

以群集分析法探討水庫優養化之動力研究

陳鴻烈⁽¹⁾ 蔡大偉⁽²⁾

摘要

本研究是以德基水庫優養化變化趨勢之相關重要參數為研究對象，運用群集分析法來進行資料最佳分類的工作，以協助水庫優養化管理策略的制定。研究中是先經過單迴歸分析篩選出的優養化重要因子為對象，透過多元複迴歸分析所形成之優養全模式為基礎來進行分析。研究中使用 TSI (Trophic State Index) 作為分群結果探討檢驗的指標，並採用階層與非階層群集分析 2 種方法來分群資料。研究結果顯示，階層分析結果中以馬氏法最符合優養指標 TSI 的變化趨勢，該方法建議將觀測組分為 4 群，而離群值修正前後會影響分群結果，且以修正後的分群效果較佳。另外，非階層分析結果顯示以 4 群為準時，其分群效果尚稱顯著。至於模式中各個變數根據 R^2 值排序結果依序為磷酸鹽、葉綠素 a、COD、懸浮固體、濁度、總磷、水溫、透明度、硝酸鹽、溶氧、氨氮及鈉，其中以磷酸鹽最為重要。若將群集分析結果與鑑別分析作一比較，可發現分群是以磷因子最為重要，中高度優養的分群效果較為明顯，但兩者分析結果的差異亦指出應該有更佳的分類指標存在。

(**關鍵字**：優養化、群集分析、水質因子、德基水庫)

Cluster Analysis of Water Quality Factors of Eutrophication in the Te-Chi Reservoir

Paris Honglay Chen⁽¹⁾ *David D.W. Tsai*⁽²⁾

Professor⁽¹⁾ and Graduate Student⁽²⁾, Department of Soil and Water Conservation, National Chung-Hsing University, R.O.C.

Abstract

The research used the cluster analysis to classify the original water quality data of eutrophication in the Te-Chi reservoir. We wished the results could enhance and help the eutrophication management of reservoir. The full eutrophication model formed from the multiple regression analysis was utilized to analyze the important trophic factors selected by the simple regression analysis. The TSI index was adopted as classification standard, as well as two methods of hierarchical and disjoint clustering were used to categorize the data. The results of hierarchical

⁽¹⁾ 國立中興大學水土保持學系教授，通訊作者

⁽²⁾ 國立中興大學水土保持學系博士班研究生

clustering showed that McQuitty's method mostly corresponded to the TSI variation. This method suggested the data should be separated into four clusters. And the cluster results after calibrating the outliers were better than that of before. Additionally, the results of disjoint clustering also indicated the effect of four clusters was acceptable. Furthermore, the R^2 values of variables in the model were phosphate, chlorophyll-a, COD, suspended solids, turbidity, total phosphorus, water temperature, secchi depth, nitrate, dissolved oxygen, ammonia nitrogen, and sodium in order. Among them, the most important variable was phosphate. To compare the results of cluster and discriminant analysis, we found phosphorus was the most important factor in the cluster studies, and better effects of cluster occurred in the middle and heavy eutrophication conditions. However, the differences between both methods also indicated that it is expectable for a better index of cluster.

(Keywords: Eutrophication, Cluster analysis, Water quality factor, Te-Chi reservoir)

前言

台灣的經濟開發與發展已逐步改善人民的物質生活，但不合理的開發亦造成生活環境的破壞，如山坡地農業開發污染水源，造成水體優養化的發生即為一例。因此，為了台灣的永續發展，合理的開發與環境資源保護已成為不可或缺的工作。台灣的經濟開發與發展已逐步改善人民的物質生活，但不合理的開發亦造成生活環境的破壞，如山坡地農業開發污染水源，造成水體優養化的發生即為一例。因此，為了台灣的永續發展，合理的開發與環境資源保護已成為不可或缺的工作。

水體優養化之研究經本團隊的努力，其動力機制已逐漸被掌握，相關研究包括時間序列預測、優養模式分析、動力機制探討等 [陳, 1997、1998、1999、2000a、2000b、2004、2005a、2005b、2006a、2006b、2006c、2006d、2007a、2007b、2007c]。為了更進一步了解台灣的優養動力，本研究將使用群集分析法來探討此一主題。

在未知優養的主要動力與變化時，可使用群集分析來探討水質的時間與空間變化 [Alberto, 2001]。若要將複雜觀測值進行分群來加以探討，以尋求最佳分群指標時，則可用群集分析法來做先導性研究 [McNeil, 2005]，以探討所用的資料中各變數的指標性與相互間的關聯 [Lambrakis, 2004]，並可檢驗既有指標的正確與適用性。

理論分析

群集分析的對象可分為兩大類 — 觀測值或是變數 [Reisenhofer, 1994]，本研究中主要是針對觀測值進行分群。研究中所採用的群集分析方法分為兩大類，第一類為階層式群集分析 (hierarchical cluster analysis)，第二類為非階層式群集分析 (nonhierarchical cluster analysis)，而主要使用的分析法則為 K-均數法 (K-means)。

階層式群集分析法又分為兩類，第一類為集結法 (agglomerative method)，第二類為分裂法 (divisive method) [彭, 1998]。

一、階層式群集分析 [林，2003；陳，2003；Cary，1999]

(一) 集結法

集結法所採用的集群方式為先計算出各觀察體間之距離或組內誤差矩陣，然後將最接近的兩個觀察體加以合併成一集群，再繼續算出合併後的觀察體間之距離或組內誤差，如此一直重複以上的步驟，直到所有觀察體合併於同一集群為止。此法共包含 11 種分析方法，將一一分述如後。

1. 單一連結法

單一連結法 (single linkage cluster analysis) 又稱為近鄰法 (nearest neighbor method)，意即在計算觀察體間之距離時，是取兩者間距離最短者，以此類推，直到全部觀察體合併成同一群體為止。

2. 完全連結法

完全連結法 (complete linkage cluster analysis) 又稱為遠鄰法 (farthest neighbor method)，其集群連結間距計算方式與單一連結法相反，亦即在計算集群間距離時，是取兩者之最遠距離。

3. 彈性 β 法

彈性 β 法 (flexible-beta method) 主要是使用特定之係數來決定集群間的距離，若以觀測組 C 與群集 CAB 結合為例，則兩者間之距離可用下式表示：

$$D(C, C_{AB}) = \alpha_A D(C, A) + \alpha_B D(C, B) + D(A, B) + \gamma |D(C, A) - D(C, B)|$$

其中

$$\alpha_A = \alpha_B = \frac{1 - \beta}{2}$$

$$\alpha_A + \alpha_B + \beta = 1$$

$$\gamma = 0, \beta < 1$$

β 建議值為 -0.25

4. 馬氏法

馬氏法 (McQuitty's method) 之新群集與其他群集距離之算法，為計算其他群集與原群集距離之平均。

5. 中數法

中數法 (median method) 之群集方式為計算兩群間之中間位置距離，亦即取兩群集之中數。

6. 平均連結法

平均連結法 (average linkage cluster analysis) 之群集方式為合併有最短間距之集群，而計算集群間距離的方式為取平均距離，然後再取間距最短者合併，一直重複此動作到合併成同一群體為止。

7. 形心連結法

形心連結法 (centroid hierarchical cluster analysis) 在計算群體間距離時是採用取形心的方式，依此計算各群體間的距離，再取有最短距離者加以合併，然後持續重複此一動作，直到全部觀察體都合併為一個群體為止。

8. 華德最小變異法

華德最小變異法 (Ward's minimum variance cluster analysis) 的群集目的是希望合併後群集內的聯合組內變異量達到最小，變異量以離均差平方和代表

$$SS = \sum_{i=1}^p \sum_{j=1}^m (x_{ij} - \bar{x}_i)^2$$

其中 i 為變項, j 為觀察體, x_{ij} 為觀測值, \bar{x}_i 為變項平均數。其集結步驟為：

(1) 將每個觀察體視為單獨一個集群, 計算其組內變異量, 此時為 0。

(2) 分別計算 N 個觀察體與其他某一個觀察體合併時的組內變異量, 然後取其中變異量最小者的兩個觀察體合併, 將之視為一個觀察體, 此時觀察體數為 N-1。

(3) 依以上方法重複合併, 直到全部合併成一個群體為止。

9. ML 法

ML 法 (maximum likelihood method) 在每一個群集結合階段時均計算其最大概似值, 其基本假設如下：

- (1) 多變數常態混合
- (2) 具相同球狀共變異矩陣
- (3) 不同抽樣機率

舉例說明之, 假設現有 A、B、C 三群集, A 與 B 群集先行合併為 CAB, 則 CAB 與 C 之距離可用下式表示：

$$D(A, B) = nv \ln(1 + \frac{B_{AB}}{P_C}) - 2(N_{C_{AB}} \ln(N_{C_{AB}}) - N_A \ln(N_A) - N_B \ln(N_B))$$

n : 觀測值個數

v : 變數個數

$$BAB = WCAB - WA - WB$$

其中

$$W_{C_{AB}} = \sum_{i \in C_{AB}} \|x_i - \bar{x}_k\|^2$$

$$W_A = \sum_{i \in A} \|x_i - \bar{x}_k\|^2$$

$$W_B = \sum_{i \in B} \|x_i - \bar{x}_k\|^2$$

$$P_C = W_C = \sum_{i \in C} \|x_i - \bar{x}_k\|^2$$

x_i 為群集內第 i 個觀測值, \bar{x}_k 為群集平均向量。

$N_{C_{AB}}$: 群集 CAB 內之觀測值個數

N_A : 群集 A 內之觀測值個數

N_B : 群集 B 內之觀測值個數

10. 密度連結法

密度連結法 (density linkage method) 代表一連串無參數群集分析的通稱, 方法為計算其密度機率, 可分為兩個步驟：

(1) 根據密度估計值與相似性來計算觀測值間的距離 d^* , 若觀測值 x_i 、 x_j 近似, 則 $d^*(x_i, x_j)$ 以中間密度來估算, 否則則為無限遠。

(2) 使用 d^* 為基礎, 用單一連結法合併群集。

其中相似性的定義可用 k 個最近鄰群集法、均質 Kernel 法、王氏雜交群集法等來估算。

11. 雙連法

雙連法 (two-stage density method) 為密

度連結法的修正，是使用兩階段的方式來確定所有觀測點在全部觀測點被合併為一群集前均被分配到各個群集。此方法分為兩個階段：

(1) 進行觀測值的分群，其方法與密度連結法相同，均使用單一連結法；不同之處在於兩個群集在進行合併時，只在其中一個群集的個數少於預設值時才進行合併。

(2) 進行群集的階層式合併，合併法亦使用單一連結法。而若在群集間有明顯差異或是平滑參數很小時，最後合併的結果可能大於 1 個群集。

(二) 分裂法

分裂法使用之群集觀念與集結法相反，其方法為依序將相異性最大者分裂為另一群集，其步驟先視全部觀察體為一個群集，然後計算哪一個觀察體與群集相異性最大或是相似性最小，將其分裂成另一群集。接著計算大群集中每個觀察體與群內及群外之平均距離，如果計算結果比較接近另一個群體，即群內平均距離減掉群外平均距離為正數者，即分到另一群集中，否則，則留在原群集。此過程一直重複至兩個群集內皆無法再分裂為止，然後繼續重複以上步驟，直到全部觀察體都被分為單一群集才停止。

二、非階層式群集分析 [林，2003；陳，2003；Cary，1999]

非階層式群集分析最常使用的方法為 k 平均法，此法需要先設定種子數 (seed) 以進行分析，而種子數的選取會影響分類的結果。在選取種子數之後，需要將觀察體分配到距離最近的種子點以進行分類。

(一) k 平均法

k 平均法之分類步驟是先假設需要分為 k 個群集，再依此設定 k 個種子點，然後計算每個觀察體到各群集形心的距離，並將其分配到最近的一群。重新分派各個觀察體到適當群集，直到達到初始假定群集數為止，此步驟持續重複到無法重新分派為止。

