를 檢討하면 道路길이( $x_4$ )가 5.1%로서 壓倒的으로 크고, 總面積( $x_2$ )·學生數( $x_8$ )·라디오台數( $x_{11}$ )·住宅數( $x_9$ )·TV台數( $x_{10}$ )·人口( $x_6$ )는 1~3%로 나타났으며, 戶當所得( $x_7$ )은 0.19%로서 다른 特性值에 비해 매우 낮다. 여기서 變動係數가 가장 큰 道路길이와 變動係數가 가장 작은 戶當所得에 대해 檢討를 해 보면 다음과 같다. 먼저 變動係數가 가장 큰 道路길이의 標準偏差는 여러 特性值 중에서 가장 작은데도 불구하고 變動係數가 壓倒的으로 큰 原因은 觀測個體가 島嶼라는 特殊性 때문이다. 즉 各 島嶼의 基本資料에 의해 島嶼의 道路網의 有無를 살펴보면, 相對的으로 島嶼面積이 큰 37個의 島嶼에만 道路網이 있고 島嶼面積이 작은 51個의 島嶼는 道路網이 없는 것으로 나타났다. 이와 같은 結果로서 道路길이의 變動係數의 分母인 平均值가 매우 작은 값을 취하게 되므로 결국 變動係數는 相對的으로 커다란 값을 갖게 된다. 다음에 戶當所得의 變動係數가 가장 작다는 것은 대체로 島嶼의 다른 여러 特性值에 비해 平均戶當所得을 중심으로 散布程度가 相對的으로 작다는 것을 意味한다. 따라서 이는 各 島嶼의 戶當所得이 다른 特性值에 비해 대체로 거의 同一水準에 가깝다고 말 할 수 있을 것이다.

3. 相關分析(Correlation Analysis)

本 論文에서 선택한 特性值는 모두 確率變數(random variable)이고 實驗者에 의하여 調節不可能하며 一般的으로 어느 特性值가 獨立變數인지 從屬變數인지 明確한 區別이 어렵다. 이와 같은 경우에 諸變數間의 線型關係의 程度를 分析하는 技法이 相關分析이다.

여기서는 本 論文에서 選定한 12個의 特性值間의 相關行列을 나타낸 <表 3>을 利用하여 諸特性值間의 相關關係를 살펴보기로 하자.

相 關 行 列

<表 3>

特性值	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$	$x_{11}$	$x_{12}$
$x_1$	1.00	0.07	0.12	-0.02	-0.11	0.11	-0.20	0.08	0.12	0.10	0.14	-0.03
$x_2$	0.07	1.00	0.50	0.26	0.35	0.50	0.07	0.48	0.46	0.47	0.48	0.47
$x_3$	0.12	0.50	1.00	0.51	0.67	0.87	0.05	0.93	0.80	0.88	0.85	0.79
$x_4$	-0.02	0.26	0.51	1.00	0.53	0.53	0.05	0.50	0.63	0.61	0.53	0.42
$x_5$	-0.11	0.35	0.67	0.53	1.00	0.67	-0.01	0.69	0.66	0.69	0.63	0.68
$x_6$	0.11	0.50	0.87	0.53	0.67	1.00	0.05	0.97	0.97	0.99	0.98	0.85
$x_7$	-0.20	0.07	0.05	0.05	-0.01	0.05	1.00	0.07	0.07	0.08	0.04	0.16
$x_8$	0.08	0.48	0.93	0.50	0.69	0.97	0.07	1.00	0.92	0.97	0.96	0.87
$x_9$	0.12	0.46	0.80	0.63	0.66	0.97	0.07	0.92	1.00	0.97	0.96	0.81
$x_{10}$	0.10	0.47	0.88	0.61	0.69	0.99	0.08	0.97	0.97	1.00	0.98	0.84
$x_{11}$	0.14	0.48	0.85	0.53	0.63	0.98	0.04	0.96	0.96	0.98	1.00	0.83
$x_{12}$	0.03	0.47	0.79	0.42	0.68	0.85	0.16	0.87	0.81	0.84	0.83	1.00

1) 内陸과의 距離  $x_1$ 과 다른 諸特性值間의 相關係數는 0.11~0.20으로서 無相關에 가깝게 나타나 있는데 이는 觀測對象이 島嶼라는 점을 考慮하면 당연한 結果라고 말할 수 있다.

2) 島嶼의 總面積과 戶當所得 및 内陸과의 距離의 相關은 거의 無相關이며 다른 特性值와의 相關은 0.26~0.50 사이에 있다.

3) 海岸線 延長距離는 人口·住宅數·學生數·TV台數·라디오台數·漁船隻數와의 相關이 0.8 이상으로 매우 높은 正相關을 가지고 있다.

