

111-2 多變量分析 期末報告

水果集群與判別分析

報告撰寫： 111354014 劉貞莉

其他組員： 111354018 陳玟儒、111354020 黃詩涵

## 目錄

壹、 研究動機 .....	2
貳、 研究方法 .....	2
一、 資料來源.....	2
二、 集群分析.....	3
三、 判別分析.....	3
參、 集群分析 .....	4
一、 平均連結法 .....	5
二、 重心方法.....	6
三、 WARD 最小變異數方法 .....	7
四、 K 平均值演算法 .....	8
五、 四種方法下之集群結果比較 .....	9
肆、 判別分析 .....	12
一、 平均連結法集群之判別結果 .....	12
二、 重心方法集群之判別結果 .....	16
三、 WARD 最小變異數方法集群之判別結果 .....	19
四、 K 平均值演算法集群之判別結果 .....	22
五、 四種方法下之判別結果比較 .....	25
伍、 結論 .....	26

## 壹、研究動機

現代社會中，健康飲食及營養狀況逐漸引起人們的重視，諸如「無麩質飲食」、「生酮飲食」、「地中海飲食」等飲食計畫開始進入大眾的視野，然而水果往往是各大計畫中不被排斥的種類，與其盲目攝取多樣水果來補充營養素，不如根據缺乏的營養對症下藥。因此本研究想利用營養成分對水果分群，提供大眾作為飲食指南參考。

## 貳、研究方法

為分析水果種類間是否存在相似的營養成分結構，本研究將自政府資料開放平臺取得水果營養成分含量，並進行集群分析與判別分析，最後根據分析結果彙整結論。詳細研究流程如下：

### 一、資料來源

資料取自政府資料開放平臺「食品營養成分資料集」(<https://data.gov.tw/dataset/8543>)，該資料集共 224,744 筆觀測值、17 個欄位，詳細欄位說明如表 一所示。

表 一、食品營養成分資料集欄位說明

欄位名稱	資料型態	資料描述	欄位名稱	資料型態	資料描述
食品分類	字串	如：水果類、魚類	分析項	字串	營養成分名稱， 如：熱量、維生素 C
資料類別	字串				
整合編號	字串		含量單位	字串	
樣品名稱	字串	如：蘋果、鯖魚	每 100 克含量	數值	
俗名	字串		樣本數	數值	
樣品英文名稱	字串		標準差	數值	
內容物描述	字串		每單位含量	數值	
廢棄率	數值		每單位重	數值	
分析項分類	字串	如：礦物質	每單位重含量	數值	

根據研究動機，本研究篩選資料集中食品分類為水果類的資料進行分析，共計 224 個水果種類，並選擇分析項為維生素（維生素 A、維生素 B1、維生素 B2、維生素 B6、維生素 C、維生素 E、維生素 K1）、礦物質（鈉、鈣、鉀、銅、鋅、錳、鎂、鐵）及其他常見營養成分（果糖、水分、熱量、粗脂肪、粗蛋白、糖質總量、總碳水化合物、膳食纖維、葉酸、葡萄糖、蔗糖、麥芽糖）共 27 項營養成分之每 100 克含量來分析，即資料維度 224 觀測值\*37 營養成分含量欄位。經處理過後之資料集型態呈現於表 二。

表 二、水果類營養成分資料集欄位說明

	欄位名稱	資料型態		欄位名稱	資料型態		欄位名稱	資料型態
1	水果種類	字串	11	鉀	數值	21	粗蛋白	數值
2	維生素 A	數值	12	銅	數值	22	糖質總量	數值
3	維生素 B1	數值	13	鋅	數值	23	總碳水化合物	數值
4	維生素 B2	數值	14	錳	數值	24	膳食纖維	數值
5	維生素 B6	數值	15	鎂	數值	25	葉酸	數值
6	維生素 C	數值	16	鐵	數值	26	葡萄糖	數值
7	維生素 E	數值	17	果糖	數值	27	蔗糖	數值
8	維生素 K1	數值	18	水分	數值	28	麥芽糖	數值
9	鈉	數值	19	熱量	數值	註：欄位編號 2-28 皆為該 營養成分之每 100 克含量		
10	鈣	數值	20	粗脂肪	數值			

## 二、集群分析

資料處理完成後，本研究將透過 SAS 軟體針對 224 種水果以表 二的資料集進行集群分析，決定這些水果合宜的分群數目。採用的集群方法共四種，包括階層式分群法中的平均連結法、重心方法與 Ward 最小變異數方法，同時也加入非階層式分群法的 K 平均值演算法進行比較。

## 三、判別分析

取得集群分析中平均連結法、重心方法、Ward 最小變異數方法與 K 平均值演算法的分群結果後，本研究將針對這四種分群結果執行判別分析，以對所得的群集做驗證，觀察不同方法下每個水果營養成分是否能準確的分群，最後根據表現最佳的驗證結果探索與彙整各群特徵來下結論。

## 參、集群分析

進入集群分析前可先透過表 三初步觀察各個變數的基本統計量，接著分別採用平均連結法、重心方法、Ward 最小變異數方法與 K 平均值演算法進行分析與比較，分別觀察其群集歷史、群集數準則與樹狀圖，準則為期望虛擬 t 平方越低越好，R 平方與虛擬 F 統計值越大越好，表示群與群之間越能區隔開來，而半偏 R 平方為每增加一個群集時 R 平方增加的幅度，當增幅明顯縮小時即可停止增加群集數目，以此決定各方法下之分群數目。

表 三、各營養成分基本統計量

變數	平均值	標準差	偏態	峰態	雙峰性
果糖	2.3156	2.3981	0.8744	0.0217	0.5762
水分	85.8308	5.4948	-1.4443	2.3095	0.5768
熱量	50.6920	20.1164	1.3293	1.9159	0.5582
粗脂肪	0.2969	0.8239	7.1795	54.6177	0.9113
粗蛋白	0.7540	0.3870	1.3488	3.3797	0.4391
糖質總量	5.8446	5.6252	0.6505	-0.2717	0.5140
維生素A總量(IU)	303.9	565.7	3.1572	12.8027	0.6923
維生素B1	0.0367	0.0248	1.9568	8.3218	0.4250
維生素B2	0.0348	0.0300	3.2028	13.6625	0.6740
維生素B6	0.1058	0.1210	2.0832	4.1296	0.7447
維生素C	23.2960	30.7948	2.9651	11.9298	0.6541
維生素E總量	0.4289	0.5389	2.5204	8.2048	0.6538
維生素K1	0.8333	6.3047	9.5123	93.5749	0.9469
總碳水化合物	12.6946	5.2135	1.3079	1.7626	0.5643
膳食纖維	1.5460	1.1658	2.2626	8.7265	0.5200
葉酸	5.9192	14.8351	4.8239	32.3778	0.6852
葡萄糖	1.8063	2.1719	1.5240	2.0856	0.6482
蔗糖	1.7152	2.8390	2.2522	5.8082	0.6862
鈉	2.7500	4.2952	2.6910	9.0540	0.6814
鈣	10.1875	8.3280	1.7623	3.0350	0.6758
鉀	184.6	82.7283	1.2793	1.3192	0.6047
銅	0.00860	0.0254	3.4659	12.3578	0.8450
鋅	0.3196	0.3732	5.4720	37.3974	0.7652
錳	0.00588	0.0429	7.9835	65.9837	0.9379
鎂	11.4821	10.0170	4.9317	38.7924	0.6053
鐵	0.2567	0.1940	0.9506	1.0691	0.4632
麥芽糖	0.00446	0.0339	7.7637	60.7416	0.9607

## 一、平均連結法

將表 二的資料集進行平均連結法集群分析，並將群集歷史、群集數準則與樹狀圖呈現於表 四與圖 一。觀察發現，群集數目為 5 時，半偏 R 平方增幅明顯變小，且 R 平方已達 0.937，虛擬 F 統計值也由 305 增加至 636，表示群與群之間已經能比分四群還區隔得開，圖 一亦能加強虛擬 F 統計值的論點，確實能觀察到其於群集數目 5 處有明顯增幅，故本研究決定平均連結法之分群數目為五群。

表 四、平均連結法群集歷史

群集歷史											
群集數目	聯結的群集		次數	半偏;R 平方	R 平方	近似;預期;R 平方	立方;群集;準則	虛擬 F;統計值	虛擬 t 平方	Norm RMS 距離	同分
10	CL18	CL14	25	0.0044	.960	.969	-3.9	574	35.7	0.3389	
9	CL42	藍寶石洋香瓜	3	0.0007	.960	.966	-2.6	637	10.0	0.3483	
8	CL11	CL20	11	0.0031	.956	.962	-2.0	678	13.5	0.4142	
7	CL13	CL12	179	0.0310	.925	.956	-8.5	449	182	0.4347	
6	CL26	CL9	7	0.0026	.923	.949	-4.7	522	12.0	0.451	
5	新疆哈密瓜	嘉寶瓜	2	0.0021	.921	.937	-2.7	636	.	0.6844	
4	CL7	CL10	204	0.1145	.806	.915	-9.7	305	338	0.8077	
3	CL8	CL6	18	0.0252	.781	.868	-6.3	394	46.2	0.8755	
2	CL3	CL4	222	0.4642	.317	.732	-14	103	471	1.8545	
1	CL2	CL5	224	0.3169	.000	.000	0.00	.	103	4.2759	

