

國立臺灣大學工學院土木工程學系

碩士論文

Department of Civil Engineering

College of Engineering

National Taiwan University

Master Thesis

資料探勘於交通事故之應用－以大客車為例

Data Mining in Traffic Accident—A Case Study for Bus

王宜慶

Wang, I-Ching

指導教授：許添本 博士

Major Professor: Hsu, Tien-Pen, Ph.D.

中華民國 98 年 6 月

June, 2009

國立臺灣大學碩士學位論文  
口試委員會審定書

資料探勘於交通事故之應用

-以大客車為例

Data Mining in Traffic Accident

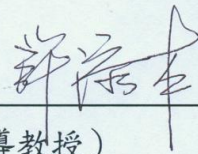
— A Case Study for Bus

本論文係王宜慶君（R96521518）在國立臺灣大學土木工程學系碩士班完成之碩士學位論文，於民國 98 年 6 月 26 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

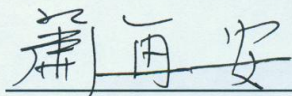
許添本

（指導教授）

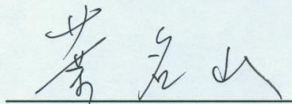


（簽名）

蕭再安



葉名山



系主任



（簽名）

## 誌謝

當我寫下誌謝時，即是 Happy Time！代表著碩士生涯將告一段落，也是另一個人生的開始。在求學過程中，經歷了許多喜怒哀樂，也因為這樣讓我一點一滴的成長，雖然無法一一細數在學生時代中幫助及支持我的師長朋友們，但透過這篇誌謝，向各位說聲「謝謝」！

首先感謝我的指導教授許添本老師，在這兩年的碩士生涯中，讓我了解什麼是作研究的態度及思考邏輯，透過老師紮實的訓練及細心的指導，使學生都謹記在心！另外，感謝口試委員葉名山老師及蕭再安老師，給予寶貴的指正與建議，使論文內容更加完善。在求學期間，感謝交通組全體師長，特別是曹壽民老師、龍天立老師、周義華老師及張學孔老師，除了課業上的指導外，也分享許多重要人生經驗與觀念。此外，我也要感謝廖久緯老師、張慧老師、黃文隆老師，從國中、五專到研究所前的階段，有你們從中的指導、幫助、支持及鼓勵且亦師亦友的關係，使我的行為及思想也有所成長，也讓我從國中的毛頭小子轉變為即將出社會的新鮮人，只能說你們對我來說都是缺一不可，特此感謝！

碩士生涯中，要特別同門師兄姐：感謝育瑞學長在口試前指導我論文上的缺失，提出您的看法，使我論文更加完備；感謝將瓊學長，當我有問題時，向您請教都會很有耐心且鉅細靡遺的說出您的看法，使我了解作研究裡的該有的細節；感謝佳紋學姐和晉瑜學姐，帶我進入作研究的領域，從機動力的計劃團隊裡，讓我知道如何作研究，也不厭其煩的糾正常犯錯的我，使我不斷的成長。另外要感謝的宏仁學長和上民學長，幫我針灸治療及貼豆，使我的身體在研究的摧殘下，依舊可以恢復該有的精氣神，另外宏仁學長在 112 無人時，會和我分享他的人生經驗談，使我有不同角度的思考，我都會銘記在心。最後，要特別感謝的學長就是駿哥，不管在課業上、生活上及做人處事的道理，都讓我在這兩年內成長了許多，我們一起打壘球、打撞球、打屁…等，有了你這個老經驗，讓我在碩士生涯裡一點也不無聊，從我心底對你有太多的感激及尊敬，只能說有你的，你真的夠屌！

我想說來台大，應該無法在打大交盃了，但有了交通組壘球隊友們：駿哥、阿維、kamin、AD、小民、kona、阿香、阿儒、阿坤、中庭、老頭、崇宇，使我可以繼續打大交盃，雖然我們被其他學校覺得是弱雞，但我們在大交盃、北交盃、台大盃時，打了一場又一場的好球，雖然在關鍵的晉級賽都會輸掉，但我只能說我們太專注於研究，要不然我們的成績絕對不只這樣而已。另外，蘇龜、乙庭、浩華、阿怡、堯堯、怡樺、格格等同學及依葶、子揚、君凱等學弟妹，感謝 96 及 97 級的各位，在台大的日子裡，都有和大家共同的回憶，謝謝你們。

台大土木系棒的立偉、映忻、三木、肇安、浩瑋、NONO、宜民、子軒、奕元、逸民、校園、祐康、小胖子、暴犯、阿國、以文、文傑及可愛的經理大龜等，使我在作研究的煩悶之餘，還有一個地方讓我發洩及嘴炮的地方。在台大，打過新生盃、台大盃、乙組賽，雖然沒有冠軍獎盃，但你們給我的回憶絕對比冠軍盃還要可貴；另外，在研究生涯煩悶無處可去時，就是淡江棒球隊給我最後的依靠，謝老師、大雄、小奕廷、小胖、錫瑋、阿松、正浩等棒球隊的各位，一起革命奮鬥多少大大小小的戰役，從大專盃預賽到決賽、春秋季聯賽等，淡江棒球隊在我學生生涯裡給我的回憶最多也最寶貴，尤其是謝老師，在私底下你就是我的兄長，照顧我就跟對待親弟弟一樣，你的照顧我點滴在心啦！另外，還有致理 ob 的學長學弟們，一起打球的時光也有六、七年了，學長們對小弟的照顧，我也是感激在心頭；再來就是洪姐姐、老大、茜雯、思凱、德川等使我碩士生涯有宣洩的對象，也謝謝你們！

最重要也是最感謝的是我的家人，外公、外婆、爸爸、媽媽及弟弟，謝謝你們長久來的支持、陪伴我成長、陪我走過人生的每個階段，希望我沒有讓你們失望，而我會更加努力成為你們一輩子的驕傲。最後，一個成功的男人，背後一定有一位偉大的女性支持，而這位女性就是我的摯愛 Ailsa，妳總是默默的陪在我身旁，在研究煩悶時，能夠靜靜的在聽我宣洩，在我快樂時能夠分享我的喜悅，妳陪著我喜也陪著我憂，能完成這篇論文，妳絕對是個幕後的推手，對妳的感謝就像我對你的愛一樣。

感謝完這麼多人，而學生生涯也告一段落了，就跟一場棒球比賽，終究是有結局，但這場比賽的結束並不代表最後一場比賽，未來的日子裡，還有很多更艱難、更崎嶇的比賽等著我去突破。研究所的兩年，真的讓我學會了許多事情，對於來到這裡，才讓我看到另一個更大的世界，更顯現我自己的渺小，所以我不會因為這樣而就此感到滿足，未來我會更努力的充實自我。

王宜慶 謹致  
2009 年 7 月於土木管 402

## 摘要

臺灣地區道路交通工程不斷改進，運具方面的結構也有也改，然而運輸問題卻日益嚴重。根據警政署的統計資料，每年交通事故不斷的攀升，平均約 2800 人死於交通事故，而大客車在每萬輛死肇肇事率都遠高於其他車種，所以一旦發生事故，其所造成的傷亡往往遠多於其他車種。

大客車事故頻傳的因素不外乎就是超速、酒醉駕車、違規駕駛、剎車失靈等，因此本研究嘗試資料探勘於交通事故之分析，探討大客車事故發生的主要因素。利用群集分析找出同質性最高的肇事集合，以此結果為基礎分析，並用卡方檢定驗證群集的正確性；然後套入判別分析中，作出其判別函數及預測其分類正確率，並找出影響大客車的肇事變數。

本研究採用警政署自民國 92 年至 96 年共六年全國大客車交通事故資料，資料總件數為 15514 件。研究結果為 92 至 96 年的訓練樣本和測試樣本的分類正確率都達 90% 以上，代表其判別率佳；影響六年的肇事變數為肇事原因、發生月份、道路類別、速限、分向設施為主要因素，其中，又以肇事原因佔的比例最高。在肇事原因中，主要發生的原因為變換車道或方向不當、未保持行車安全距離、酒醉後駕駛失控、違反號誌管制或指揮、違反特定標誌(線)標制、其他駕駛人因素、非駕駛人因素，並針對此研究結果提出改善策略。

**關鍵字：**交通安全、肇事分析、資料探勘、群集分析、判別分析、因子分析

## **Abstract**

According to the statistic data from National Police Agency, traffic accidents increase year over year. Averagely, 2,800 people die in traffic accidents. Moreover, in every thousand cars that cause deadly car accidents, bus is the main trouble maker. Thus, once a car accident happens, bus accidents usually lead to more deaths than other car accidents.

Bus accidents usually include speeding, drunk driving, driving against traffic regulations, and brake failing...etc. Therefore, in this thesis, I try to do traffic accidents analysis by using data mining, and figure out the main causes of bus accidents. In addition, I use cluster analysis to find out the most homogeneous class which causes bus accidents. As a result, I use chi-square test to verify the validity of the cluster and apply it to discriminant analysis which can determine the discriminant function and predict the accuracy of classification, so that the variable which causes bus accidents will come out.

In this research, the bus accidents data from 2003 to 2007 are included; the total number of the data is 15514, which can be found in National Police Agency. The result of this research includes the Training Samples and the Test Samples from 2003 to 2007, which's accuracy of classification are all over ninety percent. Therefore, the identification is quite excellent. The major variables which cause bus accidents from 2003 to 2007 are accident causes, which months, what kind of road, speed limit, and curb systems. Among these factors, accident causes are the most. And results for this strategy to improve.

**Keyword : Traffic Safety ; Accident Analysis ; Data Mining ; Cluster Analysis ;**

**Discriminant Analysis ; Factor Analysis**

# 目錄

誌謝 .....	I
摘要 .....	III
Abstract .....	IV
<b>第一章 緒論</b> .....	<b>1</b>
1.1 研究背景與動機 .....	1
1.2 研究目的 .....	3
1.3 研究範圍 .....	3
1.4 研究內容與流程 .....	3
<b>第二章 文獻回顧</b> .....	<b>7</b>
2.1 資料探勘相關文獻 .....	7
2.1.1 資料探勘的定義 .....	7
2.1.2 資料探勘的步驟與資料庫知識發現 .....	9
2.1.3 資料探勘的功能與技術 .....	10
2.1.4 資料探勘的應用 .....	13
2.2 交通事故相關文獻 .....	16
2.2.1 交通事故定義及分類 .....	16
2.2.2 肇事分析之應用 .....	19
2.2.3 國內道路遊覽車肇事案件分析 .....	21
2.3 綜合評析 .....	25
<b>第三章 研究方法</b> .....	<b>29</b>
3.1 群集分析 .....	29
3.2 判別分析 .....	39
3.3 因子分析 .....	43
3.4 小結 .....	48
<b>第四章 肇事資料前置處理</b> .....	<b>49</b>
4.1 分析流程 .....	49
4.2 肇事資料蒐集 .....	50
4.3 肇事資料前置處理 .....	51
4.4 肇事資料轉換 .....	52
4.5 肇事變數 .....	54
<b>第五章 資料分析</b> .....	<b>57</b>

5.1 群集方法與分析流程 .....	57
5.2 群集數目選擇 .....	58
5.3 群集結果 .....	61
5.4 群集驗證 .....	69
5.5 判別模式架構與分析流程 .....	76
5.6 判別函數的構建 .....	77
5.7 判別分析結果 .....	78
5.8 判別模式綜合分析 .....	93
<b>第六章 案例分析 .....</b>	<b>103</b>
6.1 臺灣地區 .....	103
6.2 臺北市 .....	108
6.3 南投縣 .....	114
6.4 案例比較 .....	119
<b>第七章 結論與建議 .....</b>	<b>121</b>
7.1 結論 .....	121
7.2 建議 .....	122
<b>參考文獻 .....</b>	<b>125</b>





## 圖目錄

圖 1-1	研究流程圖 .....	5
圖 2-1	資料庫知識發現的流程與步驟 .....	9
圖 3-1	群集分析之決策流程 .....	32
圖 3-2	群集間距離圖 .....	35
圖 3-3	$x_i$ 的共通性 $h_i^2 = l_{i1}^2 + \cdots + l_{iq}^2$ .....	46
圖 4-1	分析流程 .....	50
圖 5-1	群集方法分析流程 .....	57
圖 5-2	92 年肇事資料群集變化 .....	58
圖 5-3	93 年肇事資料群集變化 .....	58
圖 5-4	94 年肇事資料群集變化 .....	59
圖 5-5	95 年肇事資料群集變化 .....	59
圖 5-6	96 年肇事資料群集變化 .....	60
圖 5-7	97 年肇事資料群集變化 .....	60
圖 5-8	92-97 年肇事資料群集變化 .....	60
圖 5-9	判別模式分析流程 .....	77
圖 5-10	92 年各群集肇事之變數所佔比例 .....	94
圖 5-11	92 年各群集肇事之變數權重值 .....	94
圖 5-12	93 年各群集肇事之變數所佔比例 .....	95
圖 5-13	93 年各群集肇事之變數權重值 .....	95
圖 5-14	94 年各群集肇事之變數所佔比例 .....	96
圖 5-15	94 年各群集肇事之變數權重值 .....	96
圖 5-16	95 年各群集肇事之變數所佔比例 .....	97
圖 5-17	95 年各群集肇事之變數權重值 .....	97
圖 5-18	96 年各群集肇事之變數所佔比例 .....	98
圖 5-19	96 年各群集肇事之變數權重值 .....	98
圖 5-20	97 年各群集肇事之變數所佔比例 .....	99
圖 5-21	97 年各群集肇事之變數權重值 .....	99
圖 5-22	92-97 年各群集肇事之變數所佔比例 .....	100
圖 5-23	92-97 年各群集肇事之變數權重值 .....	100
圖 6-1	臺灣地區因素陡坡圖 .....	104
圖 6-2	臺北市因素陡坡圖 .....	109
圖 6-3	南投縣因素陡坡圖 .....	115

## 表目錄

表 1-1	道路交通事故統計.....	1
表 1-2	各車種道路交通事故發生情況.....	2
表 2-1	資料探勘的功能與技術.....	12
表 2-2	肇事因素索引.....	18
表 2-3	資料探勘應用彙整.....	25
表 2-4	肇事分析之應用彙整.....	26
表 2-5	國內道路遊覽車肇事案件分析彙整.....	26
表 4-1	肇事變數.....	54
表 5-1	92 年最後群集中心點.....	62
表 5-2	93 年最後群集中心點.....	63
表 5-3	94 年最後群集中心點.....	64
表 5-4	95 年最後群集中心點.....	65
表 5-5	96 年最後群集中心點.....	66
表 5-6	97 年最後群集中心點.....	67
表 5-7	92-97 年度最後群集中心點.....	68
表 5-8	92 年卡方檢定結果.....	69
表 5-9	93 年卡方檢定結果.....	70
表 5-10	94 年卡方檢定結果.....	71
表 5-11	95 年卡方檢定結果.....	72
表 5-12	96 年卡方檢定結果.....	73
表 5-13	97 年卡方檢定結果.....	74
表 5-14	92-97 年度卡方檢定結果.....	75
表 5-15	卡方檢定驗證結果.....	76
表 5-16	92 年肇事資料函數檢定結果.....	78
表 5-17	92 年費雪線性判別函數係數.....	79
表 5-18	92 年訓練樣本分類結果.....	79
表 5-19	92 年測試樣本分類結果.....	80
表 5-20	93 年肇事資料函數檢定結果.....	80
表 5-21	93 年費雪線性判別函數係數.....	81
表 5-22	93 年訓練樣本分類結果.....	82
表 5-23	93 年測試樣本分類結果.....	82
表 5-24	94 年肇事資料函數檢定結果.....	82
表 5-25	94 年費雪線性判別函數係數.....	83
表 5-26	94 年訓練樣本分類結果.....	84
表 5-27	94 年測試樣本分類結果.....	84

表 5-28	95 年肇事資料函數檢定結果.....	84
表 5-29	95 年費雪線性判別函數係數.....	85
表 5-30	95 年訓練樣本分類結果.....	86
表 5-31	95 年測試樣本分類結果.....	86
表 5-32	96 年肇事資料函數檢定結果.....	86
表 5-33	96 年費雪線性判別函數係數.....	87
表 5-34	96 年訓練樣本分類結果.....	88
表 5-35	96 年測試樣本分類結果.....	88
表 5-36	97 年肇事資料函數檢定結果.....	88
表 5-37	97 年費雪線性判別函數係數.....	89
表 5-38	97 年訓練樣本分類結果.....	90
表 5-39	97 年測試樣本分類結果.....	90
表 5-40	92-97 年肇事資料函數檢定結果 .....	90
表 5-41	92-97 年費雪線性判別函數係數 .....	91
表 5-42	92-97 年訓練樣本分類結果 .....	92
表 5-43	92-97 年測試樣本分類結果 .....	92
表 5-44	判別模式彙整.....	93
表 5-45	費雪函數加權值比重.....	101
表 5-46	肇事原因加權值之比重 .....	102
表 6-1	臺灣地區解說總變異量表 .....	103
表 6-2	臺灣地區各因子下分群數之判別正確率 .....	104
表 6-3	臺灣地區 3 因子轉軸後因子負荷矩陣 .....	105
表 6-4	臺灣地區 3 因子分 6 群之構面 .....	106
表 6-5	臺灣地區死亡人數之迴歸模式 .....	106
表 6-6	臺灣地區受傷人數之迴歸模式 .....	107
表 6-7	臺北市解說總變異量表 .....	108
表 6-8	臺北市在各因子下分群數之判別正確率 .....	109
表 6-9	臺北市 7 因子轉軸後因子負荷矩陣 .....	110
表 6-10	臺北市 7 因子分 7 群之構面 .....	111
表 6-11	臺北市死亡人數之迴歸模式 .....	112
表 6-12	臺北市受傷人數之迴歸模式 .....	113
表 6-13	南投縣解說總變異量表 .....	114
表 6-14	南投縣在各因子下分群數之判別正確率 .....	115
表 6-15	南投縣 3 因子轉軸後因子負荷矩陣 .....	116
表 6-16	南投縣 3 因子分 5 群之構面 .....	117
表 6-17	南投縣死亡人數之迴歸模式 .....	117
表 6-18	南投縣受傷人數之迴歸模式 .....	118
表 6-19	案例分析之彙整 .....	119

# 第一章 緒論

## 1.1 研究背景與動機

臺灣地區的道路交通工程不斷改進，運具方面的結構都有極大的改善，隨著汽車科技的進步，車輛速度、結構、性能等都不斷提昇，然而運輸問題如人為疏失、車體結構、道路設計、法令規範等日益嚴重卻無法徹底解決。根據內政部警政署的統計資料，如表 1-1 所示，截至民國 97 年底，每年的交通事故的件數不斷的攀昇，每年平均約 2800 人左右死於交通事故，這些數字尚未包括登入於資料庫中及當事人自行和解或就醫之事故紀錄，交通事故已對於民眾的生命財產造成莫大的威脅。交通事故不僅造成個人及家庭的嚴重傷害，也對社會有限醫療資源損失及社會成本增加，是以道路交通不只有思考行的效率問題，更應該把交通安全視為第一，因此如何改善運輸安全以降低交通事故為目前當務之急。

表 1-1 道路交通事故統計

年別	總計			A1 類			A2 類	
	件數	死亡	受傷	件數	死亡	受傷	件數	受傷
90	64264	3344	80612	3142	3344	1490	61122	79122
91	86259	2861	109594	2725	2861	1284	83534	108310
92	120223	2718	156303	2572	2718	1262	117651	155041
93	137221	2634	179108	2502	2634	1248	134719	177860
94	155814	2894	203087	2767	2894	1383	153047	201704
95	160897	3140	211176	2999	3140	1301	157898	209875
96	163971	2573	216927	2463	2573	1006	161508	215920
97	166288	2224	220346	2150	2224	983	164138	219363

註：A1 類係指造成人員當場或 24 小時內死亡交通事故

A2 類係指造成人員受傷之交通事故

資料來源：【1】

在大客車旅次增加的情形下，國人使用長程及短程的大客車為接駁工具都大幅提昇，使得大客車在市區、山區及城際間的交通運輸上扮演著格外重要的角色；運輸工具不僅提供快速便捷的服務，更應具備安全舒適的乘坐空間。根據內政部

警政署統計資料，以民國 97 年為例如表 1-2 所示，大客車不管是在每萬輛肇事率或每萬輛死亡率都比其他車種來的高，不難發現大客車一旦發生事故，其所造成的傷亡往往遠多於其他車種。近年來所發生的大客車事件傷亡都非常慘重，如梅嶺、九份、仰德大道等，所以大客車安全也應是被大家所重視的一項課題。

表 1-2 各車種道路交通事故發生情況

		每萬輛死亡肇事率
營業大客車		2.76
小客車	自用	0.10
	營業	0.21
大貨車	自用	0.99
	營業	1.13
小貨車	自用	0.19
	營業	0.78
機車	重型	0.07
	輕型	0.04

資料來源：【1】

臺灣大客車車輛安全發展的歷程，其實伴隨著相當多且不幸的事故，而大客車事故頻傳的因素不外乎就是人、車、路、環境四個層面，近年所發生的大客車重大事故中，從交通部或內政部警政署的車禍肇事鑑定結果，絕大多數都是歸因於人為因素。其原因如下：

1. 人：超速、酒醉駕車、違規駕駛、開車時使用行動電話...等
2. 車：車齡老舊、改造車體、使用再生胎、剎車失靈...等
3. 路：道路幅度窄小、山區路段彎度過大...等
4. 環境：天候、光線、視距、路面狀態...等

上述人、車、路、環境四個層面中有些因素或許不是造成事故最主要的因素，但卻可能是增加事故危險因素之一，所以本研究期望透過大客車的交通事故案件並用資料探勘作進一步分析。

本研究對已經發生交通事故的肇事原因了解並從交通事故資料加以分析外，希望事先了解潛在的危險因子對肇事影響且針對危險因子加以改善，進而有效降

低肇事次數及嚴重程度，因此，對於交通事故，事前的預防更勝於事後的補救。

## 1.2 研究目的

有鑑於大客車的交通事故一旦發生便容易造成重大的傷亡，交通事故發生由不同時間、環境、區域下所構成，因此發生的交通事故其嚴重性皆有所不同，其背後的影響因子絕非一語可斷定。本研究主要目的有以下幾點：