4) 船着場 길이는 内陸과의 距離 및 戶當所得과는 거의 無相關이며 다른 諸特性值間의 相關은 다소 認定된다.

5) 人口와 戶當所得과는 거의 無相關에 가깝지만, 住宅數·學生數·TV台數·라디오台數·漁船隻數와는 0.84 이상의 높은 正相關을 보여주고 있다.

6) 戶當所得과 모든 特性值間의 相關은 거의 無相關에 가깝다.

7) 住宅數·學生數·TV台數·라디오台數·漁船隻數間의 相關은 0.83 이상으로 매우 높은 正相關을 나타내고 있다. 이것은 諸特性值間의 性格을 考慮하면 豫測 가능한 結果라 할 수 있다.

이상에서 相關도가 매우 높거나 낮은 몇 개의 特性值에 局限하여 相關係數를 考慮하여 보았으나 여기서 考慮될 수 있는 두 變量間의 相關係數의 個數는 모두  ${}_{12}C_2=66$ 個나 되므로 이를 모두 個別的으로 考慮하여 綜合한다는 것은 용이하지 않다. 원래 相關係數는 2個의 特性值(變量)間의 線型關係의 程度를 測定하는데 目的이 있으므로 本 論文에서는 편의상 위에서 같이 相關도가 높거나 낮은 몇 개의 特性值에 局限하여 考察하였다. 따라서 이와 같은 相關分析의 限界를 克服하기 위해 다음에 諸特性值間의 相關關係를 考慮하여 綜合特性值를 구하는 主成分分析<sup>13)</sup>을 實施하기로 한다.

#### 4. 主成分分析(Principal Component Analysis)

本 論文에서 選定한 12個의 特性值間의 相關關係를 調査하기 위한 特性值의 相關行列로부터 主成分分析을 實施하여 <表 4>에 固有值(eigen value)와 累積寄與率, <表 5>에 固有벡터(eigen vector) <表 6>에 因子負荷量<sup>14)</sup> (factor loading)을 表示하였다.

<表 4>로 부터 固有值  $\lambda \geq 1$ 을 만족하는 것은 第2 主成分까지로 約 74%의 情報가 第1 主成分과 第2 主成分에 의해 集約된다. 따라서 12次元의 空間을 2次元으로 次元을 減少시킴으로써 發生하는 情報의 損失은 約 26%라고 말할 수 있다. 만일 第8主成分까지 고려한다면 <表 4>로 부터 約 99.46%의 情報가 說明되기 때문에 거의 情報의 損失이 없음을 알 수 있다.

本 論文에서는 有人島嶼의 12個의 特性值를 使用하여 島嶼分類를 實施하기 위하여 散布圖를 作成하여야 하므로  $\lambda \geq 1$ 인 第1主成分과 第2主成分만 考慮하기로 한다. 물론 이때 26%의 情報의 損失이 發生된다는 事實에 有意하지 않으면 안된다.

<表 5>에는 第1主成分에서 第8主成分까지의 各 特性值  $x_i'$  ( $i=1, 2, \dots, 12$ )에 대한 係數가 表示

13) 主成分分析을 위한 Computer program은 다음의 參考文獻을 活用하였다.  
淺野長一郎, 因子分析法通論, 共立出版株式會社, 1972, pp. 36~38.

14) 因子負荷量은 主成分  $z_k$ 와 特性值  $x_i'$ 와의 相關關係를 말한다. 단 여기서  $x_i'$ 는  $x_i$ 를 標準化한 값을 意味한다.

PCA에 의한 島嶼分類에 關한 研究

〈表 4〉 固有値와 累積寄與率

主成分 No.	固有値 ( $\lambda_k$ )	累積寄與率 (%)
1	7.6072	63.39
2	1.2389	73.72
3	0.9088	81.29
4	0.7283	87.36
5	0.6491	92.77
6	0.4074	96.16
7	0.2150	97.96
8	0.1806	99.46

註) 固有値  $\lambda_k$ 는 主成分  $z_k$ 의 分散을 意味하며, 主成分  $z_k$ 의 寄與率은 總分散  $p$ 에 對한  $z_k$ 의 分散( $\lambda_k$ )의 비율 즉  $\lambda_k/p$ 이다.