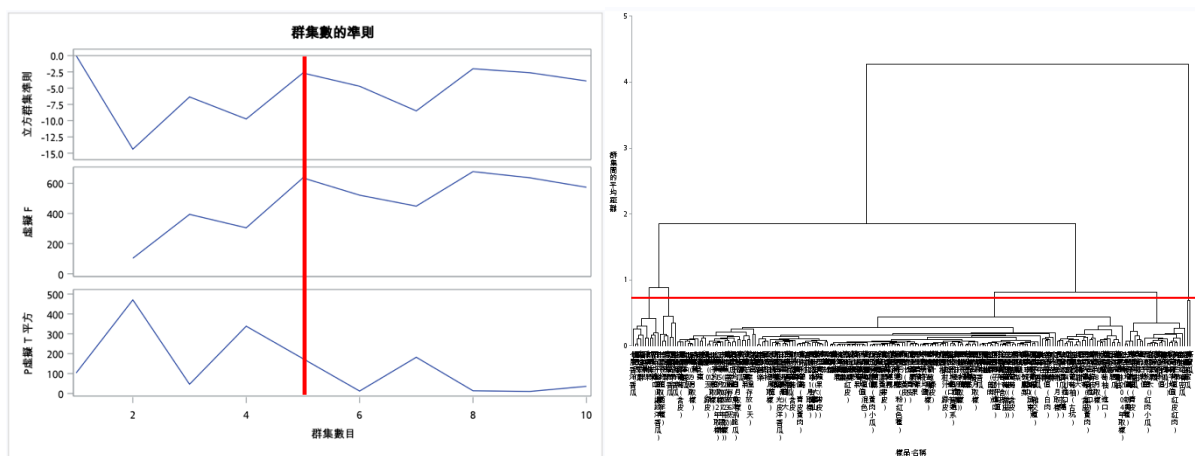


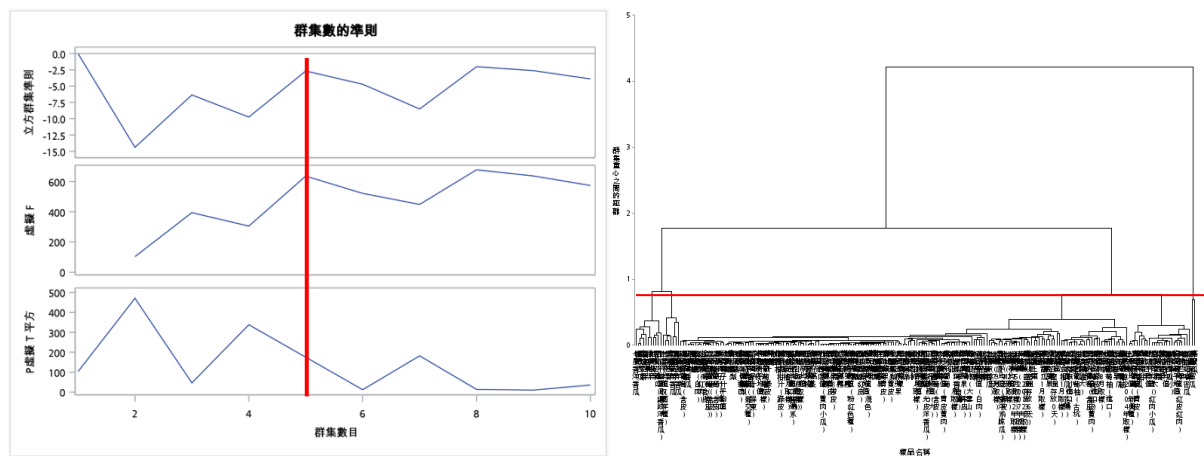
圖 一、平均連結法群集數準則與樹狀圖

## 二、重心方法

改以重心方法執行集群分析，並將群集歷史、群集數準則與樹狀圖呈現於表 五與圖二。觀察發現，重心方法的樹狀圖與平均連結法相似，而當群集數目為 5 時，半偏 R 平方增幅明顯變小，且 R 平方已達 0.921，虛擬 F 統計值也由 305 增加至 636，表示群與群之間已經能比分四群還區隔得開，圖二亦能加強虛擬 F 統計值的論點，確實能觀察到其於群集數目 5 處有明顯增幅，故本研究決定重心方法之分群數目為五群。

表 五、重心方法群集歷史

群集歷史										
群集數目	聯結的群集		次數	半偏;R 平方	R 平方	近似;預期;R 平方	立方;群集;準則	虛擬 F;統計值	虛擬 t 平方	範數重心距離
10	CL18	CL13	25	0.0044	.960	.969	-3.9	574	35.7	0.299
9	CL36	藍寶石洋香瓜	3	0.0007	.960	.966	-2.6	637	10.0	0.3426
8	CL11	CL19	11	0.0031	.956	.962	-2.0	678	13.5	0.3663
7	CL14	CL12	179	0.0310	.925	.956	-8.5	449	182	0.3885
6	CL27	CL9	7	0.0026	.923	.949	-4.7	522	12.0	0.408
5	新疆哈密瓜	嘉寶瓜	2	0.0021	.921	.937	-2.7	636	.	0.6844
4	CL7	CL10	204	0.1145	.806	.915	-9.7	305	338	0.763
3	CL8	CL6	18	0.0252	.781	.868	-6.3	394	46.2	0.8106
2	CL3	CL4	222	0.4642	.317	.732	-14	103	471	1.7689
1	CL2	CL5	224	0.3169	.000	.000	0.00	.	103	4.2218



圖二、重心方法群集數準則與樹狀圖

### 三、Ward 最小變異數方法

接著以 Ward 最小變異數方法執行集群分析，並將群集歷史、群集數準則與樹狀圖呈現於表 六與圖 三。觀察發現，Ward 最小變異數方法的樹狀圖與重心方法和平均連結法不太相同，而當群集數目為五時，半偏 R 平方增幅明顯變小，且 R 平方已達 0.937，虛擬 F 統計值也由 657 增加至 821，表示群與群之間已經能比分四群還區隔得開，圖 三亦能加強虛擬 F 統計值的論點，確實能觀察到其於群集數目 5 處有明顯增幅，虛擬 t 平方亦是較小的數值，故本研究決定 Ward 最小變異數方法之分群數目為五群。

表 六、Ward 最小變異數方法群集歷史

群集歷史										
群集 數 目	聯結的群集		次數	半偏;R 平方	R 平方	近似;預期;R 平方	立方;群集;準則	虛擬 F;統計值	虛擬 t 平方	同分
10	CL20	CL25	43	0.0025	.975	.969	3.16	918	37.5	
9	CL38	CL19	7	0.0026	.972	.966	3.23	940	12.0	
8	CL11	CL15	119	0.0070	.965	.962	1.49	856	96.5	
7	CL17	CL14	38	0.0080	.957	.956	0.27	810	55.7	
6	CL13	CL16	15	0.0082	.949	.949	-.04	812	32.6	
5	CL10	CL8	162	0.0116	.937	.937	0.03	821	89.5	
4	CL6	CL9	22	0.0379	.900	.915	-2.0	657	50.2	
3	CL5	CL7	200	0.1032	.796	.868	-5.4	432	451	
2	CL4	CL12	24	0.1236	.673	.732	-3.1	456	49.4	
1	CL2	CL3	224	0.6727	.000	.000	0.00	.	456	

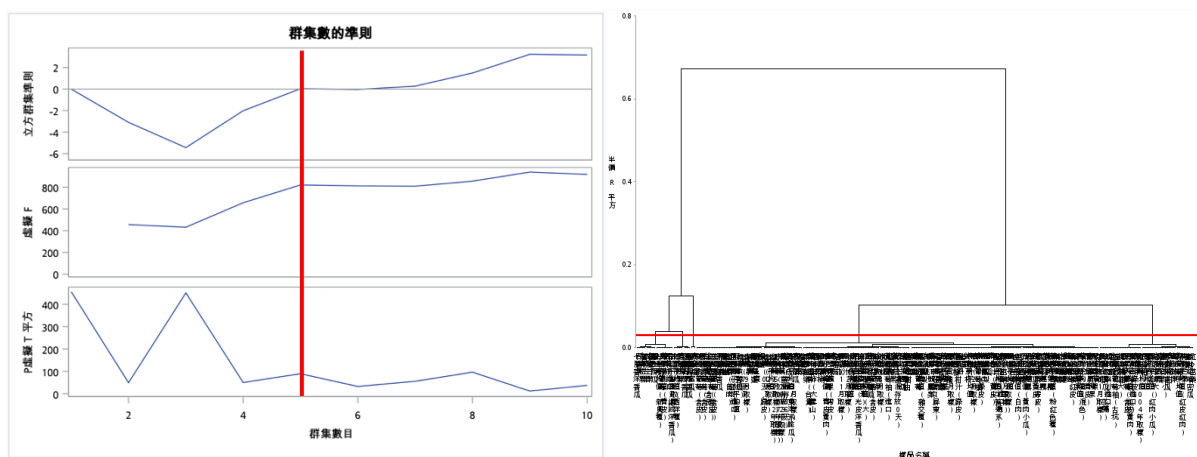


圖 三、Ward 最小變異數方法群集數準則與樹狀圖



#### 四、K 平均值演算法

不同於上述三種階層式分群法，K 平均值演算法屬於非階層式分群法，須事先決定分群數目，然而前三種方法皆決定分群數目為 5，故本研究便以 5 組來執行。根據表 七的計算結果， $RSQ/(1-RSQ)$ 表示可解釋部分與不可解釋部分的比率，僅有維生素 A 的數值大於 1，其 R 平方也是最大，表示維生素 A 在五群間的差異最大，五群幾乎都是用維生素 A 作為分類標準。