1. 種子點選擇法

k 平均法中種子點的選取會影響分類的合理性，若是選取不當，可能會導致不理想的結果，而一般種子點選取的方法有 6 種：

(1) 選擇前 k 個沒有缺失值的觀察體。

(2) 先選擇第 1 個觀察體當第 1 群集種子，其次選擇與第 1 個種子距離超過既定標準的下一個觀察體當第 2 群集種子，接著選擇與第 1、2 個種子距離超過既定標準的下一個觀察體當第 3 群集種子，依此類推，直到選出 k 個群集的種子為止。

(3) 隨機選出 k 個觀察體當種子。

(4) 先選擇 k 個種子，然後依照某種既定標準加以置換種子。

(5) 依照某種既定標準選取 k 個種子。

(6) 由研究者自訂 k 個種子。

2. 觀察體分配方法

觀察體分配方法也會影響分類的結果，通常有以下 3 種分派法：

(1) 先計算每個集群形心，然後分派觀察體到距離最近的群體中，在分派過程中，形心的使用不更新，直到全部觀察體分派完

後，再重新計算集群形心。若分派後計算的形心比分派前形心之改變超過某種收斂標準，則重新計算每個觀察體與各集群形心距離，然後再將之分派到最近的集群，直到形心的改變沒有超過標準為止。

(2) 與 (1) 的方法相似，只是在分派過程中必須計算觀察體要加入及退出集群時的形心，直到達到既定收斂標準為止。

(3) 重新分派觀察體，直到使某種統計量達到最小為止，此方法稱為爬山法 (hill-climbing method)，而一般使用的統計規則有以下 4 種：

- a. 組內 SSCP 矩陣的跡 (trace) 為最小。
- b. 組內 SSCP 矩陣的行列值 (determinant) 為最小。
- c. W-1B 矩陣的跡為最小，其中 W 與 B 分別代表組內及組間的 SSCP 矩陣。
- d. W-1B 矩陣的特徵值為最大。

三、群集數判斷方法

在使用階層式集群分析法後，群集數判斷方法可用來判斷群集分析後結果之最佳分群數為何，一般所使用的統計標準有以下 6 種：

(一) R2 (R-squared)

R2 指標代表著集群內的異質性，可經由變異數分析來計算，計算方式為：

$$R^2 = \frac{SS_B}{SS_T}$$

式中 SSB 代表組間變異平方和，SST 代

表全體變異平方和 [潘，2003]。

當觀察體持續合併的動作時，集群數目會逐漸減少，因此集群內異質性會越來越高，使得組內變異量也隨之增大，而在全體變異量固定之下，則組間變異量會因此越來越小。由於群集分析之主要目的是使組內變異量小，而組間變異量大，因此 R2 值越大代表分組效果越好，所以當合併的動作會使 R2 突然減小時，表示合併應該停止。

(二) 半淨 R2

半淨 R2 (semi-partial R-squared) 指標代表當集群加入新觀察體時所增加的異質性，其計算方法為每個集結階段比上一階段所增加的組變異量與全體變異量之比，因此半淨 R2 指標值越小代表組內所增加的異質性越小，即分組效果越佳，所以當合併動作導致半淨 R2 值突然增加時，即表示該停止合併。

(三) 均方根標準差

均方根標準差 (root-mean-square standard deviation, RMSSTD) 可代表新集群之異質性，即集群中所有觀察體之聯合標準差，其計算方式如下：

$$RMSSTD = \sqrt{\frac{(n-1) \sum_{i=1}^p \hat{S}_i^2}{P(n-1)}} = \sqrt{\frac{\sum_{i=1}^p \hat{S}_i^2}{P}}$$

式中 P 為變數個數，即分組數；n 為樣本數， \hat{S}_i 為第 i 個變數之每一群內的標準差。

因此 RMSSTD 指標值越小代表分群效果越好，所以當集群合併後使得 RMSSTD 值突然上升時，即代表應該停止合併動作。

以上 3 種指標值可以進一步繪製陡坡圖，使得其分布趨勢可以更清晰的表現，便於判斷合理的分群數。

(四) 立方分群指標

立方分群指標 (cubic clustering criterion, CCC) 評斷合理分群數之方法為找出區域最高點的地方，即數值有先升後降之處，亦即最高點處即為較佳的分群數位置。然而由於 SAS 統計軟體群集分析僅列出 20% 結果，故無法以此指標進行判斷。

(五) 偽 F 值

偽 F 值 (pseudo F, PSF) 為集群間變異均方與集群內變異均方之比例，而其判斷標準亦為找出區域最高點處，而該處所指之集群數即為合理分群數。

(六) 偽 t2 值

偽 t2 值 (pseudo t2, PST2) 指標的判斷方法為找出兩群體間有最大差異者，即指標值有突然上升之點，該點所代表之群集數即為較合理之分群結果。

研究方法

本研究是以德基水庫水質資料為基礎 [18]，並先選擇要進行群集分析之數據，所採用的數據是經過單迴歸分析後與優養化呈高度相關的水質因子，包括水溫、溶氧、懸浮固體、濁度、氨氮、硝酸鹽、總磷、磷酸鹽、鈉、COD、葉綠素 a 與透明度等 12 項，這些因子的決定係數 (R2) 如表 1 所示。

選定參數後，本研究採用統計軟體 SAS

進行分析，依序是階層與非階層群集研究，最後再按各種分類結果加以討論，然後依照評判的標準來找出最佳分類結果與方法，以作為決定最佳分群數的標準。接著將群集分析結果與鑑別分析做比較，來探討 TSI 指標的正確性與適用性，最後再根據分析數據得出最佳分類結果，研究流程如圖 1 所示。

表 1 群集分析選用參數之決定係數

Table 1 R2 of variables in the cluster analysis

排序	變數	R-Square	P-Value
1	總磷	0.7997	< 0.0001
2	懸浮固體	0.5400	< 0.0001
3	透明度	0.3956	< 0.0001
4	葉綠素 a	0.3818	< 0.0001
5	COD	0.3259	< 0.0001
6	濁度	0.2293	< 0.0001
7	水溫	0.1324	0.0034
8	鈉	0.1292	0.0194
9	溶氧	0.1290	0.0038
10	硝酸鹽	0.1055	0.0094
11	氨氮	0.1009	0.0142
12	磷酸鹽	0.0961	0.0321

結果與討論

一、階層式群集分析

本研究所採用之階層式群集分析法共有 11 種，其分析結果分別討論於後，希望藉由比較其異同來找出最佳的群集分法。

(一) 單一連結法

數據經過單一連結法分析後之結果如表 2 所示，表中第一欄代表群集的數目 (NCL)，因本研究數據為 19 組，因此群集數由 18 到 1。第二欄為群集的步驟 (Cluster Jointed)；第三欄代表各個群集步驟之群集內的觀測值數目 (FREQ)。

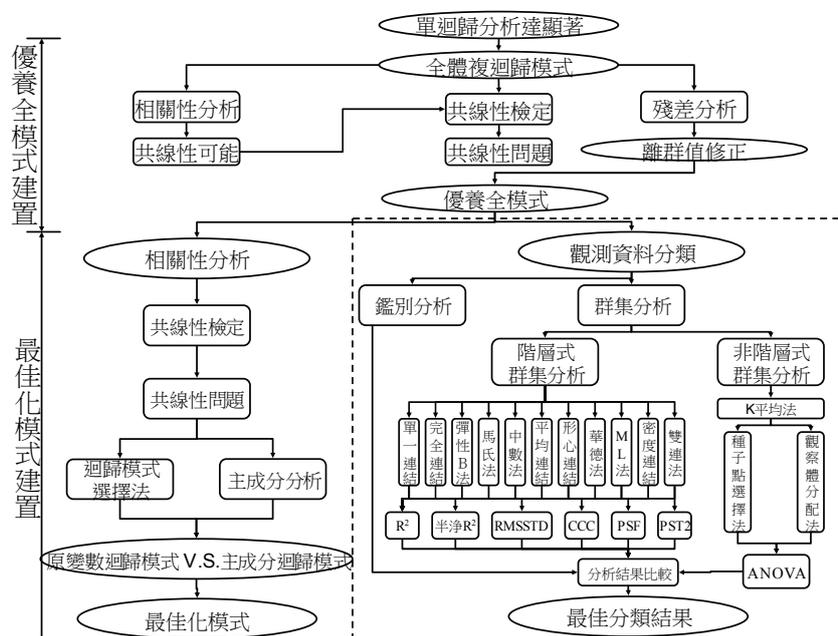


圖 1 研究流程圖

Figure. 1 Procedure diagram of investigation

第四欄表示均方根標準差 (RMSSTD)，當其值有突然增加的情形時，就表示該點為群集數之判斷依據，由表中可看出其變化不規則，而在集群數 16 時開始有較大的增加幅度，但幅度不大，仍難以判斷。

第五欄為半淨 R2 值 (SPRSQ)，可用該值突然上升之點為群集數判斷的標準，從表中可知在群集數分為 8 後，半淨 R2 有突然增加的現象 (從 0.0465 增為 0.1764)，亦即此一指標顯示將數據分為 9 群集為適當的分法。

第六欄為 R2 指標 (RSQ)，當其值突然降低時，代表群集間異質性突然減少，也就是群集內異質性增加，所以應該停止合併，其與半淨 R2 指標有相反的意義。由表中可看出在群集數分為 8 後，R2 值由 0.806 降為 0.629，有突然降低的情形，所以此一指標亦指出應將資料分為 9 個群集。為了能更清楚

看出各個指標的增減情形，以利於判斷群集數，故以均方根標準差 (RMSSTD)、半淨 R2 (SPRSQ) 與 R2 (RSQ) 等 3 個指標為基礎繪製陡坡圖 (圖 2)，由圖中可看出除了均方根標準差較看不出趨勢外，半淨 R2 與 R2 指標均顯示將數據分為 9 群組的結果。

表中第七欄為在齊一虛無假設下 R2 之近似期望值 (CCC)，第八欄為立方分群指標 (Norm min dist)，而其中由於 SAS 軟體只列出 CCC 最後 20% 的數據，因而難以作為分群的判斷。

因此，藉由單一連結分析法中各個指標數值之判斷，應將原始數據分為 9 個群集，第一群為觀測值 2、3、6、7、14、15、18，第二群為 5、8、9、13、17，剩下的 7 個觀測值 (1、4、10、11、12、16、19) 則各自成為一群。

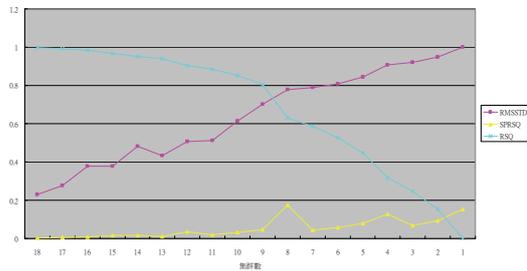


圖 2 單一連結法分析結果陡坡圖

Figure. 2 Scree plot of single linkage cluster analysis

(二) 完全連結法

完全連結法分析結果如表 3，由表中第四欄可看出均方根標準差變化情形呈現非常不規則狀態，在群集數 4 和 11 時都產生數值明顯上升的情況，所以較難做出分群的評判。第五欄數據顯示在群集數分為 3 組後，半淨 R2 指標有突然增加的現象，由 0.0652 增至 0.1581，代表分群數為 4 時即停止。另外，從第六欄可發現在群集數分為 3 後，