1. 藉由肇事分析期望找出大客車肇事主要問題，並作比較分析之間的差異。
2. 利用地區特性，找出二個不同的地區，做案例的比較分析。
3. 歸納出影響大客車肇事之原因，並提出改善策略與建議，使大客車的安全性提高。

## 1.3 研究範圍

本研究以全國大客車交通事故為研究對象，探討各年度探討影響肇事原因的實際及潛在因子，故資料蒐集時間為民國 92 年至民國 97 年共 6 年的全國大客車肇事資料。



## 1.4 研究內容與流程

依據研究動機與目的，本研究之研究流程如圖 1-1 所示，內容主要分為七部分，為緒論、文獻回顧、研究方法、肇事資料前置處理、群集分析、判別分析、結論與建議，其大致內容如下：

### 第一章 緒論

主要說明研究動機、目的、範圍、內容與流程。

### 第二章 文獻回顧

針對研究主題回顧文獻中資料探勘之定義、步驟、功能、技術及應用領域。

其次，透過交通事故相關文章及遊覽車肇事案件分析，隱含的潛在因子，

冀望透過資料探勘之方法，從資料中探尋出不明顯且隱藏的在用資訊與知識，並希望這些資訊與知識可以幫助釐清隱藏於肇事背後的危險因子與肇事原因之相關性。

### 第三章 研究方法

使用群集分析法及判別分析法，包括模式基本概念、假設、推導及模式應用。

### 第四章 肇事資料前置處理

蒐集有關全國大客車交通事故之資料，並作資料的前置處理之過程。

### 第五章 資料分析

以群集分析法，將各年度進行最佳的分數，並依照分群的結果，利用統計交叉分析的卡方檢定，在分群中的變數是否顯著，將顯著之變數套入判別分析；從分群結果之分類定義，運用判別分析，建構各年度大客車的判別預測模式及分類正確率，再從判別模式中找出影響大客車之重要變數。針對此結果提出改善策略。

### 第六章 案例分析

加入因素分析，試比較臺灣地區與臺北市、南投縣之地區特性，並作事故嚴重程度迴歸預測模式看其死亡及受傷之因子是否一致及顯著。

### 第七章 結論與建議

針對研究結果提出結論與建議。

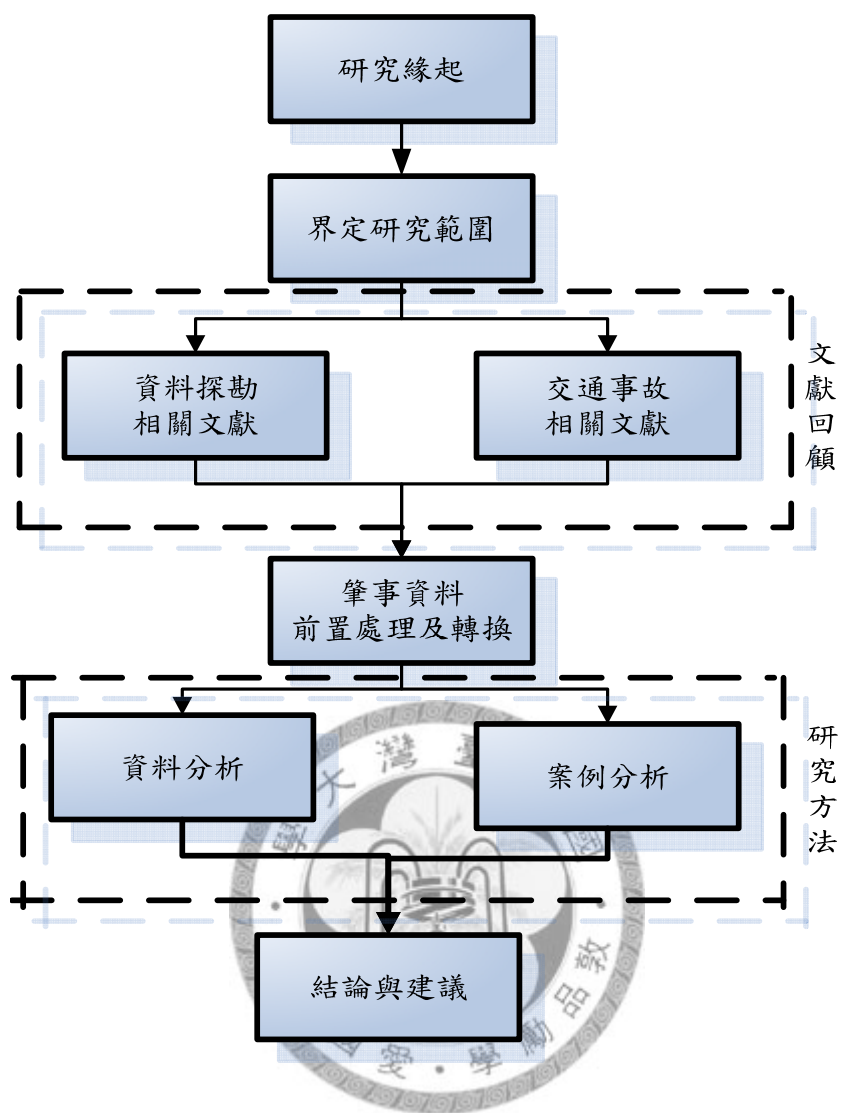


圖 1-1 研究流程圖





## 第二章 文獻回顧

本章節先介紹資料探勘之定義、步驟、功能與技術，並回顧資料探勘之相關應用，再來了解交通事故之定義並回顧其肇事之分析及應用，最後再了解國內道路的遊覽車肇事案件。

### 2.1 資料探勘相關文獻

由於資訊科技的發達，且隨著時間的累積，各領域及企業組織在資訊系統中資料儲存量也隨之增加，而資料探勘又是近年來資料庫應用領域中相當熱門的議題。資料探勘一般是指在資料庫中，利用各種演算法，將過去所累積的資料去進行分析、歸納及整合等工作，以萃取出有用的資訊，找出有意義且使用者有興趣的特徵，在進行決策時提供之參考依據。以下為資料探勘之相關文獻：

#### 2.1.1 資料探勘的定義

資料探勘由英文 Data Mining 翻譯而得，對於 Data Mining，各學者有不同的見解，如資料採礦、資料挖掘、資料考古學等。資料探勘是指找尋隱藏在資料中的訊息的過程，從資料中發掘資訊或知識，如趨勢(Trend)、特徵(Pattern)、相關性(Relationship)的過程，也就是從資料中發掘資訊或知識，目前已被許多研究人員視為結合資料庫系統、統計、人工智慧及機器學習技術的重要領域，許多業界人士也認為此一領域是一項增加企業潛能的重要指標。以下由各學者對資料探勘所提出之定義：

1. Frawley (1991)【2】：從資料庫中發掘出非顯然的、先前未知的及潛在的可能有用資訊之過程。
2. Grupe & Owrang (1995)【3】：從已經存在的資料庫中挖掘出新的及專家仍發現的新事實。

3. Fayyad (1996)【4】【5】：定義知識發掘是從大量資料中選取合適的資料，進行資料處理、轉換等工作，再進行資料探勘與結果評估的一系列過程，也就是說資料探勘只是知識發掘過程當中的一個步驟。
4. Berry & Linoff (2000)【6】：定義資料探勘為使用自動或半自動的方法，對大量資料作分析，找出有意義的關係或法則。
5. Han & Kamber (2001)【7】：從儲存在資料庫的大量資料、資料倉儲或其他資訊儲存體中去挖掘有趣的知識過程。
6. Roiger & Geatz (2003)【8】：從整個資料庫裡的資料，利用一種或多種電腦技術來自動分析或擷取知識的過程，其目的是在資料中發現趨勢與特徵。
7. 曾憲雄等人 (2005)【9】：資料探勘可解釋資料庫之知識發掘(Knowledge Discovery in Database, KDD)，也就是從一個大型資料庫裡頭所儲存的大量資料當中去萃取出一些有趣的知識，所謂的知識指的就是一些規則。
8. 尹相志 (2007)【10】：資料探勘是利用統計及機器學習的演算法，啟發性地從大量資料中找尋隱藏具有商業價值的知識與規律，以作為自動化商業策略之應用。

綜合多位學者的定義，皆認為資料探勘在龐大的資料中找出有效益的特徵、型態、趨勢及規則性，同時資料探勘可視為資料庫知識發現過程中的一個重要的步驟。

資料探勘和統計容易使人搞混，但其實兩者之間還是有差異，尹相志 (2007)【10】

強調二個重點：

- 資料探勘是啟發性而非演繹性：傳統統計檢定中，常會有所謂的統計假設，如常態分配或是變數之間是獨立等；資料探勘雖然有許多使用的技巧及方法是源自於統計，但它揚棄了許多統計上的假設，而是啟發性的透過演算法主動搜尋有意義的規則，讓資料說話。
- 資料探勘具有商業化以及行動意涵：在傳統統計或是市場調查，也揭露許多有意義規則與資訊，但是只作為角色參考，無法自動化的聯結至商業流程中。

## 2.1.2 資料探勘的步驟與資料庫知識發現

依據 Fayyad (1996)【11】的定義得知，資料探勘僅為知識發現流程之其中一個步驟也是屬於最重要的一環，整個過程包括對應領域的認知、具備相關的專業知識，確認分析目標的資料來源，接著進行資料淨化、資料整合、資料選擇、資料轉換、資料探勘、特徵評估、知識呈現，如圖 2-1 所示。

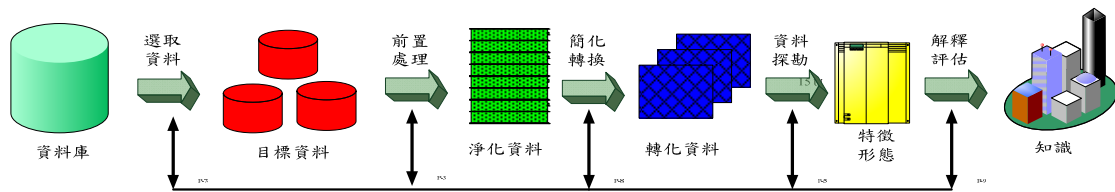


圖 2-1 資料庫知識發現的流程與步驟

資料來源：【11】

資料庫之知識發現的流程通常包含以下步驟，分別敘述如下：

### 1. 定義及分析問題

正確的定義問題，才能知道在整個資料探勘過程中所需的資料知識、運用的技術及結果該如何呈現，以解決原先設定的問題。

### 2. 資料的選取及前置處理

在確定需要的資料時，必須謹慎思索可能與進行的知識發掘任何有關資料，並瞭解需要的資料數量，以及所需的資料來源。而資料的準備，除了透過各種來源取得所需資料，主要還必須對於這些資料進行前置處理。由於各種來源取得的資料，其中可能包含許多錯誤、不需要或是不一致的資料，而資料處理主要則對於這些原始資料進行減量、清理、選擇及轉換。

### 3. 知識發現的相關設定

主要目的在選擇資料庫知識發現的分析方法，也就是資料探勘的技術，並進行相關的參數及變數的設定。

### 4. 進行資料探勘

在準備相關的資料及選定資料探勘的技術後，此步驟主要的工作就是實際的進行資料探勘，把其中所隱藏的知識萃取出來。而進行資料探勘的步驟也是整個資料庫知識發現的核心步驟。

## 5. 樣式的評估

資料探勘萃取出的知識不見得每一個都是需要的，因此，需要再經過一個樣式評估的步驟，去評估資料探勘所挖掘出來的知識是不是真的有用，因為挖掘工具可能會挖掘出很多東西，有些是有意義的，有些是沒有意義的。

## 6. 知識的呈現

此步驟就是在資料探勘的步驟進行之後所獲得的知識，透過評估知識的品質去蕪存菁，再經由可靠且妥善的方式來組織，並呈現發掘所得到的知識。

### 2.1.3 資料探勘的功能與技術

依據尹相志（2007）【10】、謝邦昌（2007）【12】及本研究的綜合整理資料探勘的功能與技術。一般而言，資料探勘的主要功能有以下幾項，茲將這些功能之意義及演算法分別敘述如下：

#### 1. 分類(Classification)

分類是指檢視新物件的特性，然後將其指定到預先定義好的類別中，利用一連串的輸入變數來預測類別變數，即描述物件屬性與類別之關係，然後再根據這些分類模型對其他未經分類或新資料進行預測。分類還可以透過分類結果的純度來提供分類機率，這些分類機率可以作為名單排序或是篩選的重要根據。例如將信用卡申請者的風險屬性區分別高度申請者、中度申請者及低度風險申請者。分類常使用的演算法包括決策樹(C4.5、CART、CHAID、Gini、ID3 等)、貝氏分類法、記基礎推論法及類神經網路。

#### 2. 推估(Estimation)

推估是善於處理連續性的數值，憑著輸入資料，可用來推估未知的連續

性變數。例如按照信用卡申請之教育程度及行為來推估信用卡消費量。推估常用的演算法包括統計上的相關分析、迴歸(線性迴歸及羅吉特迴歸)分析及類神經網路。

### 3. 預測(Prediction)

預測是推估連續變數，即推未來的數值及趨勢，以歷史資料可用來建立模型以檢視近來觀測值的資料。例如顧客過去之刷卡消費量預測未來之刷卡消費量。預測常用的演算法包括決策樹、迴歸分析、時間序列、類神經網路。

### 4. 關聯規則(Association Rule)

關聯規則就是於歷史資料中，找出哪些事件或物件總是伴隨著發生，亦稱作「購物籃分析(Market Basket Analysis, MBA)」。通常運用於商業交易的作法為蒐集一組交易資料，每一交易包括若干交易項目，關聯分析的目的是由這些交易的資料中，找出交易項目的關聯規則。關聯規則常用的演算法包括 Apriori 演算法及 FP-Growth 演算法。

### 5. 群集分析(Clustering Analysis)

群集分析又稱為資料切割，是一種多變量統計分析的技術，係指將所有物件或資料分成若干群集的過程，也就是根據物件間的相似性或不相似性。將所有物件分成若干個群集，使得每個群集內的物件具有高度的相似性，而不同群集間具有高度不相似性。群集分析的目的是要把群集與群集間的差異找出來，同時也要將群集中物件的相似性找出來。群集分析與分類的差別在於，群集分析是屬於非監督型學習方式，使用者只須供學習的學習資料組，系統必須自行發掘物件間的相似性和其他特性(如最佳群集數)，以便建立模式和群集描述，作為分類資料自動歸類的依據。群集分析常用的演算法包括 K-Means 法、K-Medoids 法及 CLARANS 法。

### 6. 序列規則(Sequential Pattern)

序列規則的重點是考慮時間的因素，利用此方法分析不同時間點上各事

件的關聯性，也就是在一長期時間區段內，事件發生的順序。其最大功能是協助我們找出生命週期序列規則與週期性規則兩種：順序性規則乃考慮事件發生之時間先後關係；週期性規則考慮時間區段的變化，分析時間區段內所發生的事情，是否其他相同時間區段內也會發生。雖然這兩種方法不同，但對使用者而言，隨著時間的多樣變化，找出有用的規則已日形重要。序列規則常用的算法包括 Aprior 演算法、決策樹、時間序列及類神經網路。

## 7. 視覺與描述

資料探勘過程中的一項重要附加價值，就是視覺與描述。在進行資料探勘的前置作業時，透過資料視覺化，能夠將資料特性呈現在資料探勘者前面，並透過良好的資料視覺化程序，可以發現許多演算法無法判別的規則型態去加以描述。

Berry & Linoff (2000)【6】將各項技術適用之功能任務做整理，如表 2-1 所示。現今資料探勘的方法一直都在被改進及改良，以符合探勘領域的變化，進而可以被使用於各領域之中。

表 2-1 資料探勘的功能與技術

技術	分類	推估	預測	關聯分組	群集化	視覺與描述
決策樹	○		○		○	○
傳統統計	○	○	○	○	○	○
群集分析					○	
連結分析	○		○		○	○
購物籃分析	○		○	○	○	
類神經網路	○	○	○		○	
基因演算法	○		○			
記憶基礎理解	○		○	○	○	

資料來源：【6】

#### 2.1.4 資料探勘的應用

Raymond 及 Jiawei (1994)【13】主要利用是空間資料探勘的自動化找到有趣的關係及特徵，使用的演算法為群集法；此研究先使用基本的 PAM 及 CLARA 的群集法，再基於隨機搜尋，制定一個新的群集法，為 CLARANS，並透過 CLARANS 法分析下，實驗空間(SD)及非空間(NSD)的資料探勘。PAM 法是找出 k 個群集並在每個群組找出一個物件，而不同特徵的資料會被分類，PAM 法主要是處理小型資料庫；CLARA 法是根據資料找出樣本範圍並應用 PAM 法找出中心點，而樣本範圍是具有隨機性，CLARA 法主要處理大型資料庫；經過分析及實驗的比較下，CLARANS 法比 PAM 法及 CLARA 法更有效率且可以分類出更好的群組。此研究的資料庫是利用溫哥華 2500 個高級住宅區的資料，而資料種類為房屋類型(大樓、獨棟、公寓)、價格及房屋大小去分類出此資料的分佈範圍。房屋類型為大樓時，價格為 150 萬至 350 萬，大小為 6000 至 10000 平方呎；房屋類型為獨棟時，價格為 80 萬至 150 萬，大小為 3000 至 7000 平方呎；房屋類型為公寓時，價格為 30 萬至 80 萬，大小為 1000 至 2500 平方呎。CLARANS 群集法去分析空間(SD)下的資料時，分類為三個群組，從三個群組中找到房屋類型分佈的範圍；在分析非空間(NSD)下的資料時，分成 12 種組合，三種不同的房屋類型下各有 4 種不同的價格及房屋大小下找到分佈的範圍。

Kao 及 Chiu (2001)【14】此研究是結合了分類與迴歸樹(CART)與類神經網路的技術，以確定信用評分模型的預測能力，其中 CART 是使用了 Gini index 來衡量、傳統類神經網路及 CART 結合類神經網路三者做比較其預測正確率。資料是臺灣的銀行，蒐集時間為 1995 年至 2000 年，樣本數約為 3000 人，資料變數包括就業狀況、個人資料、年齡、房屋及職業類型等，但為了保護個人資料的機密，此研究改變了屬性的名稱和數值，僅用象徵性的資料；資料庫中有二個混合的連續變數及七個類別變數。此研究的資料為隨機性且用了二個不同的集合，分別人訓練



集(樣本)約有 1980 個觀測值及測試集(樣本)約有 1020 個觀測值，而訓練集是使用信用評分模型去估計參數，測試集是為了測試模型的能力。基於決策樹 CART 的選擇結果，使用了六個變數(性別、職業、薪資、婚姻狀況、教育程度、信用額度)，分別為二個連續變數及四個類別變數，訓練集與測試集之估計模型觀測值的正確分類為 88.43%及 88.22%是信用良好之顧客、58.27%及 56.27%是信用不良之顧客；結合 CART 和類神經網路中，三個層級(輸入層、隱藏層、輸出層)，輸入層有七個節點包含了六個輸入值(此值是根據 CART 的結果)，訓練集與測試集之估計模型觀測值正確分類為 90.38%及 90.22%是信用良好之顧客、62.57%及 61.15%是信用不良之顧客。最後，比較 CART、傳統類神經網路、結合 CART 與類神經網路的預測正確率結果為 72.24%、67.25%、75.69%；而類神經網路的輸出值提供給 CART 在所有的情況下，比傳統的類神經網路及 CART 模型好。

Wang 等人 (2007)【15】使用決策樹的 ID3 演算法，資料取得是由第三方物流公司客戶的某一個月期間資訊系統的資料庫，樣本裡包含 19 個項目，分為二類顧客類型(會員及正常顧客)及四種屬性(運費、付款、重量及運送時間)。ID3 演算法分類出，會員有 8 個項目，正常顧客有 11 個項目，然後再把顧客類型分別放入四種屬性去算出 information gain，最後付款的 information gain 最大，因此用付款作為決策樹的樹根建立決策樹，再利用 IF-THEN 找出規則；但是 ID3 演算法仍有不足之處，所以此研究使用 information gain ratio 改善。利用決策樹的 ID3 演算法，可以算出 information gain 去分析客戶資料並發現客戶特性，並留住高價值的客戶及開發潛在客戶，以提供第三方物流公司提供高品質的服務。

Eduardo 等人 (2002)【16】應用墨西哥的道路交通事故資料庫，道路交通事故的資料是從 1995 年至 1999 年，其中共有 35 種不同的特性，包含時間、日期、道路名稱、天氣條件、發生事故的公里處等，另外還記錄著因事故發生的時間、損害道路、車輛數等主要特性而導致道路封閉及嚴重的事故中傷亡人數。1995 年時的資料包含 33 個特性，1996 年上半年的資料包含 22 個特性，1996 年下半年後至

1999 年的資料包含 17 個特性，其中 16 個特性是每年都有的；此資料庫記錄著 4380 個事故及相關之 6041 輛車。使用的技術及工具為 Bayesian network 分析各特性間的依賴性、使用關聯規則之 Apriori 演算法找出變數之間的關係、使用 Java 之 C4.8 之 J48 及 CN2 法做分類。先做簡單的分類為沒有嚴重受傷、嚴重受傷、死亡、嚴重受傷和死亡。使用 Apriori 演算法，如果在沒有天氣條件影響下及交通事故類型為一般撞擊時，就不會有嚴重的傷害；如果在沒有天氣條件影響下及交通事故類型為翻車或人被撞時，就是嚴重的車禍；如果有天氣條件的影響下發生的交通事故，大部分都為嚴重的傷害。使用 CN2 演算法，如果車禍是正常撞擊及天候條件是下雨或起霧時，則主要有嚴重的受傷；如果是連續撞擊或車轉向及沒有天候條件影響時，則沒有嚴重的受傷；如果是翻車、車輛轉向或人被撞及沒有天候條件影響時，則會有死亡；如果翻車、車輛轉向、人被撞或雙向車道及有天候條件影響時，則會有嚴重受傷和死亡。Bayesian network 法結果顯示和日期有關的資料及在斜坡、車道在嚴重事故中發現有高度依賴性。最後，三種不同的演算法發現一些主要的結果，死亡和沒使用安全帶有關、天氣狀況及道路上下坡之特性是直接影響嚴重事故的因素、特定的道路會常有事故發生、大多事故發生在下午或凌晨、車輛高速行駛會有很大的機率發生死亡事故、酒後駕車會發生嚴重的事故。