表 七、K 平均值演算法之變數統計值

變數的統計值				
變數	總 STD	STD 內	R 平方	$RSQ/(1-RSQ)$
果糖	2.39807	2.40240	0.014385	0.014595
水分	5.49482	5.52481	0.007188	0.007240
熱量	20.11645	20.21931	0.007869	0.007931
粗脂肪	0.82388	0.83050	0.002105	0.002110
粗蛋白	0.38699	0.39045	0.000270	0.000270
糖質總量	5.62522	5.62639	0.017528	0.017840
維生素A總量(IU)	565.73323	112.99273	0.960824	24.526022
維生素B1	0.02484	0.02483	0.018366	0.018709
維生素B2	0.02996	0.03017	0.004249	0.004267
維生素B6	0.12103	0.12166	0.007771	0.007832
維生素C	30.79485	31.03789	0.002375	0.002380
維生素E總量	0.53888	0.52550	0.066098	0.070776
維生素K1	6.30470	6.27070	0.028500	0.029336
總碳水化合物	5.21354	5.24270	0.006922	0.006971
膳食纖維	1.16584	1.16928	0.012122	0.012271
葉酸	14.83508	14.93678	0.004426	0.004446
葡萄糖	2.17191	2.16268	0.026264	0.026972
蔗糖	2.83904	2.81307	0.035823	0.037154
鈉	4.29516	3.98516	0.154580	0.182845
鈣	8.32795	8.36345	0.009548	0.009640
鉀	82.72834	82.39799	0.025765	0.026446
銅	0.02536	0.02521	0.029512	0.030410
鋅	0.37317	0.37471	0.009807	0.009904
錳	0.04285	0.04308	0.007388	0.007443
鎂	10.01701	10.09588	0.002412	0.002418
鐵	0.19397	0.19340	0.023653	0.024226
麥芽糖	0.03385	0.03404	0.006836	0.006883
OVER-ALL	110.35452	28.21206	0.935816	14.580140

虛擬 F 統計值 = 798.26

近似預期整體 R 平方 = 0.93711

立方群集準則 = -0.233

## 五、四種方法下之集群結果比較

此小節條列出四種方法下，各自將 224 種水果分類到何群集中（表 八）。可以發現平均連結法與重心方法的分群結果完全相同，與 Ward 最小變異數方法的結果則是有些許差異，而 K 平均值演算法則是與 Ward 最小變異數方法的結果相似。

表 八、四種方法之集群結果

樣品名稱	平均	重心	Ward	K	樣品名稱	平均	重心	Ward	K
七股香洋香瓜	3	3	3	3	美國紅葡萄	1	1	1	1
人心果	1	1	1	1	美國紅葡萄(含皮)	1	1	1	1
土芒果	2	2	3	2	美國紫葡萄(含皮)	1	1	1	1
土芭樂	1	1	1	1	美國黑葡萄(含皮)	1	1	1	1
大雪梨	1	1	1	1	美國綠葡萄	2	2	3	2
小玉西瓜	1	1	1	1	美國綠葡萄(含皮)	1	1	1	1
山竹	1	1	1	1	美濃瓜	1	1	1	1
天香洋香瓜	1	1	1	1	茂谷柑	2	2	2	2
太陽洋香瓜	1	1	1	1	香水李	1	1	1	1
文旦	1	1	1	1	香水芒果	3	3	3	3
木瓜(11 月取樣)	1	1	2	2	香瓜茄	1	1	1	1
木瓜(2 月取樣)	1	1	1	1	香吉士(進口)	1	1	1	1
木瓜(5 月取樣)	4	4	4	4	香華洋香瓜	1	1	1	1
木瓜(8 月取樣)	1	1	1	1	桃接李(大)	1	1	2	2
木瓜平均值	2	2	2	2	桃接李(小)	2	2	2	2
水蜜桃	1	1	1	1	桑葚	1	1	1	1
水蜜桃平均值	1	1	1	1	桑葚汁	1	1	1	1
牛心柿	2	2	2	2	泰國芭樂	1	1	1	1
牛奶鳳梨	1	1	1	1	海梨桶柑	1	1	2	2
牛乳芭樂	1	1	1	1	胭脂梅	1	1	1	1
世紀芭樂	1	1	1	1	脆桃平均值	1	1	1	1
加州青李	1	1	1	1	草莓	1	1	1	1
加州紅李(台灣)	1	1	1	1	荔枝平均值	1	1	1	1
加州紅李(進口)	1	1	2	2	高牆桶柑	2	2	2	2
加州蜜李(進口)	1	1	2	2	國產甜柿	1	1	1	1
加拉蘋果	1	1	1	1	將軍蜜梨	1	1	1	1
北蕉(0 天,綠皮)	1	1	1	1	梨山蜜梨	1	1	1	1
北蕉(11 月取樣)	1	1	1	1	甜瓜平均值(光皮洋香瓜)	1	1	1	1
北蕉(1 天)	1	1	1	1	甜瓜平均值(東洋系脆瓜)	1	1	1	1
北蕉(2 月取樣)	1	1	1	1	甜瓜平均值(網紋洋香瓜)	3	3	3	3
北蕉(3 天)	1	1	1	1	甜柿(進口)	3	3	3	3

北蕉(5月取樣)	1	1	1	1	甜柿平均值	2	2	2	2
北蕉(7天)	1	1	1	1	甜蜜桃	1	1	1	1
北蕉(9月取樣)	1	1	1	1	甜蜜蜜鳳梨	1	1	1	1
北蕉平均值	1	1	1	1	甜橙平均值(普遍系)	1	1	1	1
可可椰子汁(屏東)	1	1	1	1	蛋黃果	1	1	2	2
可可椰子汁(進口)	1	1	1	1	凱特芒果	2	2	2	2
可可椰子汁平均值	1	1	1	1	富士蘋果	1	1	1	1
台灣土棗	1	1	1	1	富士蘋果(帶皮)	1	1	1	1
四周柿	3	3	3	3	富有甜柿(大雪山)	1	1	1	1
四季芒果	3	3	3	3	椪柑	2	2	2	2
巨峰葡萄	1	1	1	1	無子西瓜	2	2	2	2
玉文芒果	1	1	1	2	無子芭樂	1	1	1	1
玉荷芭荔枝	1	1	1	1	無子紅葡萄(含皮)	1	1	1	1
甘蔗鳳梨	1	1	1	1	無花果	1	1	1	1
白油桃	1	1	1	1	筆柿	2	2	2	2
白柚	1	1	1	1	菠蘿蜜	1	1	1	1
白櫻桃	1	1	1	1	華寶大西瓜(花蓮)	1	1	2	2
印度棗(長形)	1	1	1	1	萊姆	1	1	1	1
印度棗平均值(大)	1	1	1	1	萊姆汁	1	1	1	1
安石榴	1	1	1	1	黃皮葡萄柚	1	1	1	1
百香果	3	3	3	3	黃皮葡萄柚(進口)	1	1	1	1
百香果汁	1	1	2	2	黃肉李	1	1	1	1
西瓜平均值(紅肉小瓜)	2	2	2	2	黃金果	1	1	1	1
西瓜平均值(黃肉小瓜)	1	1	1	1	黃香瓜	1	1	1	1
西施蜜柚	1	1	1	1	黑李	1	1	1	1
西洋梨(青皮)	1	1	1	1	黑金剛蓮霧	1	1	1	1
西洋梨(紅皮)	1	1	1	1	黑珍珠蓮霧	1	1	1	1
西洋梨(黃皮)	1	1	1	1	黑美人西瓜	1	1	1	1
西洋梨平均值	1	1	1	1	黑美人葡萄(含皮)	1	1	2	2
冷凍荔枝	1	1	1	1	黑香芒果	4	4	4	4
李子平均值(青皮黃肉)	1	1	1	1	黑葉仔荔枝	1	1	1	1
李子平均值(紅皮紅肉)	2	2	2	2	圓果金柑	1	1	1	1
李子平均值(紅皮黃肉)	1	1	2	2	圓果金柑汁(綠皮)	1	1	1	1
李林蕉(2007年取樣)	1	1	1	1	意大利葡萄	1	1	1	1
李林蕉(2012年取樣)	1	1	1	1	愛文芒果	4	4	4	4
芒果平均值(西洋種)	4	4	4	4	新世紀哈密瓜	2	2	2	2
芒果平均值(新興種)	2	2	3	2	新世紀梨	1	1	1	1
奇異果	1	1	1	1	新興梨	1	1	1	1

幸水梨	1	1	1	1	新疆哈密瓜	5	5	5	5
枇杷	3	3	3	3	楊桃(11月取樣)	1	1	1	1
狀元瓜	1	1	1	1	楊桃(2月取樣)	1	1	1	1
玫瑰桃	1	1	2	2	楊桃(4月取樣)	1	1	1	1
肯特芒果	2	2	2	2	楊桃(5月取樣)	1	1	1	1
芭樂平均值(白肉)	1	1	1	1	楊桃(9月取樣)	1	1	1	1
金冠蘋果	1	1	1	1	楊桃平均值	1	1	1	1
金香葡萄(含皮)	1	1	1	1	聖心芒果	2	2	2	2
金黃奇異果	1	1	1	1	酪梨(室溫存放0天)	1	1	1	1
金煌芒果	3	3	3	3	酪梨(室溫存放3天)	1	1	1	1
金蜜芒果	1	1	1	1	酪梨(室溫存放6天)	1	1	1	1
金興芒果	3	3	3	3	酪梨(綠皮)	1	1	2	2
金蘭西瓜	1	1	1	1	嘉玉甜瓜	1	1	1	1
金鑽鳳梨	1	1	1	1	嘉寶瓜	5	5	5	5
長果金柑	1	1	2	2	榴槤	1	1	1	1
青木瓜	1	1	1	1	福壽桃	1	1	1	1
青皮葡萄柚	1	1	1	1	綠葡萄平均值	2	2	2	2
青龍蘋果	4	4	4	4	翠玉青蘋果	1	1	1	1
南華蕉	1	1	1	1	翠玉青蘋果(帶皮)	1	1	1	1
柳橙	1	1	1	1	翡翠李	1	1	1	1
津經蜜蘋果	1	1	1	1	蜜世界洋香瓜	1	1	1	1
珍珠芭樂	1	1	1	1	蜜棗(圓形)	1	1	1	1
珍珠蜜棗	1	1	1	1	蜜棗李(進口)	1	1	1	1
皇宮西瓜	1	1	2	2	鳳光西瓜	3	3	3	3
秋香洋香瓜	2	2	2	2	鳳梨平均值(雜交種)	1	1	1	1
秋香蘋果	1	1	1	1	鳳梨釋迦	1	1	1	1
突目1號鳳梨	1	1	1	1	蓮花芒果	2	2	2	2
紅心芭樂	2	2	2	2	蓮霧平均值(粉紅色種)	1	1	1	1
紅毛丹	1	1	1	1	橫山梨	1	1	1	1
紅甘蔗汁	1	1	1	1	澳洲甜橙(進口)	1	1	1	1
紅皮蕉(2004年取樣)	1	1	2	2	龍眼	1	1	1	1
紅皮蕉(2012年取樣)	1	1	1	1	檸檬	1	1	1	1
紅肉李(大)	2	2	2	2	檸檬汁(黃皮)	1	1	1	1
紅肉李(小)	1	1	2	2	檸檬汁(綠皮)	1	1	1	1
紅柑	1	1	1	1	檸檬汁平均值	1	1	1	1
紅鈴西瓜	2	2	2	2	檸檬柑	1	1	1	1
紅龍果(白肉)	1	1	1	1	藍寶石洋香瓜	4	4	4	4
紅龍果(紅肉)	1	1	1	1	蟠桃	1	1	1	1