表 2 單一連結法分析結果

Table 2 Results of single linkage cluster analysis

The data have been standardized to mean 0 and variance 1								
Root-Mean-Square Total-Sample Standard Deviation = 1								
Mean Distance Between Observations = 4.674706								
NCL	Cluster Joined		FREQ	RMSSTD	SPRSQ	RSQ	CCC	Norm Min dist
18	OB14	OB18	2	0.2296	0.0029	0.997	-	0.2406
17	CL18	OB15	3	0.2759	0.0055	0.992	-	0.2851
16	OB6	OB7	2	0.3776	0.0079	0.984	-	0.3957
15	OB3	CL17	4	0.3783	0.0154	0.968	-	0.4014
14	OB2	CL16	3	0.4810	0.0178	0.950	-	0.4437
13	OB5	OB13	2	0.4324	0.0104	0.940	-	0.4531
12	CL14	CL15	7	0.5055	0.0356	0.904	-	0.4701
11	CL13	OB9	3	0.5114	0.0187	0.886	-	0.5207
10	CL11	OB8	4	0.6132	0.0336	0.852	-	0.5238
9	CL10	OB17	5	0.7009	0.0465	0.806	-	0.5407
8	CL12	CL9	12	0.7789	0.1764	0.629	-	0.6095
7	CL8	OB12	13	0.7893	0.0446	0.585	-	0.6770
6	CL7	OB19	14	0.8095	0.0579	0.527	-	0.6777
5	CL6	OB4	15	0.8435	0.0801	0.447	-	0.7255
4	OB1	CL5	16	0.9056	0.1300	0.317	-	0.7467
3	CL4	OB16	17	0.9204	0.0696	0.247	-1.40	0.7620
2	CL3	OB11	18	0.9473	0.0944	0.153	-0.74	0.8635
1	CL2	OB10	19	1.0000	0.1525	0.000	0.00	1.0270

R2 值有突然降低的現象，由 0.599 降至 0.441，意味著將此群組資料分為 4 個群集為合理的分法。同樣地，將均方根標準差 (RMSSTD)、半淨 R2 (SPRSQ) 與 R2 (RSQ) 等 3 個指標繪成陡坡圖 (圖 3) 後，可看出其趨勢分布情形。

由完全連結分析法中各個指標結果顯示，應該將原數據分為 4 個群組，分群結果為將第 1、10、11 觀測值分為一組，第 4、5、13、16、17 與第 8、9、12、19 觀測值為另外

的兩組，剩下的第 2、3、6、7、14、15、18 觀測值則各自成為一組。

若與單一連結法相比較，可發現完全連結法將數據分為 4 個群組，而單一連結法則分為 9 群，顯示以完全連結法進行分析會得到較佳的分群結果，亦即分群結果較為精簡。

(三) 彈性 β 法

使用彈性 β 法分析後，由表 4 可知均方根標準差的主要驟升點在集群數 4 的地方，由 0.7437 增為 0.9625，表示此時聯合標準差

突然上升，因此應於集群數為 3 時停止合併。而半淨 R2 指標則是在分群成 2 時有突然增加的現象，由 0.1044 上升到 0.1951，代表此時組內變異增加過多，因此合理分群數為 3 群。至於 R2 指標的主要變化點亦同樣位於集群數為 2 群的地方，數值由 0.495 降到 0.3，所以 R2 指標指出資料應分為 3 群為合理。3 個指標之趨勢分布情形如圖 4 所示。

根據 3 個分群指標判斷後，判定此原始資料應分為 3 群，亦即第 1、10、11 觀測值為一組，第 2、3、6、7、14、15、18 為一組，另一組則是第 4、5、8、9、12、13、14、17、19 觀測值。

將彈性 β 法分析結果與單一連結法、完全連結法相比較，其中，單一連結法分群數為 9，完全連結法為 4，而彈性 β 法則為 3，顯示彈性 β 法分群結果最為精簡。

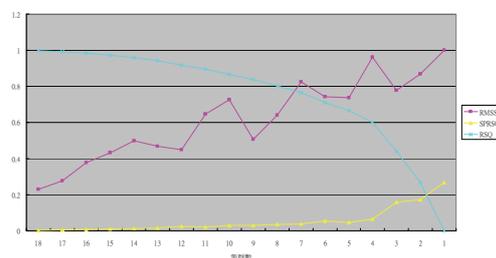


圖 3 完全連結法分析結果陡坡圖

Figure. 3 Scree plot of complete linkage cluster analysis.

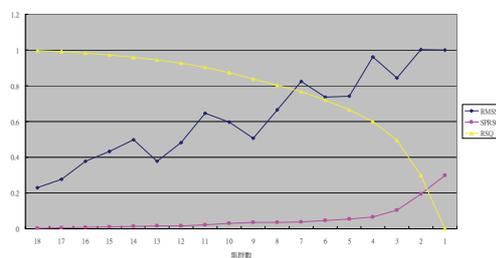


圖 4 彈性 β 法分析結果陡坡圖

Figure. 4 Scree plot of flexible-beta cluster analysis

表 3 完全連結法分析結果

Table 3 Results of complete linkage cluster analysis

The data have been standardized to mean 0 and variance 1								
Root-Mean-Square Total-Sample Standard Deviation = 1								
Mean Distance Between Observations = 4.674706								
NCL	Cluster Joined		FREQ	RMSSTD	SPRSQ	RSQ	CCC	Norm Max Dist
18	OB14	OB18	2	0.2296	0.0029	0.997	-	0.2406
17	CL18	OB15	3	0.2759	0.0055	0.992	-	0.3341
16	OB6	OB7	2	0.3776	0.0079	0.984	-	0.3957
15	OB5	OB13	2	0.4324	0.0104	0.973	-	0.4531
14	OB8	OB9	2	0.4998	0.0139	0.959	-	0.5238
13	OB3	CL16	3	0.4694	0.0166	0.943	-	0.5363
12	CL13	CL17	6	0.4500	0.0233	0.919	-	0.6089
11	OB12	OB19	2	0.6466	0.0232	0.896	-	0.6777
10	OB16	OB17	2	0.7271	0.0294	0.867	-	0.7620
9	OB2	CL12	7	0.5055	0.0289	0.838	-	0.7696
8	OB4	CL15	3	0.6400	0.0351	0.803	-	0.7862
7	OB1	OB11	2	0.8240	0.0377	0.765	-	0.8635
6	CL14	CL11	4	0.7437	0.0551	0.710	-	0.9324
5	CL8	CL10	5	0.7358	0.0454	0.665	-	1.0343
4	CL7	OB10	3	0.9625	0.0652	0.599	-	1.0807
3	CL9	CL6	11	0.7770	0.1581	0.441	4.12	1.1547
2	CL3	CL5	16	0.8682	0.1724	0.269	2.85	1.3691
1	CL4	CL2	19	1.0000	0.2689	0.000	0.00	1.7819

表 4 彈性 β 法分析結果

Table 4 Results of flexible-beta cluster analysis

Beta = -0.25								
The data have been standardized to mean 0 and variance 1								
Root-Mean-Square Total-Sample Standard Deviation = 1								
Mean Distance Between Observations = 4.674706								
NCL	Cluster Joined		FREQ	RMSSTD	SPRSQ	RSQ	CCC	Norm Flex Dist
18	OB14	OB18	2	0.2296	0.0029	0.997	-	0.2406
17	CL18	OB15	3	0.2759	0.0055	0.992	-	0.3269
16	OB6	OB7	2	0.3776	0.0079	0.984	-	0.3957
15	OB5	OB13	2	0.4324	0.0104	0.973	-	0.4531
14	OB8	OB9	2	0.4998	0.0139	0.959	-	0.5238
13	OB3	CL17	4	0.3783	0.0154	0.944	-	0.5334
12	OB2	CL16	3	0.4810	0.0178	0.926	-	0.5780
11	OB12	OB19	2	0.6466	0.0232	0.903	-	0.6777
10	CL15	OB17	3	0.5959	0.0291	0.874	-	0.7371
9	CL12	CL13	7	0.5055	0.0356	0.838	-	0.7561
8	OB4	CL10	4	0.6653	0.0343	0.804	-	0.8206
7	OB1	OB11	2	0.8240	0.0377	0.766	-	0.8635
6	CL8	OB16	5	0.7358	0.0465	0.720	-	0.9848
5	CL14	CL11	4	0.7437	0.0551	0.665	-	0.9949
4	CL7	OB10	3	0.9625	0.0652	0.599	-	1.1265
3	CL6	CL5	9	0.8445	0.1044	0.495	5.98	1.2960
2	CL4	CL3	12	1.0032	0.1951	0.300	3.90	1.8647
1	CL2	CL9	19	1.0000	0.2999	0.000	0.00	2.4132

(四) 馬氏法

馬氏法分析結果為表 5，由表中均方根標準差指標可看出在集群數為 10 或 13 時均有較大的增幅，因此較難給予確切的判定。至於半淨 R2 指標則明顯在資料分為 3 群時有突然上升的情形，數值由 0.0617 升到 0.219，所以此指標顯示分為 4 群時即應停止合併。另外，R2 指標亦同樣於集群數為 3 時有突然改變的現象，R2 由 0.553 降到 0.334，所以同樣指出將資料分為 4 群是合理的。3 個指標之趨勢分布情形如圖 5 所示。

馬氏法分析結果的 4 個族群分別是第 1、11 觀測值為一組，第 4、5、8、9、13、16、17 為一組，第 2、3、6、7、12、14、15、18、19 為一組，最後剩下的第 10 觀測值則自成一組。

與前三種方法相比較，馬氏法建議分為 4 群，此結果較單一連結法精簡，較彈性 β 法則不夠精簡，而與完全連結法相同，但兩者

間分群結構差異甚大，馬氏法將資料主要分為兩大族群，完全連結法則分群較為平均。

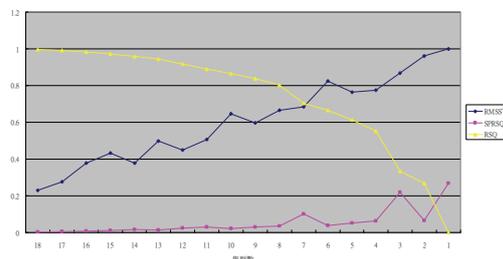


圖 5 馬氏法分析結果陡坡圖

Figure. 5 Scree plot of McQuitty's similarity analysis

(五) 中數法

由表 6 數據可知，均方根標準差變化起伏並不明顯，在分群數為 2、7、10、16 時均有相似而相對較大的起伏，因此不易判斷正確分群數。至於半淨 R2 指標則明顯在分為 8 群時有顯著的增加，由 0.0232 升至 0.1006，所以可判定資料分為 9 群時即應停止合併動作。而 R2 指標之明顯變化處同樣為分為 8 群

時，數值由 0.829 降到 0.728，所以 R2 指標同樣指出 9 群為合理的分群數。3 個指標之趨勢分布情形可繪成圖 6 的陡坡圖。

中數法分析結果建議將資料分為 9 個群體，包括第 2、3、6、7、14、15、18 觀測值為一組，第 5、8、9、13 為一組，第 12、19 為一組，最後剩下的第 1、4、10、11、16、17 觀測值則自成一組。

與前面其他分群法相比較，中數法分析結果為 9 群，分群數較完全連結法、彈性 β 法與馬氏法等 3 個方法分析結論還要多，卻與單一連結法分群數相同，且兩者間之分群結構也相當類似，差別僅在於對第 12、17、19 等 3 個觀測值的處理。

(六) 平均連結法

平均連結法分群結果如表 7 所示，由表中第四欄可看出均方根標準差指標在集群數等於 10 與 13 時，其計算結果均明顯上升，

表示這些分群數都是可能合理的結果，因此無法做出確切的判定。而由第五欄半淨 R2 值分析結果可看出在集群數分為 3 群後，其值有明顯增加的現象，由 0.1044 升至 0.2261，亦即此一指標指出應將數據分為 4 群。至於第六欄的 R2 值，當集群數分為 3 群後，其數值有明顯減少的現象，由 0.56 減至 0.334，因此亦指出將資料分為 4 群是合理的結果。圖 7 為 3 個指標的陡坡圖。