Marukatat (2007)【17】從框架(framework)內選擇並提出分類、挑選及篩選，根據規則結果作關聯分析，利用 Apriori 演算法找出高的支持度(support)及信心水準(confidence)。資料蒐集是從泰國當地的 Nakorn Pathom 省的交通事故，時間為 2003 年 1 月 1 日至 2006 年 3 月 31 日，總共有 1007 筆記錄。資料庫中有 20 個二元(binary)變數包括車子類型、事故種類及人為損失等，3 個名義(nominal)變數包括時間、地點及道路特性。使用框架選擇包含語義(semantic)規則分類及排列分析二部分，語義規則分類有 abundant、strongly abundant 及 weakly abundant rule 三種類型，而 abundant 及 strongly abundant rule 在此研究是篩選的作用，然後在用 weakly abundant rule 找出候選規則，而排列分析可能會多次使用不同的參數產生的規則，

再 confidence 與 life 從中選出重要的做比較。此研究的關聯規則執行了八次，產生了 3042 個關聯規則，然後篩選掉重覆或排列規則，結果包含 105 個 candidate rule 及 294 個 weakly abundant rule。candidate rule 產生四種規則，一、中午十二點到晚上六點、地方道路、交叉路口→沒有讓右轉車輛，二、中午十二點到晚上六點，直線道、死亡→卡車，三、凌晨十二點到早上六點、卡車→死亡，四、曲線道、腳踏車→地方道路；weakly abundant rule 產生二個規則，一、上坡、行人→沒有機車、死亡，二、卡車→公路、沒有汽車、沒有超速。

吳冠宏等人（2006）【18】將利用資料探勘技術中的Two-Step 法、自組織映射圖 (Self-Organizing Maps)與K平均值演算法(K-Means Method)來分析大量的肇事資料以進行分群工作，幫助我們從其中找出有用的資訊提供建議或決策，以減少車禍的發生。把這三種演算法都分成二群去分析平均快車道數目、速限、道路類別及平均損失金額四個變數，其分群結果發現K-Means演算法在肇事變數中具有比較好的分群能力，能有效的區分出不同的群集。

## 2.2 交通事故相關文獻

本節了解交通事故的定義及分類，交通事故發生由人、車、路、環境四個層面影響，再由過去相關的肇事文獻了解由什麼肇事原因對交通事故影響最深。

### 2.2.1 交通事故定義及分類

根據道路交通事故處理辦法第二條：道路交通事故，指因汽車或動力機械在道路上行駛，致有人傷亡或車輛財物損壞之事故。其中對交通事故定義的內容說明如下【19】：

1. 肇事的主體須有「汽車」或「動力機械」。
2. 肇事地點須發生在「道路」上，而「道路」的定義係依道路交通管理處罰條例第三條第一項第一款之規定：「道路係指公路、街道、巷弄、廣場、騎樓、

走廊或其他供公眾通行之地方」。

3. 汽車或動力機械須有「行駛」的狀態。
4. 因肇事而致使「有人傷亡(體傷)或車輛則物損壞(財損)」。
5. 須出於「過失或無過失的行為」。

交通部運輸研究所與內政部警政署於民國八十五年十一月合編之「道路交通事故調查表」中，將一般事故傷亡統計依事故受傷及死亡人數統計，可分為三類：

A1：造成人員當場或 24 小時內死亡之車輛交通事故案件。

A2：造成人員受傷之車輛交通事故案件。

A3：無人傷亡，但造成財物損失的交通事故案件。

依照「道路交通事故調查表」中，囊括了事故發生之人、事、物各項肇因，並區分出人、車、路及環境等四大因素，分述如下：

- 人：性別、受傷程度、保護裝置、當事者行動狀態、駕駛者資格情形、駕駛者執照種類、飲酒情形、肇事逃逸、職業、旅次目的、年齡。
- 車：車輛用途、車種、車輛牌照號碼、車輛撞擊部位。
- 路：道路類別、速限、道路型態、路面狀況、車道劃分設施(分道及分向設施)。
- 環境：時間、天候、光線、事故位置、路面狀況、道路障礙、號誌。

目前警察處理事故所記錄之肇事影響因素，係依據交通部運輸研究所民國 92 年「易肇事地點改善作業技術參考手冊」【20】及「易肇事地點改善作業手冊之教育訓練計劃」【21】，分為「車輛駕駛人因素」及「非車輛駕駛人因素」兩大類，其中，第 1 至 43 項為車輛駕駛人因素，第 44 至 64 項為非車輛駕駛人因素，共計有 64 項肇事因素，茲彙整如表 2-2 所示。

表 2-2 肇事因素索引

車輛駕駛因素	(一) 駕駛人	01 違規超車、02 爭(搶)道行駛、03 它行及方向不定、04 逆向行駛、05 未靠右行駛、06 未依規定讓車、07 變換車道或方向不當、08 左轉彎未依規定、09 右轉彎未依規定、10 迴轉未依規定、11 橫越道路不慎、12 倒車未依規定、13 超速失控、14 未依規定減速、15 搶越行人穿越道、16 未保持行車安全距離、17 未保持行車安全間隔、18 停車操作時，未注意其他車(人)安全、19 起步未注意其他車(人)安全、20 吸食違禁物後駕駛失控、21 酒醉(後)駕駛失控、22 疲勞(患病)駕駛失控、23 未注意車前狀態、24 搶(闖)越平交道、25 違反號誌管制或指揮、26 違反特定標誌(線)禁制
	(二) 燈光	27 未依規定使用燈光、 28 暗處停車無燈光及標識
	(三) 裝載	29 裝載貨物不穩妥、30 載貨超重而失控、31 超載人員而失控、32 貨物超長、超寬、超高而肇事、33 裝卸貨不當、34 裝載未盡安全措施、35 未待乘客安全上下開車、36 其他裝載不當肇事
	(四) 其他	37 違規停車或暫停不當而肇事、38 拋錨未採取安全措施、39 開啟車門不當而肇事、40 使用手持行動電話失控、41 其他引起事故之違規或不當行為、42 不明原因肇事
	(五) 無	43 尚未發現肇事因素
非車輛駕駛因素	(六) 機件	44 煞車失靈、45 方向操縱系統故障、46 燈光系統故障、47 車輪脫落或輪胎爆裂、48 其他引起事故之故障
	(七) 行人 (或乘客)	49 未依規定行走行人穿越道、地下道或天橋而穿越道路、50 未依標誌、標線、號誌或手勢指揮穿越道路、51 穿越道路未注意左右來車、52 在道路上嬉戲或奔走不定、53 未待車輛停盪而上下車、54 上下車輛未注意安全、55 頭手伸出車外而肇事、56 乘坐不當而跌落、57 在路上工作未設適當標識、58 當其他引起事故之疏失
	(八) 交通管制 (設施)	59 路況危險無無全(警告)設施、60 交通管制設施失靈或損毀、61 交通指揮不當、62 平交道看守 疏失或未放柵欄、63 其他交通管制不當
	(九) 無	64 尚未發現肇事因素

資料來源：【20】【21】

### 2.2.2 肇事分析之應用

Al-Ghamdi (2003)【22】主要分析在沙烏地阿拉伯首都利雅德在交岔路口發生交通事故傷害，瞭解在交岔路口或是在非交岔路口對於交通事故的問題，所以能提出預防事故發生，改善交通問題以作為對策。因此資料中總共有1774位意外事故，時間是1997-1998年，就事故發生的因素有事故型態，碰撞型態，原因和時間，駕駛者有速度和其它重要的行為因素，道路因素有交岔路口及非交岔路口分析。碰撞固定物因子於研究中對於嚴重事故與僅財產損失事故類型z檢定皆不顯著，但相對於事故地點方面的卡方分析中，發現交叉路口與非交叉路口下之事故比率皆頗為明顯，故此路旁固定物之潛藏因子當納入利雅德有關當局於道路幾何設計時需加以再度審慎考量。最後顯示50%嚴重事故都涉及行人這一部份且路段之死亡率24.89%明顯地高於路口死亡率8.16%；其中，最主要是發生在非十字路口上。不適當的駕駛行為是要為這多數交通事故負責任，像是超速、無讓路、闖紅燈比其它所有事故因素的機率都高於一半以上。

Amoros等學者(2003)【23】為探討各地區交通安全標準訂定時所需考量的因素以避免交通意外的發生，以法國東南方阿爾卑斯地區8個郡的交通路網與肇事資料為研究對象，其考量的因素為道路等級包括國道、高速公路、郡道、一般道路等，且將道路特性分成郊區與市區兩種，以及事故發生時間與社經變數。社經變數包含各地區人口平均年齡、肇事者駕照持有時間等，觀測期間為1986~1993年。肇事率依不同種類道路有不同的曝光率計算方式，國道、快速公路的曝光率是駕駛哩程數，郡道、一般道路則為駕駛的行駛公里數，以油耗量來估計；在嚴重程度方面，分成死亡與受傷事件，為瞭解各地區之間肇事率與嚴重程度的差異，在肇事率以相關風險來比較，是以一地區的肇事率為基準，與其他地區肇事率之比率；在嚴重程度以損益比來比較，研究方法是負二項迴歸與羅吉斯迴歸來分析。其結果顯示，在肇事率比較中，郡道與一般道路下，以Rhone區為基準，Drome

區的風險會高於Rhone區2.29倍，而7個地區在國道與高速公路風險都低於Rhone區；嚴重程度比較發現，同樣在高速公路上，7個地區的嚴重程度都高於Rhone區，而在郡道卻是低於此區。考量社經變數的結果顯示，新手的肇事率較一般駕駛高出0.43倍，但持學習駕照者卻少0.58倍，各地區考量社經變數與未考量時並無顯著差異；在嚴重程度中，新手死亡機率較低，反而持學習駕照者死亡率較高。所以各地區會因為道路種類的不同而肇事率與嚴重程度也有所差異，因此各地區交通安全指標的制定，需由不同的道路等級來區分較適合。社經資料卻在各地區並無差異，表示社經資料是所有共同的特性，所以在管理上各地區應該相同。

Lee 等學者（2002）【24】為降低澳洲年輕駕駛人的肇事率，探討年經駕駛發生事故的潛在因素，由於年輕駕駛肇事率逐年增高，所以更需瞭解肇事發生的原因，才能加以宣導與管理。本研究以 1174 持有駕照滿 12 個月的年輕駕駛為研究對象，透過問卷調查方式，測得駕駛特性、習慣、肇事紀錄等。研究方法以卜瓦松迴歸、負二項迴歸與 Zero-Inflated Poisson Model(ZIP)來分析，透過概似比檢定結果發現負二項迴歸與 ZIP 均比卜瓦松迴歸更適合，而負二項迴歸與 ZIP 透過 Pearson goodness-of-fit 檢定結果相同，而變數在模式結果發現自我駕駛技術因素並不顯著影響事故發生，駕駛性別並與喜愛冒險之特性無關，而持有學習駕照者會具有冒險特性，拿到駕照後的一年內有較高的事故發生機率，若先由學習駕照而駕駛至獲得駕照者，其發生事故機率比過去沒有學習駕照者來的高。

楊思瑜（2003）【25】探討桃竹苗地區小型車事故嚴重程度預測，以因子分析及群集分析將人、車、路進行分群；以多元羅吉特模式建構各群集事故嚴重程度的預測模式來探討事故嚴重程度，分析影響各集群嚴重程度的主成分與影響程度；結果可概分為「外在環境構面」（道路等級、當地速限）、「個人背景構面」（教育程度、年紀、駕駛資格）、「時間環境構面」（季節、天色、例假日）、「碰撞型態構面」（碰撞方向、本車動作、他車動作）等，而各事故嚴重程度模式中，年紀愈長愈嚴重、教育程度愈低愈嚴重、駕駛資格完整愈嚴重、碰撞方向

愈偏向前方愈嚴重。

邱裕鈞等人（2004）【26】在肇事案例關聯性方面，以SAS統計軟體的階層式群集分析方法，將所有案例分為40個群落，分群數的決定採用判定係數作為依據，通常判定係數愈大，表示分群的結果愈好。在第107群時所求出之判定係數為最大，但觀察107群到41群間，均發生某幾群僅有一個樣本數之狀況，直到降至40群時，每群均包含兩個以上案例方停止。因此以40群為分群組數，將538筆案例加以分群，重要變數包括車種、超速、飲酒、道路類別、道路型態、行向、車損部位、筆錄速率、是否預見、駕駛死亡、路權。再依分群結果，以判別分析找出判別函數，其加權值為判別函數係數，以便鑑定案例之分群索引使用。如發生兩車碰撞事故，將其變數資料分別代入40群所屬之判別函數，加以計算40個判別值，判別值最高者，即代表新的案例屬於該群組。

### 2.2.3 國內道路遊覽車肇事案件分析

近年來，一旦大客車或遊覽車發生交通事故後，都會造成慘重的傷亡，後果也會不堪設想，因此本研究蒐集從民國九十年後國內發生的遊覽車肇事案件，其整理分析如下：

#### 一、清境遊覽車車禍

民國 91 年 10 月 6 日早上 10 時，一載搭載 41 人遊覽車行經南投縣仁愛鄉清境農場附近，因煞車失靈又適逢人車擁擠路段，駕駛將車輛擦撞山壁減速，但山路曲折，因而翻覆，導致 1 死 43 傷。

肇車遊覽車為使用車身、車體與安全結構未經核可之車輛，也因牌照早被註銷，卻另掛他車車牌上路。調查發現，肇事車資未依規定定期檢驗，且逾期 6 個月，早在 86 年 9 月 1 日被高雄市監理處註銷牌照，但事發時所掛車牌竟是另一輛 30 人座的車型巴士車牌。

#### 二、縣 102 線九份路段遊覽車車禍



民國 93 年 10 月 18 日下午 2 時 40 分，一輛搭載 38 名來自香港觀光客的遊覽車，在臺北縣瑞芳鎮九份風景區汽車路與濱二路口，發生翻覆路側的重大意外，造成 5 死 32 輕重傷的交通事故。

此次事故發生在九份派出所附近的一個連續轉彎下坡處，肇事原因經鑑定為酒後駕車，事故發生時天雨路滑，山區濃霧大，遊覽車煞車不及，在九份派出所附近整輛車翻覆，墜落約 8 公尺深的濱二路。

### 三、縣 124 線苗栗遊覽車墜陡坡

民國 94 年 5 月 2 日下午 5 時 17 分，一輛遊覽車行經苗栗縣獅潭鄉 124 縣道一處大彎時，失控衝毀路旁水泥護欄，翻落約 50 公尺深的邊坡，結果造成 4 死 25 傷。肇事原因是彎度將近 180 度且超過 40 度的陡峭路段且疑路況不熟所致。事後，交通部清查出事遊覽車車籍資料時，發現該遊覽車 1990 年 3 月出廠，車齡逾 15 年。

### 四、台 21 線 92 豐丘路段遊覽車車禍

民國 95 年 10 月 2 日晚間 7 時 50 分，一輛搭載大陸觀光團的遊覽車，在南投縣東埔鄉往水里鄉的台 21 線 92K 路段發生失控墜落下邊坡，導致 6 死 15 傷的嚴重車禍。

研判肇事遊覽車從阿里山下山，行駛到豐丘明隧道附近數十公里的長下坡路段，未用低速檔控制速度，加上煞車氣墊爆破，煞車失靈，肇事前已無法換檔，雖然駕駛曾採取撞山壁緩衝措施，但仍然沒有避免遊覽車失控翻落葡萄園。此次發生意外的遊覽車駕駛，並不具備駕駛遊覽車的資格，而且行車記錄器顯示，該車在速限 50 公里的路段上，曾以高達 104 公里的時速行駛。

### 五、縣 188 線梅嶺遊覽車車禍

民國 95 年 12 月 3 日下午 4 時 40 分，一輛遊覽車搭載 46 人於臺南縣楠西鄉鄉道南 188 線梅嶺路段失控翻車，墜入 30 公尺深的溪谷，造成 22 人死亡、24 人輕重傷。本事故車輛 1988 年出廠，車齡 18 年，事故肇因在於下坡

路段的車速過快，疑似煞車失靈，在左彎路段直接衝出邊坡護欄，肇事鑑定時一度懷疑車輛使用再生胎，惟經公路總局調查，證實該車所有輪胎皆非再生胎。因為此次車禍是近年來死傷最慘重的交通事故，所以事故發生後，主管機關實施以下措施：

- 研擬並實施國道客運及遊覽車禁止使用再生胎。
- 修法將胎紋深度納入遊覽車定期檢驗項目之一。
- 事故發生後3個月完成修法，強制大客車裝置安全帶。
- 車齡12年以上之大客車限經營固定路線交通車，不得經營遊覽車，同時需在車身明顯部位標示出廠時間。

#### 六、台18線阿里山公路64K車輛向前滑動

民國96年1月14日，一輛遊覽車在台18線阿里山公路64公里處發生一起遊覽車事故。事故原因是車輛故障暫停在路旁修理時，車輛突然往前滑動，車內部份乘客受到驚嚇，緊急跳下車，結果造成九人受到輕重傷。

#### 七、陽明山仰德大道大客車車禍

民國96年6月24日下午6點40分左右，一輛有36人之遊覽車從陽明山仰德大道下山，行經永公路口時失控翻覆墜落100公尺下邊坡之山谷，造成8人死亡、25人輕重傷之事故。事故發生前，遊覽車下山時車速過快、煞車失靈，與前方小客車追撞，失控翻落山谷，以臺北市交通局初研判煞車失靈的原因很可能跟駕駛人習慣有關。

#### 八、台7甲線武陵農場遊覽車車禍

民國96年10月31日下午6點40分，一輛載滿43人的遊覽車在往武陵農場方向的台7甲線49K+400附近，翻覆掉落到約3公尺的下邊坡，造成28名乘客輕重傷。本事故車輛2004年出廠，車齡3年，肇事原因疑似遊覽車失控打滑，撞及山壁後車輛衝出路外，有幸下邊坡高程落差不大。

#### 九、台14線埔霧公路遊覽車與拖板車對撞

民國 97 年 5 月 17 日上午 10 時 25 分，一輛遊覽車行經南投縣埔里鎮台 14 線埔霧公路 62.3 公里處與拖板車對撞，結果造成 1 死 23 傷。肇車原因從事故現場初步發現，拖板車有明顯的煞車輪胎痕，且煞車痕未越過對向道，而遊覽車則無煞車痕，所以疑似為不當超車。

#### 十、臺北市仰德大道二段遊覽車與小客車擦撞

民國 97 年 10 月 11 日下午 2 時，一輛小客車行經臺北市仰德大道二段下坡路段，因車速過快，閃避不及與對向車道上的遊覽車擦撞，小客車 5 人輕重傷，遊覽車上 40 多名乘客無人傷亡。

本事故責任仍調查釐清中，初步勘查事故遊覽車在上山途中，下坡方向的自小客車疑似車速過快且跨越分向限制線，對向擦撞遊覽車，小客車車頭全毀，兩事故車輛均未翻覆或駛離路面。

#### 十一、國道 3 號北上 66.5K 遊覽車車禍

民國 97 年 11 月 9 日上午 8 時，一輛遊覽車行經國道 3 號北上 66.5 公里處龍潭交標道附近，肇車原因疑似天雨路滑、視線不佳、駕駛未保持安全距離，以致先擦撞前方貨車後，失控衝出護欄，翻落約 10 公尺深的產業道路，結果造成 3 死 39 傷。

#### 十二、中山高南下 22K 遊覽車車禍

民國 97 年 11 月 26 日中午 12 時 18 分，一輛遊覽車行經中山高南下 22 公里處，撞上一部正在施工中且隸屬國道高速公路局北區工程處的工程車，車上僅 21 名乘客受到輕傷，車禍原因仍調查釐清中。

#### 十三、台 18 線阿里山公路 67.3K 遊覽車與轎車對撞

民國 98 年 1 月 30 日，一輛搭載中國旅遊團 20 人的遊覽車在台 18 線阿里山公路 67.3 公里處發生一起遊覽車與小客車對撞車禍，結果造成 7 人受傷(遊覽車 2 人、小客車 5 人)。肇車原因為小客車下坡時，越過雙黃線、偏入對向車道，撞擊上山的遊覽車。

## 2.3 綜合評析

經由前兩節討論資料探勘及交通事故之相關文獻，彙整其資料探勘的應用，如表 2-3 所示；肇事分析之應用，如表 2-4 所示；國內道路遊覽車肇事案件分析如表 2-5 所示。

表 2-3 資料探勘應用彙整

作者(年份)	研究領域	研究對象	研究方法	研究結果
Raymond & Jiawei (1994)	資訊管理 資訊工程	溫哥華高級住宅區	群集分析	在群集方法分析下，不管是空間或非空間的實驗下，皆能找到各類型變數的分佈範圍。
Kao & Chiu (2001)	金融業	臺灣地區之銀行	CART、類神經網路	在三種模型預測正確率結果比較下，結合 CART 與類神經網路比 CART 及傳統類神經模型好。
Wang et al. (2007)	運輸物流	第三方物流公司之客戶	ID3 演算法	利用 ID3 演算法發現客戶特性，留住高價值客戶並開發潛在客戶，以提供高品質之服務。
Eduardo et al. (2002)	交通事故	墨西哥的道路交通事故資料庫	Bayesian network、Apriori 演算法、C4.8 及 CN2 法	先做簡單的分類並從交通事故的變數裡，找出發生交通事故的因素，然後判斷哪些因素主要影響傷亡人數。
Rangsipan (2007)	交通事故	泰國 Nakorn Pathom 省的道路交通事故資料	Apriori 演算法	找出各變數之間的關聯性及規則，提供泰國當地政府參考來改善
吳冠宏等人 (2006)	交通事故	臺灣地區之肇事變數	Two-Step 法、自組織映射圖與 K-Means	其分群結果發現 K-Means 演算法在肇事變數中具有比較好的分群能力，能有效的區分出不同的群集