紅寶石洋香瓜	4	4	4	4	豐水梨	1	1	1	1
紅寶石葡萄柚(古坑)	1	1	2	2	蘋果平均值(青皮)	2	2	3	2
紅寶石蓮霧	1	1	1	1	蘋果平均值(混色)	1	1	1	1
美國五爪蘋果	1	1	1	1	釋迦	1	1	1	1

註：K 平均值演算法原分類標籤與另外三種方法較不一致，故此表有經過標籤號碼之替換，將原 K 平均值演算法群集標籤 1 替換成 3、2 替換成 1、5 替換成 2、3 替換成 5，以便與其他三種方法比較。

## 肆、判別分析

根據前一節取得的四種分群結果執行線性判別分析，以對群集做驗證，觀察不同方法下的分類效果如何。本研究會採用相等的先驗機率執行計算，呈現線性判別函數，並且觀察多變量統計值檢定結果是否通過、哪些單變量檢定統計值呈現顯著、交叉驗證前與交叉驗證後之分類錯誤表，最後以整體分類錯誤最低的比例來決定採用哪個分群結果為結論。

### 一、平均連結法集群之判別結果

平均連結法的集群資訊列示於表 九，第一群的水果種類最多，有 179 種，第五群內僅兩種水果最少，據此資料計算出的判別函數呈現於表 十一。在設定顯著水準為 0.05 下，表 十的檢定結果皆呈現顯著，顯示此判別函數在不同群集之間的差異是顯著的，能夠有效的區分不同群集。

表 九、平均連結法集群之類別層級資訊

類別層級資訊					
CLUSTER	變數名稱	次數	加權	比例	先驗機率
1	1	179	179.0000	0.799107	0.200000
2	2	25	25.0000	0.111607	0.200000
3	3	11	11.0000	0.049107	0.200000
4	4	7	7.0000	0.031250	0.200000
5	5	2	2.0000	0.008929	0.200000

表 十、平均連結法集群之多變量檢定結果

多變量統計值和 F 近似					
S=4 M=11 N=95.5					
統計值	值	F 值	分子自由度	分母自由度	Pr > F
Wilks 的 Lambda	0.03496623	9.45	108	768.68	<.0001
Pillai 追蹤	1.24012521	3.26	108	784	<.0001
Hotelling-Lawley 追蹤	20.41924373	36.22	108	645.54	<.0001
Roy 最大根	20.09677945	145.89	27	196	<.0001
附註: Roy 最大根的 F 統計值是上限。					

表 十一、平均連結法集群之線性判別函數

CLUSTER的線性判別函數					
變數	1	2	3	4	5
常數	-458905	-459253	-459324	-459911	-460098
果糖	70.35976	71.07723	72.66359	66.92575	59.04438
水分	9181	9184	9185	9190	9188
熱量	617.07033	616.53464	615.17368	615.18095	610.53776
粗脂肪	4158	4165	4177	4183	4222
粗蛋白	7305	7306	7306	7308	7314
糖質總量	-78.71780	-79.38314	-81.31719	-74.97518	-69.88463
維生素A總量(IU)	0.22155	0.26085	0.30319	0.34739	0.46104
維生素B1	-7051	-7039	-7044	-7089	-7099
維生素B2	4612	4642	4663	4660	4693
維生素B6	-1016	-1015	-1018	-1015	-1018
維生素C	5.46395	5.47884	5.47946	5.51135	5.54414
維生素E總量	-82.23178	-84.45961	-86.68489	-90.66646	-99.08917
維生素K1	-3.22574	-3.21587	-3.34778	-3.39945	-3.53242
總碳水化合物	6983	6988	6994	6999	7015
膳食纖維	144.97277	145.02021	145.37058	144.43250	143.65502
葉酸	-8.82571	-8.79089	-8.75746	-8.71143	-8.65894
葡萄糖	150.95084	151.70632	154.00489	146.94961	144.85393
蔗糖	97.53509	98.30103	100.27604	94.14049	88.92112
鈉	43.97463	44.08342	43.99242	44.02957	44.85204
鈣	12.37186	12.38121	12.38116	12.56732	12.57101
鉀	12.83125	12.83998	12.86143	12.89554	12.88680
銅	-1505	-1503	-1443	-1457	-1421
鋅	288.29842	287.96615	289.37994	288.56944	289.55418
錳	-485.26817	-497.25274	-515.97258	-523.28157	-544.68604
鎂	24.52587	24.50740	24.50857	24.44200	24.49326
鐵	-110.90955	-112.12187	-113.28349	-116.72019	-121.35675
麥芽糖	1998	2001	1994	2006	2020

若觀察單變量檢定統計值（表 十二），發現所有變數僅維生素 A、蔗糖、鈉呈現顯著，顯示這三個營養成分在不同群集之間存在顯著的平均值差異，將是區分不同群集的重要指標。

表 十二、平均連結法集群之單變量檢定統計值

單變量檢定統計值							
F 統計值、Num DF=4、Den DF=219							
變數	總標準差	集區標準差	標準差之間	R 平方	R 平方 / (1-RSq)	F 值	Pr > F
果糖	2.3981	2.4004	0.3390	0.0161	0.0163	0.89	0.4687
水分	5.4948	5.5089	0.6957	0.0129	0.0131	0.71	0.5828
熱量	20.1164	20.1541	2.6796	0.0143	0.0145	0.79	0.5315
粗脂肪	0.8239	0.8291	0.0684	0.0055	0.0056	0.30	0.8747
粗蛋白	0.3870	0.3886	0.0428	0.0098	0.0099	0.54	0.7040
糖質總量	5.6252	5.6205	0.8779	0.0196	0.0200	1.09	0.3609
維生素A總量(IU)	565.7332	133.6784	613.5493	0.9452	17.2373	943.74	<.0001
維生素B1	0.0248	0.0249	0.003511	0.0161	0.0163	0.89	0.4687
維生素B2	0.0300	0.0302	0.002056	0.0038	0.0038	0.21	0.9338
維生素B6	0.1210	0.1216	0.0126	0.0087	0.0088	0.48	0.7496
維生素C	30.7948	31.0396	1.6358	0.0023	0.0023	0.12	0.9736
維生素E總量	0.5389	0.5348	0.1089	0.0328	0.0339	1.86	0.1189
維生素K1	6.3047	6.3280	0.7259	0.0107	0.0108	0.59	0.6706
總碳水化合物	5.2135	5.2343	0.5841	0.0101	0.0102	0.56	0.6935
膳食纖維	1.1658	1.1676	0.1588	0.0149	0.0151	0.83	0.5083
葉酸	14.8351	14.9518	0.8145	0.0024	0.0024	0.13	0.9702
葡萄糖	2.1719	2.1724	0.3205	0.0175	0.0178	0.98	0.4219
蔗糖	2.8390	2.8009	0.6655	0.0441	0.0462	2.53	0.0416
鈉	4.2952	3.9780	1.9023	0.1576	0.1871	10.24	<.0001
鈣	8.3280	8.3619	0.9255	0.0099	0.0100	0.55	0.7001
鉀	82.7283	82.0395	17.0724	0.0342	0.0354	1.94	0.1048
銅	0.0254	0.0253	0.004219	0.0222	0.0228	1.25	0.2926
鋅	0.3732	0.3739	0.0498	0.0143	0.0145	0.79	0.5299
錳	0.0429	0.0431	0.003293	0.0047	0.0048	0.26	0.9026
鎂	10.0170	10.0692	0.9788	0.0077	0.0077	0.42	0.7917
鐵	0.1940	0.1943	0.0260	0.0144	0.0146	0.80	0.5250
麥芽糖	0.0339	0.0341	0.002503	0.0044	0.0044	0.24	0.9146

平均 R 平方	
未加權	0.0539342
根據變異數加權	0.9207588

接著觀察交叉驗證前與後之分類錯誤表（表 十三與表 十四），交叉驗證前的結果皆為第一群出現分類錯誤的情形，共有 6 個被分類到第二群；交叉驗證後的結果同樣為第一群的水果出現分類錯誤，乘上先驗機率之分類錯誤率 0.0134。