〈表 5〉 固有 vector

主成分 特性值	$z_1$	$z_2$	$z_3$	$z_4$	$z_5$	$z_6$	$z_7$	$z_8$
$x_1'$	0.030	0.737	0.447	0.317	0.156	0.340	0.068	0.088
$x_2'$	0.200	0.025	0.330	-0.754	0.525	-0.057	0.078	-0.036
$x_3'$	0.330	0.045	0.039	-0.054	-0.070	0.141	-0.830	-0.065
$x_4'$	0.225	-0.093	-0.345	0.406	0.739	-0.138	-0.119	0.220
$x_5'$	0.273	-0.129	-0.359	-0.089	0.037	0.806	0.247	-0.236
$x_6'$	0.354	0.052	0.040	0.027	-0.138	-0.163	0.125	-0.172
$x_7'$	0.029	-0.634	0.658	0.334	0.094	0.175	0.014	-0.112
$x_8'$	0.353	0.019	0.031	-0.008	-0.198	-0.048	-0.198	-0.089
$x_9'$	0.348	0.038	-0.003	0.141	0.0	-0.223	0.352	-0.125
$x_{10}'$	0.357	0.025	0.006	0.106	-0.066	-0.130	0.060	-0.161
$x_{11}'$	0.350	0.077	0.050	0.056	-0.133	-0.228	0.165	-0.128
$x_{12}'$	0.321	-0.128	0.064	-0.098	-0.240	0.121	0.121	0.880

註)  $x_i'$ 는  $x_i$ 를 標準化한 값임.

되어 있다. 〈表 5〉의 固有벡터를 利用하면 第1主成分  $z_1$  (綜合特性值)과 第2主成分  $z_2$  (綜合特性值)은 次式에 의해 얻어진다.

$$z_1 = 0.030 x_1' + 0.200 x_2' + 0.330 x_3' + 0.225 x_4' + 0.273 x_5' + 0.354 x_6' + 0.029 x_7' + 0.353 x_8' + 0.348 x_9' + 0.357 x_{10}' + 0.350 x_{11}' + 0.321 x_{12}' \dots \textcircled{3}$$

$$z_2 = 0.737 x_1' + 0.025 x_2' + 0.045 x_3' - 0.093 x_4' - 0.129 x_5' + 0.052 x_6' - 0.634 x_7' + 0.019 x_8' + 0.038 x_9' + 0.025 x_{10}' + 0.077 x_{11}' - 0.128 x_{12}' \dots \textcircled{4}$$

$$\text{단 } x_i' = (x_i - \bar{x}_i) / s_i$$

$x_i$ ; 島嶼의  $i$ 번째의 特性值( $i=1, 2, \dots, 12$ )

수 산 경 영 론 집

$\bar{x}_i$ ;  $x_i$ 의 平均值( $i=1, 2, \dots, 12$ )

$s_i$ ;  $x_i$ 의 標準偏差( $i=1, 2, \dots, 12$ )

이제 ③式과 ④式에 各 島嶼의 特性值  $x_i'$ 를 代入하면 各 島嶼에 對해 2個의 綜合特性值인  $z_1$ 과  $z_2$ 를 算出할 수 있다.

〈表 6〉 因 子 荷 負 量

主成分 特性值	$z_1$	$z_2$	寄 與 率 (%)
$x_1'$	0.082	0.820	67.93
$x_2'$	0.551	0.028	30.39
$x_3'$	0.911	0.050	83.26
$x_4'$	0.621	-0.103	39.68
$x_5'$	0.754	-0.143	58.83
$x_6'$	0.978	0.058	95.91
$x_7'$	0.081	-0.706	50.47
$x_8'$	0.973	0.021	94.73
$x_9'$	0.959	0.042	92.17
$x_{10}'$	0.984	0.028	96.96
$x_{11}'$	0.966	0.085	94.04
$x_{12}'$	0.885	-0.142	80.25

註) 여기서의 寄與率은 因子負荷量의 自乘合으로서 計算되며 이는 第1主成分  $z_1$ 과 第2主成分  $z_2$ 에 의해서 特性值  $x_i'$ 가 설명되어지는 程度를 나타낸다.

다음에 主成分分析에 의해 算出한 固有值·固有벡터·〈表 6〉의 因子負荷量을 利用하여 第1主成分과 第2主成分의 意味를 考察하면 다음과 같다.

1) 主成分의 意義

i) 第1主成分  $z_1$ ;

第1主成分  $z_1$ 의 係數(固有벡터)는 〈表 5〉에서 보는 바와 같이 모두 陽이며 그 크기는 特性值 “內陸과의 距離 및 戶當所得”을 除外하면 대체로 0.2~0.35의 範圍에 있다. 여기서 각 特性值의 값이 모두 陽이므로 각 特性值의 값이 增減됨에 따라서 第1主成分  $z_1$ 의 得點이 增減됨을 알 수 있으므로 第1主成分은 크기(size)를 나타내는 因子(factor)라고 말할 수 있다. 단일 88個의 島嶼를  $z_1$ 의 得點  $z_{\alpha 1}(\alpha=1, 2, \dots, 88)$ 의 크기順으로 配列하면 이는 島嶼의 綜合的인 크기의 順으로 配列될 것이다.