資料來源：本研究整理

表 2-4 肇事分析之應用彙整

作者(年份)	研究對象	研究方法	研究結果
Al-Ghamdi (2003)	路口位置	成對樣本比較與 卡方獨立性檢定	叉路口與非交叉路口下之 事故比率皆為顯著
Amoros et al. (2003)	各級道路	負二項迴歸與羅 吉特迴歸	道路等級與肇事次數有關
Lee et al. (2002)	年輕駕駛肇事 次數	卜瓦松迴歸、負二 項迴歸與ZIP	剛拿到駕照一年內有較高 的肇事次數
楊思瑜 (2003)	桃竹苗地區小 型車	因子分析、群集分 析、多元羅吉特	年紀長、教育程度低、駕 駛資格完整、碰撞方向偏 向前方時最嚴重
邱裕鈞等人 (2004)	臺灣地區的肇 事案件	群集分析與判別 分析	根據最佳分群套入判別模 式，並以新資料再帶入判 別函數，其值最高者屬於 其群組

資料來源：本研究整理

表 2-5 國內道路遊覽車肇事案件分析彙整

時間	事件	傷亡情況	肇事型態	肇事原因	肇事地點
91.10.06	清境遊覽 車車禍	1 死 43 傷	翻覆	煞車失靈	長下坡
93.10.18	九份路段遊 覽車車禍	5 死 32 傷	翻覆墜落 8 公尺深之 下邊坡	酒後駕車	連續彎路 下坡處
94.05.02	苗栗遊覽 車墜陡坡	4 死 25 傷	翻覆墜落 50 公尺深 之下邊坡	駕駛路況不 熟、陡彎路段	長下坡 急彎處
95.10.02	豐丘路段遊 覽車車禍	6 死 15 傷	衝出路 外、翻覆	車速過快、煞 車失靈、違規	長下坡

表 2-5 國內道路遊覽車肇事案件分析彙整(續)

				行駛禁行路段	
95.12.03	梅嶺遊覽車車禍	22 死 24 傷	翻覆墜落 30 公尺深之下邊坡	車速過快、煞車失靈	長下坡急彎處
96.01.14	阿里山公路 64K 車輛向前滑動	9 傷	人為疏失	車輛滑動	長下坡
96.06.24	陽明山仰德大道大客車車禍	8 死 25 傷	翻覆墜落 100 公尺深之下邊坡	車速過快、煞車失靈	長下坡
96.10.31	武陵農場遊覽車車禍	28 傷	翻覆墜落 3 公尺深之下邊坡	車速過快、失控	長下坡
97.05.17	埔霧公路遊覽車與拖板車對撞	1 死 23 傷	對向擦撞	不當超車	連續彎路
97.10.11	仰德大道二段遊覽車與小客車擦撞	小客車 5 傷	對向擦撞	小客車違反分向限制標線禁制	長下坡
97.11.09	國道 3 號北上遊覽車車禍	3 死 39 傷	翻落約 10 公尺深的產業道路	天雨路滑、視線不佳、未保持安全距離	直線道
97.11.26	中山高南下遊覽車與工程車擦撞	21 傷	正向擦撞	尚調查釐清中	直線道
98.01.30	阿里山公路 67.3K 遊覽車與轎車對撞	7 人受傷 (遊覽車 2 人、小客車 5 人)	對向擦撞	小客車違反分向限制標線禁制	小客車下坡 遊覽車上坡

資料來源：本研究整理

運用資料探勘技術對於交通事故之研究或許還不夠成熟及完備，但資料探勘為新的方法且結合多個領域之技術，其技術包含統計學、機器學習、人工智慧及資料庫技術等，或許可以在交通領域上得到意想不到的結果。利用資料庫的技術，使資料的龐大及複雜性，可以加以整理，並利用統計、機器學習及人工智慧中的

演算法去分析，從龐大的資料中萃取出不明顯、隱藏的有用資訊與知識。正因為資料探勘本身並非單一技術或方法，其分析內容牽涉的層面也相當廣泛，所以為了達到有效率的探勘過程與有價值的結果，掌握研究的問題本質，對原始的資料進行前置處理，選擇正確的演算法才能獲得預期之結果。

從過去文獻中，資料探勘應用的領域有資訊、金融業、運輸物流，也有研究交通事故方面。但其實應用領域最廣泛的是在商業行為上，如銀行業、保險業、零售業等，但近年來也逐漸應用到醫療業、製造業、生化業等，其運輸安全的領域雖然少見，但也有研究嘗試分析交通事故在運輸安全領域上；從肇事分析之應用大多以統計方法做為評估，如負二項迴歸、羅吉特迴歸、卜瓦松迴歸等，依照分析路段或路口等來考慮所用之統計方法，也有用群集方法再套入多元羅吉特去分析，找出哪些肇事因素對傷亡嚴重程度影響最大。

近年來，一旦大客車或遊覽車發生交通事故後，都造成嚴重的傷亡，後果都不堪想。經由彙整的結果，都是發生於山區道路中，其肇事型態大致為翻覆及墜落至下坡邊、正面及對向擦撞；肇事原因為煞車失靈、酒後駕車、車速過快、不當超車、違反分向限制標線禁制、天雨路滑、視線不佳等因素，主要都是人為因素；肇事地點主要都在長下坡、連續彎路長下坡等。

發生在 95 年 12 月 3 日的梅嶺車禍及 96 年 6 月 24 日的陽明山仰德大道車禍，分別死傷人數為 22 死 24 傷及 8 死 25 傷，都為近年來嚴重的車禍。尤其在梅嶺車禍發生後，才受到政府相關單位的重視，因為發生事故的車輛不但車齡老舊且又使用再生胎，有鑑於此，才立即修法，並實施措施來防治車輛所造成交通事故的因素。但又發生陽明山車禍，或許有車輛、道路及環境因素的成份，但主要還是人為因素的成份居高，因此本研究希望透過資料探勘的方法，分析哪些肇事因子與人、車、路、環境等層面之間嚴重程度之相關性。

### 第三章 研究方法

本章根據前述的研究目的與過去相關文獻的探討，提出本研究之研究方法做為資料探勘方法之理論。本章研究方法為群集分析、判別分析及因素分析，而筆事資料是先使用群集分析並將此結果再套入判別分析中，此一流程作為後續筆事資料分析的基礎。

#### 3.1 群集分析

群集分析主要目的是分析資料彼此間的相似程度，藉由分析找到群集結果，推論出有用或隱含的特性和現象。

##### 一、基本概念

群集分析是根據多變量之間不同的相互關聯程度，而將比較相近的變量或樣本聚集在一起，因而形成所謂的群集。群集分析的主要目標是依照收集的  $p$  個群集變數( $x_1 \sim x_p$ )將  $n$  個個體分成幾個群，使群體內個體間離的近(相似性高)，不同群的個體離的遠(相似性低)，所以群集分析的群內差異小，群間差異大。

##### 二、統計假設

許多數量方法都是在作統計推論的工作，也就是說，想藉由樣本推論至母體，然而，集群分析則是完全不同，集群分析不作統計推論的工作，而是將觀察值的結構予以量化，所以適用於一般數量方法的統計假設，例如：線性、常態性和變異數相等性(Homoscedasticity)都不適用於群集分析。雖然如此，群集分析的統計假設尚須考慮二點：

- 樣本的代表性：由於群集分析無法藉由樣本推論至母體，因此研究人員必須確定取得的樣本是足以代表母體的。
- 共線性的問題：共線性發生時，具有共線性的變數會有加權的情形產生，會影響計算結果，解決共線性的問題，可以使用減少變數至相同的數目或使用



馬氏距離(Mahalanobis Distance)方法計算距離，以避免共線性的影響。

### 三、決策流程

群集分析的決策流程由圖 3-1 所示，包括研究問題與目的、選擇變數、衡量相似性、選擇群集方法、決定群集數、解釋群集等步驟。

#### 1. 研究問題與目的

首先要確定群集分析的主要目的是探索性或是驗證性，傳統的群集分析最常用於探索性的用途，即將觀察值作客觀的分類，目的是要將觀察值分群，使得每群內的元素對群集變數而言，同質性高。

#### 2. 選擇變數

不論研究的目的是探索性或驗證性，群集分析的結果都會受到所選變數的限制，變數的選擇必須兼顧理論、觀念和實務的考量，不應該毫無選擇的把不相關的變數納進來，因為群集分析技術無法區分相關和不相關的變數。如果選擇不相關的變數，可能會增加離群值(Outliers)的產生機會，對於群集分析結果會有重影之影響。

#### 3. 衡量相似性

衡量相似性就是量測觀察值之間的相似性，以作為分群的基礎，群集分析將觀察值間相似性高的集合在一起，以形成一個個的群體。量測觀察值間相似程度的方法有很多種，常用的有關連衡量(Association Measures)、相關衡量(Correlation Measures)、距離衡量(Distance Measures)，其中距離衡量有歐幾里得距離(Euclidean Distance)、城市街道距離(City Block Distance)、馬氏距離(Mahalanobis Distance)三種方法。

#### 4. 選擇群集方法

將成對的事物間的相似性加以衡量後，接著就可以利用群集方法將各樣本歸入群集中。群集方法主要有三大類：

##### I. 階層式(Hierarchical)：分為集結法(Agglomerative Method)及分離法

(Divisive Method)，其中集結法有單一連結法(Single Linkage Method)又稱最近法、完全連結法(Complete Linkage Method)又稱最遠法、平均連結法(Average Linkage Method)、中心法(Centroid Method)及華德法(Ward's Method)又稱最小變異數法(Minimum Variance Method)等五種方法。

II. 非階層式(Non Hierarchical)：為 K 平均法(K-Means Method)。

III. 二階段法：結合階層法和非分層法。

## 5. 決定群集數

到目前為止，沒有一定的標準可以決定群集的數量，可理解的是群集數量的增加，群集內觀察值的相似性也會增加，解讀資料的能力下降；群集數量減少，群集內觀察值的相似性也會減少，解讀資料的能力上升。因此，決定群集數目端看研究的取捨。

雖然，沒有一定的標準可以決定數量，但依據蕭文龍(2007)【27】整理，如下表示至少有些方針可以遵循：

- 理論上支持：由理論的延伸到可以說明群集數目。
- 實務上的考量：為了解決某些問題，在實務上必須決定群集數目才能說明實務上的現象，有時候在實務上會決定群集數的範圍，再逐一評估最佳的群集數。
- 群集數距離突增時：使用階層式和非階層式的群集程序時，發現群集步驟突然增加很多時，代表群體間有較大的差異，表示可以決定群集數的所在。

## 6. 解釋群集

在解釋每個群集的目的是希望能找出正確代表該群集的名稱，以說明該群集的代表意義，一般情形下，會檢視群集的平均值和重心，以描述該群集的表徵，找出合理的解釋。

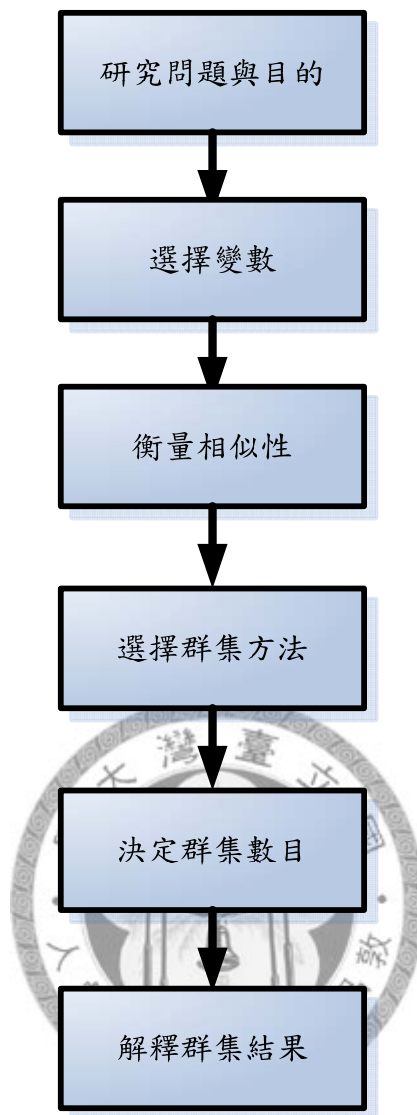


圖 3-1 群集分析之決策流程

#### 四、相似性衡量

選擇使用的相關變數、惕除異常值後，就要衡量觀察值間的相似性，即確定觀察值與觀察值間的相似程度。衡量方法分為關聯衡量、相關衡量及距離衡量等三大類，分別敘述如下：

##### 1. 關連衡量

關連衡量適用於非量化的變數，例如：名目尺度或順序尺度，常用的方式是回答百分比來計算。

## 2. 相關衡量

相關衡量是將代表觀察值用矩陣的方式計算成對觀察值間的相關係數，然後以各觀察值間相關係數的大小來判定觀察值間的相似性，高的相關係數代表有高的相似性。當對變數做分群而不是對個體做分群時， $x_i$  和  $x_j$  兩個變數間的距離可以用  $1-r_{ij}$  來測量，其中  $r_{ij}$  為  $x_i$  和  $x_j$  的相關係數，其公式為：

$$d(i, j) = 1 - r_{ij} \quad (\text{式 3.1})$$

## 3. 距離衡量

距離衡量主要是以點與點間的距離為代表，有歐幾里德距離、馬氏距離、城市街道距離三種。

- 歐幾里德距離：為最常使用的方法，此法是多維空間兩點的幾何距離，其公式如下：


$$d(x, y) = \left( \sum_{i=1}^p (x_i - y_i)^2 \right)^{1/2} \quad (\text{式 3.2})$$

由於歐幾里德距離有根號計算的麻煩，因此也經常改為歐幾里德距離平方做分析，其公式如下：

$$d(x, y) = \sum_{i=1}^p (x_i - y_i)^2 \quad (\text{式 3.3})$$

其中  $x = (x_1, \dots, x_p)$  和  $y = (y_1, \dots, y_p)$  為兩個群體在  $p$  種量測的數據。

- 城市街道距離：為各個變數差距之和，在很多情形以此種距離做分群所得結果與歐幾里德距離做分群相似；但城市街道距離會將變數差距大的效應減弱，因為歐幾里德距離在單一維度上是以差距平方為考量。另外，使用城市街道距離是有條件的，那就是變數間不能有相關性，若變數間有相關性，則需使用馬氏距離。其公式如下：

$$d(x, y) = \sum_{i=1}^p |x_i - y_i| \quad (\text{式 3.4})$$

- 馬氏距離：為歐幾里得距離平方的延伸使用，它不只是標準化資料且加總組內變數矩陣以調整變數間的相互關係，特別適用於變數有相關性時使用，其公式如下：

$$d(x, y) = \frac{\sum_{i=1}^p x_i y_i}{\sqrt{(\sum_{i=1}^p x_i^2)(\sum_{i=1}^p y_i^2)}} \quad (\text{式 3.5})$$

## 五、階層式群集方法

階層式群集方法分為集結法和分離法。集結法的方式，係統計算出各觀察值間的距離或組內誤差矩陣，然後將最近的兩個觀察值加以合併成一群集，接著再算出合併後的觀察體間距離或組內誤差，並重覆以上的程序，直到所有觀察值合併成同一群集。較常使用的方法有單一連結法、完全連結法、平均連結法、中心法及華德法。

- 單一連結法(又稱最近法)

單一連結法對 A、B 兩群距離計算是以 A 群內每一點到 B 群內每一點的距離中之最小值當做 A、B 兩群距離，其公式如下：

$$d_{A,B} = \min_{\substack{i \in A \\ j \in B}} d_{ij} \quad (\text{式 3.6})$$

- 完全連結法(又稱最遠法)

完全連結法對 A、B 兩群距離計算是以 A 群內每一點到 B 群內每一點的距離中之最大值當做 A、B 兩群距離，其公式如下：

$$d_{A,B} = \max_{\substack{i \in A \\ j \in B}} d_{ij} \quad (\text{式 3.7})$$

➤ 平均連結法

平均連結法的距離計算是以 A 群內每一點到 B 群內每一點的距離之平均，其公式如下：

$$d_{A,B} = \frac{\sum_{i \in A} \sum_{j \in B} d_{ij}}{n} \quad (\text{式 3.8})$$

其中  $n$  表示全部距離的個數。

如下圖所示，單一連結法、完全連結法及平均連結法的群集間距離

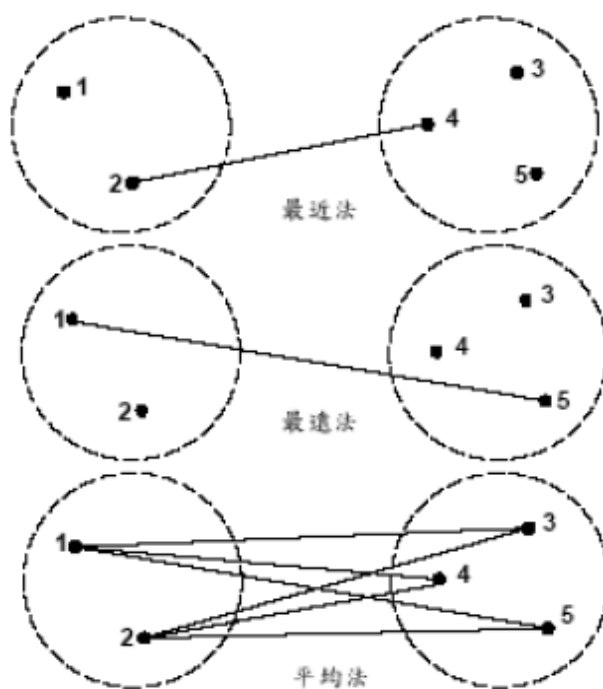


圖 3-2 群集間距離圖

➤ 中心法

A、B 兩群距離以 A 群的中心點  $\bar{x}_A$  到 B 群的中心點  $\bar{x}_B$  的距離，其公式如下：

$$d_{A,B} = d(\bar{x}_A, \bar{x}_B) = \left\| \bar{x}_A - \bar{x}_B \right\|^2 \quad (\text{式 3.9})$$

➤ 華德法(又稱最小變異數法)

計算 A、B 兩群距離是以 A 群中心點  $\bar{x}_A$  到兩群合併中心點  $\bar{x}$  距離平方乘以 A 群的個數與 B 群中心點  $\bar{x}_B$  到兩群合併中心點  $\bar{x}$  距離平方乘以 B 群的個數之和，其公式如下：

$$d_{A,B} = n_A \cdot \left\| \bar{x}_A - \bar{x} \right\|^2 + n_B \cdot \left\| \bar{x}_B - \bar{x} \right\|^2 \quad (\text{式 3.10})$$

分裂法的概念與集結法相反，它是依序將相異性最大的觀察值分裂成另一集群，步驟如下：

1. 首先將所有觀察值當成一個群集。
2. 接著計算哪一個觀察值與群集相異性最大或相似性最小，通常取此觀察值與群集內其他觀察值的平均距離，然後將其分裂成另一群集。
3. 分別計算大群集中每個觀察值與群集內及群集外的平均距離，如果比較接近另一個群集，則將此觀察值分到另一個群集，否則就留在群集內。此過程一直到兩個群集內的觀察體都不能再分裂為止。
4. 重覆第 2 步及第 3 步兩個步驟，將所有觀察值一直分裂到單獨一個群集為止。

#### 六、非階層式群集方法

非階層式群集分析中最常使用的方法為 K-Means 群集分析法。K-Means 群集分析法是將  $n$  個資料分至  $k$  個群集中，不斷的運算，直到達到目標函數的最小值，即得到最終的群集結果，也就是每一群之間的資料是最相似的，而群與群之間的相似度則是最小的。其衡量相似度的基準在於求出每一群資料的平均值，之後以此平均值為基準，來衡量平均值與資料間的相似度。

目標函數通常採用平方誤差準則，其公式如下：

$$E = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2 \quad (\text{式 3.11})$$

其中， $E$  表示所有群集觀察值的平方誤差和， $p$  是群集對象， $m_i$  是  $C_i$  類別的各群集觀察值之平均值，即：

$$m_i = \frac{\sum_{p \in C_i} p}{|C_i|} \quad (\text{式 3.12})$$

其中， $|C_i|$  表示  $C_i$  類別的群集觀察值數目。

因為在每一次處理中，每一個點要計算與各群集中心的距離，並將距離最近的群集作為該點所屬類別。K-Means 群集分析法是解決群集問題的一種典型演算法，此演算法簡單且快速。然而，K-Means 群集分析法對初始值敏感，對於不同的初始值，可能會導致不同的群集結果。

因此，K-Means 群集分析法的集群過程分為以下四步驟：

步驟一：指定要形成的集群數，對樣本進行初始分群並計算每一群的重心。

步驟二：調整分群，計算每個觀察值到各群重心的距離，把每個觀察值歸入距離重心最近的那一群。

步驟三：重新計算每一群的重心。

步驟四：重覆步驟二~三，直到沒有觀察值可以再調整為止。

## 七、群集數評量指標

到目前為止，還沒有討論過如何確定群集數，群集分析的目的是要對研究對象進行分群，因此如何選擇分群數成為集群方法中的主要問題之一。在選擇非層次群集法需要先設定分群數，所以陳順宇（民 94）【28】提出三個評量指標，分別為 RMSSTD(Root-Mean-Square Standard Deviation)、 $R^2$ 、SPR(Semipartial R-Square)。

### 1. RMSSTD

新結合群體中所有變數的綜合標準差，即

$$RMSSTD = \sqrt{\frac{\sum_{i=1}^p s_i^2}{p}} \quad (\text{式 3.13})$$



其中  $s_i^2$  為新結合群內第  $i$  個變數的變異數， $p$  為變數個數， $RMSSTD$  愈小表示群體內個體的相似性愈高。

## 2. 判定係數( $R^2$ )

$$\text{做 ANOVA 的 } R^2, \text{ 即 } R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (\text{式 3.14})$$

當  $R^2$  愈大表示群體間的相異性愈高，也間接地顯示分群的效果愈好。

## 3. SPR

是以兩個群體組成新結合群體，群內相似性損失比例為指標，即

$$SPR_i = R_{i+1}^2 - R_i^2 = (\text{新結合群內變異} - \text{未結合前群內變異}) \div \text{所有樣的本群內變異} \quad (\text{式 3.15})$$

其中  $R_i^2$  分成  $i$  群時的  $R^2$ ， $R_{i+1}^2$  為分成  $i+1$  群時的  $R^2$ 。SPR 愈小表示形成新結合群體時，群內相似性損失比例愈小，即新結合群體的群內相似性愈高。