表 十三、平均連結法集群之交叉驗證前分類錯誤表

分類為 CLUSTER 的觀測值數目和百分比						
來自 CLUSTER	1	2	3	4	5	總計
1	173 96.65	6 3.35	0 0.00	0 0.00	0 0.00	179 100.00
2	0 0.00	25 100.00	0 0.00	0 0.00	0 0.00	25 100.00
3	0 0.00	0 0.00	11 100.00	0 0.00	0 0.00	11 100.00
4	0 0.00	0 0.00	0 0.00	7 100.00	0 0.00	7 100.00
5	0 0.00	0 0.00	0 0.00	0 0.00	2 100.00	2 100.00
總計	173 77.23	31 13.84	11 4.91	7 3.13	2 0.89	224 100.00
事前	0.2	0.2	0.2	0.2	0.2	

CLUSTER 的誤差計數估計值						
	1	2	3	4	5	總計
比率	0.0335	0.0000	0.0000	0.0000	0.0000	0.0067
事前	0.2000	0.2000	0.2000	0.2000	0.2000	

表 十四、平均連結法集群之交叉驗證後分類錯誤表

分類為 CLUSTER 的觀測值數目和百分比						
來自 CLUSTER	1	2	3	4	5	總計
1	167 93.30	12 6.70	0 0.00	0 0.00	0 0.00	179 100.00
2	0 0.00	25 100.00	0 0.00	0 0.00	0 0.00	25 100.00
3	0 0.00	0 0.00	11 100.00	0 0.00	0 0.00	11 100.00
4	0 0.00	0 0.00	0 0.00	7 100.00	0 0.00	7 100.00
5	0 0.00	0 0.00	0 0.00	0 0.00	2 100.00	2 100.00
總計	167 74.55	37 16.52	11 4.91	7 3.13	2 0.89	224 100.00
事前	0.2	0.2	0.2	0.2	0.2	

CLUSTER 的誤差計數估計值						
	1	2	3	4	5	總計
比率	0.0670	0.0000	0.0000	0.0000	0.0000	0.0134
事前	0.2000	0.2000	0.2000	0.2000	0.2000	



## 二、重心方法集群之判別結果

重心方法的集群資訊列示於表 十五，與平均連結法相同，第一群的水果種類最多，有 179 種，第五群內僅兩種水果最少，據此資料計算出的判別函數呈現於表 十六。在設定顯著水準為 0.05 下，表 十七的檢定結果皆呈現顯著，顯示此判別函數在不同群集之間的差異是顯著的，能夠有效的區分不同群集。

表 十五、重心方法集群之類別層級資訊

類別層級資訊					
CLUSTER	變數名稱	次數	加權	比例	先驗機率
1	1	179	179.0000	0.799107	0.200000
2	2	25	25.0000	0.111607	0.200000
3	3	11	11.0000	0.049107	0.200000
4	4	7	7.0000	0.031250	0.200000
5	5	2	2.0000	0.008929	0.200000

表 十六、重心方法集群之線性判別函數

CLUSTER的線性判別函數					
變數	1	2	3	4	5
常數	-458905	-459253	-459324	-459911	-460098
果糖	70.35976	71.07723	72.66359	66.92575	59.04438
水分	9181	9184	9185	9190	9188
熱量	617.07033	616.53464	615.17368	615.18095	610.53776
粗脂肪	4158	4165	4177	4183	4222
粗蛋白	7305	7306	7306	7308	7314
糖質總量	-78.71780	-79.38314	-81.31719	-74.97518	-69.88463
維生素A總量(IU)	0.22155	0.26085	0.30319	0.34739	0.46104
維生素B1	-7051	-7039	-7044	-7089	-7099
維生素B2	4612	4642	4663	4660	4693
維生素B6	-1016	-1015	-1018	-1015	-1018
維生素C	5.46395	5.47884	5.47946	5.51135	5.54414
維生素E總量	-82.23178	-84.45961	-86.68489	-90.66646	-99.08917
維生素K1	-3.22574	-3.21587	-3.34778	-3.39945	-3.53242
總碳水化合物	6983	6988	6994	6999	7015
膳食纖維	144.97277	145.02021	145.37058	144.43250	143.65502
葉酸	-8.82571	-8.79089	-8.75746	-8.71143	-8.65894
葡萄糖	150.95084	151.70632	154.00489	146.94961	144.85393
蔗糖	97.53509	98.30103	100.27604	94.14049	88.92112
鈉	43.97463	44.08342	43.99242	44.02957	44.85204
鈣	12.37186	12.38121	12.38116	12.56732	12.57101
鉀	12.83125	12.83998	12.86143	12.89554	12.88680
銅	-1505	-1503	-1443	-1457	-1421
鋅	288.29842	287.96615	289.37994	288.56944	289.55418
錳	-485.26817	-497.25274	-515.97258	-523.28157	-544.68604
鎂	24.52587	24.50740	24.50857	24.44200	24.49326
鐵	-110.90955	-112.12187	-113.28349	-116.72019	-121.35675
麥芽糖	1998	2001	1994	2006	2020

表 十七、重心方法集群之多變量檢定結果

多變量統計值和 F 近似					
S=4 M=11 N=95.5					
統計值	值	F 值	分子自由度	分母自由度	Pr > F
Wilks 的 Lambda	0.03496623	9.45	108	768.68	<.0001
Pillai 追蹤	1.24012521	3.26	108	784	<.0001
Hotelling-Lawley 追蹤	20.41924373	36.22	108	645.54	<.0001
Roy 最大根	20.09677945	145.89	27	196	<.0001
附註: Roy 最大根的 F 統計值是上限。					

若觀察單變量檢定統計值（表 十八），發現所有變數僅維生素 A、蔗糖、鈉呈現顯著，顯示這三個營養成分在不同群集之間存在顯著的平均值差異，將是區分不同群集的重要指標，其結果與平均連結法相同。

表 十八、重心方法集群之單變量檢定統計值

單變量檢定統計值							
F 統計值、Num DF=4、Den DF=219							
變數	總標準差	集區標準差	標準差之間	R 平方	R 平方 / (1-RSq)	F 值	Pr > F
果糖	2.3981	2.4004	0.3390	0.0161	0.0163	0.89	0.4687
水分	5.4948	5.5089	0.6957	0.0129	0.0131	0.71	0.5828
熱量	20.1164	20.1541	2.6796	0.0143	0.0145	0.79	0.5315
粗脂肪	0.8239	0.8291	0.0684	0.0055	0.0056	0.30	0.8747
粗蛋白	0.3870	0.3886	0.0428	0.0098	0.0099	0.54	0.7040
糖質總量	5.6252	5.6205	0.8779	0.0196	0.0200	1.09	0.3609
維生素A總量(IU)	565.7332	133.6784	613.5493	0.9452	17.2373	943.74	<.0001
維生素B1	0.0248	0.0249	0.003511	0.0161	0.0163	0.89	0.4687
維生素B2	0.0300	0.0302	0.002056	0.0038	0.0038	0.21	0.9338
維生素B6	0.1210	0.1216	0.0126	0.0087	0.0088	0.48	0.7496
維生素C	30.7948	31.0396	1.6358	0.0023	0.0023	0.12	0.9736
維生素E總量	0.5389	0.5348	0.1089	0.0328	0.0339	1.86	0.1189
維生素K1	6.3047	6.3280	0.7259	0.0107	0.0108	0.59	0.6706
總碳水化合物	5.2135	5.2343	0.5841	0.0101	0.0102	0.56	0.6935
膳食纖維	1.1658	1.1676	0.1588	0.0149	0.0151	0.83	0.5083
葉酸	14.8351	14.9518	0.8145	0.0024	0.0024	0.13	0.9702
葡萄糖	2.1719	2.1724	0.3205	0.0175	0.0178	0.98	0.4219
蔗糖	2.8390	2.8009	0.6655	0.0441	0.0462	2.53	0.0416
鈉	4.2952	3.9780	1.9023	0.1576	0.1871	10.24	<.0001
鈣	8.3280	8.3619	0.9255	0.0099	0.0100	0.55	0.7001
鉀	82.7283	82.0395	17.0724	0.0342	0.0354	1.94	0.1048
銅	0.0254	0.0253	0.004219	0.0222	0.0228	1.25	0.2926
鋅	0.3732	0.3739	0.0498	0.0143	0.0145	0.79	0.5299
錳	0.0429	0.0431	0.003293	0.0047	0.0048	0.26	0.9026
鎂	10.0170	10.0692	0.9788	0.0077	0.0077	0.42	0.7917
鐵	0.1940	0.1943	0.0260	0.0144	0.0146	0.80	0.5250
麥芽糖	0.0339	0.0341	0.002503	0.0044	0.0044	0.24	0.9146

平均 R 平方	
未加權	0.0539342
根據變異數加權	0.9207588

接著觀察交叉驗證前與後之分類錯誤表（表 十九與表 二十），亦與平均連結法相同，交叉驗證前的結果皆為第一群出現分類錯誤的情形，共有 6 個被分類到第二群；交叉驗證後的結果同樣為第一群的水果出現分類錯誤，乘上先驗機率之分類錯誤率 0.0134。

表 十九、重心方法集群之交叉驗證前分類錯誤表

分類為 CLUSTER 的觀測值數目和百分比						
來自 CLUSTER	1	2	3	4	5	總計
1	173 96.65	6 3.35	0 0.00	0 0.00	0 0.00	179 100.00
2	0 0.00	25 100.00	0 0.00	0 0.00	0 0.00	25 100.00
3	0 0.00	0 0.00	11 100.00	0 0.00	0 0.00	11 100.00
4	0 0.00	0 0.00	0 0.00	7 100.00	0 0.00	7 100.00
5	0 0.00	0 0.00	0 0.00	0 0.00	2 100.00	2 100.00
總計	173 77.23	31 13.84	11 4.91	7 3.13	2 0.89	224 100.00
事前	0.2	0.2	0.2	0.2	0.2	

CLUSTER 的誤差計數估計值						
	1	2	3	4	5	總計
比率	0.0335	0.0000	0.0000	0.0000	0.0000	0.0067
事前	0.2000	0.2000	0.2000	0.2000	0.2000	