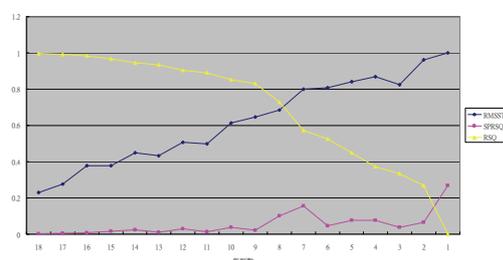


圖 6 中數法分析結果陡坡圖

Figure. 6 Scree plot of median hierarchical cluster analysis

表中第九欄是偽 F 值 (pseudo F)，代表集群間與集群內變異均方的比例，其數值之區域

表 5 馬氏法分析結果

Table 5 Results of McQuitty's similarity analysis

The data have been standardized to mean 0 and variance 1								
Root-Mean-Square Total-Sample Standard Deviation = 1								
Mean Distance Between Observations = 4.674706								
NCL	Cluster Joined		FREQ	RMSSTD	SPRSQ	RSQ	CCC	Norm McQ
18	OB14	OB18	2	0.2296	0.0029	0.997	-	0.2406
17	CL18	OB15	3	0.2759	0.0055	0.992	-	0.3096
16	OB6	OB7	2	0.3776	0.0079	0.984	-	0.3957
15	OB5	OB13	2	0.4324	0.0104	0.973	-	0.4531
14	OB3	CL17	4	0.3783	0.0154	0.958	-	0.4579
13	OB8	OB9	2	0.4998	0.0139	0.944	-	0.5238
12	CL14	CL16	6	0.4500	0.0245	0.919	-	0.5330
11	OB2	CL12	7	0.5055	0.0289	0.891	-	0.5964
10	OB12	OB19	2	0.6466	0.0232	0.867	-	0.6777
9	CL15	OB17	3	0.5959	0.0291	0.838	-	0.6803
8	OB4	CL9	4	0.6653	0.0343	0.804	-	0.7661
7	CL11	CL10	9	0.6857	0.1006	0.703	-	0.8376
6	OB1	OB11	2	0.8240	0.0377	0.666	-	0.8635
5	CL13	OB16	3	0.7631	0.0508	0.615	-	0.9015
4	CL8	CL5	7	0.7750	0.0617	0.553	-	0.9188
3	CL7	CL4	16	0.8682	0.2190	0.334	0.90	1.0660
2	CL6	OB10	3	0.9625	0.0652	0.269	2.85	1.0739
1	CL2	CL3	19	1.0000	0.2689	0.000	0.00	1.3857

表 6 中數法分析結果

Table 6 Result of median hierarchical cluster analysis

The data have been standardized to mean 0 and variance 1								
Root-Mean-Square Total-Sample Standard Deviation = 1								
Root-Mean-Square Distance Between Observations = 4.898979								
NCL	Cluster Joined		FREQ	RMSSTD	SPRSQ	RSQ	CCC	Norm Median Dist
18	OB14	OB18	2	0.2296	0.0029	0.997	-	0.2296
17	CL18	OB15	3	0.2759	0.0055	0.992	-	0.2733
16	OB6	OB7	2	0.3776	0.0079	0.984	-	0.3776
15	OB3	CL17	4	0.3783	0.0154	0.968	-	0.4110
14	CL15	CL16	6	0.4500	0.0245	0.944	-	0.4116
13	OB5	OB13	2	0.4324	0.0104	0.933	-	0.4324
12	OB2	CL14	7	0.5055	0.0289	0.904	-	0.4958
11	OB8	OB9	2	0.4998	0.0139	0.891	-	0.4998
10	CL13	CL11	4	0.6132	0.0384	0.852	-	0.5879
9	OB12	OB19	2	0.6466	0.0232	0.829	-	0.6466
8	CL12	CL9	9	0.6857	0.1006	0.728	-	0.6632
7	CL8	CL10	13	0.8001	0.1551	0.573	-	0.6767
6	CL7	OB17	14	0.8095	0.0465	0.527	-	0.7179
5	CL6	OB16	15	0.8419	0.0779	0.449	-	0.7005
4	CL5	OB4	16	0.8682	0.0769	0.372	-	0.7825
3	OB1	OB11	2	0.8240	0.0377	0.334	0.90	0.8240
2	CL3	OB10	3	0.9625	0.0652	0.269	2.85	0.9383
1	CL2	CL4	19	1.0000	0.2689	0.000	0.00	0.9663

最高點即可能為合理集群數之結果。由表中可看出偽 F 值在分為 5 群時有區域最高點，但並不是很明顯，亦即此一指標判定將資料分為 5 群是合理的結果。另外，表中第十欄代表偽 t2 值 (pseudo t2)，當其值有突然上升的情形時，表示該點數據為合理的集群數。由表中數據可發現當資料分為 3 群後，數值由 3.4 升至 7.9，這代表此一指標顯示 4 群為合理的結果。

因此，平均連結法分析結果指出應將原始數據分為 4 群，第一群是將觀測值第 2、3、6、7、14、15、18 納入，第二群為第 4、5、8、9、12、13、16、17、19 觀測值，第三群包括第 1 與 11 觀測值，最後剩下的第 10 觀測值自成一群。

與前述五種方法相比較，可知單一連結法與中數法分為 9 群，完全連結法與馬氏法分為 4 群，而彈性 β 法為 3 群，由此可知，平均連結法分為 4 群的結果與完全連結法和馬氏法相近，表示三者分群效果相似。而其

中平均連結法與馬氏法分類結果最為相近，兩者間的差異僅在於對第 12、19 兩個觀測值的分類方式。

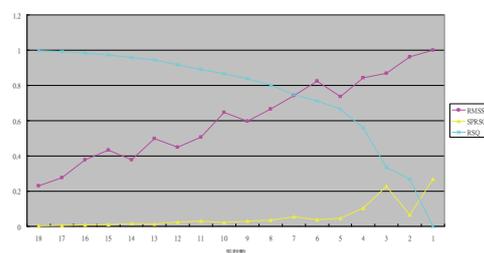


圖 7 平均連結法分析結果陡坡圖

Figure. 7 Scree plot of average linkage cluster analysis

(七) 形心連結法

形心連結法分析結果如表 8 所示，由表中均方根標準差的數值分布情形可看出，在集群數分為 8、10、16 後，數值均有較明顯上升的情形，因此無法確定真正合理的分群數。若以半淨 R2 值指標來分析，結果顯示分群數為 7 以後，數值明顯上升，由 0.0725 增至 0.1832，由此判斷當資料分為 8 群時有較合理的結果。

表 7 平均連結法分析結果

Table 7 Results of average linkage cluster analysis

The data have been standardized to mean 0 and variance 1										
Root-Mean-Square Total-Sample Standard Deviation =1										
Root-Mean-Square Distance Between Observations = 4.898979										
NCL	Cluster Joined		FREQ	RMSSTD	SPRSQ	RSQ	CCC	PSF	PST2	Norm RMS Dist
18	OB14	OB18	2	0.2296	0.0029	0.997	-	20.0	-	0.2296
17	CL18	OB15	3	0.2759	0.0055	0.992	-	14.7	1.9	0.2964
16	OB6	OB7	2	0.3776	0.0079	0.984	-	12.0	-	0.3776
15	OB5	OB13	2	0.4324	0.0104	0.973	-	10.4	-	0.4324
14	OB3	CL17	4	0.3783	0.0154	0.958	-	8.7	3.6	0.4583
13	OB8	OB9	2	0.4998	0.0139	0.944	-	8.4	-	0.4998
12	CL14	CL16	6	0.4500	0.0245	0.919	-	7.3	3.1	0.5045
11	OB2	CL12	7	0.5055	0.0289	0.891	-	6.5	2.6	0.6229
10	OB12	OB19	2	0.6466	0.0232	0.867	-	6.5	-	0.6466
9	CL15	OB17	3	0.5959	0.0291	0.838	-	6.5	2.8	0.6627
8	OB4	CL9	4	0.6653	0.0343	0.804	-	6.4	1.7	0.7282
7	CL13	CL10	4	0.7437	0.0551	0.749	-	6.0	3.0	0.8141
6	OB1	OB11	2	0.8240	0.0377	0.711	-	6.4	-	0.8240
5	CL8	OB16	5	0.7358	0.0465	0.665	-	6.9	1.9	0.8304
4	CL5	CL7	9	0.8445	0.1044	0.560	-	6.4	3.4	0.9203
3	CL11	CL4	16	0.8682	0.2261	0.334	0.90	4.0	7.9	0.9712
2	CL6	OB10	3	0.9625	0.0652	0.269	2.85	6.3	1.7	1.0248
1	CL2	CL3	19	1.0000	0.2689	0.000	0.00	-	6.3	1.2728

再者，若以 R2 指標來判斷，亦顯示集群數分為 8 群會有較合理的結果。上述 3 個指標的陡坡圖如圖 8 所示。

此外，偽 F 值分析結果發現分群數為 12 時有區域最高點 17.3，故此一指標判別合理的分群數為 12。當使用偽 t2 值來進行分析時，結果發現資料分為 7 群後，其值由 3.4 增至 8.3，為數值增加幅度最大的點，代表分成 8 集群是最合理的結果。

形心連結法分析結果顯示將資料分為 8 群可能是較合理的結果，其中，第一群為觀測值 2、3、6、7、14、15、18，第 2 群是 5、8、9、12、13、19 觀測值，而剩下的 1、4、10、11、16、17 等 6 個觀測值則自成一類。

與前 6 種群集分析法相比較，此法將資料分為 8 群，與前述方法均不同，僅與單一連結法及中數法之結果較為近似，但較兩者分群結果更為精簡。至於分群內容，與中數法幾乎完全相同，差別僅在第 12 與 19 觀測值的分群，而與單一連結法則相差較多。

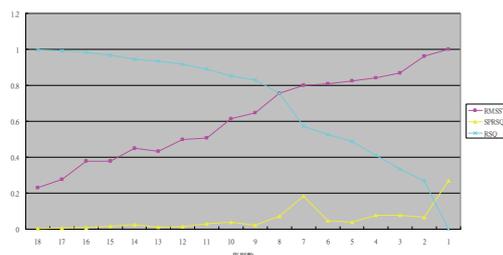


圖 8 形心連結法分析結果陡坡圖

Figure. 8 Scree plot of centroid hierarchical cluster analysis

(八) 華德法

華德法分析結果如表 9 所示，從表中均方根標準差指標可知，集群數分為 7 群後之數值有大幅增加的情況，由 0.5055 升到 0.824，代表合理的分群數為 8。若以半淨 R2 指標值來看，集群數分為 2 後就有數值明顯上升的情形發生，由 0.1044 增至 0.1951，意味著應將資料分為 3 群為最合理的結果。另外，以 R2 指標值來分析時，可發現在分為 2 群後，數值由 0.495 降至 0.3，表示資料分群在分為 3 群後即應停止，和半淨 R2 指標有相

同的結果。圖 9 為以上 3 個指標的陡坡圖。

此外，若以偽 F 值來做分析時，數值之局部最高點發生在資料分為 3 群時，但數值轉折變化並不够明顯，因此只能判斷最後分群結果在 3 時有相對最大的組間變異。至於偽 t2 數值之變化，可發現當分群步驟到達 8 群後，其值由 1.7 上升至 3.6，此時有較大的增幅，代表分群應在分為 9 群後停止。綜合以上 5 個指標的判定，決定分群數 3 是最後的結果。

因此，以華德法進行分群時，得到應該將資料分為 3 群的結果。第一群為觀測值 2、3、6、7、14、15、18，第二群為 4、5、8、9、12、13、16、17、19，第 3 群則包括 1、10、11 等 3 個觀測值。

華德法分群結果與彈性 β 法一致，都指出將原始數據分為 3 群是最合理的結果，且分群內容亦完全相同。

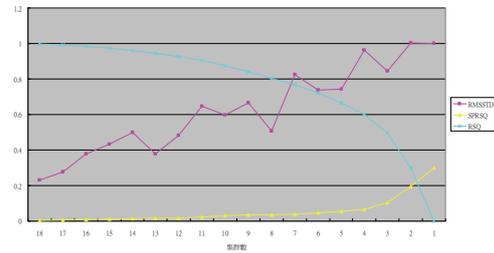


圖 9 華德連結法分析結果陡坡圖

Figure. 9 Scree plot of Ward's minimum variance cluster analysis

(九) ML 法

ML 法分析結果如表 10，表中均方根標準差指標顯示資料合併至 4 群時，其值有明顯上升跡象，數值由 0.7437 升到 0.9625，代表此一指標判定 5 群是合理的結果。若以半淨 R2 指標的變化來看，當分群數達到剩下 2 群時，指標值有突然上升的現象，由 0.1044 增至 0.1951，所以此一指標判別分群動作應於分成 3 群後停止。