한편 〈表 6〉으로 부터 第1主成分  $z_1$ 과 각 特性值間의 相關關係를 表示한 因子負荷量에 의하면, 第1主成分  $z_1$ 은 海岸線 延長距離( $x_3'$ )·人口( $x_6'$ )·住宅數( $x_8'$ )·學生數( $x_9'$ )·TV 台數( $x_{11}'$ )·漁船隻數( $x_{12}'$ )와는 높은 正相關을 갖고 있다.

따라서 第1主成分  $z_1$ 은 이들 特性值의 性格을 考慮해 볼 때 島嶼의 綜合的인 크기를 나타낸다고



PCA에 의한 島嶼分類에 관한 研究

할 수 있다. 그러나 第1主成分과 内陸과의 距離( $x_1$ ) 및 戶當所得과는 거의 無相關에 가깝기 때문에 泉型關係의 程度가 없음을 알 수 있다.

ii) 第2主成分  $z_2$ ;

第2主成分의 경우 特性值  $x_i'$  ( $i=1, 2, \dots, 12$ )의 係數의 符號는 <表 5>로부터 알 수 있는 바와 같이 +와 -가 混在되어 있으며, 内陸과의 距離 $x_1'$  ( $x_1'$ 의 第2主成分의 係數는 0.737임)와 戶當所得  $x_7'$  ( $x_7'$ 의 第2主成分의 係數는 -0.634임)를 除外하고는 거의 대부분의 係數가 0에 가깝다. 또한 <表 5>으로부터 第2主成分과 각 特性值  $x_i'$ 間의 相關係數를 살펴보면, 第2主成分  $z_2$ 와 内陸과의 距離 ( $x_1'$ ) 및 戶當所得 ( $x_7'$ )의 相關係數는 각각 0.820, -0.706이고 그 밖에 다른 모든 特性值과 第2主成分과의 相關關係는 없는것을 거의 알 수 있다. 이상으로 보아 第2主成分의 合成得點(綜合特性值)은 内陸으로부터 멀리 멀어지고 戶當所得이 相對적으로 작은 島嶼일수록 그 값이 커지고, 반대로 内陸과 인접한 島嶼로서 戶當所得이 높은 島嶼는 第2主成分의 合成得點인 綜合特性值가 작아질 것이다. 그러므로 第2主成分은 内陸과의 距離 및 戶當所得에 의해 그 값이 거의 決定된다고 말할 수 있다. 이상으로 보아 第2主成分  $z_2$ 는 島嶼의 位置에 따른 戶當所得의 分布形態(shape)를 나타내는 因子라고 볼 수 있다.

2) 特性值의 分類

여기서는 本 論文에서 選定한 12個의 特性值의 變動이 第1主成分  $z_1$  과 第2主成分  $z_2$ 로서 說明되어지는 程度에 따라서 이들 特性值를 分類하는 것이 目的이다. <그림 1>은 半徑이 1인 圓으로 <表 6>에 表示되어 있는 각 特性值  $x_i'$  ( $i=1, 2, \dots, 12$ )의 第1主成分  $z_1$ 과 第2主成分  $z_2$ 에 대한 因子負荷量  $r(z_k, x_i')$ 을  $z_1$ 과  $z_2$ 軸에 대해 表示한 것이다.

<그림 1>에 의하면 島嶼의 綜合的 크기와 島嶼의 戶當所得과 位置에 대한 分布形態를 나타내는 特性值는 다음과 같이 分類된다.

島嶼의 綜合的 크기;  $x_3', x_6', x_8', x_9', x_{10}', x_{11}', x_{12}'$

島嶼의 分布形態;  $x_1', x_7'$

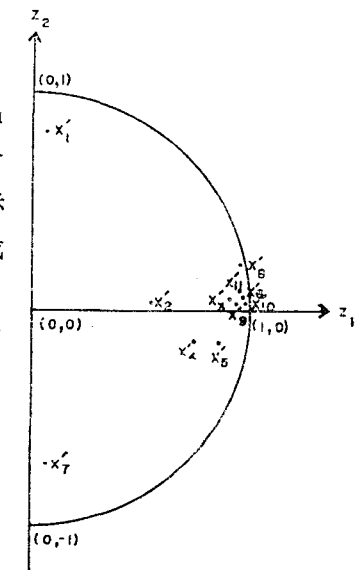
여기서 特性值  $x_i'$ 에 대한  $z_1$  및  $z_2$ 의 寄與率 $C_i$ <sup>15)</sup>는 次式