表 二十、重心方法集群之交叉驗證後分類錯誤表

分類為 CLUSTER 的觀測值數目和百分比						
來自 CLUSTER	1	2	3	4	5	總計
1	167 93.30	12 6.70	0 0.00	0 0.00	0 0.00	179 100.00
2	0 0.00	25 100.00	0 0.00	0 0.00	0 0.00	25 100.00
3	0 0.00	0 0.00	11 100.00	0 0.00	0 0.00	11 100.00
4	0 0.00	0 0.00	0 0.00	7 100.00	0 0.00	7 100.00
5	0 0.00	0 0.00	0 0.00	0 0.00	2 100.00	2 100.00
總計	167 74.55	37 16.52	11 4.91	7 3.13	2 0.89	224 100.00
事前	0.2	0.2	0.2	0.2	0.2	

CLUSTER 的誤差計數估計值						
	1	2	3	4	5	總計
比率	0.0670	0.0000	0.0000	0.0000	0.0000	0.0134
事前	0.2000	0.2000	0.2000	0.2000	0.2000	

### 三、Ward 最小變異數方法集群之判別結果

Ward 最小變異數方法的集群資訊列示於表 二十一，第一群的水果種類最多，有 162 種，第五群內僅兩種水果最少，據此資料計算出的判別函數呈現於表 二十二。在設定顯著水準為 0.05 下，表 二十三的檢定結果皆呈現顯著，顯示此判別函數在不同群集之間的差異是顯著的，能夠有效的區分不同群集。

表 二十一、Ward 最小變異數方法集群之類別層級資訊

類別層級資訊					
CLUSTER	變數名稱	次數	加權	比例	先驗機率
1	1	162	162.0000	0.723214	0.200000
2	2	38	38.0000	0.169643	0.200000
3	3	15	15.0000	0.066964	0.200000
4	4	7	7.0000	0.031250	0.200000
5	5	2	2.0000	0.008929	0.200000

表 二十二、Ward 最小變異數方法集群之線性判別函數

CLUSTER的線性判別函數					
變數	1	2	3	4	5
常數	-460690	-460478	-459842	-459887	-458608
果糖	44.00255	48.71357	51.54374	50.54961	51.39145
水分	9218	9215	9208	9208	9190
熱量	632.60219	632.13816	630.85846	630.52405	625.79185
粗脂肪	4058	4059	4064	4068	4094
粗蛋白	7356	7355	7349	7348	7344
糖質總量	-65.41093	-69.80430	-71.49858	-69.51436	-71.29595
維生素A總量(IU)	-0.72979	-0.68421	-0.61839	-0.54697	-0.38528
維生素B1	-7322	-7307	-7293	-7332	-7317
維生素B2	4366	4377	4375	4360	4346
維生素B6	-936.16221	-940.51679	-950.60921	-955.15373	-974.83113
維生素C	5.38206	5.38101	5.37383	5.39171	5.39207
維生素E總量	-47.85853	-48.03217	-49.52593	-52.29495	-57.13465
維生素K1	-2.96447	-2.84072	-2.88354	-2.83634	-2.69414
總碳水化合物	6958	6958	6956	6957	6958
膳食纖維	153.86301	153.29448	152.40695	150.77716	147.77497
葉酸	-10.37857	-10.32680	-10.20162	-10.11198	-9.92980
葡萄糖	152.80402	156.61069	156.79300	153.27438	155.12416
蔗糖	86.18744	90.25984	91.63023	89.40725	89.98564
鈉	42.72418	42.80318	42.80447	42.82419	43.69505
鈣	12.12340	12.15411	12.17280	12.38501	12.44705
鉀	12.70435	12.69415	12.70029	12.71994	12.66738
銅	-1764	-1765	-1730	-1736	-1721
鋅	275.27217	275.69399	277.60550	277.88393	280.79093
錳	-58.83974	-78.52536	-113.43604	-137.41459	-193.13928
鎂	25.30639	25.31780	25.26568	25.21943	25.26974
鐵	-93.59480	-93.64002	-94.30527	-97.00827	-99.50737
麥芽糖	2120	2112	2095	2095	2079

表二十三、Ward 最小變異數方法集群之多變量檢定結果

多變量統計值和 F 近似					
S=4 M=11 N=95.5					
統計值	值	F 值	分子自由度	分母自由度	Pr > F
Wilks 的 Lambda	0.02296988	11.30	108	768.68	<.0001
Pillai 追蹤	1.29587553	3.48	108	784	<.0001
Hotelling-Lawley 追蹤	30.02554933	53.27	108	645.54	<.0001
Roy 最大根	29.64990606	215.24	27	196	<.0001
附註: Roy 最大根的 F 統計值是上限。					

若觀察單變量檢定統計值（表二十四），發現所有變數僅維生素 A、維生素 E、蔗糖、鈉呈現顯著，顯示這四個營養成分在不同群集之間存在顯著的平均值差異，將是區分不同群集的重要指標。

表二十四、Ward 最小變異數方法集群之單變量檢定統計值

單變量檢定統計值							
F 統計值、Num DF=4、Den DF=219							
變數	總標準差	集區標準差	標準差之間	R 平方	R 平方 / (1-RSq)	F 值	Pr > F
果糖	2.3981	2.3911	0.4114	0.0236	0.0242	1.33	0.2612
水分	5.4948	5.5238	0.5320	0.0075	0.0076	0.42	0.7974
熱量	20.1164	20.2157	2.0354	0.0082	0.0083	0.45	0.7693
粗脂肪	0.8239	0.8305	0.0415	0.0020	0.0020	0.11	0.9782
粗蛋白	0.3870	0.3900	0.0216	0.0025	0.0025	0.14	0.9683
糖質總量	5.6252	5.6115	0.9456	0.0227	0.0232	1.27	0.2818
維生素A總量(IU)	565.7332	110.3694	619.1888	0.9626	25.7539	1410.03	<.0001
維生素B1	0.0248	0.0248	0.003836	0.0192	0.0195	1.07	0.3721
維生素B2	0.0300	0.0302	0.002040	0.0037	0.0037	0.20	0.9356
維生素B6	0.1210	0.1217	0.0115	0.0073	0.0073	0.40	0.8077
維生素C	30.7948	31.0007	2.3713	0.0048	0.0048	0.26	0.9020
維生素E總量	0.5389	0.5280	0.1436	0.0570	0.0605	3.31	0.0117
維生素K1	6.3047	6.2522	1.3014	0.0342	0.0355	1.94	0.1047
總碳水化合物	5.2135	5.2419	0.4942	0.0072	0.0073	0.40	0.8098
膳食纖維	1.1658	1.1695	0.1412	0.0118	0.0119	0.65	0.6256
葉酸	14.8351	14.9205	1.3441	0.0066	0.0066	0.36	0.8344
葡萄糖	2.1719	2.1641	0.3831	0.0250	0.0256	1.40	0.2337
蔗糖	2.8390	2.8087	0.6239	0.0388	0.0404	2.21	0.0688
鈉	4.2952	3.9828	1.8899	0.1556	0.1842	10.09	<.0001
鈣	8.3280	8.3453	1.0931	0.0138	0.0140	0.77	0.5466
鉀	82.7283	82.5646	13.6326	0.0218	0.0223	1.22	0.3026
銅	0.0254	0.0253	0.004304	0.0231	0.0237	1.30	0.2721
鋅	0.3732	0.3756	0.0298	0.0051	0.0051	0.28	0.8897
錳	0.0429	0.0431	0.004064	0.0072	0.0073	0.40	0.8096
鎂	10.0170	10.1007	0.4263	0.0015	0.0015	0.08	0.9885
鐵	0.1940	0.1940	0.0291	0.0181	0.0184	1.01	0.4041
麥芽糖	0.0339	0.0340	0.003088	0.0067	0.0067	0.37	0.8309

接著觀察交叉驗證前與後之分類錯誤表（表 二十五與表 二十六），交叉驗證前的結果僅第一群有兩種水果被錯分到第二群，為目前表現最佳的集群；交叉驗證後的結果在第一群跟第三群各有一種水果被分類錯誤，乘上先驗機率之分類錯誤率 0.0146，約略高於平均連結法與重心方法。

表 二十五、Ward 最小變異數方法集群之交叉驗證前分類錯誤表

分類為 CLUSTER 的觀測值數目和百分比						
來自 CLUSTER	1	2	3	4	5	總計
1	161 99.38	1 0.62	0 0.00	0 0.00	0 0.00	162 100.00
2	0 0.00	38 100.00	0 0.00	0 0.00	0 0.00	38 100.00
3	0 0.00	0 0.00	15 100.00	0 0.00	0 0.00	15 100.00
4	0 0.00	0 0.00	0 0.00	7 100.00	0 0.00	7 100.00
5	0 0.00	0 0.00	0 0.00	0 0.00	2 100.00	2 100.00
總計	161 71.88	39 17.41	15 6.70	7 3.13	2 0.89	224 100.00
事前	0.2	0.2	0.2	0.2	0.2	

CLUSTER 的誤差計數估計值						
	1	2	3	4	5	總計
比率	0.0062	0.0000	0.0000	0.0000	0.0000	0.0012
事前	0.2000	0.2000	0.2000	0.2000	0.2000	

表 二十六、Ward 最小變異數方法集群之交叉驗證後分類錯誤表

分類為 CLUSTER 的觀測值數目和百分比						
來自 CLUSTER	1	2	3	4	5	總計
1	161 99.38	1 0.62	0 0.00	0 0.00	0 0.00	162 100.00
2	0 0.00	38 100.00	0 0.00	0 0.00	0 0.00	38 100.00
3	0 0.00	0 0.00	14 93.33	1 6.67	0 0.00	15 100.00
4	0 0.00	0 0.00	0 0.00	7 100.00	0 0.00	7 100.00
5	0 0.00	0 0.00	0 0.00	0 0.00	2 100.00	2 100.00
總計	161 71.88	39 17.41	14 6.25	8 3.57	2 0.89	224 100.00
事前	0.2	0.2	0.2	0.2	0.2	