表 8 形心連結法分析結果

Table 8 Results of centroid hierarchical cluster analysis

The data have been standardized to mean 0 and variance 1										
Root-Mean-Square Total-Sample Standard Deviation = 1										
Root-Mean-Square Distance Between Observations = 4.898979										
NCL	Cluster Joined		FREQ	RMSSTD	SPRSQ	RSQ	CCC	PSF	PST2	Norm Cent Dist
18	OB14	OB18	2	0.2296	0.0029	0.997	-	20.0	-	0.2296
17	CL18	OB15	3	0.2759	0.0055	0.992	-	14.7	1.9	0.2733
16	OB6	OB7	2	0.3776	0.0079	0.984	-	12.0	-	0.3776
15	OB3	CL17	4	0.3783	0.0154	0.968	-	8.7	3.6	0.4297
14	CL15	CL16	6	0.4500	0.0245	0.944	-	6.5	3.1	0.4065
13	OB5	OB13	2	0.4324	0.0104	0.933	-	7.0	-	0.4324
12	OB8	OB9	2	0.4998	0.0139	0.919	-	17.3	-	0.4998
11	OB2	CL14	7	0.5055	0.0289	0.891	-	6.5	2.6	0.5510
10	CL13	CL12	4	0.6132	0.0384	0.852	-	5.8	3.2	0.5879
9	OB12	OB19	2	0.6466	0.0232	0.829	-	6.1	-	0.6466
8	CL10	CL9	6	0.7550	0.0725	0.756	-	4.9	3.4	0.6993
7	CL11	CL8	13	0.8001	0.1832	0.573	-	2.7	8.3	0.7144
6	CL7	OB17	14	0.8095	0.0465	0.527	-	2.9	1.3	0.6716
5	OB1	OB11	2	0.8240	0.0377	0.489	-	3.3	-	0.8240
4	CL6	OB16	15	0.8419	0.0779	0.411	-	3.5	2.1	0.8669
3	CL4	OB4	16	0.8682	0.0769	0.334	0.90	4.0	2.0	0.8595
2	CL5	OB10	3	0.9625	0.0652	0.269	2.85	6.3	1.7	0.9383
1	CL2	CL3	19	1.0000	0.2689	0.000	0.00	-	6.3	0.9787

表 9 華德連結法分析結果

Table 9 Results of Ward's minimum variance cluster analysis

The data have been standardized to mean 0 and variance 1									
Root-Mean-Square Total-Sample Standard Deviation = 1									
Root-Mean-Square Distance Between Observations = 4.898979									
NCL	Cluster Joined		FREQ	RMSSTD	SPRSQ	RSQ	CCC	PSF	PST2
18	OB14	OB18	2	0.2296	0.0029	0.997	-	20.0	-
17	CL18	OB15	3	0.2759	0.0055	0.992	-	14.7	1.9
16	OB6	OB7	2	0.3776	0.0079	0.984	-	12.0	-
15	OB5	OB13	2	0.4324	0.0104	0.973	-	10.4	-
14	OB8	OB9	2	0.4998	0.0139	0.959	-	9.1	-
13	OB3	CL17	4	0.3783	0.0154	0.944	-	8.4	3.6
12	OB2	CL16	3	0.4810	0.0178	0.926	-	8.0	2.2
11	OB12	OB19	2	0.6466	0.0232	0.903	-	7.4	-
10	CL15	OB17	3	0.5959	0.0291	0.874	-	6.9	2.8
9	OB4	CL10	4	0.6653	0.0343	0.840	-	6.5	1.7
8	CL12	CL13	7	0.5055	0.0356	0.804	-	6.4	3.6
7	OB1	OB11	2	0.8240	0.0377	0.766	-	6.6	-
6	CL9	OB16	5	0.7358	0.0465	0.720	-	6.7	1.9
5	CL14	CL11	4	0.7437	0.0551	0.665	-	6.9	3.0
4	CL7	OB10	3	0.9625	0.0652	0.599	-	7.5	1.7
3	CL6	CL5	9	0.8445	0.1044	0.495	5.98	7.8	3.4
2	CL4	CL3	12	1.0032	0.1951	0.300	3.90	7.3	4.6
1	CL2	CL8	19	1.0000	0.2999	0.000	0.00	-	7.3

再以 R2 指標來看，可發現在分群數為 2 時，指標值突然大幅下降，由 0.495 掉到 0.3，所以 R2 指標判斷分群數結果與半淨 R2 指標一致，均以 3 群為最佳分群結果。上述 3 個指標的陡坡圖如圖 10 所示。

ML 法建議將原始資料分為 3 群，其分群結構為 2、3、6、7、14、15、18 觀測值為一群，4、5、8、9、12、13、16、17、19 為一群，最後剩下的 1、10、11 為一群。

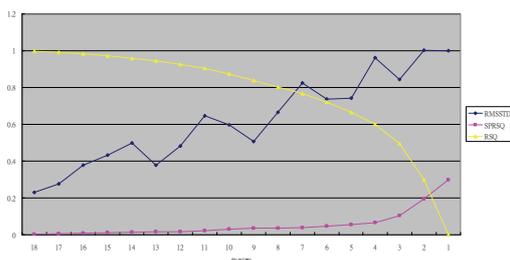


圖 10 ML 法分析結果陡坡圖

Figure. 10 Scree plot of maximum likelihood method

將 ML 法分群結果與前面 8 種方法相比

較，可發現建議分為 3 群的結果與彈性 β 法和華德法相同，且在群組的結構方面也完全相同，故此 3 種分群法可歸納為相似的方法，有相似的分群結果。

(十) 密度連結法

表 11 為密度連結法分析結果，由表中均方根標準差指標顯示資料合併成 16 與 10 群時，有相同而相對較大的增幅，但變化差距小，因而不易判斷合理的分群數。至於半淨 R2 指標，可觀察到在資料分群數為 8 時，指標值有明顯增加的情形，由 0.0563 上升至 0.1323，此時組內變異有較大的增幅，因此應停止合併，故以 9 個分群數較為合理。最後再使用 R2 指標來判斷，發現指標值在分群數為 8 時有明顯下降的情況，從 0.771 降到 0.638，所以此一指標與半淨 R2 同樣顯示應以 9 為合理分群數。上述 3 個指標若繪成陡坡圖，則如圖 11 所示。

密度連結法顯示資料應分為 9 群，亦即

第 2、3、6、7、12、14、15、17、18 觀測值為一群，5、9、13 為一群，剩下的 1、4、8、10、11、16、19 等 7 個觀測值則自成一群。

密度連結法建議分為 9 群之結論，若與前面 9 種分群法相比較，可知與單一連結法

和中數法相同，但族群結構則大不相同。其中，密度連結法較為特別，其組成有相當多的單一觀測值連結，而另外兩種方法則是有較大的部分先將觀測值做一初步的連結，然後再予以合併各個小族群。

表 10 ML 法分析結果

Table 10 Results of maximum likelihood method

The data have been standardized to mean 0 and variance 1									
Root-Mean-Square Total-Sample Standard Deviation = 1									
Root-Mean-Square Distance Between Observations = 4.898979									
NCL	Cluster Joined		FREQ	RMSSTD	SPRSQ	RSQ	CCC	LNLR	LNLIKE
18	OB14	OB18	2	0.2296	0.0029	0.997	-	.	673.8
17	CL18	OB15	3	0.2759	0.0055	0.992	-	238.1	435.7
16	OB6	OB7	2	0.3776	0.0079	0.984	-	147.9	287.8
15	OB5	OB13	2	0.4324	0.0104	0.973	-	109.2	178.6
14	OB8	OB9	2	0.4998	0.0139	0.959	-	92.46	86.17
13	OB3	CL17	4	0.3783	0.0154	0.944	-	68.70	17.47
12	OB2	CL16	3	0.4810	0.0178	0.926	-	59.04	-41.57
11	OB12	OB19	2	0.6466	0.0232	0.903	-	59.61	-101.2
10	CL15	OB17	3	0.5959	0.0291	0.874	-	55.92	-157.1
9	CL12	CL13	7	0.5055	0.0356	0.838	-	47.14	-204.2
8	OB4	CL10	4	0.6653	0.0343	0.804	-	39.38	-243.6
7	OB1	OB11	2	0.8240	0.0377	0.766	-	37.34	-281.0
6	CL8	OB16	5	0.7358	0.0465	0.720	-	36.39	-317.4
5	CL14	CL11	4	0.7437	0.0551	0.665	-	35.36	-352.7
4	CL7	OB10	3	0.9625	0.0652	0.599	-	36.70	-389.4
3	CL6	CL5	9	0.8445	0.1044	0.495	5.98	40.46	-429.9
2	CL4	CL3	12	1.0032	0.1951	0.300	3.90	60.97	-490.8
1	CL2	CL9	19	1.0000	0.2999	0.000	0.00	56.27	-547.1

表 11 密度連結法分析結果

Table 11 Results of density linkage cluster analysis

K = 3											
Root-Mean-Square Total-Sample Standard Deviation = 1											
NCL	Cluster Joined		FREQ	RMSSTD	SPRSQ	RSQ	CCC	Normalized Fusion Density	Maximum in Each Lesser	Density Cluster Greater	Tie
18	OB14	OB15	2	0.2721	0.0041	0.996	-	25.959	14.9153	100.0	T
17	CL18	OB18	3	0.2759	0.0043	0.992	-	25.959	14.9153	100.0	-
16	CL17	OB3	4	0.3783	0.0154	0.976	-	0.341	0.1708	100.0	-
15	CL16	OB7	5	0.4476	0.0207	0.955	-	0.1398	0.0702	100.0	-
14	CL15	OB6	6	0.4500	0.0117	0.944	-	0.1230	0.4969	100.0	-
13	OB9	OB13	2	0.4969	0.0137	0.930	-	0.0702	0.0678	0.0727	-
12	CL13	OB5	3	0.5114	0.0153	0.915	-	0.0565	0.0462	0.0727	-
11	CL14	OB2	7	0.5055	0.0289	0.886	-	0.0356	0.0184	100.0	-
10	CL11	OB12	8	0.6086	0.0589	0.827	-	0.0061	0.00308	100.0	-
9	CL10	OB17	9	0.6714	0.0563	0.771	-	0.0038	0.00192	100.0	-
8	CL9	CL12	12	0.7693	0.1323	0.638	-	0.0037*	0.0727	100.0	-
7	CL8	OB19	13	0.7953	0.0600	0.578	-	0.0025	0.00132	100.0	-
6	CL7	OB4	14	0.8341	0.0808	0.498	-	0.0012	0.000603	100.0	-
5	CL6	OB16	15	0.8628	0.0765	0.421	-	0.0008	0.000396	100.0	-
4	CL5	OB8	16	0.8682	0.0492	0.372	-	0.0007	0.000370	100.0	-
3	CL4	OB1	17	0.9204	0.1248	0.247	-1.4	0.0003	0.000168	100.0	-
2	CL3	OB11	18	0.9473	0.0944	0.153	-0.74	0.002	0.000129	100.0	-
1	CL2	OB10	19	1.0000	0.1525	0.000	0.00	265E-7	0.000013	100.0	-

* indicates fusion of two modal or multimodal clusters 3 modal clusters have been formed.