$$C_i = \sum_{k=1}^2 r^2(z_k, x_i'); \quad i=1, 2, \dots, 12$$

으로 表示되며 또한 다음과 같은 關係式

$$\sum_{k=1}^p r^2(z_k, x_i') = 1; \quad i=1, 2, \dots, p^{16)}$$

이 成立하므로 <그림 1>에서의 各 特性值의 位置는 항상 半徑이 1인 圓內에 있게 된다. 따라서 <그림 1>에서 第1主成分  $z_1$ 과 第2主成分  $z_2$ 만으로 그 變動이 갖는 情報가 거의 說明되는 特性值는 圓周에 가까운 位置에 占하게 될 것이며, 반대로  $z_1$ 과  $z_2$ 로서 그 變動이 갖는 情報가 說明되어 지지



<그림 1>  $z_1$ 과  $z_2$ 에 대한 因子負荷量

15) 각 特性值에 대한 第1·第2 主成分의 寄與率은 <表-6>을 參照.

16) 여기서  $p$ 는 特性值(變量)의 個數를 意味한다.

않는 特性値는 圓의 中心에 가까운 位置에 있게 될 것이다. 이제 <그림 1>로부터 圓周에 가까이 있는 特性値는  $x_3', x_6', x_8', x_9', x_{10}', x_{11}', x_{12}'$ 이며 이들 特性値의 寄與率을 <表 6>으로부터 살펴보면 모두 80% 이상이 됨을 알 수 있다. 반대로 <그림 1>에서 圓의 中心에 비교적 가까운 位置에 있는 特性値는  $x_2', x_4', x_5', x_7'$ 로서 이들 特性値들은  $z_1$ 과  $z_2$ 로서는 그다지 說明이 되지 않는다는 것을 알 수 있다. 이는 <表 6>에서 이들 特性値의 寄與率을 보면 約 30~50%로서  $z_1, z_2$ 의 寄與率이 매우 낮음을 알 수 있다. 또한 이들 4個의 特性値는 <表 5>의  $z_3$ 에 대한 特性値의 固有벡터로부터 第3主成分  $z_3$ 의 合成得點을 決定하는 特性値라는 것을 알 수 있다. 따라서 이들 特性値는 第3主成分  $z_3$ 에 의해서 說明이 되어질 것이다.

### 5. 主成分의 合成得點에 의한 島嶼分類의 節次

本 論文에서 島嶼의 各 特性値에 대해 主成分分析을 實施한 目的은 綜合特性値  $z_1, z_2$ 의 合成得點을 使用하여 慶南地域內에 散在해 있는 島嶼를 分類하고자 하기 위함이었다. 여기서 分類對象인 島嶼의 數  $n=88$ 個 島嶼이며 選定된 特性値의 數  $p=12$ 個의 變量이다. 따라서 각각의 島嶼는 12次元 空間上의 한 點에 該當될 것이다. 이제 이와 같은 12次元空間上에 位置해 있는 각 點을 第1主成分  $z_1$ 과 第2主成分  $z_2$ 에 의한 2次元空間으로 次元을 減少시켜 各 點의 散布狀態를 考慮하여 島嶼의 分類를 實施하는 節次를 要約하면 다음과 같다.

step 1. 주어진 資料(各 島嶼의 特性値)를 利用하여 各 特性値의 相關行列을 구한다.

step 2. 主成分分布을 實施하여 크기가 가장 큰 2個의 固有值  $\lambda_1, \lambda_2(\lambda_1 > \lambda_2)$ 를 구하고 이에 對應하는 固有 벡터를  $(l_{11}, l_{12}, \dots, l_{1p}), (l_{21}, l_{22}, \dots, l_{2p})$ 라 하면 第1主成分  $z_1$  및 第2主成分  $z_2$ 는 各島嶼의 特性値를 式③과 ④에 代入하여 구한다.

step 3. step 2.에서 구한 各 島嶼의 第1主成分  $z_1$  및 第2主成分  $z_2$ 의 得點을 使用하여 散布圖(scatter diagram)를 作成한다.

step 4. step 3.에서 구한 散布圖로부터 散布狀態를 考慮하여 島嶼分類를 實施한다.

以上과 같은 節次에 따라 各 島嶼別로 第1, 第2主成分의 合成得點을 구하여 表示한 <表 7>의 島嶼別 主成分得點을 利用하여 <그림 2>와 같은 散布圖를 作成하였다.

<그림 2>에 表示된 숫자는 島嶼番號를 意味하며 이는 島嶼番號 1부터 88까지로 되어 있다.