CLUSTER 的誤差計數估計值						
	1	2	3	4	5	總計
比率	0.0062	0.0000	0.0667	0.0000	0.0000	0.0146
事前	0.2000	0.2000	0.2000	0.2000	0.2000	



#### 四、K 平均值演算法集群之判別結果

K 平均值演算法的集群資訊列示於表 二十七，第二群的水果種類最多，有 161 種，第三群內僅兩種水果最少，據此資料計算出的判別函數呈現於表 二十八。在設定顯著水準為 0.05 下，表 二十九的檢定結果皆呈現顯著，顯示此判別函數在不同群集之間的差異是顯著的，能夠有效的區分不同群集。惟此處的群集標籤與表 八不同，表 八有重新定義標籤，此處則否，第一群至第五群皆為原始編號。

表 二十七、K 平均值演算法集群之類別層級資訊

類別層級資訊					
CLUSTER	變數名稱	次數	加權	比例	先驗機率
1	1	11	11.0000	0.049107	0.200000
2	2	161	161.0000	0.718750	0.200000
3	3	2	2.0000	0.008929	0.200000
4	4	7	7.0000	0.031250	0.200000
5	5	43	43.0000	0.191964	0.200000

表 二十八、K 平均值演算法集群之線性判別函數

CLUSTER的線性判別函數					
變數	1	2	3	4	5
常數	-457560	-457426	-457881	-457996	-457532
果糖	106.14100	88.43821	119.92027	107.73881	97.00141
水分	9152	9151	9150	9155	9152
熱量	618.99162	620.95356	614.30050	619.00659	620.51622
粗脂肪	4117	4100	4157	4121	4105
粗蛋白	7304	7300	7315	7306	7302
糖質總量	-115.10483	-97.13715	-131.02201	-116.08756	-105.64513
維生素A總量(IU)	0.08786	-0.02428	0.29563	0.14383	0.02120
維生素B1	-7136	-7144	-7192	-7182	-7134
維生素B2	4377	4349	4360	4360	4360
維生素B6	-1066	-1049	-1092	-1070	-1056
維生素C	5.37423	5.35655	5.44107	5.40607	5.36889
維生素E總量	-57.46530	-58.03422	-60.65241	-58.87515	-57.09498
維生素K1	-2.77693	-2.91049	-2.50658	-2.70670	-2.76989
總碳水化合物	6948	6940	6964	6952	6943
膳食纖維	142.39743	143.03935	138.82099	140.95708	142.53336
葉酸	-9.15004	-9.16525	-9.15017	-9.13200	-9.16667
葡萄糖	185.55627	167.75531	202.63582	185.52104	176.00741
蔗糖	131.20939	113.93589	145.70126	131.99030	122.08419
鈉	43.00638	43.05891	43.72804	43.00186	43.10418
鈣	12.38110	12.34426	12.61904	12.57980	12.36565
鉀	12.72498	12.72318	12.69870	12.74484	12.71497
銅	-1416	-1496	-1365	-1423	-1485
鋅	296.32552	293.46831	299.70665	296.39115	294.14331
錳	-463.44366	-418.23066	-516.67709	-476.84025	-434.81283
鎂	25.23032	25.01574	25.63093	25.27635	25.12265
鐵	-92.03496	-94.51580	-91.30575	-93.05418	-92.89915
麥芽糖	1927	1950	1919	1930	1942

表 二十九、K 平均值演算法集群之多變量檢定結果

多變量統計值和 F 近似					
S=4 M=11 N=95.5					
統計值	值	F 值	分子自由度	分母自由度	Pr > F
Wilks 的 Lambda	0.02485891	10.94	108	768.68	<.0001
Pillai 追蹤	1.29735906	3.48	108	784	<.0001
Hotelling-Lawley 追蹤	27.57268898	48.91	108	645.54	<.0001
Roy 最大根	27.19222335	197.40	27	196	<.0001
附註: Roy 最大根的 F 統計值是上限。					

若觀察單變量檢定統計值（表 三十），發現所有變數僅維生素 A、維生素 E、鈉呈現顯著，顯示這三個營養成分在不同群集之間存在顯著的平均值差異，將是區分不同群集的重要指標。

表 三十、K 平均值演算法集群之單變量檢定統計值

單變量檢定統計值							
F 統計值、Num DF=4、Den DF=219							
變數	總標準差	集區標準差	標準差之間	R 平方	R 平方 / (1-RSq)	F 值	Pr > F
果糖	2.3981	2.4024	0.3208	0.0144	0.0146	0.80	0.5269
水分	5.4948	5.5248	0.5197	0.0072	0.0072	0.40	0.8111
熱量	20.1164	20.2193	1.9906	0.0079	0.0079	0.43	0.7838
粗脂肪	0.8239	0.8305	0.0422	0.0021	0.0021	0.12	0.9769
粗蛋白	0.3870	0.3905	0.007093	0.0003	0.0003	0.01	0.9996
糖質總量	5.6252	5.6264	0.8308	0.0175	0.0178	0.98	0.4211
維生素A總量(IU)	565.7332	112.9927	618.6102	0.9608	24.5260	1342.80	<.0001
維生素B1	0.0248	0.0248	0.003755	0.0184	0.0187	1.02	0.3956
維生素B2	0.0300	0.0302	0.002179	0.0042	0.0043	0.23	0.9192
維生素B6	0.1210	0.1217	0.0119	0.0078	0.0078	0.43	0.7878
維生素C	30.7948	31.0379	1.6740	0.0024	0.0024	0.13	0.9712
維生素E總量	0.5389	0.5255	0.1545	0.0661	0.0708	3.88	0.0046
維生素K1	6.3047	6.2707	1.1873	0.0285	0.0293	1.61	0.1738
總碳水化合物	5.2135	5.2427	0.4839	0.0069	0.0070	0.38	0.8216
膳食纖維	1.1658	1.1693	0.1432	0.0121	0.0123	0.67	0.6122
葉酸	14.8351	14.9368	1.1010	0.0044	0.0044	0.24	0.9134
葡萄糖	2.1719	2.1627	0.3926	0.0263	0.0270	1.48	0.2102
蔗糖	2.8390	2.8131	0.5994	0.0358	0.0372	2.03	0.0907
鈉	4.2952	3.9852	1.8838	0.1546	0.1828	10.01	<.0001
鈣	8.3280	8.3634	0.9078	0.0095	0.0096	0.53	0.7154
鉀	82.7283	82.3980	14.8133	0.0258	0.0264	1.45	0.2192
銅	0.0254	0.0252	0.004860	0.0295	0.0304	1.66	0.1592
鋅	0.3732	0.3747	0.0412	0.0098	0.0099	0.54	0.7049
錳	0.0429	0.0431	0.004109	0.0074	0.0074	0.41	0.8031
鎂	10.0170	10.0959	0.5488	0.0024	0.0024	0.13	0.9704
鐵	0.1940	0.1934	0.0333	0.0237	0.0242	1.33	0.2611
麥芽糖	0.0339	0.0340	0.003122	0.0068	0.0069	0.38	0.8250



接著觀察交叉驗證前與後之分類錯誤表（表 三十一與表 三十二），交叉驗證前的結果竟沒有任何分類錯誤的情形；交叉驗證後的結果則分別在第一群與第五群水果出現分類錯誤，共有 4 個分錯，乘上先驗機率之分類錯誤率 0.0321。

表 三十一、K 平均值演算法集群之交叉驗證前分類錯誤表

分類為 CLUSTER 的觀測值數目和百分比						
來自 CLUSTER	1	2	3	4	5	總計
1	11 100.00	0 0.00	0 0.00	0 0.00	0 0.00	11 100.00
2	0 0.00	161 100.00	0 0.00	0 0.00	0 0.00	161 100.00
3	0 0.00	0 0.00	2 100.00	0 0.00	0 0.00	2 100.00
4	0 0.00	0 0.00	0 0.00	7 100.00	0 0.00	7 100.00
5	0 0.00	0 0.00	0 0.00	0 0.00	43 100.00	43 100.00
總計	11 4.91	161 71.88	2 0.89	7 3.13	43 19.20	224 100.00
事前	0.2	0.2	0.2	0.2	0.2	

CLUSTER 的誤差計數估計值						
	1	2	3	4	5	總計
比率	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
事前	0.2000	0.2000	0.2000	0.2000	0.2000	

表 三十二、K 平均值演算法集群之交叉驗證後分類錯誤表

分類為 CLUSTER 的觀測值數目和百分比						
來自 CLUSTER	1	2	3	4	5	總計
1	10 90.91	0 0.00	0 0.00	1 9.09	0 0.00	11 100.00
2	0 0.00	161 100.00	0 0.00	0 0.00	0 0.00	161 100.00
3	0 0.00	0 0.00	2 100.00	0 0.00	0 0.00	2 100.00
4	0 0.00	0 0.00	0 0.00	7 100.00	0 0.00	7 100.00
5	2 4.65	1 2.33	0 0.00	0 0.00	40 93.02	43 100.00
總計	12 5.36	162 72.32	2 0.89	8 3.57	40 17.86	224 100.00
事前	0.2	0.2	0.2	0.2	0.2	

CLUSTER 的誤差計數估計值						
	1	2	3	4	5	總計
比率	0.0909	0.0000	0.0000	0.0000	0.0698	0.0321
事前	0.2000	0.2000	0.2000	0.2000	0.2000	

## 五、四種方法下之判別結果比較

經上述分析得知，四種方法集群結果進行判別分析之多變量檢定結果皆呈現顯著，代表四種判別函數在不同群集之間的差異皆是顯著的，都能夠有效的區分不同群集，差異僅在於單變量檢定統計值呈現顯著的營養成分不同，以及交叉驗證前與後之分類錯誤率各有優劣。