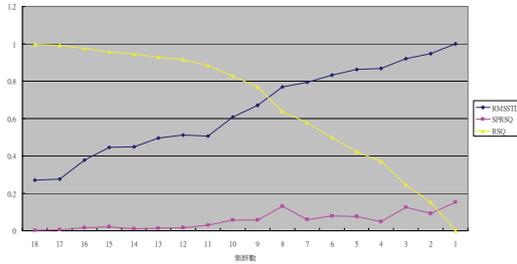


圖 11 密度連結法分析結果陡坡圖

Figure. 11 Scree plot of density linkage cluster analysis

(十一) 雙連法

雙連法群集分析結果如表 12，在均方根標準差指標方面，在資料分為 10 與 16 個群體時均有相對較大的增幅，但兩者幅度相當接近，因而無法明顯判斷哪一個結果最為正確。接著在使用半淨 R2 指標來判斷時，發現當全部資料合併為 1 個群體時，指標值突然大增，由 0.0972 升至 0.3403，由此可判斷雙連法應將資料分為兩個族群。至於 R2 指標則與半淨 R2 有相同的結果，亦即同樣在剩下 1

個族群時有最大的改變，由 0.34 降至 0。綜合以上 3 種指標之判斷，雙連法建議合併為 2 個群體為合理。圖 12 則為 3 種指標的陡坡圖。

雙連法中 2 大族群的第一群是由 2、3、6、7、12、14、15、17、18、19 觀測值所組成，第二群則包括 1、4、5、8、9、10、11、13、16。

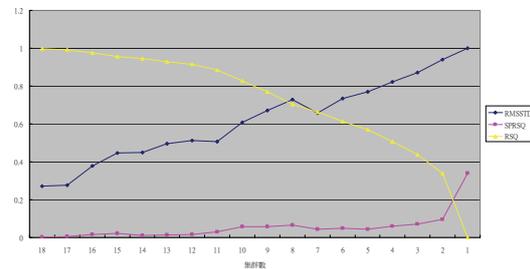


圖 12 雙連法分析結果陡坡圖

Figure. 12 Scree plot of two-stage density linkage clustering

將雙連法分析結果與前述 10 種群集分析法來比較，可發現雙連法是分群數最少的方法，亦即建議將原始資料分為 2 群，

表 12 雙連法分析結果

Table 12 Results of two-stage density linkage clustering

K = 3											
Root-Mean-Square Total-Sample Standard Deviation = 1											
NCL	Cluster Joined		FREQ	RMSSTD	SPRSQ	RSQ	CCC	Normalized Fusion Density	Maximum in Each Lesser	Density Cluster Greater	Tie
18	OB14	OB15	2	0.2721	0.0041	0.996	-	25.959	14.9153	100.0	T
17	CL18	OB18	3	0.2759	0.0043	0.992	-	25.959	14.9153	100.0	-
16	CL17	OB3	4	0.3783	0.0154	0.976	-	0.341	0.1708	100.0	-
15	CL16	OB7	5	0.4476	0.0207	0.955	-	0.1398	0.0702	100.0	-
14	CL15	OB6	6	0.4500	0.0117	0.944	-	0.123	0.4969	100.0	-
13	OB9	OB13	2	0.4969	0.0137	0.930	-	0.0702	0.0678	0.0727	-
12	CL13	OB5	3	0.5114	0.0153	0.915	-	0.0565	0.0462	0.0727	-
11	CL14	OB2	7	0.5055	0.0289	0.886	-	0.0356	0.0184	100.0	-
10	CL11	OB12	8	0.6086	0.0589	0.827	-	0.0061	0.00308	100.0	-
9	CL10	OB17	9	0.6714	0.0563	0.771	-	0.0038	0.00192	100.0	-
8	CL9	OB19	10	0.7299	0.0661	0.705	-	0.0025	0.00132	100.0	-
7	CL12	OB4	4	0.6562	0.0427	0.662	-	0.0012	0.000603	0.0727	-
6	CL7	OB16	5	0.7350	0.0483	0.614	-	0.0008	0.000396	0.0727	-
5	CL6	OB8	6	0.7693	0.0444	0.569	-	0.0007	0.000370	0.0727	-
4	CL5	OB1	7	0.8225	0.0611	0.508	-	0.0003	0.000168	0.0727	-
3	CL4	OB11	8	0.8726	0.0706	0.438	4.00	0.0002	0.000129	0.0727	-
2	CL3	OB10	9	0.9407	0.0972	0.340	5.34	265E-7	0.000013	0.0727	-
1	CL8	CL2	19	1.0000	0.3403	0.000	0.00	0.0037	0.0727	100.0	-

若與分析結果最相似的彈性 β 法、華德法與 ML 法相比較，可知分群數雖然僅差 1，但群體組成卻大不相同。相反的，雙連法的組成反而是與密度連結法 (9 群) 的分群結構極為相似，表示此二者之分群方式類似。

(十二) 群集法比較

將本研究 11 種階層式群集分析法之結果摘要如表 13，由表中可知雙連法建議分為 2 群，彈性 β 法、華德法、ML 法為 3 群，完全連結法、馬氏法、平均連結法為 4 群，形心連結法為 8 群，單一連結法、中數法、密度連結法為 9 群，共有 5 種結論。其中，除了 2 與 8 群僅一種分析法支持外，其他均有 3 種分析法支持。

表 14 是使用離群值修正前之資料進行群集分析的結果，若與表 13 相比較，可發現除了中數法與形心連結法外，去除離群值後再進行集群分析 (表 13) 時，會得到較為精簡的分群數，有助於指標適用性與動力機制的討論。

在群集的組成結構方面，由表 13 中可看出 11 種分群法中均有相同的分群趨勢，亦即將第 2、3、6、7、14、15、18 等 7 個觀測值分為一群，這也是各分析法之主要群體組成，表示這 7 個觀測值之性質最為接近，因此均分在同一個群體內。另外，若以群體組成的樹狀圖 (圖 13) 來比較，大致上亦可分為兩大類，一類為將各個性質相近的觀測值先予以初步合併，接著再一一合併性質較為接近的小群體，本研究所採用的階層式群集分析法中，除了密度連結法與雙連法外，大部分均以此方式做合併動作。至於密度連結法與雙連法，則是先將性質最為接近的觀測

值合併之後，再開始與性質較相近的個別單一觀測值進行合併，因此其組成中較無小群體的成分存在。

一般的研究常用平均連結法 [Heessen, 1996] 或是華德法作為最佳群集分析之結果，但為了判斷哪一個群集分析法之結果最適合應用於本研究之優養化資料的分群，必須以分群的結果配合原始資料來判斷，以 TSI (Trophic State Index) 來作為判斷的憑據，看哪一個分析結果最符合 TSI 的分群。首先，先去除分群過多的分析法，因為過多的分群，尤其是單一觀測值所構成的群體是較不符合實際情形的結果，故我們先去除單一連結法、中數法、形心連結法與密度連結法，只取分為 4 個群體以下的結果。剩下的方法中，彈性 β 法、華德法、ML 法之結果一致，均分為 3 個群體，分群的內容也相同，但其分群在高度優養情形下無法明顯分群，因此本研究不採用該分群結果。至於完全連結法、平均連結法、馬氏法雖然都將資料分做 4 群，但其分群內容均不相同，在對照原始資料並進行差異顯著性檢定後，發現以馬氏法所分的群組間差異性最大，所以馬氏法分析結果最符合本研究之需求。再者，雙連法將所有觀測值分為兩群，且兩群亦有顯著差異。因此，在本研究中以經離群值修正的觀測值進行集群分析時，是以馬氏法與雙連法分析結果較適合進行優養化的分群，亦即建議將該資料分為 2 或 4 群。

而以未經離群值修正的觀測值進行集群分析後，亦將其分析結果與原始資料比對以進行判斷，首先同樣先刪除分群效果較差的結果，包括單一連結法、中數法與密度連結法，只留下分群 6 組以下的結果。剩下的方

法中，完全連結法與形心連結法雖然均將資料分為 6 群，但其分群之間並不顯著，故不予採用。至於馬氏法與平均連結法，其分析結果均為 5 群，然而分群之間同樣多無顯著差異，故未符合本研究的要求。再者，彈性 β 法、華德法、ML 法之結果一致，均分為 3 群，其中，華德法和 ML 法之分群內容相同，而彈性 β 法則不同；在分群間差異性比較方面，彈性 β 法之分群結果雖較為平均，但群間均未達顯著差異，而華德法、ML 法之分群則是在中高度優養情形下分群效果顯著，然而在高度優養情形下則無法顯著分群。另外，雙連法是將觀測值分為兩群，且兩群體間有顯著差異。總結這部份分析結果，建議應分為 2 或 3 群，亦即在本研究中是以華德法、ML 法與雙連法分群效果較佳。

因此，將離群值修正前後之資料進行集群分析後之結果來比較，發現離群值修正前，分析結果建議分為 2 或 3 群，而離群值修正後則變為 2 或 4 群。兩者主要差別在於高度優養化的情形，亦即經離群值修正後，

群集分析結果建議在高度優養下必須予以再分群。在分群內容方面，經過比對後發現相同的分群法在不考慮離群值的情形下，其分群內容均相同，但在分群結果的差異顯著性上，仍以經離群分析後的資料所做的群集分析結果較顯著。

二、非階層式群集分析

經過階層分析後，接著再進行非階層分析，以檢驗階層分析的結果。因非階層分析必須先自訂分群數，故本研究採用階層分析的結果，亦即 2 與 4 群為預定群數進行分析。至於集群中心點的設定方面，設定標準距離 Radius = 0，中心點取代法為依據集群數與標準距離來決定，而重複計算集群中心點次數 = 1，且所有集群的中心點均予以保留。

在經過兩組分析結果比較後，發現當集群數為 4 時，其整體解釋力較強，平均 $R^2 = 0.5531$ ，大於集群數為 2 群時的 R^2 值 (= 0.2804)，因此本研究將就集群數為 4 時的分析結果加以詳細探討。

表 13 各群集分析法之結果比較
Table 13 Comparisons of all cluster analyses

集群分析法	集群數	集群組成				
		第1群	第2群	第3群	第4群	其他族群
單一連結法	9	2.3.6.7.14.15.18	5.8.9.13.17	1	10	4.11.12.16.19
完全連結法	4	2.3.6.7.14.15.18	8.9.12.19	4.5.13.16.17	1.10.11	-
彈性 β 法	3	2.3.6.7.14.15.18	4.5.8.9.12.13.16.17.19	1.10.11	-	-
馬氏法	4	2.3.6.7.12.14.15.18.19	4.5.8.9.13.16.17	1.11	10	-
中數法	9	2.3.6.7.14.15.18	5.8.9.13	12.19	1	4.10.11.16.17
平均連結法	4	2.3.6.7.14.15.18	4.5.8.9.12.13.16.17.19	1.11	10	-
形心連結法	8	2.3.6.7.14.15.18	5.8.9.12.13.19	1	10	4.11.16.17
華德法	3	2.3.6.7.14.15.18	4.5.8.9.12.13.16.17.19	1.10.11	-	-
ML法	3	2.3.6.7.14.15.18	4.5.8.9.12.13.16.17.19	1.10.11	-	-
密度連結法	9	2.3.6.7.12.14.15.17.18	5.9.13	1	10	4.8.11.16.19
雙連法	2	2.3.6.7.12.14.15.17.18.19	1.4.5.8.9.10.11.13.16	-	-	-
共同組成		2.3.6.7.14.15.18	9	-	-	-

表 14 離群值修正前各群集分析法之結果比較
Table 14 Comparisons of all cluster analyses before calibrating outlier

