



# 運用常見蔬菜與水果的資料屬性之分群技術應用

指導教授: 廖強棋 老師

組員: 杜易軒、溫翔杰、邱建勝、邱建順

## 一、摘要

這個研究是蔬菜與水果的資料分析，首先我們蒐集了許多網路上的資料，然後我們統計分析結果如礦物質是大部分蔬果都有的，且有一半的水果具有抗癌效果。接著我們使用資料探勘中的分群方法，分析結果第0群維生素C跟維生素B還有礦物質佔的比例較多，有檸檬。第1群維生素C跟預防心臟病的比例較高，有西瓜。第2群維生素A跟高血壓佔的比例較多，有紅蜜芒果。再來是第3群膳食纖維跟預防癌症比例較多，有文旦。最後是第4群維生素C跟心血管疾病還有預防膽固醇過高的比例較多，有覆盆子。

## 二、技術說明：

我們用weka軟體[1]對蔬果進行大數據分析，蔬果大全[2]則是找尋一些蔬果得資料，維基百科[3]則是把蔬果資料弄完整，食品營養成分[4]則是搜尋常見水果的養標示、預防疾病[5]則是查詢蔬果可預防的病種類。

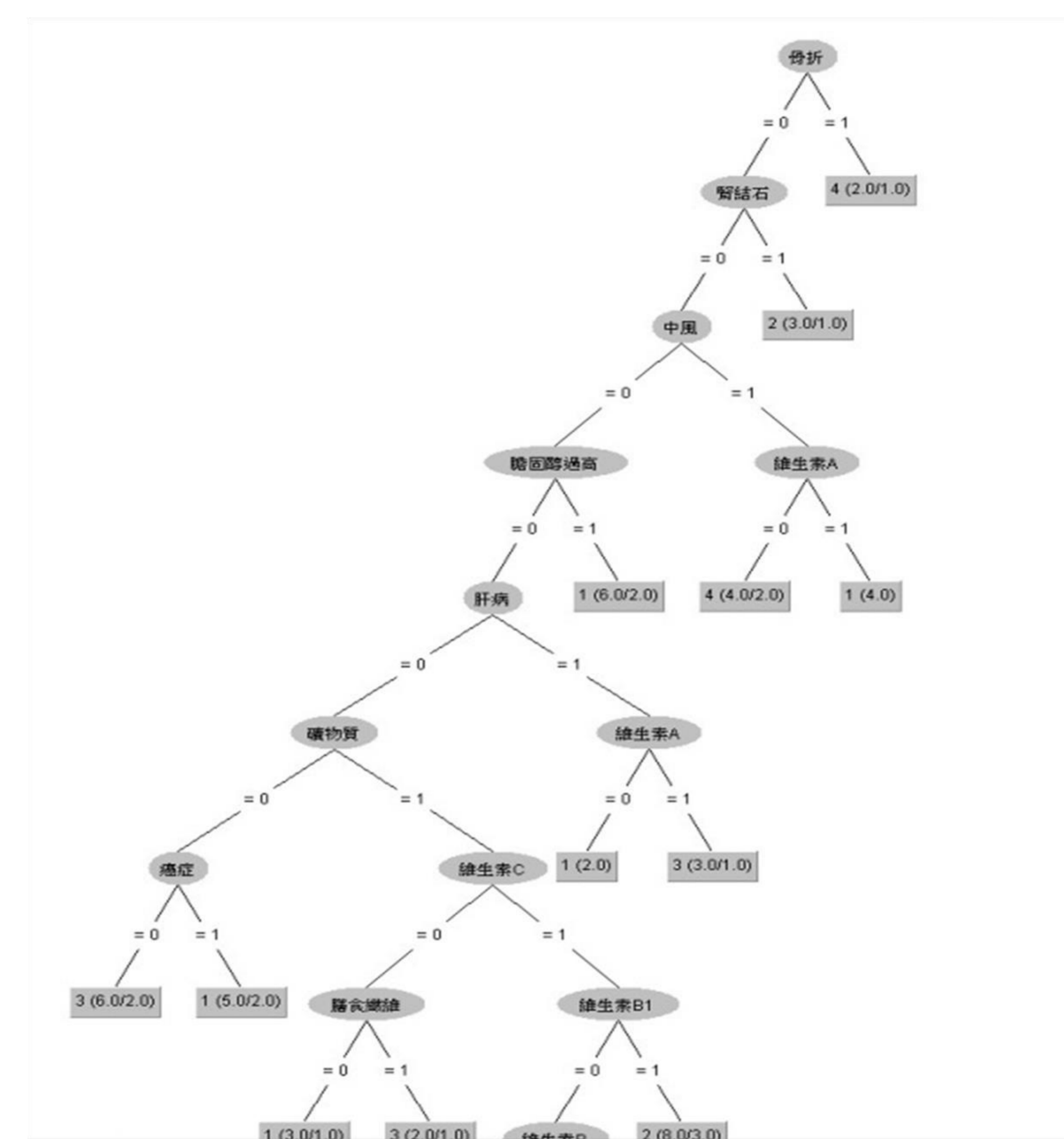
## 三、創新性說明：

運用資料探勘的方法區分析蔬菜水果的資料  
第一點我們利用資料探勘的技術來分析  
第二點蒐集蔬菜果的資料來製作大數據  
第三點我們用網頁呈現分群結果

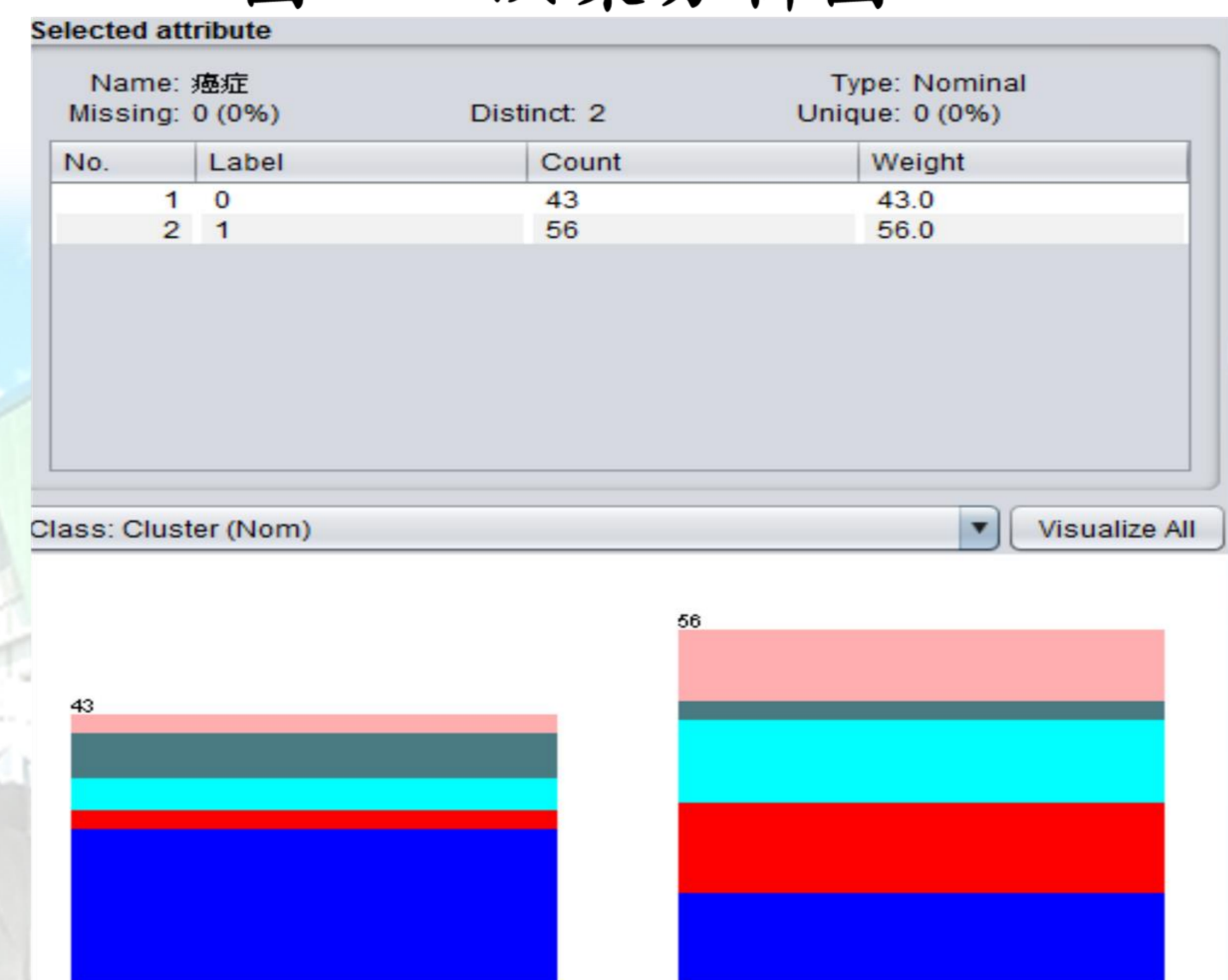
Final cluster centroids:

Attribute	Full Data (98.0)	Cluster#				
		0 (10.0)	1 (31.0)	2 (38.0)	3 (17.0)	4 (2.0)
aa	白柚	柑橘	白柚	檸檬	西瓜	麻豆文旦
維生素A	0.3673	0.3	0.8065	0.1579	0.1176	0
維生素B	0.6224	0.8	0.6774	0.7895	0.1176	0
維生素B1	0.2347	0.8	0.0645	0.0263	0.5882	1