## 八、群集結果的驗證

由於群集數的選定後，需要由卡方檢定來驗證此群集數和各變數是否相關。統計假設為

$$\begin{cases} H_0: \text{行變數與列變數有關(群集數和各變數有關)} \\ H_1: \text{行變數與列變數無關(群集數和各變數無關)} \end{cases}$$

若變數和群集數有關就保留，無關之變數予以剔除。

基於上述之理論方法，本研究使用群集分析之作法流程為：在進行分群前，先將資料標準化，以使得分群時所有變數均可發揮相同的效力，不會有某一個變數因其數值分佈大小而主導分群結果。接著試用 2-10 個群集數，然後以 ANOVA 的判定係數( $R^2$ )去決定群集的數目。決定群集的數目後，再由交分析之卡方檢定了解群集數和各筆事變數是否相關，相關變數保留，不相關變數予以剔除。

## 3.2 判別分析

判別分析主要目的是了解群體間的差異，是先利用判別變數建立判別規則(判別函數)，然後再由判別規則對個體做分類，預測每個個體屬於各群組的可能機率。

### 一、基本概念

判別分析是一種判別和分類的技術，是由一個分分類(grouping)變數當因變數與多個計量的區別(discriminant)變數當自變數的技術。判別分析會建立此分類變數與其它多個計量變數間的對應關係而建立判別函數。判別分析的基本模型就是判別函數，亦即判別變數的線性函數關係，其公式如下：

$$D = d_0 + d_1x_1 + d_2x_2 + \cdots + d_ix_i + \cdots + d_nx_n \quad (\text{式 3.16})$$

其中， $D$  為判別函數值， $x_i$  為各判別變數， $d_i$  為相對應的判別係數。

根據黃俊英（1998）【29】提到，判別分析有七個基本假定：

1. 兩個或兩個以上的群體。
2. 每個群體至少有兩個觀察值。
3. 預測變數項應少於總觀察值減 2。
4. 預測變數為區間尺度。
5. 任何預測變數都不是其他預測變數的線性組合。
6. 每一組群體的組內共變異數矩陣大致相等，否則，判別函數就不能使各組的差異達到最大。然而，如果樣本數不多，而各組內共變異數矩陣差異性也不大，則使用一般區別分析仍是恰當的。
7. 每一組群體都是從多變量常態分配的母體中抽選出來的。不過，隨著樣本數增加，這個假設經常會違反。如果資料嚴重違反多變量常態分配的假設，可以改用羅吉特迴歸或是無母數判別分析。

### 二、判別函數

建立判別函數的方法有很多，比較常用的方法為費雪法(Fisher's Method)及馬氏距離法(Mahalanobis'  $D^2$  Method)，分別敘述如下：

### 1. 費雪判別函數

費雪判別函數的重點在於使組間變異對組內變異之比值為最大，

即  $T = W + A$ 。

其中， $T$  為總樣本的平方與交叉乘積和(SSCP)矩陣

$W$  為聯合組內平方與交叉乘積和(Pooled within-groups SSCP)矩陣

$A$  為組間平方與交叉乘積和(Between-groups SSCP)矩陣

在判別分析中，也須找出一組權重( $K$ )，以求得判別軸( $D$ )，即  $D = XK$ 。

$$\text{又 } K = \begin{bmatrix} K_1 \\ K_2 \\ \vdots \\ K_m \end{bmatrix}, \text{ 則 } D = XK = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1m} \\ X_{21} & X_{22} & \cdots & X_{2m} \\ \vdots & \vdots & & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nm} \end{bmatrix} \begin{bmatrix} K_1 \\ K_2 \\ \vdots \\ K_m \end{bmatrix} = \begin{bmatrix} D_1 \\ D_2 \\ \vdots \\ D_n \end{bmatrix} \quad (\text{式 3.17})$$

將原始資料矩陣  $X$  轉化為以總樣本平均修正的資料矩陣  $X_d$ ，則

$$D = X_d K = \begin{bmatrix} X_{d11} & X_{d12} & \cdots & X_{d1m} \\ X_{d21} & X_{d22} & \cdots & X_{d2m} \\ \vdots & \vdots & & \vdots \\ X_{dn1} & X_{dn2} & \cdots & X_{dnm} \end{bmatrix} \begin{bmatrix} K_1 \\ K_2 \\ \vdots \\ K_m \end{bmatrix} = \begin{bmatrix} D_1 \\ D_2 \\ \vdots \\ D_n \end{bmatrix} \quad (\text{式 3.18})$$

所以費雪的線性判別函數為：

$$D_i = X_{di} K = X_{di1} K + X_{di2} K + \cdots + X_{dim} K \quad (i = 1, 2, \cdots, n) \quad (\text{式 3.19})$$

利用此公式，可求得所有樣本在判別軸  $D$  的分數，然後比較  $\bar{d}$  (假設 A 與 B 之平均數)與  $d_i$ ，即：

$d_i > \bar{d}$ ，將第  $i$  個樣本數指派到群體 A 或 B；

$d_i < \bar{d}$ ，將第  $i$  個樣本數指派到群體 B 或 A。

## 2. 馬氏距離法

馬氏距離法是將傳統的歐幾里德距離公式用聯合組內共變數矩陣( $C_w$ )來加以調整。兩個平均數修正之重心  $k$  和  $h$  間的馬氏距離( $D^2$ )，其公式如下：

$$\begin{aligned} D^2 &= (X_k - X_h)' C_w^{-1} (X_k - X_h) \\ &= (X_{k1} - X_{h1}, X_{k2} - X_{h2}, \dots, X_{km} - X_{hm}) C_w^{-1} \begin{bmatrix} X_{k1} - X_{h1} \\ X_{k2} - X_{h2} \\ \vdots \\ X_{km} - X_{hm} \end{bmatrix} \end{aligned} \quad (\text{式 3.20})$$

如果第  $i$  個樣本點與群體 A 的重心  $a$  二者間的  $D^2$  較該點到群體 B 之重心  $b$  的  $D^2$  為小，則該樣本數  $i$  即為群體 A；反之，則為群體 B。即：

$D_{ia}^2 < D_{ib}^2$  為 A 群體； $D_{ia}^2 > D_{ib}^2$  為 B 群體。

費雪判別函數和馬氏距離法的主要差異在於前者是將原有空間投影在區別軸  $Y$  上，後者是直接在預測變數的原有空間中計算距離。不過，這兩種方法所獲得的判別結果是完全一樣的。作為一種劃分群體的技術而言，費雪判別函數與馬氏距離具有相同的效果。

另外，費雪判別(又稱典型判別分析)，主要目的是將判別變數經線性組合變成典型變量(canonical variable)，是以  $p$  個判別變數來區分  $g$  個群體，所以在空間上會改成  $\text{Min}(p, g-1)$ 。費雪判別不需假設資料是常態分配就可以得到的判別函數，並使群體間差異最大(即  $F$  值最大)的線性組合。

## 三、統計顯著性檢定

在統計的顯著性檢定，有 Hotelling 的  $T^2$  檢定、Bartlett 的  $\chi^2$  檢定及 Wilks' Lambda 檢定，分別敘述如下：

### 1. Hotelling 的 $T^2$ 檢定

$$T^2 = \frac{n_a n_b}{n_a + n_b} d' C_w^{-1} d \quad (\text{式 3.21})$$

$T^2$  與  $F$  分配的關係如下：

$$\frac{(n-m-1)}{m(n-2)} T^2 \sim F(m, n-m-1) \quad (\text{式 3.22})$$

其中， $n = n_a + n_b$  (總樣本數)， $m$  為預測變數數目。

## 2. Bartlett 的 $\chi^2$ 檢定

$$\chi_m^2 = -(n-1 - \frac{m+g}{2}) \ln \Lambda_k \sim \chi^2(p-k)(g-k-1) \quad (\text{式 3.23})$$

其中， $n$  為總樣本數， $m$  為預測變數數目， $g$  為群體數目。

## 3. Wilks' Lambda( $\Lambda$ )檢定

$$\text{Wilks' Lambda}(\Lambda) = \prod_{i=k+1}^q \frac{1}{1 + \lambda_i} \quad (\text{式 3.24})$$

其中， $q$  為判別函數數目， $\lambda$  為特徵值， $k$  為代表刪除的函數數目。

## 三、逐步判別分析

對群體做判別分析時，會尋找許多有判別能力的變數，但這些變數的選取是必須經過篩選，將沒有判別能力或與其他變數有共線性的變數刪除，因此用來篩選判別變數的方式就是逐步判別分析。利用逐步判別分析找出有判別能力的變數後，再用這些有判別能力的變數做線性(或典型)判別。

逐步判別分別向前選取(Forward stepwise)及向後選取(Backward stepwise)兩種。向前選取法是將最有判別能力的變數選入，也就是在每一步驟將評估選取最有判別能力的變數進入模式內，然後再進行下一步驟，選取剩下的變數中最有判別能力之變數，如此依序進行；向後選取法是將所有變數都選入模式內，接著第一步驟將最沒有判別能力的變數刪除，再依序將沒有判別能力(其偏  $F$  值小於門檻)的變數一一去除。

逐步判別分析的程序是以指定的  $F_{in}$  值與  $F_{out}$  值做為門檻，以便決定何時停止選取，對一變數的偏  $F$  值是判別群體間的顯著性指標。向前逐步選取區別分析，在每一步驟當變數的偏  $F$  值大於指定的  $F_{in}$  值時就選入變數，否則就停止選取；向

後逐步判別分析，在每一步驟當變數的偏 F 值小於指定的  $F_{out}$  值就去除，否則就停止選取。

#### 四、正確判別率

判別分析的另一個目的是用來對每一個觀察值做分類，當建立判別規則後，用分類表(為分類矩陣)來當作判別效果的指標，分類表中顯示有多少個觀察值被正確分類(在矩陣的對角線上)，以及有多少觀察值被錯誤分類。分類表是建立判別規則，然後再以此判別規則預測資料內每一個觀察值屬於那一群體，這樣做會得到較少比例的誤判。因此，本研究將肇事資料總樣本數的三分之二當作訓練樣本，三分之一當作測試樣本，比較訓練樣本和測試樣本其分類正確率。

基於上述之理論方法，本研究卻使用判別分析之作法流程為：群集分析後，將分類好的群集和各肇事變數作費雪(典型)判別分析，並以 *Wilks' Lambda* ( $\Lambda$ ) 值及卡方值並檢定其判別函數是否顯著。而本研究是用逐步判別分析的向前選取法將影響各年度重要的變數選取出來並比較訓練樣本和測試樣本的分類正確率。

### 3.3 因子分析

因子分析是數學中的一種精簡作法，它能夠將眾多的變數濃縮成個數較少的幾個變數，這些精簡的變數就是因子，因此此種統計處理就稱為因子分析。因子分析理論是假定樣本在變數上之得分，係由兩個部分組成，一個是各變數共有的成份，即共通因子，另一個是各變數所獨有的成份，即獨特因子，共同因子可能是僅有一個、兩個或好幾個。而萃取出變數間的共同因子，便是因子分析的主要目的。

為使原來的變數有較大的負荷量在主要因子上，而讓各主要因子的意義更為清晰和明顯，需利用因子旋轉使各新變數具有較大的獨立性，即在每個因子中的

部分變數具有較大的權重，並利用這少數的因子來代表原多變數。最後可獲得因子矩陣及其他有關的資料，並可由各主要因子所代表的意義加以闡釋，以了解變數與因子間的相關性。

### 一、模式架構

因子分析是討論如何將 $p$ 個變數 $x_1 \sim x_p$ 的每一個變數 $x_i$ 分解成少數幾個( $q$ 個且 $q \leq p$ )共同因子 $f_j, j=1, 2, \dots, q$ ，與獨特因子 $\varepsilon_i$ 的線性組合。因子分析的領域包括甚廣，如：區位、交通、成本的控管等，都可以利用其固有的方法進行解析研究。而因子分析的模型可將實際變動的數據，利用一次線性結合式來表示，稱之為因子模式，如下列各式：

$$\begin{aligned} x_1 &= l_{11}f_1 + l_{12}f_2 + \dots + l_{1q}f_q + \varepsilon_1 \\ x_2 &= l_{21}f_1 + l_{22}f_2 + \dots + l_{2q}f_q + \varepsilon_2 \\ &\vdots \\ x_p &= l_{p1}f_1 + l_{p2}f_2 + \dots + l_{pq}f_q + \varepsilon_p \end{aligned} \quad (\text{式3.25})$$

其中 $f_1, \dots, f_q$ 是共同因子，它們在每一個變數 $x_i$ 中都共同擁有，而 $\varepsilon_i$ 是獨特因子，只有在第 $i$ 個變數中才擁有。 $l_{ij}$ 為第 $i$ 個變數 $x_i$ 在第 $j$ 個共同因子 $f_j$ 的因子負荷。此因子分析模式，有下列幾個基本假設：

1. 獨特因子 $\varepsilon_1, \dots, \varepsilon_p$ 是互相獨立且是常態分配， $\varepsilon_i$ 的平均數為0，而變異數為 $\Psi_i$ ，即

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_p \end{bmatrix} \sim N \left[ \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \Psi = \begin{bmatrix} \Psi_1 & 0 & \dots & 0 \\ 0 & \Psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Psi_p \end{bmatrix} \right] \quad (\text{式3.26})$$

其中， $\Psi$ 是對角矩陣，表示獨特因子 $\varepsilon_i$ 之間是獨立的。

2. 共同因子 $f_1, \dots, f_q$ 間的共變異矩陣為 $\Phi$ ，即：

$$f = \begin{bmatrix} f_1 \\ \vdots \\ f_q \end{bmatrix}, \text{Cov}(f) = E(ff') = \Phi = \begin{bmatrix} \Phi_{11} & \Phi_{12} & \cdots & \Phi_{1p} \\ \Phi_{21} & \Phi_{22} & \cdots & \Phi_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_{p1} & \Phi_{p2} & \cdots & \Phi_{pp} \end{bmatrix} \quad (\text{式3.27})$$

一般要求 $\Phi$ 對角線上的元素 $\Phi_{ii}=1$ ，而當 $i \neq j$ 時（即對角線外） $\Phi_{ij}=0$ ，也就是 $\Phi=I$ ，它表示因子間是獨立的且變異數皆為1。

3. 共同因子與獨特因子皆為獨立，即 $\text{Cov}(f_j, \varepsilon_i)=0$ ，對所有 $ij$ ，因子分析模式也可以寫成矩陣表示 $X=Lf+\varepsilon$ 。其中，

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}, \mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix}, L = \begin{pmatrix} l_{11} & \cdots & l_{1q} \\ l_{21} & \cdots & l_{2q} \\ \vdots & \ddots & \vdots \\ l_{p1} & \cdots & l_{pq} \end{pmatrix}, f = \begin{pmatrix} f_1 \\ \vdots \\ f_q \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{pmatrix} \quad (\text{式3.28})$$

而 $E(f)=0$ ， $\text{Cov}(f)=\Phi$ ， $E(\varepsilon)=0$ ， $\text{Cov}(\varepsilon)=\Psi$ ， $\text{Cov}(f, \varepsilon)=0$ 。

二、共通性與獨特性之因子模式(式3.25)可寫成

$$x_i = \sum_{j=1}^q l_{ij} f_j + \varepsilon_i = c_i + \varepsilon_i \quad i=1, 2, \dots, p \quad (\text{式3.29})$$

稱為 $x_i$ 的「共通」部份，即變數 $x_i$ 分解成二個獨立的部份，一為共通部份，一為獨特部份，而 $x_i$ 的變異數也分成共通部份的變異數與獨特部份的變異數(由於 $c_i$ 與 $\varepsilon_i$ 獨立)，即

$$\text{Var}(x_i) = \text{Var}(c_i) + \text{Var}(\varepsilon_i) \quad (\text{式3.30})$$

由(式3.29)及 $\text{Var}(f_i)=1$ ，得

$$\text{Var}(c_i) = \sum_{j=1}^q l_{ij}^2 = h_i^2 \quad (\text{式3.31})$$

$\text{Var}(c_i)$  稱為變數 $x_i$ 的「共通性」，而 $\Psi_i = \text{Var}(\varepsilon_i)$  稱為「獨特性」。



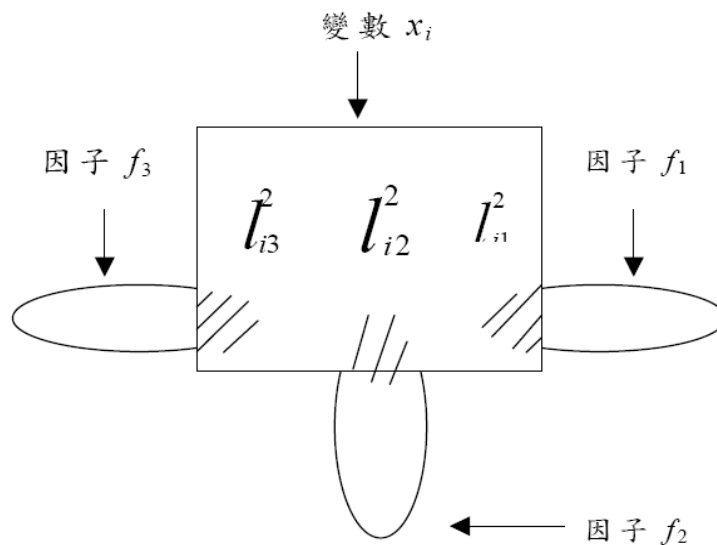


圖 3-3  $x_i$  的共通性  $h_i^2 = l_{i1}^2 + \dots + l_{iq}^2$

而獨特性就是  $x_i$  的變異沒有被所有共同因子解釋的比例為  $\Psi_i = 1 - h_i^2$ ，且每個

變數  $x_i$  的變異數  $\sigma_{ii}$  可表成共通性  $h_i^2 = \sum_{j=1}^q l_{ij}^2$  與獨特性  $\Psi_i$  的和，即

$$\sigma_{ii} = l_{i1}^2 + l_{i2}^2 + \dots + l_{iq}^2 + \Psi_i \quad (\text{式3.32})$$

### 三、共通因子數目的取決

因子分析的重要任務在於抽取  $q < p$  個共通因子，使能以少數  $q$  個向度空間便可適當地代表  $p$  個變數。所以在使用主成份法時，由於其反覆地求解，目的即為了使共通因子的數目減少，而就選取因子的標準有下列各項：

1. Kaiser 準則（1960）：特徵值大於1者即選取。此法為電子計算機程式中最流行的一種，因特徵值小於1 的共通因子對總變異數之貢獻被視為微不足道，因此放棄其參考價值，因為每一個變數的變異數為1，而它的貢獻並未超過1。
2. Gutman's 準則：特徵值大於0者即選取，其主要為防止重要的共通因子被忽略。此法稱為「古特曼最強下限」標準，旨在將特徵值為「負」的所有共通因子予以放棄，這是較為保守的方法。
3. 陡坡考驗法（scree test）：運用數值曲線圖，其特徵值可經由圖示予以取

捨（本法趨於主觀）。因特徵值通常由大至小順序出現，設定以大小為橫座標，以數值大小為縱座標，其低點即出現在右下角，參考價值低。

4. 經驗法則（1970）：再統計分析時常常出現統計與實質的意義不符合或不能做合理的解釋，所以抽取到此一共通因子反而是一種困擾。因此在使用以上任何方法求解時，尚需加入理智的判斷，如此方不致失去因子分析的真正意義。

#### 四、因子旋轉

一般求出之因子負荷，往往不能顯著地解釋元變量與因子間的關係，故必須對因子負荷進行旋轉，使其平方值向0與1極化，以便於解釋原變量與因子之關係。因子旋轉有許多方法，可分為直交轉軸法和斜交轉軸法兩種。使用直交轉軸法時，各因子間均保持90度的關係，即因子與因子間彼此是互相獨立的。而斜交轉軸法則應用於當因子間確定存在於某種關係時。

#### 五、因子得分

因子分析目的是以較低維度的變數組合，即數個主成份因子來代替原始龐大的資料結構。通常是以因子負荷量較大的變數作為主成份因子的代表，並重新計算每個樣本在主成份因子上的總合因子得分，以取代原始變數的集合，作為後續研究之用。

#### 六、適合度檢定

對於分析的資料是否適用於因子分析，可用Kaiser-Meyer-Okin (KMO) 的抽樣正確量，由偏相關係數來檢定，其定義為：

$$KMO = \frac{\sum_{i \neq j} \sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} \sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} \sum_{i \neq j} a_{ij}^2} \quad (\text{式3.33})$$

其中 $r_{ij}$ 為變數 $i$ 與 $j$ 之簡單相關係數， $a_{ij}$ 為此兩變數之偏相關係數。若所有 $a_{ij}^2$ 之和遠小於所有 $r_{ij}^2$ 之和時，則KMO值接近1，表示這些變數有共同因子存在，可

使用因子分析；一般來說，KMO 值低於0.4 是不可接受的，其值愈大愈好，0.75以上可視為極佳。

## 七、因子命名

因子分析為探索性研究的重要工具。一般探索性研究是應用因子分析精簡資料，將原始變數歸屬於各主要因子並賦予適當的名字，以作為進一步統計分析之輸入資訊。在正交因子中，是以因子負荷量之大小，也就是顯著性來選取主要因子，而因子負荷之顯著性，通常以因子負荷之絕對值大於0.3就被認為顯著的，大於0.5則非常顯著。由主要因子中負荷量大的變數之相關描述，可給予主要因子一適當的名字。

### 3.4 小結

從資料分析中，群集分析對於觀察值分類或分組都是未知的，而判別分析在作分析之前就已經把觀察值分類或群組。本研究方法是想透過群集分析法把大客車的各年度及全部年度的肇事資料內相似性的樣本分組，再把各肇事變數利用卡方檢定去檢定變數是否有顯著性，有顯著性的變數就代表此變數在分類好的群組是有相關性高，不顯著就予以剔除；然後再把分類群組的肇事變數做判別函數並預測其分類正確率及判別函數是否顯著，透過逐步判別分析選取重要之肇事變數，最後比較全部年度及各年度的肇事變數之關係及差別。