여기서 <그림 2>의 散布圖로부터 散布狀態를 보면 88個의 對象島嶼는 明確하게 異質的인 對象島嶼群임을 알 수 있다. 따라서 이 對象島嶼의 散布圖로부터 散布狀態를 고려하면 대체로 對象島嶼群은 A, B, C, D, E, F群으로 分割可能하다. 여기서 F群은 <그림 2>에 나타나 있는 A, B, C, D, E群과는 異質的인 特性을 갖는 島嶼이다. 이와 같이 分類를 할 경우에는 물론 12次元空間을 2次元空間으로 次元을 減少시킴으로써 發生되는 情報의 損失을 가능한 最少로 하지 않으면 안된다.

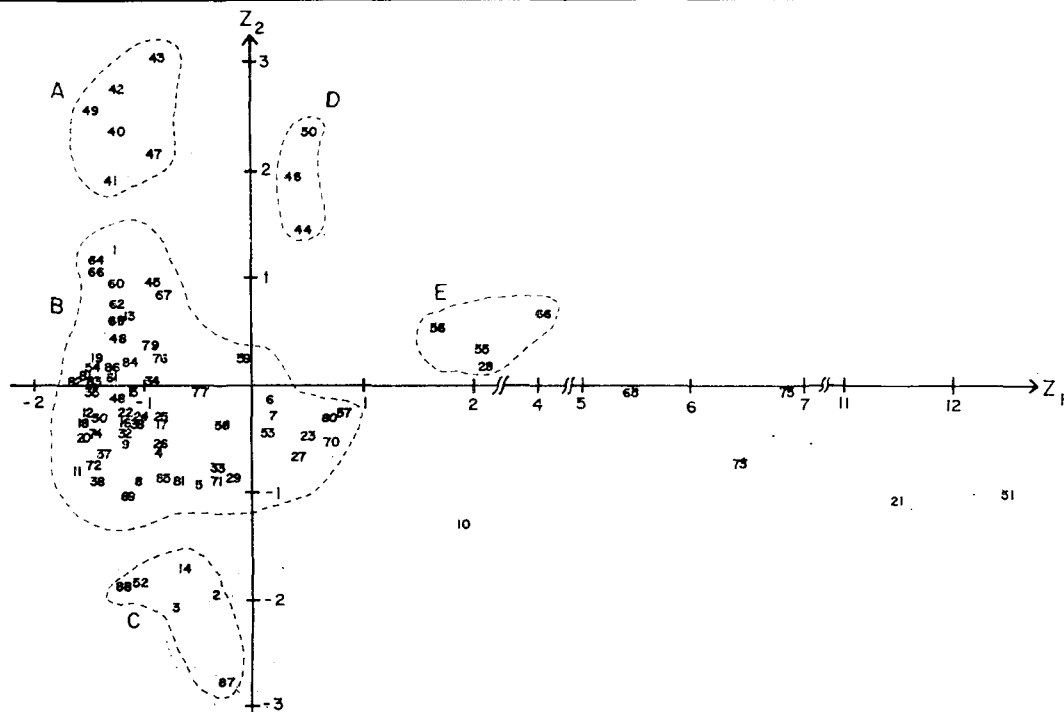
本 論文의 경우 次元減少로 인한 情報의 損失은 約 26%이나 대부분의 경우 程度의 차이는 있으나 이와 같은 情報의 損失은 必然的으로 發生하게 된다. 따라서 이와 같은 情報의 損失은 人間의 思考力이 機能을 發揮하는 3次元空間 以下로 次元을 減少시키는 경우에 生成되는 必要惡的 存在로서 理解되어야 할 것이다.