集群分析法	集群數	集群組成				
		第1群	第2群	第3群	第4群	其他族群
單一連結法	10	2.3.4.7.8.9.17.18.21.22	6.10.11.16.20	1	13	5.12.14.15.19.23
完全連結法	6	2.3.4.7.8.9.17.18.21.22	5.6.10.11.16.19.20	1.13	14.23	12.15
彈性β法	3	2.3.4.5.6.7.8.9.10.11.15.16.17.18.19.20.21.22	1.12.13	14.23	-	-
馬氏法	5	2.3.4.7.8.9.14.17.18.21.23	5.6.10.11.16.19.20.22	1.13	12	15
中數法	8	2.3.4.7.8.9.17.18.21	6.10.11.16.19.20.22	14.23	1	5.12.13.15
平均連結法	5	2.3.4.7.8.9.14.17.18.21.22.23	5.6.10.11.16.19.20	1.13	12	15
形心連結法	6	2.3.4.7.8.9.17.18.21.22	5.6.10.11.14.16.19.20.23	1	13	12.15
華德法	3	2.3.4.7.8.9.17.18.21.22	5.6.10.11.14.15.16.19.20.23	1.12.13	-	-
ML法	3	2.3.4.7.8.9.17.18.21.22	5.6.10.11.14.15.16.19.20.23	1.12.13	-	-
密度連結法	9	2.3.4.7.8.9.14.17.18.20.21.22	6.10.11.16	1	13	5.12.15.19.23
雙連法	2	2.3.4.7.8.9.14.17.18.20.21.22.23	1.5.6.10.11.12.13.15.16.19	-	-	-
共同組成		2.3.4.7.8.9.17.18.21	-	-	-	-

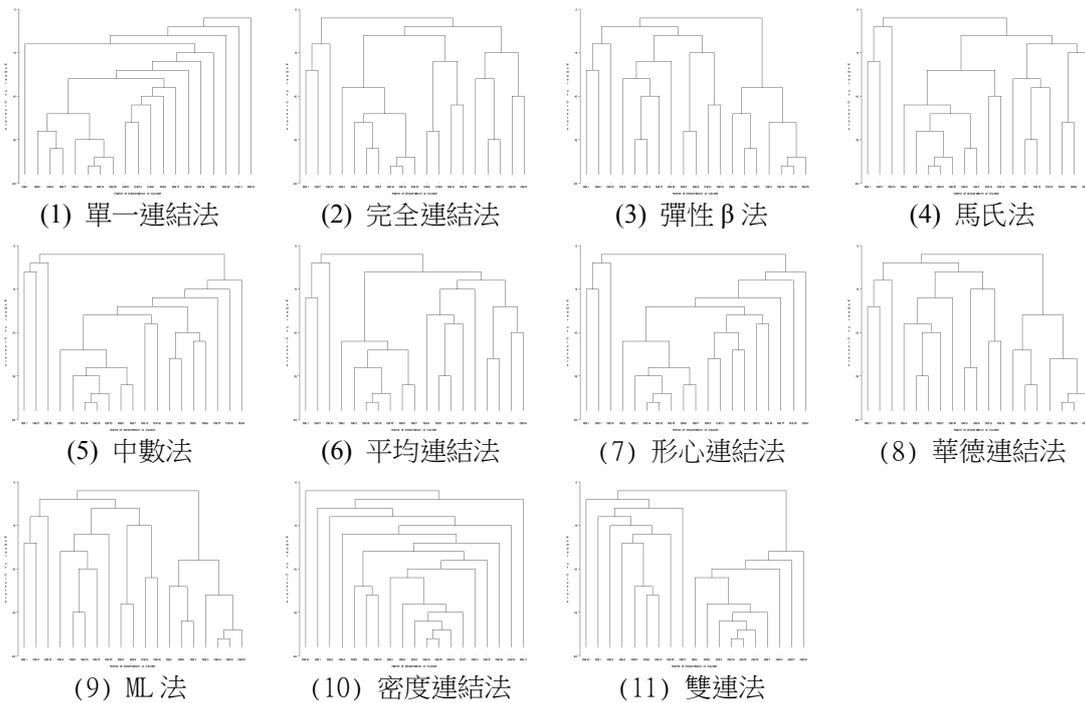


圖 13 所有群集分析法分析結果樹型圖

Figure. 13 Dendrogram of all cluster analyses

首先，各觀測值分群結果與種子點的距離如表 15，由表中可知各單一觀測值的分群情形。至於分群結果之摘要則如表 16，由表中數據可看出分到第 1 群者共有 2 個觀測值、第 2 群 9 個、第 3 群 1 個、第 4 群則有 7 個，各群之標準差相距不遠，約為 0.75 左右，表示各族群離散程度差異不大。至於群體中離種子點有最大距離的為第 4 群，距離為 3.545。另外，群集中心的平均距離亦相差不遠，約為 4 左右。每一群體間的中心距離可整理如表 17，由表中得知第 2 與 4 群組間差距最小，第 2 與 3 群最大，但實際差距間差異並不大。由上述討論可知，群集之大小相似，且群集間距離差異不大，表示分群甚為平均，但亦表示群集間無特別明顯的差異。

表 15 非階層分析之分群結果

Table 15 Results of disjoint cluster analysis

Obs.	Cluster	Distance from Seed
1	1	2.0183
2	2	1.9748
3	2	1.4271
4	4	2.8063
5	4	1.6323
6	2	1.6981
7	2	1.9629
8	4	3.5450
9	4	2.3243
10	3	0.0000
11	1	2.0183
12	2	2.9974
13	4	1.7610
14	2	2.1426
15	2	1.9710
16	4	2.8240
17	4	2.3352
18	2	2.1924
19	2	3.4092
Criterion Based on Final Seeds = 0.6614		

經過非階層分析後，模式中各個變數之統計結果摘要如表 18，由表中可看出以磷酸

鹽之解釋能力最強，其次為葉綠素 a、COD、懸浮固體等。這表示在分群的結果中，以磷酸鹽之差異性最大，最能辨別出不同群體間的差異。而表中最後一欄 RSQ/(1-RSQ)值代表可解釋變異度與不可解釋變異度的比值，其值越大代表解釋能力越強，同樣以磷酸鹽值最高。

為了更進一步了解各個變數的情形，本研究再將所有變數逐一與依照分群結果所得之虛擬變項進行單變量分析，結果如表 19。

由表中可看出各個變數的統計顯著程度，其中，達顯著者有水溫、懸浮固體、濁度、總磷、磷酸鹽、COD、葉綠素 a 等 7 個變數，表示這些變數在群組間有顯著的差異，若以單一變數來判斷群組時，這些變數將有顯著的結果。若分群結果以 TSI 來加以判斷，可發現第 2 與 4 分群甚為合理，能有效將中高度優養情形區分開來；但第 1、3 族群則不符合 TSI 趨勢，其中，第 1 群中兩個中高度優養的觀測值被分為同一群，查看原始數據發現兩組數據相當一致，僅營養鹽磷的觀測數據有顯著差異，而中度優養的觀測組卻有葉綠素 a 值呈現相當高的現象，推測因此才分為同一群組。第 3 群中僅有 1 個觀測組，以 TSI 來看應分為高度優養組，查看原始數據後發現其懸浮固體觀測值特高，與高度優養群組差異甚大，推測因此而被獨立自成一組。根據單變量分析結果，因大部分變數均達統計顯著水準，故此分群結果尚可以接受。

表 16 非階層分析分群結果摘要
Table 16 Summary of disjoint cluster analysis

Cluster	Frequency	RMS Std Deviation	Maximum Distance from Seed to Observation	Radius Exceeded	Nearest Cluster	Distance Between Cluster Centroids
1	2	0.8240	2.0183	-	4	3.8250
2	9	0.6857	3.4092	-	4	3.4660
3	1	-	0.0000	-	1	4.5969
4	7	0.775	3.545	-	2	3.4660

表 17 各群組間的中心距離

Table 17 Distance between cluster centroids

Nearest Cluster	1	2	3	4
1	-	5.752904	4.596861	3.825016
2	5.752904	-	7.256442	3.466013
3	4.596861	7.256442	-	5.296459
4	3.825016	3.466013	5.296459	-

表 18 分析結果中各變數之統計摘要

Table 18 Statistics summary of variables

Variable	Total STD	Within STD	R-Square	RSQ/(1-RSQ)
水溫	1	0.79941	0.467459	0.87779
溶氧	1	0.92559	0.286073	0.400704
懸浮固體	1	0.48017	0.807864	4.204639
濁度	1	0.58248	0.717261	2.536836
氨氮	1	0.92633	0.284934	0.398473
硝酸鹽	1	0.90899	0.311445	0.452317
總磷	1	0.64704	0.651114	1.866268
磷酸鹽	1	0.19356	0.968779	31.02949
COD	1	0.45184	0.829864	4.877642
鈉	1	1.05795	0.067281	0.072134
葉綠素a	1	0.35708	0.893747	8.411469
透明度	1	0.88224	0.351384	0.541745
OVER-ALL	1	0.73231	0.5531	1.237639

Pseudo F Statistic = 6.19

表 19 單變量分析結果

Table 19 Results of univariate analysis

Dependent Variable -- 水溫						
Source	DF	Sum of Squares	Mean Square	F Value	Pr>F	R-Square
Model	3	8.4143	2.8048	4.39	0.0209	0.4675
Error	15	9.5857	0.6390	-	-	-
Corrected Total	18	18	-	-	-	-
Dependent Variable -- 溶氧						
Source	DF	Sum of Squares	Mean Square	F Value	Pr>F	R-Square
Model	3	5.1493	1.7164	2	0.1568	0.2861
Error	15	12.8507	0.8567	-	-	-
Corrected Total	18	18	-	-	-	-
Dependent Variable -- 懸浮固體						
Source	DF	Sum of Squares	Mean Square	F Value	Pr>F	R-Square
Model	3	14.5415	4.8472	21.02	<.0001	0.8079
Error	15	3.4585	0.2306	-	-	-
Corrected Total	18	18	-	-	-	-

表 19 單變量分析結果 (續)
Table 19 Results of univariate analysis (continue)

Dependent Variable -- 濁度						
Source	DF	Sum of Squares	Mean Square	F Value	Pr>F	R-Square
Model	3	12.9107	4.3036	12.68	0.0002	0.7173
Error	15	5.0893	0.3393	-	-	-
Corrected Total	18	18	-	-	-	-
Dependent Variable -- 氨氮						
Source	DF	Sum of Squares	Mean Square	F Value	Pr>F	R-Square
Model	3	5.1288	1.7096	1.99	0.1585	0.2849
Error	15	12.8712	0.8581	-	-	-
Corrected Total	18	18	-	-	-	-
Dependent Variable -- 硝酸鹽						
Source	DF	Sum of Squares	Mean Square	F Value	Pr>F	R-Square
Model	3	5.6060	1.8687	2.26	0.1232	0.3114
Error	15	12.3940	0.8263	-	-	-
Corrected Total	18	18	-	-	-	-
Dependent Variable -- 總磷						
Source	DF	Sum of Squares	Mean Square	F Value	Pr>F	R-Square
Model	3	11.7201	3.9067	9.33	0.0010	0.6511
Error	15	6.2799	0.4187	-	-	-
Corrected Total	18	18	-	-	-	-
Dependent Variable -- 磷酸鹽						
Source	DF	Sum of Squares	Mean Square	F Value	Pr>F	R-Square
Model	3	17.4380	5.8127	155.15	<.0001	0.9688
Error	15	0.5620	0.0375	-	-	-
Corrected Total	18	18	-	-	-	-
Dependent Variable -- COD						
Source	DF	Sum of Squares	Mean Square	F Value	Pr>F	R-Square
Model	3	14.9375	4.9792	24.39	<.0001	0.8299
Error	15	3.0625	0.2042	-	-	-
Corrected Total	18	18	-	-	-	-
Dependent Variable -- 鈉						
Source	DF	Sum of Squares	Mean Square	F Value	Pr>F	R-Square
Model	3	1.2111	0.4037	0.36	0.7823	0.0673
Error	15	16.7889	1.1193	-	-	-
Corrected Total	18	18	-	-	-	-
Dependent Variable -- 葉綠素a						
Source	DF	Sum of Squares	Mean Square	F Value	Pr>F	R-Square
Model	3	16.0874	5.3625	42.06	<.0001	0.8937
Error	15	1.9126	0.1275	-	-	-
Corrected Total	18	18	-	-	-	-
Dependent Variable -- 透明度						
Source	DF	Sum of Squares	Mean Square	F Value	Pr>F	R-Square
Model	3	6.3249	2.1083	2.71	0.0822	0.3514
Error	15	11.6751	0.7783	-	-	-
Corrected Total	18	18	-	-	-	-