從案例分析中，擬採用主成份分析法並使特徵值大於 1 之取決因子最大數目的標準，即採取 Kaiser 準則及配合陡坡圖考驗法進行分析。從群集分析的群集數目套入判別分析中，選定最佳判別正確率當作選定的因子數目。因子分析用直交轉軸之最大變異法，為了使轉軸後每一個共通因子本身之內的因子負荷量大小相差盡量達最大，利用最大變異法轉軸後所得之因子結構較為簡單且容易解釋，最後將因子命名。最後，做事故嚴重程度迴歸預測模式。

## 第四章 筆事資料前置處理

本章說明筆事資料的分析流程及蒐集，並做筆事資料的前置處理，作為後續套入群集分析及判別分析之準備。

### 4.1 分析流程

本研究分析流程如圖 4-1 所示，包括資料選取、資料前置處理、執行群集分析並解釋結果、建構筆事判別分類模式，分別敘述如下：

#### 一、資料選取

資料探勘的首先步驟，為選定要探勘及分析之資料，並選擇資料來源，但是不一定為特定的資料庫中所有的資料，可能從一個或多個資料庫中結合成所以執行探勘的資料庫。因此，使用者可以根據資料庫所需要的條件自己篩選。

#### 二、資料前置處理

為確認資料正確與完整性並順利進行資料探勘，須先將已收集的資料進行整理、清除重覆或無效的資料記錄，確認所需資料的完整。並將資料轉換為探勘軟體所能讀取之存取格式，以利進行資料庫之構建。

#### 三、群集分析

依照研究目的，本研究選用群集分析的 K-Means 演算法，利用群集分析的 K-Means 演算法從龐大的筆事資料庫中發掘存在的多種特徵及資訊並進行分類。

#### 四、群集結果

經過群集分析後，可以得到最佳之分類群組，此分類群組作為判別模式完成前的重要基礎。

## 五、建構判別模式

以群集分析結果並驗證各預測變數之顯著性，作為輸入判別模式的分群預測變數，建立全部年度及各年度之費雪(典型)線性別函數模式，了解哪些預測變數在全部年度及各年度的重要性，並比較訓練樣本和測試樣本的分類正確率。

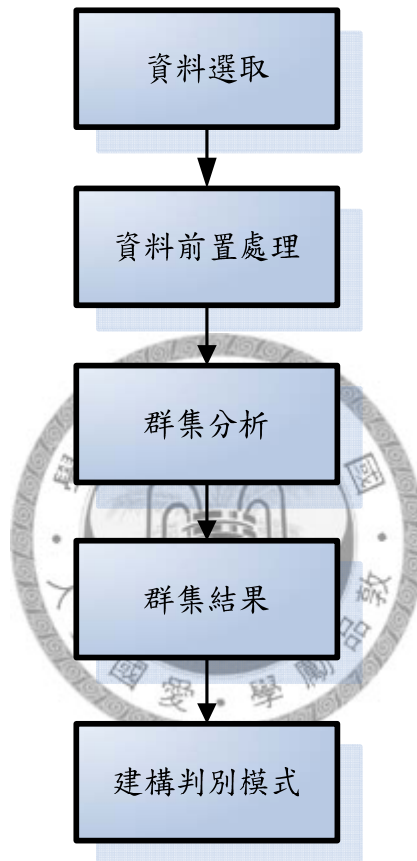


圖 4-1 分析流程

### 4.2 肇事資料蒐集

本研究採用內政部警政署自民國 92 年至民國 97 年大客車道路交通事故資料，並依據資料檔案格式界定及釐清各項資料格式與內容，資料之敘述如下：

- 分析資料來自內政部警政署之構建的大客車交通事故資料庫。
- 資料期間為民國 92 年至民國 97 年共 6 年的全國大客車肇事資料。
- 道路交通事故資料期間，經過處理與清除無效資料後，92 年 2615 件、93 年

2740 件、94 年 2779 件、95 年 2737 件、96 年 2365 件、97 年 2278 件，共計肇事件數為 15514 件。

- 根據交通道路事故資料表內容之變數定義，本研究大客車資料共有 16 個欄位及 22 個屬性資料。

### 4.3 肇事資料前置處理

前置處理為資料探勘前的一個重要處理過程，不同的探勘領域所需的前置處理技術也不盡相同。本研究之資料前置處理部分主要有下列幾個步驟：

#### 1. 資料選取

於已蒐集的資料中，將各項資料所包含之項目，依據本研究篩選所需及刪除並整合所需項目。

#### 2. 資料雜訊清除

將各項資料中含有亂碼或是有部分遺漏之資料予以清除，並且查核字元之正確性，以及刪除分析所不需要的資料。

#### 3. 資料屬性欄位合併

交通事故資料表裡，詳細記錄交通道路肇事案件，使得資料檔案格式之欄位屬性過於複雜，為符合研究目的，依照交通肇事特性將各欄位之細節屬性加以整合。

#### 4. 資料屬性及格式轉換

將原始肇事資料屬性欄位在合理化條件下進行資料屬性定義轉換；資料格式為 Access 檔案，但以後後續資料庫之構建，將資料格式統一轉換為 Excel 格式。

#### 5. 資料庫構建

將原始肇事資料經合併、轉換、篩選與整理完成後之資料，再建本研究所需要之資料庫，才能進行後續分析。

#### 4.4 肇事資料轉換

由於道路交通事故調查報告表某些屬性過於詳細且對於本研究預期分析結果與目的並無實質意義，因此減低資料的複雜度，本研究將屬性資料依肇事特性分類合理的合併。

##### 1. 發生時間

將發生時間分成 0-3(不含)時、3-6(不含)時、6-9(不含)時、9-12(不含)時、12-15(不含)時、15-18(不含)時、18-21(不含)時、21-24(不含)時共八個時段。

##### 2. 速限

將速限欄位分成小於 40 公里、40-50(不含)公里、50-60(不含)公里、60-70(不含)公里、大於 70 公里共五大類。

##### 3. 道路型態

將道路型態分成平交道、交岔道、單路部份、圓環廣場共四大類。

##### 4. 分向設施

將分向設施分成中央分隔島、雙向禁止超車線、單向禁止超車線、行車分向線、無分向設施共五大類。

##### 5. 發生地區

依地理區域將發生地區分成北部、中部、南部、東部、離島共五大類。北部包含縣市為基隆市、台北縣市、桃園縣、新竹縣市、苗栗縣；中部包含縣市為台中縣市、彰化縣、雲林縣、南投縣；南部包含縣市為嘉義縣市、台南縣市、高雄縣市、屏東縣；東部包含縣市為宜蘭縣、花蓮縣、台東縣；離島包含縣市為澎湖縣、金門縣、連江縣

##### 6. 天候

將天候分成暴雨、強風、風沙、霧或煙、雨、陰、晴共七大類。

7. 視距

將視距分成不良、良好共二大類。

8. 事故位置

將事故位置分成交岔路口、路段、交流道、其他事故位置共四大類。

9. 事故類型及型態

將事故類型及型態分成人與汽(機)車、車與車、汽(機)本身、平交道事故共四大類。

10. 肇事原因

將肇事原因分成未依規定讓車、變換車道或方向不當、未保持行車安全距離、未保持行車安全間隔、未注意車前狀態、酒醉駕駛失控、違反號制管制或指揮、違反特定標誌(線)標制、其他駕駛人因素、非駕駛人因素共十大類。





## 4.5 肇事變數

為了符合研究方法所能分析的資料，將原始資料屬性重新進行整理與定義並轉為類別變數。經過資料前置處理與轉換的步驟，肇事變數如表 4-1 所示。

表 4-1 肇事變數

屬性	屬性代號	變數	變數代號	屬性	屬性代號	變數	變數代號
發生月份	A1	一月	X <sub>a1</sub>	速限	D1	小於40公里	X <sub>d1</sub>
	A2	二月	X <sub>a2</sub>		D2	40-50公里	X <sub>d2</sub>
	A3	三月	X <sub>a3</sub>		D3	50-60公里	X <sub>d3</sub>
	A4	四月	X <sub>a4</sub>		D4	60-70公里	X <sub>d4</sub>
	A5	五月	X <sub>a5</sub>		D5	大於70公里	X <sub>d5</sub>
	A6	六月	X <sub>a6</sub>	道路型態	E1	平交道	X <sub>e1</sub>
	A7	七月	X <sub>a7</sub>		E2	交岔道	X <sub>e2</sub>
	A8	八月	X <sub>a8</sub>		E3	單路部份	X <sub>e3</sub>
	A9	九月	X <sub>a9</sub>		E4	圓環廣場	X <sub>e4</sub>
	A10	十月	X <sub>a10</sub>	路面缺陷	F1	路面鬆軟	X <sub>f1</sub>
	A11	十一月	X <sub>a11</sub>		F2	高低不平	X <sub>f2</sub>
	A12	十二月	X <sub>a12</sub>		F3	有坑洞	X <sub>f3</sub>
發生時間	B1	0-3時	X <sub>b1</sub>		F4	無缺陷	X <sub>f4</sub>
	B2	3-6時	X <sub>b2</sub>	分向設施	G1	中央分隔島	X <sub>g1</sub>
	B3	6-9時	X <sub>b3</sub>		G2	雙向禁止超車線	X <sub>g2</sub>
	B4	9-12時	X <sub>b4</sub>		G3	單向禁止超車線	X <sub>g3</sub>
	B5	12-15時	X <sub>b5</sub>		G4	行車分向線	X <sub>g4</sub>
	B6	15-18時	X <sub>b6</sub>		G5	無	X <sub>g5</sub>
	B7	18-21時	X <sub>b7</sub>	快車道間	H1	禁止變換車道線 (附標記)	X <sub>h1</sub>
	B8	21-24時	X <sub>b8</sub>		H2	禁止變換車道線 (無標記)	X <sub>h2</sub>
道路類別	C1	國道	X <sub>c1</sub>		H3	車道線(附標記)	X <sub>h3</sub>
	C2	省道	X <sub>c2</sub>		H4	車道線(無標記)	X <sub>h4</sub>
	C3	縣道	X <sub>c3</sub>		H5	未繪設車道線	X <sub>h5</sub>
	C4	鄉道	X <sub>c4</sub>	快慢 車道間	I1	寬式快慢車道分隔島 (50公分以上)	X <sub>i1</sub>
	C5	市區道路	X <sub>c5</sub>		I2	窄式快慢車道分隔島 (附柵欄)	X <sub>i2</sub>
	C6	村里道路	X <sub>c6</sub>		I3	窄式快慢車道分隔島 (無柵欄)	X <sub>i3</sub>
	C7	專用道路	X <sub>c7</sub>		I4	快慢車道分隔線	X <sub>i4</sub>
	C8	其他	X <sub>c8</sub>		I5	未繪設快慢車道分隔線	X <sub>i5</sub>

表 4.1 肇事變數(續)

屬性	屬性代號	變數	變數代號	屬性	屬性代號	變數	變數代號
路面邊緣	J1	有	X <sub>j1</sub>	障礙物	P4	其他障礙物	X <sub>p4</sub>
	J2	無	X <sub>j2</sub>		P5	無障礙物	X <sub>p5</sub>
發生地區	K1	北部	X <sub>k1</sub>	視距	Q1	不良	X <sub>q1</sub>
	K2	中部	X <sub>k2</sub>		Q2	良	X <sub>q2</sub>
	K3	南部	X <sub>k3</sub>	號誌種類	R1	行車管制號誌	X <sub>r1</sub>
	K4	東部	X <sub>k4</sub>		R2	行車管制號誌 (附設行人專用號誌)	X <sub>r2</sub>
	K5	離島	X <sub>k5</sub>		R3	閃光號誌	X <sub>r3</sub>
天候	L1	暴雨	X <sub>l1</sub>		R4	無號誌	X <sub>r4</sub>
	L2	強風	X <sub>l2</sub>	號誌動作	S1	正常	X <sub>s1</sub>
	L3	風沙	X <sub>l3</sub>		S2	不正常	X <sub>s2</sub>
	L4	霧或煙	X <sub>l4</sub>		S3	無動作	X <sub>s3</sub>
	L5	雨	X <sub>l5</sub>		S4	無號誌	X <sub>s4</sub>
	L6	陰	X <sub>l6</sub>	事故位置	T1	岔路口	X <sub>t1</sub>
	L7	晴	X <sub>l7</sub>		T2	路段	X <sub>t2</sub>
光線	M1	日間自然光線	X <sub>m1</sub>		T3	交流道	X <sub>t3</sub>
	M2	晨或暮光	X <sub>m2</sub>		T4	其他	X <sub>t4</sub>
	M3	夜間有照明	X <sub>m3</sub>	事故類型 及型態	U1	人與汽(機)車	X <sub>u1</sub>
	M4	夜間無照明	X <sub>m4</sub>		U2	車與車	X <sub>u2</sub>
路面鋪裝	N1	柏油	X <sub>n1</sub>		U3	汽(機)車本身	X <sub>u3</sub>
	N2	水泥	X <sub>n2</sub>		U4	平交道事故	X <sub>u4</sub>
	N3	碎石	X <sub>n3</sub>	肇事原因	V1	未依規定讓車	X <sub>v1</sub>
	N4	其他鋪裝	X <sub>n4</sub>		V2	變換車道或方向不當	X <sub>v2</sub>
	N5	無鋪裝	X <sub>n5</sub>		V3	未保持行車安全距離	X <sub>v3</sub>
路面狀況	O1	冰雪	X <sub>o1</sub>		V4	未保持行車安全間隔	X <sub>v4</sub>
	O2	油滑	X <sub>o2</sub>		V5	未注意車前狀態	X <sub>v5</sub>
	O3	泥濘	X <sub>o3</sub>		V6	酒醉(後)駕駛失控	X <sub>v6</sub>
	O4	濕潤	X <sub>o4</sub>		V7	違反號誌管制或指揮	X <sub>v7</sub>
	O5	乾燥	X <sub>o5</sub>		V8	違反特定標誌(線)標制	X <sub>v8</sub>
障礙物	P1	道路工程中	X <sub>p1</sub>		V9	其他駕駛人因素	X <sub>v9</sub>
	P2	有堆積物	X <sub>p2</sub>		V10	非駕駛人因素	X <sub>v10</sub>
	P3	路上有停車	X <sub>p3</sub>				

資料來源：本研究整理



## 第五章 資料分析

經由前述章節將肇事資料做前置處理及設定肇事變數，本章分為群集方法與分析流程、群集數目選擇、群集結果、群集驗證、判別模式架構與分析流程、函數構建、分析結果、模式綜合分析共八節，由逐步判別選出重要肇事變數，並探討影響大客車之肇事變數。

### 5.1 群集方法與分析流程

依據第三章研究方法所述，群集分析方法有階層式群集分析法及非階層式群集分析法(K-Means 法)。由於實際應用階層式群集分析法非常耗費資源與時間，所以在實務上常應用非階層式群集分析法。林岩震【30】提到觀察值的個數超過 200 筆以上，採用 K-Means 群集分析法較為適宜。由於本研究總肇事資料高達 15514 筆，經過實際處理分析，階層式群集分析方法確實無法處理，所以採用 K-Means 群集分析法，其群集分析流程如圖 5-1 所示。本研究在進行分群前，先將分群變數標準化，使得分群時的所有變數均可以發揮相同的效力，不會有某一個變數因其數值大小而主導分群結果，並以歐幾里德距離為相似性衡量的基準來計算距離。

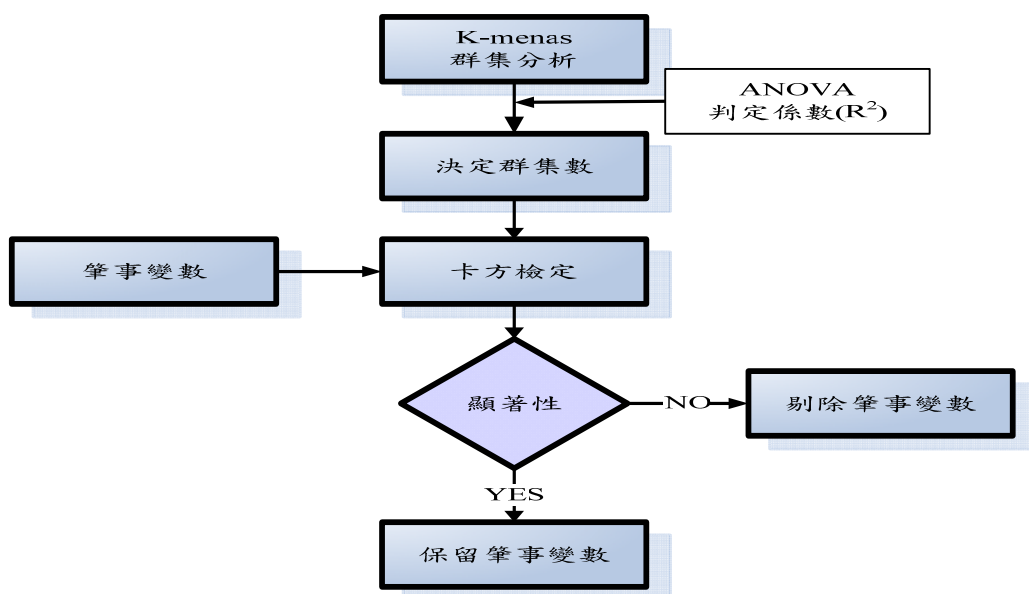


圖 5-1 群集方法分析流程

## 5.2 群集數目選擇

根據 K-Means 群集分析法的演算規則，要事先訂定群集個數。本研究根據肇事資料性質且為了分析應用時能夠方便清楚，選訂 2-10 個群集數目，並依照群集數評量指標的 ANOVA 之判定係數( $R^2$ )作為選定群數的數目，當判定係數( $R^2$ )愈大表示群體間的相異性愈高，也顯示分群效果愈好。本研究將大客車各年度及全部年度的肇事資料進行分群，其群集數目選擇結果如下：