PCA에 의한 島嶼分類에 關한 研究

<表 7>

島嶼別主成分得點

島嶼No.	Z <sub>1</sub>	Z <sub>2</sub>	島嶼No.	Z <sub>1</sub>	Z <sub>2</sub>	島嶼No.	Z <sub>1</sub>	Z <sub>2</sub>	島嶼No.	Z <sub>1</sub>	Z <sub>2</sub>
1	-1.23	1.25	23	0.47	-0.45	45	-0.93	0.98	67	-0.86	0.85
2	-0.38	-1.95	24	-1.04	-0.30	46	0.31	1.99	68	-1.26	0.60
3	-0.76	-2.04	25	-0.82	-0.29	47	-0.93	2.14	69	-1.15	-1.01
4	-0.83	-0.57	26	-0.82	-0.51	48	-1.24	0.41	70	0.69	-0.51
5	-0.54	-0.93	27	0.40	-0.63	49	-1.46	2.57	71	-0.36	-0.87
6	0.12	-0.15	28	2.13	0.15	50	0.44	2.35	72	-1.50	-0.74
7	0.19	-0.26	29	-0.20	-0.93	51	12.44	-1.00	73	6.20	-0.76
8	-1.05	-0.86	30	-1.42	-0.26	52	-1.11	-1.85	74	-1.48	-0.42
9	-1.19	-0.43	31	-0.67	-0.89	53	0.09	-0.44	75	6.88	-0.02
10	1.86	-1.32	32	-1.17	-0.42	54	-1.48	0.09	76	-0.86	0.21
11	-1.59	-0.79	33	-0.35	-0.78	55	2.04	0.34	77	-0.52	-0.04
12	-1.48	-0.21	34	-0.97	0.03	56	1.66	0.53	78	-1.26	-0.10
13	-1.19	0.62	35	-1.47	-0.02	57	0.79	-0.27	79	-0.97	0.38
14	-0.63	-1.65	36	-1.04	-0.37	58	-0.31	-0.37	80	0.76	-0.27
15	-1.08	-0.05	37	-1.40	-0.63	59	-0.02	0.22	81	-1.50	0.06
16	-1.16	-0.32	38	-1.41	-0.91	60	-1.25	0.98	82	-1.59	0.05
17	-0.85	-0.36	39	11.85	2.76	61	-1.30	0.11	83	-1.50	0.06
18	-1.57	-0.32	40	-1.22	2.35	62	-1.22	0.79	84	-1.14	0.21
19	-1.45	0.22	41	-1.29	1.88	63	-1.44	1.07	85	-0.82	-0.84
20	-1.54	-0.46	42	-1.22	2.75	64	-1.44	1.18	86	-1.32	0.16
21	11.45	-1.11	43	-0.92	3.07	65	5.51	-0.02	87	-0.26	-2.74
22	-1.17	-0.25	44	0.43	1.46	66	4.04	0.66	88	-1.15	-1.84



<그림 2> 88個 島嶼에 대한 合成得點의 散布圖

수 산 경 영 론 집

〈表 8〉 島 嶼 分 類 結 果

群	No.	島 嶼 名	No.	島 嶼 名	No.	島 嶼 名	No.	島 嶼 名	No.	島 嶼 名	No.	島 嶼 名
A群 (6個島嶼)	40	內 草 島	41	蓬 島	42	國 島	43	葛 島	47	下 老 大 島	49	外 巨 七 里 島
B群 (62個島嶼)	1	叢 島	4	友 島	5	馬 島	6	楮島 (상천군)	7	勒 島	8	草 養 島
	9	辰 島	11	竹島(양산)	12	椿 島	13	眞 友 島	15	實 里 島	16	松 島
	17	羊 島	18	弓 島	19	水 牛 島	20	竹島(의창)	22	海 民 島	23	紙 島
	24	水島(통영)	25	於 義 島	26	鳥 飛 島	27	昆 里 島	29	鶴 林 島	30	松島 (통영 산장면)
	31	楮島 (통영 산장면)	32	晚 地 島	33	烟 台 島	34	烏 各 島	35	鳶 島	36	邑 島
	37	笠 島	38	楮島 (통영 광도면)	45	牛 島	48	納 島	53	佐 島	54	松島 (통영 한산면)
	57	比 珍 島	58	竹島(통영)	59	每 勿 島	60	悔 勿 島	61	加 王 島	62	長 蛇 島
	63	大 德 島	64	漁 遊 島	67	樹 牛 島	68	只 心 島	69	內 島	70	山 蓬 島
	71	花 島	72	放 火 島	74	高 介 島	76	黃 德 島	77	利 水 島	78	臥 島
	79	차 란 도	80	飛 兎 島	81	辰 島	82	別 鶴 島	83	月 登 島	84	槽 島
	85	鳥 島	86	虎 島								
	C群 (6個島嶼)	2	水島(진해)	3	椽 島	14	猪 島	52	飛 山 島	87	大 島	88
D群 (3 " )	44	蓮 花 島	46	上 老 大 島	50	頭 尾 島						
E群 (4 " )	28	楸 島	55	七 川 島	56	龍 虎 島	66	下 島				
F群 (7 " )	10	新 樹 島	21	加 德 島	39	欲 知 島	51	開 山 島	65	上 島	73	加 助 島
	75	秋 蜂 島										

註) 〈表 7〉의 No.는〈그림 2〉에 나타나 있는 숫자로서 島嶼名의 code임.

Ⅲ. 結 論

主成分의 合成得點에 의한 島嶼의 分類節次에 따라 慶南地域에 散在해 있는 88個의 有人島嶼를 〈그림 2〉의 散布圖의 散布狀態에 따라 分類한 結果를 表示하면 〈表 8〉과 같다.