維生素B2	0.2245	0.8	0.0645	0.0789	0.4118	1
維生素B6	0.0102	0	0	0.0263	0	0
維生素B12	0.0204	0	0	0	0.1176	0
維生素C	0.9184	0.7	0.9677	0.9474	0.8824	1
維生素D	0.0408	0.1	0.0968	0	0	0
維生素E	0.1837	0	0.3871	0.1053	0.1176	0
維生素K	0.0204	0	0.0323	0.0263	0	0
維生素P	0.1122	0.1	0.129	0.1316	0	0.5
紅蘿蔔素	0.3571	0.7	0.3548	0.3421	0.1765	0.5
礦物質	0.8469	0.9	0.9032	0.9474	0.5882	0
脂肪	0.0102	0	0	0.0263	0	0
膳食纖維	0.3878	0.6	0.5806	0.2632	0.1176	1
蛋白質	0.102	0.1	0.2581	0	0.0588	0
糖尿病	0.102	0.2	0.0968	0.1316	0	0
高血壓	0.4694	0.5	0.7419	0.1842	0.5882	0.5
心臟病	0.3061	0.3	0.3871	0.1053	0.6471	0
氣喘	0	0	0	0	0	0
中風	0.0918	0.1	0.129	0.0789	0.0588	0
貧血	0.0408	0	0.0645	0.0263	0.0588	0
腎結石	0.0306	0.1	0.0323	0	0.0588	0
尿失禁	0.0102	0	0.0323	0	0	0
動脈硬化	0.2449	0.4	0.4194	0.0789	0.2353	0
血管阻塞	0.5408	1	0.6190	0.5	0.2041	0

圖一、分群分析圖



圖二、決策分析圖



圖三、癌症統計圖