### 一、92 年肇事資料

由圖 5-2 所示，當群集個數在 5 時，其判定係數( $R^2$ )達 88.04%，此時分群效果最佳，因此 92 年肇事資料分成 5 群。

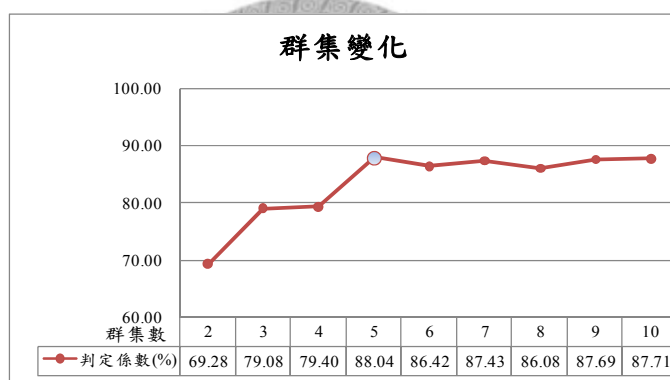


圖 5-2 92 年肇事資料群集變化

### 二、93 年肇事資料

由圖 5-3 所示，當群集個數在 5 時，其判定係數( $R^2$ )達 87.48%，此時分群效果最佳，因此 93 年肇事資料分成 5 群。

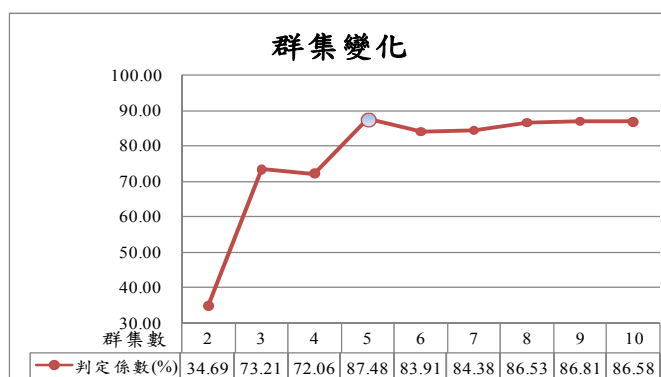


圖 5-3 93 年肇事資料群集變化

### 三、94 年肇事資料

由圖 5-4 所示，當群集個數在 6 時，其判定係數( $R^2$ )達 86.39%，此時分群效果最佳，因此 94 年肇事資料分成 6 群。

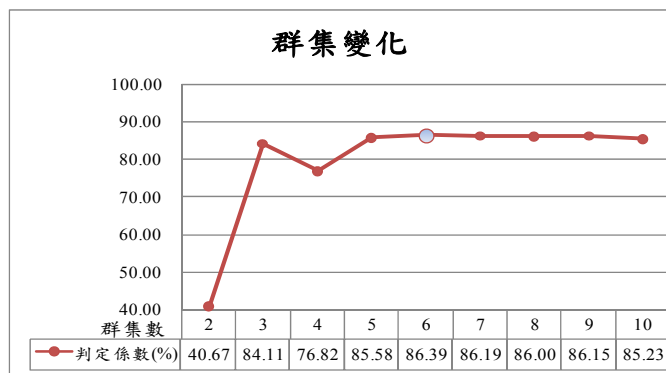


圖 5-4 94 年肇事資料群集變化

### 四、95 年肇事資料

由圖 5-5 所示，當群集個數在 5 時，其判定係數( $R^2$ )達 86.86%，此時分群效果最佳，因此 92 年肇事資料分成 5 群。

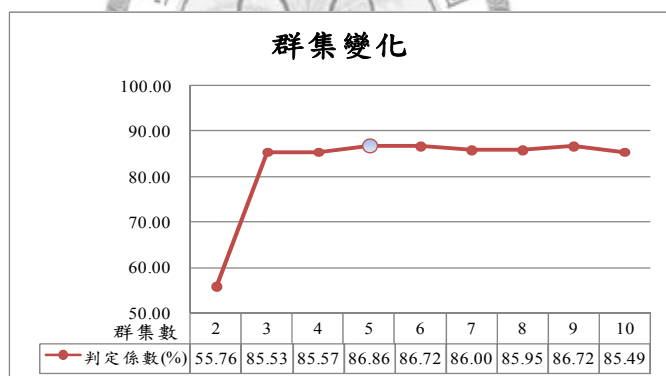


圖 5-5 95 年肇事資料群集變化

### 五、96 年肇事資料

由圖 5-6 所示，當群集個數在 7 時，其判定係數( $R^2$ )達 85.40%，此時分群效果最佳，因此 96 年肇事資料分成 7 群。

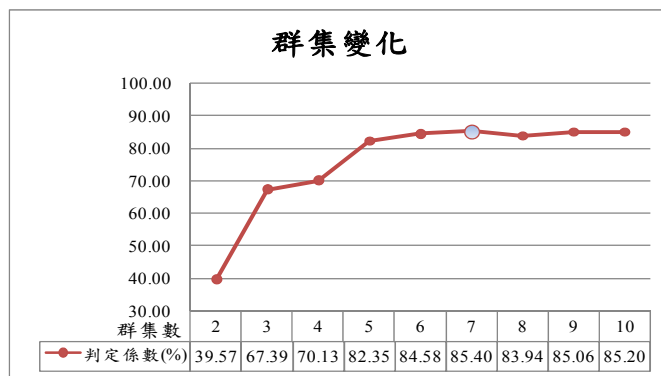


圖 5-6 96 年肇事資料群集變化

#### 六、97 年肇事資料

由圖 5-7 所示，當群集個數在 7 時，其判定係數( $R^2$ )達 88.45%，此時分群效果最佳，因此 97 年肇事資料分成 7 群。

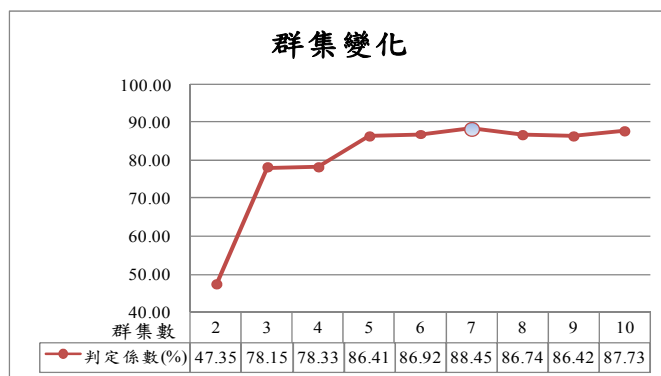


圖 5-7 97 年肇事資料群集變化

#### 七、92-97 年肇事資料

由圖 5-8 所示，當群集個數在 6 時，其判定係數( $R^2$ )達 92.68%，此時分群效果最佳，因此總肇事資料分成 6 群。

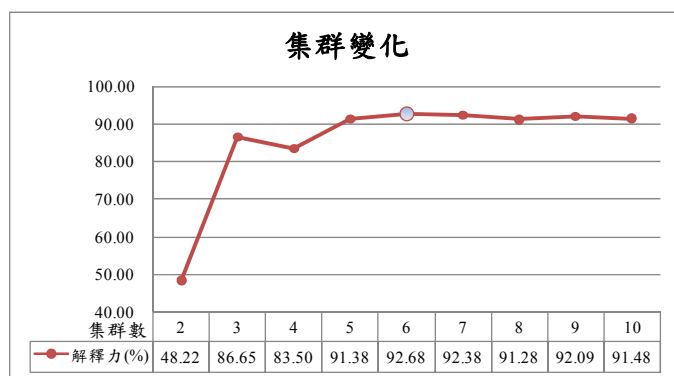


圖 5-8 92-97 年肇事資料群集變化

### 5.3 群集結果

利用非階層群集分析法的 K-Means 法，其群集過程的步驟如下：

1. 將觀察值分割成 K 個原始群落。
2. 計算每群觀察值到各群落中心的距離，採用歐基里德距離，然後將觀察值分派到最近的群落，並重新計算得到及喪失觀察值的群落之重心。
3. 重覆步驟 2，直到沒有觀察值可以再調整為止。
4. 最後群集在各筆事變數的中心點位置。

依照上節將最佳群集數的判定結果，各年度及 92-97 年的筆事資料分群數目為 5、5、6、5、7、7、6 群之原始群落，即  $K=5\sim 7$  群。另外，由於各群集中的筆事變數條件太多且都有所不同，因此命名上困難，所以群集就為 1~7 命名。其群集結果如下：





## 一、92 年肇事資料

K-Means 法將初始 5 群的中心定出，然後計算各觀察值到各群集的重心，經過 10 次中心點的變更後，沒有觀察值需要再調整，表 5-1 為群集分析的最後群集的結果，為 92 年肇事資料的群集中心點，即所有肇事資料已根據相似性被分派至各群集中，也是各群集中最相近的變數集合。92 年的肇事資料總共有 2615 筆，群集 1、2、3、4、5 分別有 712 筆、440 筆、526 筆、434 筆、503 筆肇事資料。

表 5-1 92 年最後群集中心點

92年	群集				
	1	2	3	4	5
發生月份	三月	九月	三月	九月	十月
發生時間	0-3時	15-18時	15-18時	3-6時	18-21時
道路類別	縣道	國道	鄉道	鄉道	鄉道
車速	50-60公里	大於70公里	40-50公里	40-50公里	40-50公里
道路型態	單路部份	單路部份	單路部份	交岔道	單路部份
路面缺陷	無缺陷	無缺陷	無缺陷	無缺陷	無缺陷
分向設施	雙向禁止超車線	中央分隔島	單向禁止超車線	單向禁止超車線	單向禁止超車線
快車道	車道線(無標記)	車道線(附標記)	車道線(無標記)	車道線(無標記)	車道線(無標記)
快慢車道	未繪設快慢車道分隔線	未繪設快慢車道分隔線	未繪設快慢車道分隔線	快慢車道分隔線	快慢車道分隔線
路面標誌	有	有	無	無	無
發生地點	北部	南部	北部	北部	南部
天氣	晴	晴	晴	晴	晴
光線	日間自然光線	夜間無照明	日間自然光線	日間自然光線	夜間有照明
路面鋪裝	柏油	柏油	柏油	無鋪裝	柏油
路面狀態	乾燥	乾燥	乾燥	乾燥	乾燥
障礙物	無障礙物	無障礙物	道路工程中	無障礙物	無障礙物
視距	良	良	良	良	良
號誌種類	閃光號誌	無號誌	閃光號誌	閃光號誌	閃光號誌
號誌動作	無動作	無號誌	無動作	無動作	無動作
事故位置	路段	路段	路段	交岔路口	路段
事故類型及型態	車與車	車與車	車與車	車與車	車與車
肇事原因	未保持行車安全距離	未注意車前狀態	其他駕駛人因素	其他駕駛人因素	未保持行車安全距離

## 二、93 年肇事資料

K-Means 法將初始 5 群的中心定出，然後計算各觀察值到各群集的重心，經過 27 次中心點的變更後，沒有觀察值需要再調整，表 5-2 為群集分析的最後群集的結果，為 93 年肇事資料的群集中心點，即所有肇事資料已根據相似性被分派至各群集中，也是各群集中最相近的變數集合。93 年的肇事資料總共有 2740 筆，群集 1、2、3、4、5 分別有 659 筆、226 筆、416 筆、578 筆、861 筆肇事資料。

表 5-2 93 年最後群集中心點

93年	群集				
	1	2	3	4	5
發生月份	十月	六月	五月	九月	三月
發生時間	12-15時	3-6時	21-24時	0-3時	21-24時
道路類別	鄉道	國道	國道	市區道路	市區道路
速限	50-60公里	大於70公里	大於70公里	40-50公里	40-50公里
道路型態	單路部份	單路部份	單路部份	交岔道	交岔道
路面缺陷	無缺陷	無缺陷	無缺陷	無缺陷	無缺陷
分向設施	單向禁止超車線	中央分隔島	中央分隔島	單向禁止超車線	單向禁止超車線
快車道同	車道線(無標記)	車道線(附標記)	車道線(附標記)	車道線(無標記)	車道線(無標記)
快慢車道同	未繪設快慢車道分隔線	未繪設快慢車道分隔線	未繪設快慢車道分隔線	快慢車道分隔線	快慢車道分隔線
路面邊線	有	有	有	無	無
發生地區	中部	南部	南部	南部	北部
天候	晴	陰	晴	晴	晴
光線	日間自然光線	夜間無照明	夜間無照明	日間自然光線	夜間有照明
路面鋪裝	柏油	柏油	柏油	無鋪裝	柏油
路面狀態	乾燥	乾燥	乾燥	乾燥	乾燥
障礙物	無障礙物	無障礙物	無障礙物	無障礙物	無障礙物
視距	良	良	良	良	良
號誌種類	閃光號誌	無號誌	無號誌	閃光號誌	閃光號誌
號誌動作	無動作	無號誌	無號誌	無動作	無動作
事故位置	路段	路段	路段	交岔路口	交岔路口
事故類型及型態	車與車	車與車	車與車	車與車	車與車
肇事原因	未保持行車安全距離	非駕駛人因素	未保持行車安全距離	其他駕駛人因素	酒醉(後)駕駛失控

### 三、94 年肇事資料

K-Means 法將初始 6 群的中心定出，然後計算各觀察值到各群集的重心，經過 18 次中心點的變更後，沒有觀察值需要再調整，表 5-3 為群集分析的最後群集的結果，為 94 年肇事資料的群集中心點，即所有肇事資料已根據相似性被分派至各群集中，也是各群集中最相近的變數集合。94 年的肇事資料總共有 2779 筆，群集 1、2、3、4、5、6 分別有 538 筆、219 筆、495 筆、563 筆、518 筆、446 筆肇事資料。

表 5-3 94 年最後群集中心點

94年	群集					
	1	2	3	4	5	6
發生月份	三月	七月	十月	十月	六月	三月
發生時間	6-9時	0-3時	15-18時	18-21時	21-24時	21-24時
道路類別	鄉道	國道	市區道路	鄉道	國道	市區道路
車速	40-50公里	大於70公里	40-50公里	40-50公里	大於70公里	40-50公里
道路型態	單路部份	單路部份	岔路	單路部份	單路部份	岔路
路面狀況	無缺陷	無缺陷	無缺陷	無缺陷	無缺陷	無缺陷
分向設施	單向禁止超車線	中央分隔島	單向禁止超車線	單向禁止超車線	中央分隔島	單向禁止超車線
快車道	車道線(無標記)	車道線(附標記)	車道線(無標記)	車道線(無標記)	車道線(附標記)	車道線(無標記)
快慢車道	快慢車道 分隔線	未增設快慢 車道分隔線	快慢車道 分隔線	快慢車道 分隔線	未增設快慢 車道分隔線	快慢車道 分隔線
路面標誌	無	有	無	無	有	有
發生地區	離島	中部	中部	北部	北部	北部
天候	晴	陰	晴	晴	晴	晴
光線	日間自然光線	日間自然光線	日間自然光線	夜間無照明	夜間有照明	夜間無照明
路面鋪裝	柏油	柏油	柏油	無鋪裝	柏油	柏油
路面乾濕	乾燥	乾燥	乾燥	乾燥	乾燥	乾燥
障礙物	無障礙物	無障礙物	無障礙物	無障礙物	無障礙物	無障礙物
視距	良	良	良	良	良	良
標誌種類	閃光號誌	無號誌	閃光號誌	閃光號誌	無號誌	閃光號誌
標誌動作	無動作	無號誌	無動作	無動作	無號誌	無動作
事故位置	路段	路段	岔路口	路段	路段	岔路口
事故類型及型態	車與車	車與車	車與車	車與車	車與車	車與車
肇事原因	未保持 行車安全距離	非駕駛人因素	其他駕駛人因素	未保持 行車安全間隔	未保持 行車安全距離	其他駕駛人因素

#### 四、95 年肇事資料

K-Means 法將初始 5 群的中心定出，然後計算各觀察值到各群集的重心，經過 18 次中心點的變更後，沒有觀察值需要再調整，表 5-4 為群集分析的最後群集的結果，為 95 年肇事資料的群集中心點，即所有肇事資料已根據相似性被分派至各群集中，也是各群集中最相近的變數集合。95 年的肇事資料總共有 2737 筆，群集 1、2、3、4、5 分別有 650 筆、354 筆、488 筆、601 筆、644 筆肇事資料。

表 5-4 95 年最後群集中心點

95年	群集				
	1	2	3	4	5
發生月份	三月	九月	十月	四月	九月
發生時間	0-3時	0-3時	12-15時	0-3時	18-21時
道路類別	縣道	國道	鄉道	縣道	鄉道
速限	50-60公里	大於70公里	40-50公里	50-60公里	40-50公里
道路型態	單路部份	單路部份	岔岔道	單路部份	單路部份
路面缺陷	無缺陷	無缺陷	無缺陷	無缺陷	無缺陷
分向設施	雙向禁止超車線	中央分隔島	單向禁止超車線	單向禁止超車線	單向禁止超車線
快車道同	車道線(無標記)	車道線(附標記)	車道線(無標記)	車道線(無標記)	車道線(無標記)
快慢車道同	未繪設快慢車道分隔線	未繪設快慢車道分隔線	未繪設快慢車道分隔線	未繪設快慢車道分隔線	未繪設快慢車道分隔線
路面邊緣	有	有	無	無	無
發生地區	南部	北部	北部	北部	北部
天候	陰	晴	晴	雨	晴
光線	日間自然光線	夜間無照明	日間自然光線	日間自然光線	夜間有照明
路面鋪裝	柏油	柏油	柏油	無鋪裝	柏油
路面狀態	乾燥	乾燥	乾燥	乾燥	乾燥
障礙物	無障礙物	無障礙物	無障礙物	無障礙物	無障礙物
視距	良	良	良	良	良
標誌種類	閃光號誌	無號誌	閃光號誌	閃光號誌	閃光號誌
標誌動作	無動作	無號誌	無動作	無動作	無動作
事故位置	路段	路段	岔岔路口	路段	路段
事故類型及型態	車與車	車與車	車與車	車與車	車與車
肇事原因	未保持行車安全距離	未注意車前狀態	其他駕駛人因素	其他駕駛人因素	未保持行車安全距離

## 五、96 年肇事資料

K-Means 法將初始 7 群的中心定出，然後計算各觀察值到各群集的重心，經過 23 次中心點的變更後，沒有觀察值需要再調整，表 5-5 為群集分析的最後群集的結果，為 96 年肇事資料的群集中心點，即所有肇事資料已根據相似性被分派至各群集中，也是各群集中最相近的變數集合。96 年的肇事資料總共有 2365 筆，群集 1、2、3、4、5、6、7 分別有 251 筆、250 筆、536 筆、365 筆、300 筆、517 筆、146 筆肇事資料。

表 5-5 96 年最後群集中心點

96年	群集						
	1	2	3	4	5	6	7
發生月份	四月	四月	三月	十月	七月	十月	八月
發生時間	6-9時	18-21時	0-3時	15-18時	21-24時	6-9時	3-6時
道路類別	市區道路	鄉道	鄉道	市區道路	國道	市區道路	國道
車速	40-50公里	40-50公里	40-50公里	40-50公里	大於70公里	40-50公里	大於70公里
道路型態	岔道	單路部份	單路部份	岔道	單路部份	單路部份	單路部份
路面狀況	無缺陷	無缺陷	無缺陷	無缺陷	無缺陷	無缺陷	無缺陷
方向標誌	單向禁止超車線	雙向禁止超車線	單向禁止超車線	單向禁止超車線	中央分隔島	單向禁止超車線	中央分隔島
快車道	車道線(無標記)	車道線(無標記)	車道線(無標記)	車道線(無標記)	車道線(附標記)	車道線(無標記)	車道線(附標記)
快慢車道	未繪設快慢車道分隔線	未繪設快慢車道分隔線	未繪設快慢車道分隔線	未繪設快慢車道分隔線	未繪設快慢車道分隔線	未繪設快慢車道分隔線	未繪設快慢車道分隔線
路面標誌	無	有	無	無	有	無	有
發生地區	北部	中部	北部	南部	北部	南部	南部
天候	陰	晴	晴	晴	晴	晴	晴
光線	日間自然光線	夜間無照明	夜間無照明	夜間有照明	夜間有照明	日間自然光線	夜間無照明
路面濕度	柏油	柏油	柏油	柏油	柏油	柏油	柏油
路面坡度	乾燥	乾燥	乾燥	乾燥	乾燥	乾燥	乾燥
障礙物	無障礙物	無障礙物	無障礙物	無障礙物	無障礙物	無障礙物	無障礙物
視距	良	良	良	良	良	良	良
標誌種類	行車管制號誌	無號誌	閃光號誌	閃光號誌	無號誌	閃光號誌	無號誌
標誌動作	正常	無號誌	無動作	不正常	無號誌	無動作	無號誌
事故位置	岔路口	路段	路段	岔路口	路段	路段	路段
事故類型及型態	車與車	車與車	車與車	車與車	車與車	車與車	車與車
肇事原因	違反特定標誌(線)標制	其他駕駛人因素	未保持行車安全距離	其他駕駛人因素	未保持行車安全距離	未保持行車安全距離	非駕駛人因素



## 六、97 年肇事資料

K-Means 法將初始 7 群的中心定出，然後計算各觀察值到各群集的重心，經過 39 次中心點的變更後，沒有觀察值需要再調整，表 5-6 為群集分析的最後群集的結果，為 97 年肇事資料的群集中心點，即所有肇事資料已根據相似性被分派至各群集中，也是各群集中最相近的變數集合。97 年的肇事資料總共有 2278 筆，群集 1、2、3、4、5、6、7 分別有 209 筆、386 筆、312 筆、447 筆、302 筆、302 筆、320 筆肇事資料。

表 5-6 97 年最後群集中心點

97年	群集						
	1	2	3	4	5	6	7
發生月份	九月	十月	九月	三月	八月	三月	四月
發生時間	18-21時	15-18時	9-12時	15-18時	0-3時	18-21時	21-24時
道路類別	國道	市區道路	鄉道	鄉道	市區道路	縣道	市區道路
車速	大於70公里	40-50公里	50-60公里	40-50公里	40-50公里	60-70公里	40-50公里
道路型態	單路部份	單路部份	單路部份	單路部份	交岔道	單路部份	交岔道
路面狀況	無缺陷	無缺陷	無缺陷	無缺陷	無缺陷	無缺陷	無缺陷
分向標誌	中央分隔島	單向禁止超車線	雙向禁止超車線	雙向禁止超車線	行車分向線	雙向禁止超車線	行車分向線
快車道	車道線(附標記)	車道線(無標記)	車道線(無標記)	車道線(無標記)	車道線(無標記)	車道線(附標記)	車道線(無標記)
快車道	未繪設快慢車道分隔線	未繪設快慢車道分隔線	未繪設快慢車道分隔線	未繪設快慢車道分隔線	未繪設快慢車道分隔線	未繪設快慢車道分隔線	未繪設快慢車道分隔線
路面標誌	有	無	有	無	無	有	無
發生地區	中部	北部	南部	中部	北部	北部	南部
天氣	晴	晴	陰	晴	陰	雨	晴
光線	夜間無照明	日間自然光線	日間自然光線	日間自然光線	日間自然光線	夜間有照明	夜間無照明
路面濕度	柏油	柏油	柏油	柏油	柏油	柏油	柏油
路面溫度	乾燥	乾燥	乾燥	乾燥	乾燥	乾燥	乾燥
障礙物	無障礙物	無障礙物	無障礙物	無障礙物	無障礙物	無障礙物	無障礙物
視距	良	良	良	良	良	良	良
標誌種類	無號誌	閃光號誌	無號誌	閃光號誌	行車管制號誌	無號誌	行車管制號誌
標誌動作	無號誌	無動作	無號誌	無動作	正常	無號誌	不正常
事故位置	路段	路段	路段	路段	交岔路口	路段	交岔路口
事故類型及型態	車與車	車與車	車與車	車與車	車與車	車與車	車與車
肇事原因	未保持行車安全距離	未保持行車安全距離	其他駕駛人因素	其他駕駛人因素	其他駕駛人因素	未保持行車安全間隔	未保持行車安全距離

## 七、92-97 年肇事資料

K-Means 法將初始 6 群的中心定出，然後計算各觀察值到各群集的重心，經過 13 次中心點的變更後，沒有觀察值需要再調整，表 5-7 為群集分析的最後群集的結果，為總肇事資料的群集中心點，即所有肇事資料已根據相似性被分派至各群集中，也是各群集中最相近的變數集合。92-97 年度的肇事資料共有 15514 筆，群集 1、2、3、4、5、6 分別有 2275 筆、3205 筆、2659 筆、3452 筆、1161 筆、2762 筆肇事資料。

表 5-7 92-97 年度最後群集中心點

變數	群集					
	1	2	3	4	5	6
發生月份	五月	四月	三月	十月	七月	十月
發生時間	21-24時	0-3時	6-9時	21-24時	0-3時	18-21時
道路類別	國道	鄉道	市區道路	鄉道	國道	市區道路
車速	大於70公里	40-50公里	40-50公里	50-60公里	大於70公里	40-50公里
道路型態	單路部份	單路部份	岔路	單路部份	單路部份	岔路
路面狀況	無缺陷	無缺陷	無缺陷	無缺陷	無缺陷	無缺陷
分向設施	中央分隔島	單向禁止超車線	單向禁止超車線	單向禁止超車線	中央分隔島	單向禁止超車線
快車道	車道線(附標記)	車道線(無標記)	車道線(無標記)	車道線(無標記)	車道線(附標記)	車道線(無標記)
快慢車道	未繪設快慢車道分隔線	未繪設快慢車道分隔線	未繪設快慢車道分隔線	未繪設快慢車道分隔線	未繪設快慢車道分隔線	未繪設快慢車道分隔線
發生地區	北部	中部	中部	北部	北部	中部
路面濕度	有	無	無	有	有	無
天候	晴	晴	雨	陰	晴	晴
光線	晨或暮光	夜間有照明	日間自然光線	夜間有照明	夜間無照明	日間自然光線
路面積雪	柏油	柏油	柏油	柏油	柏油	柏油
路面積冰	乾燥	乾燥	乾燥	乾燥	乾燥	乾燥
障礙物	無障礙物	無障礙物	無障礙物	無障礙物	無障礙物	無障礙物
視距	良	良	良	良	良	良
標誌種類	無標誌	閃光標誌	閃光標誌	閃光標誌	無標誌	閃光標誌
標誌動作	無標誌	無動作	無動作	無動作	無標誌	無動作
事故位置	路段	路段	岔路口	路段	路段	岔路口
事故類型及型態	車與車	車與車	人與汽(機)車	車與車	車與車	車與車
肇事原因	未保持行車安全距離	未保持行車安全距離	其他駕駛人因素	未保持行車安全距離	非駕駛人因素	其他駕駛人因素