여기서 〈表 8〉의 島嶼分類 結果에 의한 各 島嶼群의 特性에 關係 島嶼群別로 考察하면 다음과 같다.

A 群 :

A群에 屬하는 島嶼는 總 6個 島嶼로서 이들 島嶼는 대체적으로 島嶼의 綜合的 規模<sup>17)</sup>가 작고 內陸으로부터 매우 멀리 位置해 있으며 戶當所得이 매우 낮은 特性을 갖는 島嶼群이다.

B 群 :

B群에 屬하는 島嶼는 總 62島嶼로서 가장 많으며 이 島嶼들이 갖는 特性을 보면 다음과 같다.

17) 여기서 綜合的 規模라 함은 島嶼面積, 海岸線 延長距離, 人口, 學生數, TV台數, 라디오台數, 漁船隻數 등의 綜合的인 크기를 말한다.

### PCA에 의한 島嶼分類에 관한 研究

B群의 島嶼는 島嶼의 綜合的인 規模面에서는 A群과 비슷하거나 약간 큰 島嶼이며, 內陸으로부터 비교적 가까운 位置에 있으면서 戶當所得이 그다지 높지 않는 島嶼와 內陸으로부터 비교적 멀리 떨어져 있으면서 戶當所得이 비교적 높은 島嶼이다.

#### C 群 :

C群은 總 6個島嶼로서 이들 島嶼의 綜合的인 規模는 A群과 비슷하다. 그러나 內陸과 가까운 位置에 있으면서 戶當所得이 매우 높은 島嶼들로 構成되어 있다.

#### D 群 :

D群에 屬하는 島嶼는 3個島嶼로서 島嶼의 綜合的 規模는 A群보다 약간 크며, 內陸으로부터 떨어져 있으며 戶當所得이 낮은 島嶼이다.

#### E 群 :

E群에 屬하는 島嶼는 4個島嶼이며 이 群에 屬하는 島嶼의 綜合的 規模는 상당히 크며, 內陸에서 比較的 먼 位置에 있으면서 戶當所得이 보통인 島嶼群이다.

#### F 群 :

F群의 島嶼는 總 7個島嶼로서 A, B, C, D, E群과는 매우 異質的인 島嶼들로서 島嶼의 綜合的인 規模가 매우 큰 島嶼群이다. 이 島嶼들은 內陸과의 距離 및 戶當所得은 各 島嶼別로 매우 相異하다. 따라서 F群은 A, B, C, D, E群에 屬하지 않는 島嶼들로서 各 島嶼가 제각기 特異한 特性을 갖고 있는 島嶼群이라 할 수 있다.

### 參 考 文 獻

- 1) 內務部, 島嶼·落島現況, 第1號, 1981, pp.136~145.
- 2) 松原 望, 意思決定의 基礎, 朝倉書店, 1978, p.155.
- 3) 奥野忠一, 外3人, 多變量分析法, 日科技連, 1978, p.6, pp.192~208.
- 4) 山田文道, 經營指標による企業評價のための新しい試み-多變量解析によるアプローチ, engineers, No.262.
- 5) OR事典編集委員會編, OR事典, 日科技連, 1975, pp.167~168.
- 6) 淺野長一郎, 因子分析法通論, 共立出版株式會社, 1972, pp.36~38.
- 7) 高根芳雄·柳井晴夫, 多變量解析法, 朝倉書店, 1978, p.84.
- 8) 內務部, 島嶼카드(慶南地域), 1981.
- 9) Donald F. Morrison, Multivariate statistical methods, McGraw-Hill, 1976, pp.266~299.
- 10) G. Wiegand, Some applications of principle components analysis in vegetation: Ecological research of aquatic communities, Advances in vegetation science 2, 1980, pp.67~74.

## A Study on the Classification of Islands by PCA

Kang-Woo Lee

### Summary

This paper considers a classification of the 88 islands located at Kyong-nam area in Korea, using by examples of 12 components of the islands. By means of principal component analysis 2 principle components were extracted, which explained a total of 73.7% of the variance. Using an eigen variable criterion ( $\lambda > 1$ ), no further principle components were discussed. Principal component 1 and 2 explained 63.4% and 10.3% of the total variance respectively.

The representation of the unrotated factor scores along the first and second principal axes produced a new information with respect to the classification of the islands. Based upon the representation, 88 islands were classified into 6 groups i. e. A, B, C, D, E, and F according to similarity of the components among them in this paper.

The "Group F" belongs to a miscellaneous assortment that does not fit into the logical category.