表 三十三彙整四種方法集群結果之單變量檢定統計值呈現顯著的營養成分，可以發現維生素 A 與鈉皆是重要的區別變數，部分方法下會包含維生素 E 或蔗糖。

表 三十三、集群結果之單變量檢定統計值顯著營養成分彙整

方法	維生素 A	維生素 E	蔗糖	鈉
平均連結法	V		V	V
重心方法	V		V	V
Ward 最小變異數方法	V	V	V	V
K 平均值演算法	V	V		V

接著彙整四種方法集群結果之判別分析交叉驗證前與後分類錯誤率（表 三十四），以列表表示不同分群方法，以欄表示五群交叉驗證前後分類錯誤的比例，能夠觀察到多數為第一群出現分類錯誤，整體表現可以著重在最後兩欄——執行交叉驗證前，所有方法以 K 平均值演算法表現最佳，完全沒有分類錯誤，Ward 最小變異數方法以分類錯誤率 0.0012 次之；執行交叉驗證後，平均連結法與重心方法最佳，分類錯誤率 0.0134，Ward 最小變異數方法以 0.1% 左右的差距緊追在後，分類錯誤率 0.0146，K 平均值演算法此時反倒最差。因此，本研究認為 Ward 最小變異數方法整體表現最好，後續將以其分群結果觀察各群特徵並作結論。

表 三十四、判別分析之交叉驗證前與後分類錯誤率彙整

CLUSTER	1		2		3		4		5		總計	
交叉驗證	前	後	前	後	前	後	前	後	前	後	前	後
平均連結法	0.0335	0.067	0	0	0	0	0	0	0	0	0.0067	0.0134
重心方法	0.0335	0.067	0	0	0	0	0	0	0	0	0.0067	0.0134
Ward	0.0062	0.0062	0	0	0	0.0667	0	0	0	0	0.0012	0.0146
K 平均值	0	0.0909	0	0	0	0	0	0	0	0.0698	0	0.0321
事前	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2		

## 伍、結論

本研究欲將 224 種水果進行分群，嘗試了平均連結法、重心方法、Ward 最小變異數方法與 K 平均值演算法共四種集群分析，並分別根據集群分析給定的群集做判別分析驗證，認為 Ward 最小變異數方法的分群效果最佳，因此，以下將針對此方法的分群結果做探索與總結。

首先，表 三十五列示第一群至第五群的所有水果，但水果種類眾多、不易觀察，本研究另外繪製各個群集水果名稱的文字雲（圖 四），以便根據水果名稱的字體大小了解其出現頻率。

表 三十五、五群水果列表

第一群			第二群	第三群	第四群	第五群
共 162 種			共 38 種	共 15 種	共 7 種	共 2 種
人心果	芭樂平均值	無花果	木瓜(11 月取樣)	七股香洋香瓜	木瓜(5 月取樣)	新疆哈密瓜
土芭樂	(白肉)	菠蘿蜜				嘉寶瓜
大雪梨	金冠蘋果	萊姆	木瓜平均值	土芒果	芒果平均值	
小玉西瓜	金香葡萄(含皮)	萊姆汁	牛心柿	四周柿	(西洋種)	
山竹	金黃奇異果	黃皮葡萄柚	加州紅李(進口)	四季芒果	青龍蘋果	
天香洋香瓜	金蜜芒果	黃皮葡萄柚(進口)	加州蜜李(進口)	百香果	紅寶石洋香瓜	
太陽洋香瓜	金蘭西瓜	黃肉李	百香果汁	芒果平均值(新興種)	黑香芒果	
文旦	金鑽鳳梨	黃金果	西瓜平均值(紅肉小瓜)	枇杷	愛文芒果	
木瓜(2 月取樣)	青木瓜	黃香瓜	李子平均值(紅皮紅肉)	金煌芒果	藍寶石洋香瓜	
木瓜(8 月取樣)	青皮葡萄柚	黑李	李子平均值(紅皮黃肉)	金興芒果		
水蜜桃	南華蕉	黑金剛蓮霧	玫瑰桃	美國綠葡萄		
水蜜桃平均值	柳橙	黑珍珠蓮霧	肯特芒果	香水芒果		
牛奶鳳梨	津經蜜蘋果	黑美人西瓜	長果金柑	甜瓜平均值(網紋洋香瓜)		
牛乳芭樂	珍珠芭樂	黑葉仔荔枝	皇宮西瓜	甜柿(進口)		
世紀芭樂	珍珠蜜棗	圓果金柑	秋香洋香瓜	鳳光西瓜		
加州青李	秋香蘋果	圓果金柑汁(綠皮)	紅心芭樂	蘋果平均值(青皮)		
加州紅李(台灣)	突目 1 號鳳梨	意大利葡萄	紅皮蕉(2004 年取樣)			
加拉蘋果	梨	新世紀梨				
	紅毛丹	新興梨				
	紅甘蔗汁	楊桃(11 月取樣)				
	紅皮蕉(2012 年取樣)					

北蕉(0 天,綠皮)	紅柑	楊桃(2 月取樣)	紅肉李(大)			
北蕉(11 月取樣)	紅龍果(白肉)	楊桃(4 月取樣)	紅肉李(小)			
北蕉(1 天)	紅龍果(紅肉)	楊桃(5 月取樣)	紅鈴西瓜			
北蕉(2 月取樣)	紅寶石蓮霧	楊桃(9 月取樣)	紅寶石葡萄			
北蕉(3 天)	美國五爪蘋果	楊桃平均值	柚(古坑)			
北蕉(5 月取樣)	美國紅葡萄(含皮)	酪梨(室溫存放 0 天)	茂谷柑			
北蕉(7 天)	美國紅葡萄(含皮)	酪梨(室溫存放 3 天)	桃接李(大)			
北蕉(9 月取樣)	美國紫葡萄(含皮)	酪梨(室溫存放 6 天)	桃接李(小)			
北蕉平均值	美國黑葡萄(含皮)	嘉玉甜瓜	海梨桶柑			
可可椰子汁(屏東)	美國綠葡萄(含皮)	榴槤	高牆桶柑			
可可椰子汁(進口)	美濃瓜	福壽桃	甜柿平均值			
可可椰子汁平均值	香水李	翠玉青蘋果(帶皮)	蛋黃果			
台灣土棗	香瓜茄	翡翠李	凱特芒果			
巨峰葡萄	香吉士(進口)	蜜世界洋香瓜	極柑			
玉文芒果	香華洋香瓜	蜜棗(圓形)	無子西瓜			
玉荷包荔枝	桑葚	蜜棗李(進口)	筆柿			
甘蔗鳳梨	桑葚汁	鳳梨平均值(雜交種)	華寶大西瓜(花蓮)			
白油桃	泰國芭樂	鳳梨釋迦	黑美人葡萄(含皮)			
白柚	胭脂梅	蓮霧平均值(粉紅色種)	新世紀哈密瓜			
白櫻桃	脆桃平均值	橫山梨	聖心芒果			
印度棗(長形)	草莓	澳洲甜橙(進口)	酪梨(綠皮)			
印度棗平均值(大)	荔枝平均值		綠葡萄平均值			
安石榴	國產甜柿		蓮花芒果			
西瓜平均值(黃肉小瓜)	將軍蜜梨					
	梨山蜜梨					
	甜瓜平均值(光皮洋香瓜)					
	甜瓜平均值(東洋系脆瓜)					

西施蜜柚	甜蜜桃	龍眼				
西洋梨(青皮)	甜蜜蜜鳳梨	檸檬				
西洋梨(紅皮)	甜橙平均值	檸檬汁(黃皮)				
西洋梨(黃皮)	(普遍系)	檸檬汁(綠皮)				
西洋梨平均值	富士蘋果	檸檬汁平均值				
冷凍荔枝	富士蘋果(帶皮)	檸檬柑				
李子平均值	富有甜柿(大雪山)	蟠桃				
(青皮黃肉)	無子芭樂	豐水梨				
李林蕉(2007年取樣)	無子紅葡萄(含皮)	蘋果平均值(混色)				
李林蕉(2012年取樣)		釋迦				
奇異果		櫻桃				
幸水梨		鶯哥桃				
狀元瓜						

圖四第一列由左至右的圖為第一群至第三群水果的文字雲，第二列由左至右的圖則為第四與第五群，可以觀察到第一群有許多漿果類的水果，諸如香蕉、鳳梨、楊桃等；二三四群則是注意到皆有出現芒果，如第四群的香水芒果與第五群的愛文芒果，表示不同品種的芒果富含不同的營養成分，這樣的差異甚至能將它們分成不同群；第五群只有嘉寶瓜與新疆哈密瓜兩種水果，可見他們的營養成分與其他水果相比特別不同。



圖四、五群水果文字雲

本研究想進一步了解各群營養成分差異，故繪製 **Error! Reference source not found.** 的盒鬚圖（未呈現所有分析之營養成分，只呈現其中 12 個），每張圖的灰色窗格代表該營養成分在第一群至第五群的分佈，其中，左上角的維生素 A 分布差異最明顯，與表 三十三列示之單變量檢定統計值顯著之營養成分相同，其他營養成分有出現較明顯差異的群集，將以紅色箭頭標示出來。



圖 五、五群水果部分營養成分盒鬚圖

綜上所述，本研究認為各個群集水果特色分別為：第一群維生素 A 最低、葡萄糖高、漿果類水果多；第二群維生素 A 次低、蔗糖偏低、鈉偏低；第三群維生素 A 居中、維生素 B1 多、維生素 B6 多、碳水化合物多；第四群維生素 A 次高、維生素 E 多、蔗糖多、鈣多、鎂多；第五群維生素 A 最高、維生素 B1 少、碳水化合物少、鉀少、種類少。期望能提供食用水果的大眾參考。