將分群結果中各個變數之平均數計算出來如表 20，由表中可比較各個群組間的差異。其中，第 4 群為高度優養的觀測組，可看出其水溫最高與透明度最低，而其溶氧值亦為所有分群中最高的一群，由此可見其高優養的程度。當優養程度極高時，水體中的溶氧會因藻類白天行光合作用而大量增加，促使水體的溶氧值接近飽和。第 2 群組為中低優養的觀測組，由表中可知其水溫、懸浮固體、濁度、氨氮、總磷、磷酸鹽、COD、葉綠素 a 觀測值均為最低，而透明度則是最高的群組。第 3 群體雖屬於高度優養情形，但其懸浮固體與磷酸鹽觀測值特高，因而被獨立出來。至於第 1 群，因呈現濁度特別高的現象，所以亦被歸為同一群。由此可知，分群結果顯示以中低度優養之歸類效果較佳，且較符合優養的一般變化，至於高度優養群組則較不符合。此外，第 1、3 群組的分群推論是因觀測值中有特殊的情形才被獨立出來，故與優養變化趨勢不一致。最後，將分群結果與階層分析結果相比較，可發現其結果與馬氏法之分群結果完全相同，由此可印證採用馬氏法分析結果的正確性。

三、與鑑別分析之比較

群集分析結果之結論是以馬氏法為準，若將其分群結果與鑑別分析的分群標準相比較，可發現兩者的內容相差甚多，但同樣可以將大部分中高度優養的觀測组分群。鑑別分析的分組是以 TSI 為指標來分群，若同樣以 TSI 來檢驗群集分析的結果，則可以發現第 1、3 組較不符合 TSI 的趨勢，主要是因這

些觀測組中有特殊的觀測值，故導致獨立分群的結果。

若就單變量分析的結果來比較，可知兩方法所指出的變數重要程度並不相同。將各個變數依 R² 值大小來排序 (表 21)，由表中數據可發現兩者間僅有懸浮固體與鈉的重要排序是相同的，其他均不同，但卻同樣指出磷是最重要的因子，只是鑑別分析著重在總磷含量的不同，而群集分析則重視無機磷酸鹽有顯著的不同。另外，在群集分析中葉綠素 a 的重要性提升不少，表示在該分群結果中除了磷酸鹽外，藻類族群數量亦是重要的分類依據。

若以鑑別分析的分組為標準，用來檢驗群集分析的結果，其結果如表 22。從群集分析的分群結果來看，原始鑑別分析的第 1 組並沒有適合的群組，第 2 組則與第 2 群相似，第 3 組應歸為第 4 群，所以正確符合的觀測組共有 15 組，正確率為 78.95%，亦即兩组分群的結果大致吻合，但由於分群的標準不同，因此會有相異之處。同時，兩者的不同似乎也暗示了以總磷為基礎的 TSI 指標並無法完全區分本研究所使用的觀測組。

結論與建議

由階層分析的結果來看，可以得到各種分群法的建議分群數，包括單一連結法建議分為 9 群、完全連結法 4 群、彈性 β 法 3 群、馬氏法 4 群、中數法 9 群、平均連結法 4 群、形心連結法 8 群、華德最小變異法 3 群、ML 法 3 群、密度連結法 9 群，及雙連法的 2 群。

表 20 各群集之平均重心

Table 20 Cluster means in all clusters

Cluster	水溫	溶氧	懸浮固體	濁度	氨氮	硝酸鹽
1	0.3817208	-0.5738864	1.3741251	2.0493125	0.8873410	-1.2283330
2	-0.6927525	-0.3821306	-0.8056040	-0.5680766	-0.5205734	0.4548411
3	0.4514582	-0.1180826	2.0090149	1.2641244	0.1419746	-0.7899377
4	0.7171247	0.6721472	0.3561673	-0.0357229	0.3955006	-0.1209952

Cluster	總磷	磷酸鹽	COD	鈉	葉綠素a	透明度
1	0.3575399	0.0642514	1.7759880	0.2676650	1.8922984	0.3064300
2	-0.7879840	-0.3963670	-0.8076375	0.0120665	-0.7868589	0.5177704
3	1.6301986	3.9964210	1.4148517	-1.0227202	1.7332449	-0.1922818
4	0.6780825	-0.0796601	0.3288443	0.0541130	0.2234126	-0.7257875

表 21 鑑別與群集分析結果之
單變量分析比較

Table 21 Comparisons of univariate
analysis between discriminant and
cluster analysis

水質項目	R ²	
	鑑別單變量	群集單變量
總磷	0.7159	0.6511
水溫	0.4901	0.4675
透明度	0.4631	0.3514
懸浮固體	0.4438	0.8079
COD	0.3538	0.8299
氨氮	0.3178	0.2849
葉綠素a	0.2788	0.8937
溶氧	0.1965	0.2861
濁度	0.1526	0.7173
磷酸鹽	0.1163	0.9688
硝酸鹽	0.1011	0.3114
鈉	0.0056	0.0673

若將分群的結果逐一對照原始 TSI 值來檢驗，可發現以馬氏法與雙連法的結果最符合優養變化趨勢，因此總結是建議將觀測值分為 2 或 4 群；而若使用未經離群值修正的觀測值進行集群分析時，結果以彈性 β 法、華德法、ML 法與雙連法分群效果較佳，且建議分為 2 或 3 群。最後將經過離群值修正前後的分析結果相比較，發現仍是以經離群值修正後的分析結果較佳。

表 22 鑑別與群集分析結果之比較

Table 22 Comparisons between
discriminant and cluster analysis

Cluster	TSI			
	1	2	3	Total
1	0	1	1	2
	0	5.26	5.26	10.53
	0	50	50	-
	0	11.11	11.11	-
2	1	8	0	9
	5.26	42.11	0	47.37
	11.11	88.89	0	-
	100	88.89	0	-
3	0	0	1	1
	0	0	5.26	5.26
	0	0	100	-
	0	0	11.11	-
4	0	0	7	7
	0	0	36.84	36.84
	0	0	100	-
	0	0	77.78	-
Total	1	9	9	19
	5.26	47.37	47.37	100

在經過非階層分析後，發現以分群數為 4 時，其鑑別效果較為顯著，而在 12 個變數中，R2 值排序依次為磷酸鹽 → 葉綠素 a → COD → 懸浮固體 → 濁度 → 總磷 → 水溫 → 透明度 → 硝酸鹽 → 溶氧 → 氨氮 → 鈉，其中以磷酸鹽最為重要。

將群集與鑑別分析的結果相比較後，可發現兩者分群大致相同，若以鑑別分析分組為標準，則群集分析的正確率可達 78.95%，表示均可區分大部分的中高度優養情形，且同時指出分群中具有最顯著差異的為磷因子，表示若要區分出優養程度，仍是以磷最為重要。

總磷指標包括了有機與無機磷，由於有機磷必須經過細菌分解後才可被植物吸收利用，所以對優養化而言，總磷指標屬於一潛力指標，可用來評估優養的潛勢。然而，在該水體之優養現狀方面，總磷指標之敏感度可能就有所不及，可能是因為使用總磷為指標時會產生模糊空間所致，故在本研究結果中才會發現磷酸鹽為分群最重要的指標，或許更適合用於現狀的分析。

在群集分析結果中，有許多與鑑別分析不同之處，探討其原因可能是因在許多觀測組中存有較特殊的觀測值，因此被獨立分群出來，這表示群集分析結果可能具有尋找特殊觀測組的功能，可藉此找出不同的數據而加以檢查與討論。另外，鑑別與群集分析結果的差異亦表示以總磷為基礎的 TSI 指標之區別能力有限，必須進一步蒐集不足資料或探討更適合解釋優養化的指標。

參考文獻

1. 林師模、陳范欽。2003。多變量分析--管理上的應用。雙葉書廊有限公司。台北市。
2. 陳正昌、程炳林、陳新豐、劉子鍵。2003。多變量分析方法：統計軟體應用。五南圖書出版股份有限公司。台北市。
3. 陳鴻烈、鄭慧玲。1997。台中縣政府大甲溪及德基水庫水質監測研究方案。台中縣環境保護局委託研究計畫期末報告。pp.94-117。
4. 陳鴻烈、鄭慧玲。1998。水庫優養化之時間數列分析研究。水土保持學報 30(4)：331-337。
5. 陳鴻烈、梁家柱、鄭慧玲、王久泰。1999。利用相加性季節變動模式之德基水庫優養化時間數列分析。水土保持學報 31(3)：139-144。
6. 陳鴻烈、梁家柱、王久泰、鄭慧玲。2000a。以多變量統計主因子分析法探討德基水庫水質之變異性研究。水土保持學報 32(1)：5-10。
7. 陳鴻烈、梁家柱、王久泰、鄭慧玲。2000b。德基水庫優養化之空間分析研究。水土保持學報 32(3)：117-124。
8. 陳鴻烈、梁家柱、羅惠芬、鄭慧玲。2004。水庫優養化時間序列模式分析比較研究。水土保持學報 36(2)：169-178。
9. 陳鴻烈、蔡大偉。2005a。不同模式之預測能力研究。水土保持學報 37(2)。
10. 陳鴻烈、蔡大偉。2005b。不同預測時距對模式預測能力影響之研究。水土保持學報 37(3)。
11. 陳鴻烈、蔡大偉。2006a。德基水庫優養水質研究。水土保持學報 38(1)。
12. 陳鴻烈、蔡大偉。2006b。不同模式因子

- 對優養最佳化模式預測之研究。農林學報 55(1)。
13. 陳鴻烈、蔡大偉。2006c。不同模式時距對模式預測能力影響之研究。水土保持學報 38(2)。
14. 陳鴻烈、蔡大偉。2006d。以複迴歸分析法探討水質因子與優養化全模式之研究。水土保持學報 38(3)。
15. 陳鴻烈、蔡大偉。2007a。不同優養水質共線性分析及模式選擇之研究。水土保持學報。
16. 陳鴻烈、蔡大偉。2007b。以主成分分析法探討水庫優養化之動力研究。水土保持學報。
17. 陳鴻烈、蔡大偉。2007c。以鑑別分析法探討水庫優養化之動力研究。農林學報。
18. 經濟部水利署德基水庫集水區管理委員會。1983-1996。德基水庫集水區整體治理計畫水質監測與管理模式研究。
19. 彭昭英。1998。SAS 與統計分析。儒林圖書有限公司。台北市。
20. 潘南飛編譯 (Lpain 原著)。2003。工程統計。全威圖書有限公司。台北縣。
21. Alberto, W. D., D. A. M. D. Pilar, A. M. Valeria, P. S. Fabiana, H. A. Cecilia and B. M. D. L. Ngeles. 2001. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: suqui'a river basin (co'rdoba-argentina). Water Research. 35: 2881-2894.
22. Cary. 1999. The CLUSTER procedure, SAS/STAT user's guide, version 8, chapter 23. SAS Institute Inc.
23. Reisenhofer, E., A. Picciotto and D. Li. 1994. A factor analysis approach to the study of the eutrophication of a shallow, temperate lake (San Daniele, North Eastern Italy). Analytica Chimica Acta. 306: 99-106.
24. Heessen, H. J. L. and N. Daan. 1996. Long-term trends in ten non-target North Sea fish species. ICES Journal of Marine Science. 5: 1063-1078.
25. Lambrakis, N., A. Antonakos and G. Panagopoulos. 2004. The use of multi-component statistical analysis in hydrogeological environmental research. Water Research. 38: 1862-1872.
26. McNeil, V. H., M. E. Cox and M. Preda. 2005. Assessment of chemical water types and their spatial variation using multi-stage cluster analysis, Queensland, Australia. Journal of Hydrology. 310: 181-200.

97年09月05日 收稿

97年09月26日 修改

97年11月07日 接受