## 5.4 群集驗證

對於各年度之肇事變數的最佳分群結果，要保證肇事變數在不同群集的集合是否存在顯著的差異，因此以卡方檢定來驗證此群集數和肇事變數是否相關。其

統計假設為

$$\begin{cases} H_0: \text{行變數與列變數有關(群集數和肇事變數有關)} \\ H_1: \text{行變數與列變數無關(群集數和肇事變數無關)} \end{cases}$$

若肇事變數和群集數有關就保留，無關之變數就予以剔除，所以檢定結果為顯著的肇事變數，則選入判別模式進行判別分析。其群集驗證結果如下：

### 一、92 年肇事資料

由表 5-8 所示，除了路面缺陷、路面鋪裝、障礙物外，其餘 19 個肇事變數皆為顯著，因此選入判別模式進行分析。

表 5-8 92 年卡方檢定結果

92 年	卡方值	自由度	P 值	顯著性
發生月份	2431.063	44	0.00	**
發生時間	61.070	28	0.00	**
道路類別	1199.109	28	0.00	**
速限	1503.644	16	0.00	**
道路型態	381.419	12	0.00	**
路面缺陷	7.468	12	0.83	—
分向設施	558.971	16	0.00	**
快車道間	888.653	16	0.00	**
快慢車道間	257.375	16	0.00	**
路面邊線	489.059	4	0.00	**
發生地區	61.285	16	0.00	**
天候	32.608	16	0.01	*
光線	129.241	12	0.00	**
路面鋪裝	13.268	12	0.35	—
路面狀態	43.061	16	0.00	**
障礙物	21.369	16	0.17	—
視距	25.928	4	0.00	**
號誌種類	352.019	12	0.00	**
號誌動作	335.941	12	0.00	**
事故位置	458.329	12	0.00	**
事故類型及型態	190.889	12	0.00	**
肇因原因	2732.866	36	0.00	**

註：「\*」為  $0.01 \leq P \text{ 值} \leq 0.05$ ，「\*\*」為  $P \text{ 值} < 0.01$



## 二、93 年肇事資料

由表 5-9 所示，除了路面缺陷、路面鋪裝、視距外，其餘 19 個肇事變數皆為顯著，因此選入判別模式進行分析。

表 5-9 93 年卡方檢定結果

93 年	卡方值	自由度	P 值	顯著性
發生月份	2157.379	44	0.00	**
發生時間	61.541	28	0.00	**
道路類別	1749.999	28	0.00	**
速限	1865.453	16	0.00	**
道路型態	566.220	12	0.00	**
路面缺陷	7.423	12	0.83	—
分向設施	774.177	16	0.00	**
快車道間	1278.861	16	0.00	**
快慢車道間	295.125	16	0.00	**
路面邊線	801.712	4	0.00	**
發生地區	32.278	16	0.01	*
天候	71.970	16	0.00	**
光線	164.843	12	0.00	**
路面鋪裝	17.824	12	0.33	—
路面狀態	36.693	16	0.00	**
障礙物	25.427	16	0.01	*
視距	6.403	4	0.17	—
號誌種類	445.238	12	0.00	**
號誌動作	427.003	12	0.00	**
事故位置	562.313	12	0.00	**
事故類型及型態	288.658	12	0.00	**
肇因原因	2771.186	36	0.00	**

註：「\*」為  $0.01 \leq P \text{ 值} \leq 0.05$ ，「\*\*」為  $P \text{ 值} < 0.01$

### 三、94 年肇事資料

由表 5-10 所示，除了路面缺陷、路面鋪裝、障礙物外，其餘 19 個肇事變數皆為顯著，因此選入判別模式進行分析。

表 5-10 94 年卡方檢定結果

94 年	卡方值	自由度	P 值	顯著性
發生月份	2122.386	55	0.00	**
發生時間	94.313	35	0.00	**
道路類別	2108.243	35	0.00	**
速限	2487.849	20	0.00	**
道路型態	594.713	15	0.00	**
路面缺陷	11.893	10	0.29	—
分向設施	889.905	20	0.00	**
快車道間	1694.078	20	0.00	**
快慢車道間	391.521	20	0.00	**
路面邊線	872.271	5	0.00	**
發生地區	58.221	20	0.00	**
天候	54.253	25	0.00	**
光線	288.474	15	0.00	**
路面鋪裝	13.636	15	0.55	—
路面狀態	38.193	20	0.01	*
障礙物	26.804	20	0.14	—
視距	24.656	5	0.00	**
號誌種類	448.331	15	0.00	**
號誌動作	447.621	15	0.00	**
事故位置	600.570	15	0.00	**
事故類型及型態	401.784	15	0.00	**
肇因原因	4109.670	45	0.00	**

註：「\*」為  $0.01 \leq P \text{ 值} \leq 0.05$ ，「\*\*」為  $P \text{ 值} < 0.01$

#### 四、95 年肇事資料

由表 5-11 所示，除了路面缺陷、路面鋪裝、障礙物、視距外，其餘 18 個肇事變數皆為顯著，因此選入判別模式進行分析。

表 5-11 95 年卡方檢定結果

95 年	卡方值	自由度	P 值	顯著性
發生月份	2594.512	44	0.00	**
發生時間	66.712	28	0.00	**
道路類別	1115.075	28	0.00	**
速限	1272.101	16	0.00	**
道路型態	350.386	8	0.00	**
路面缺陷	14.874	12	0.25	—
分向設施	464.575	16	0.00	**
快車道間	887.124	16	0.00	**
快慢車道間	93.476	16	0.00	**
路面邊緣	405.375	4	0.00	**
發生地區	59.729	16	0.01	*
天候	45.122	16	0.00	**
光線	120.520	12	0.00	**
路面鋪裝	15.636	16	0.48	—
路面狀態	34.101	16	0.01	*
障礙物	10.832	12	0.54	—
視距	8.609	4	0.07	—
號誌種類	260.646	12	0.00	**
號誌動作	277.310	12	0.00	**
事故位置	380.472	12	0.00	**
事故類型及型態	144.721	8	0.00	**
肇因原因	2865.618	36	0.00	**

註：「\*」為  $0.01 \leq P \text{ 值} \leq 0.05$ ，「\*\*」為  $P \text{ 值} < 0.01$

## 五、96 年肇事資料

由表 5-12 所示，除了路面缺陷、路面鋪裝、路面狀態、視距外，其餘 18 個肇事變數皆為顯著，因此選入判別模式進行分析。

表 5-12 96 年卡方檢定結果

96 年	卡方值	自由度	P 值	顯著性
發生月份	2058.225	66	0.00	**
發生時間	137.108	42	0.00	**
道路類別	1572.860	42	0.00	**
速限	1814.816	24	0.00	**
道路型態	707.669	12	0.00	**
路面缺陷	18.206	18	0.44	—
分向設施	632.273	24	0.00	**
快車道間	1184.394	24	0.00	**
快慢車道間	150.071	24	0.00	**
路面邊線	492.689	6	0.00	**
發生地區	61.991	24	0.00	**
天候	37.063	24	0.04	*
光線	148.511	18	0.00	**
路面鋪裝	26.287	18	0.09	—
路面狀態	18.539	24	0.78	—
障礙物	49.076	24	0.00	**
視距	11.338	6	0.08	—
號誌種類	884.522	18	0.00	**
號誌動作	876.905	18	0.00	**
事故位置	691.469	18	0.00	**
事故類型及型態	237.271	12	0.00	**
肇因原因	3388.463	54	0.00	**

註：「\*」為  $0.01 \leq P \text{ 值} \leq 0.05$ ，「\*\*」為  $P \text{ 值} < 0.01$

## 六、97 年肇事資料

由表 5-13 所示，除了路面缺陷外，其餘 21 個肇事變數皆為顯著，因此選入判別模式進行分析。

表 5-13 97 年卡方檢定結果

97 年	卡方值	自由度	P 值	顯著性
發生月份	2241.780	66	0.00	**
發生時間	74.389	42	0.00	**
道路類別	1007.065	42	0.00	**
速限	1007.942	24	0.00	**
道路型態	838.273	18	0.00	**
路面缺陷	19.700	18	0.35	—
分向設施	728.847	24	0.00	**
快車道間	875.363	24	0.00	**
快慢車道間	83.562	24	0.00	**
路面邊線	434.891	6	0.00	**
發生地區	47.516	24	0.00	**
天候	99.475	36	0.04	*
光線	105.109	18	0.00	**
路面鋪裝	31.177	18	0.03	*
路面狀態	43.868	24	0.01	*
障礙物	34.833	18	0.01	*
視距	18.617	6	0.01	*
號誌種類	1050.627	18	0.00	**
號誌動作	1036.699	18	0.00	**
事故位置	808.678	18	0.00	**
事故類型及型態	166.843	12	0.00	**
肇因原因	3074.085	54	0.00	**

註：「\*」為  $0.01 \leq P \text{ 值} \leq 0.05$ ，「\*\*」為  $P \text{ 值} < 0.01$

## 七、總肇事資料

由表 5-14 所示，除了路面缺陷外，其餘 21 個肇事變數皆為顯著，因此選入判別模式進行分析。

表 5-14 92-97 年度卡方檢定結果

總年度	卡方值	自由度	P 值	顯著性
發生月份	12690.885	55	0.00	**
發生時間	194.687	35	0.00	**
道路類別	10263.846	35	0.00	**
速限	11760.990	20	0.00	**
道路型態	3006.322	15	0.00	**
路面缺陷	14.212	15	0.51	—
分向設施	4271.966	20	0.00	**
快車道間	7848.041	20	0.00	**
快慢車道間	1208.312	20	0.00	**
路面邊線	3914.355	5	0.00	**
發生地區	125.840	20	0.00	**
天候	171.542	30	0.00	**
光線	992.815	15	0.00	**
路面鋪裝	48.292	20	0.00	**
路面狀態	127.451	20	0.00	**
障礙物	107.466	20	0.00	**
視距	47.625	5	0.00	**
號誌種類	2443.011	15	0.00	**
號誌動作	2353.433	15	0.00	**
事故位置	3067.149	15	0.00	**
事故類型及型態	1889.964	15	0.00	**
肇因原因	21746.066	45	0.00	**

註：「\*」為  $0.01 \leq P \text{ 值} \leq 0.05$ ，「\*\*」為  $P \text{ 值} < 0.01$

根據上述各年度及 92-97 年，不顯著變數有路面缺陷、路面鋪裝、障礙物、視距及路面狀態。而路面缺陷在各年度及總年度皆不顯著，代表此肇事變數在不同的最佳群集數下都不相關，如表 5-15 將整理最佳群集數下，顯著之肇事變數之個數及不顯著且予以剔除之肇事變數。

表 5-15 卡方檢定驗證結果

年度	最佳群集數	顯著之肇事變數個數	剔除之肇事變數
92	5	19	路面缺陷、路面鋪裝、障礙物
93	5	19	路面缺陷、路面鋪裝、視距
94	6	19	路面缺陷、路面鋪裝、障礙物
95	5	18	路面缺陷、路面鋪裝、障礙物、視距
96	7	18	路面缺陷、路面鋪裝、路面狀態、視距
97	7	21	路面缺陷
92-97	6	21	路面缺陷

其餘顯著之肇事變數有發生月份、發生時間、道路類別、速限、道路型態、分向設施、快車道間、快慢車道間、路面邊線、發生地區、天候、光線、號誌種類、號誌動作、事故位置、事故類型及型態及肇事原因等選入判別模式進行後續分析。

## 5.5 判別模式架構與分析流程

判別分析所建立的判別模式可提供研究者對於一群未知的母體，透過判別分析將其變數投入判別模式後，得到分群分類之目的。而交通事故的發生絕非是單一的因素所影響，應該為某些因素下對肇事之發生有顯著的影響，並且由多種因素之權重值的組合，但也有可能為單一因素所佔的權重值比例很高。

因此本研究將影響各年度及總年度之顯著肇事變數投入判別分析，並結合群集分析所得之結果，以構建完成各年度及總年度下的肇事判別模式並比較訓

練樣本和測試樣本之分類正確率，其分析流程如圖 5-9 所示。

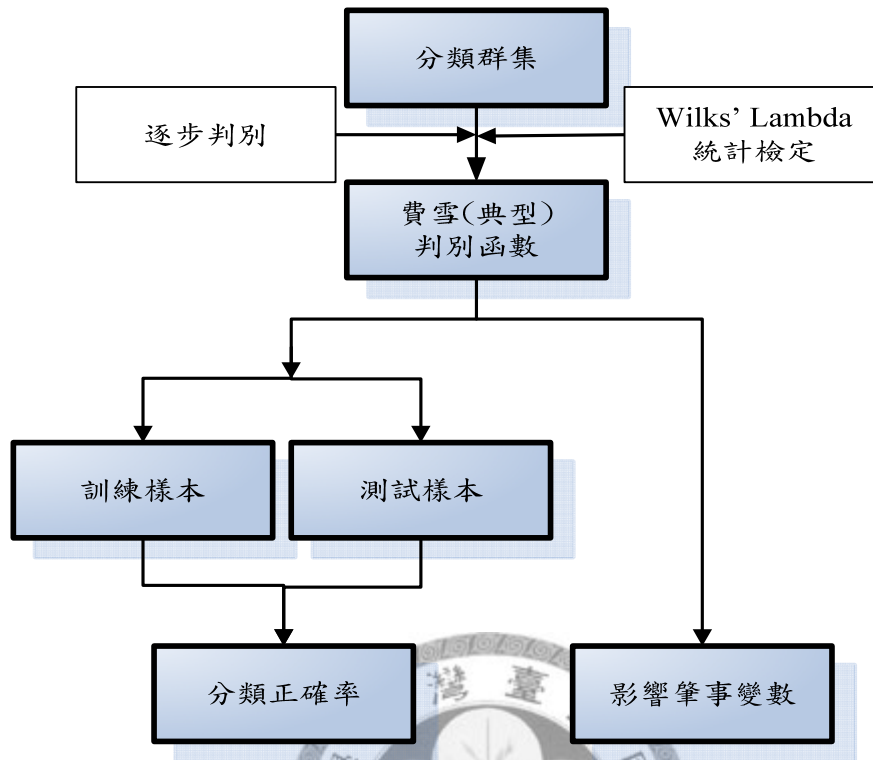


圖 5-9 判別模式分析流程

## 5.6 判別函數的構建

本研究之判別模式架構，以各年度及 92-97 年的群集為判別函數，各項的顯著筆事變數為預測變數，其公式如下所示：

$$D = f(X) = d_0 + d_1x_{aj} + d_2x_{bj} + \cdots + d_{22}x_{vj} \quad (j=1,2,\cdots,12) \quad (\text{式 6.1})$$

其中， $D$  為各年度及 92-97 年群集判別函數， $X$  為顯著的筆事變數 $(x_a, x_b, \cdots, x_v)$

根據先前卡方顯著性檢定結果可知，92 年、93 年、94 年、95 年、96 年、97 年、92-97 年顯著的變數分別為 19 個、19 個、19 個、18 個、18 個、21 個、21 個的預測變數。但預測變數為類別變數時，則必須將類別變數轉換成虛擬變數才套入判別模式才能分析，因此筆事變數共有 118 個虛擬變數，故依照各年度的不同，所帶入的預測變數也有所不同。



## 5.7 判別分析結果

經由 K-Menas 分析法進行分群後，所得最佳群集結果進行判別分析的函數驗證與建構判別模式。由判別分析建立判別函數為費雪(典型)線性判別函數，以 *Wilks' Lambda(A)* 值與卡方值檢定其函數是否能夠有效分群，使得群集間是否有顯著的差異。最後，比較訓練樣本和測試樣本的分類正確率。其判別分析結果如下：

### 一、92 年肇事資料

將 92 年肇事資料的群集結果經由逐步判別分析所建立的判別函數及係數及 *Wilks' Lambda(A)* 值與卡方值檢定結果，如表 5-16 及表 5-17 所示：

表 5-16 92 年肇事資料函數檢定結果

函數檢定	Wilks' Lambda 值	卡方值	自由度	P 值	顯著性
1 到 4	0.005	13795.226	76	0.000	**
2 到 4	0.054	7617.880	54	0.000	**
3 到 4	0.427	2213.093	34	0.000	**
4	0.980	51.567	16	0.000	**

註：「\*\*」為 P 值 < 0.01

表 5-17 92 年費雪線性判別函數係數

92 年	群集觀察值個數				
	1	2	3	4	5
(常數)	-32.246	-16.871	-50.960	-27.510	-9.800
X <sub>a1</sub>	45.710	5.446	47.370	5.552	3.210
X <sub>a2</sub>	45.010	4.893	47.961	5.494	3.157
X <sub>a3</sub>	45.121	4.860	47.468	5.509	2.880
X <sub>a4</sub>	44.988	4.482	47.094	5.289	2.995
X <sub>a5</sub>	39.617	6.569	42.951	5.319	2.170
X <sub>a6</sub>	27.465	5.633	30.030	4.395	1.747
X <sub>c1</sub>	-0.010	-0.526	-1.470	-2.503	0.002
X <sub>c2</sub>	1.599	2.440	1.846	1.165	0.087
X <sub>d5</sub>	-3.169	9.279	-3.792	-1.582	-3.330
X <sub>r4</sub>	3.931	3.163	3.639	2.585	2.514
X <sub>u2</sub>	15.444	13.630	14.111	13.098	14.668
X <sub>u3</sub>	14.400	11.098	8.671	6.696	13.445
X <sub>v2</sub>	2.368	2.505	5.290	6.070	2.163
X <sub>v3</sub>	2.906	3.704	6.710	7.087	2.385
X <sub>v6</sub>	1.854	5.199	12.638	13.139	2.255
X <sub>v7</sub>	5.406	11.626	42.330	41.371	2.038
X <sub>v8</sub>	2.774	9.167	39.908	38.168	-0.070
X <sub>v9</sub>	6.396	12.470	43.182	42.067	3.570
X <sub>v10</sub>	6.230	10.065	38.747	36.132	2.962

依據所得之判別函數進行分類時，將每一觀察值帶入上述五組函數，分類函數大者即屬於該群集。再經由群集分析所得到之分群結果與判別分析進行交叉比對，可得到整體預測分群正確率，表 5-18 訓練樣本判別率 95.8%及表 5-19 測試樣本判別率 93.5%。

表 5-18 92 年訓練樣本分類結果

群集觀察值個數 N=1743		預測各組成員					總合
		1	2	3	4	5	
個數	1	582	0	0	0	0	582
	2	4	160	0	43	14	221
	3	8	0	406	0	0	414
	4	0	0	0	243	4	247
	5	0	0	0	0	279	279
%	1	100.0	0.0	0.0	0.0	0.0	100.0
	2	1.8	72.4	0.0	19.5	6.3	100.0
	3	1.9	0.0	98.1	0.0	0.0	100.0
	4	0.0	0.0	0.0	98.4	1.6	100.0
	5	0.0	0.0	0.0	0.0	100.0	100.0

註：正確判別率 95.8%=(582+160+406+243+279)/1743

表 5-19 92 年測試樣本分類結果

群集觀察值個數 N=872		預測各組成員					總合
		1	2	3	4	5	
個數	1	130	0	0	0	0	130
	2	0	214	0	0	5	219
	3	4	12	96	0	0	112
	4	1	0	15	168	3	187
	5	17	0	0	0	207	224
%	1	100.0	0.0	0.0	0.0	0.0	100.0
	2	0.0	97.7	0.0	0.0	2.3	100.0
	3	3.6	10.7	85.7	0.0	0.0	100.0
	4	0.5	0.0	8.0	89.8	1.6	100.0
	5	7.6	0.0	0.0	0.0	92.4	100.0

註：正確判別率 $93.5\%=(130+214+96+168+207)/872$

## 二、93 年肇事資料

將 93 年肇事資料的群集結果經由逐步判別分析所建立的判別函數及係數

及 *Wilks' Lambda*(*A*)值與卡方值檢定結果，如表 5-20 及表 5-21 所示：

表 5-20 93 年肇事資料函數檢定結果

函數檢定	Wilks' Lambda 值	卡方值	自由度	P 值	顯著性
1 到 4	0.018	10948.947	96	0.000	**
2 到 4	0.084	6760.465	69	0.000	**
3 到 4	0.283	3436.757	44	0.000	**
4	0.830	508.800	21	0.000	**

註：「\*\*」為 P 值 $<0.01$

表 5-21 93 年費雪線性判別函數係數

93 年	群集觀察值個數				
	1	2	3	4	5
(常數)	-12.533	-21.853	-13.352	-18.003	-4.722
$X_{a6}$	9.068	7.181	4.715	10.583	2.017
$X_{a7}$	12.691	8.566	5.366	13.430	1.665
$X_{a8}$	14.535	8.788	5.530	15.477	1.408
$X_{a9}$	15.618	9.057	5.412	16.126	1.450
$X_{a10}$	17.657	7.536	2.682	18.072	1.537
$X_{a11}$	18.181	7.536	2.683	18.643	1.965
$X_{a12}$	18.200	6.382	2.048	18.228	1.654
$X_{c1}$	2.706	8.510	5.328	0.772	-0.165
$X_{c2}$	2.004	5.785	5.101	0.290	0.258
$X_{d3}$	1.489	3.197	3.084	0.438	0.858
$X_{d4}$	0.941	6.518	5.777	-1.253	-0.396
$X_{d5}$	-2.125	8.039	9.818	-4.282	-1.281
$X_{g1}$	2.531	2.788	2.930	1.723	2.079
$X_{m4}$	0.472	1.233	-0.666	0.599	0.604
$X_{s1}$	1.537	0.756	0.568	2.020	2.210
$X_{u3}$	-0.237	5.218	0.812	-0.535	0.480
$X_{v1}$	7.275	7.230	6.395	5.944	4.656
$X_{v2}$	5.025	-0.079	5.356	4.684	3.897
$X_{v3}$	4.137	-1.416	4.714	3.930	3.599
$X_{v6}$	4.386	4.539	3.929	7.914	5.560
$X_{v7}$	1.683	9.471	1.806	12.722	6.436
$X_{v8}$	3.078	12.850	3.084	15.164	7.885
$X_{v9}$	3.897	12.764	3.146	15.614	7.673
$X_{v10}$	3.969	18.229	2.690	16.658	7.698

依據所得之判別函數進行分類時，將每一觀察值帶入上述五組函數，分類函數大者即屬於該群集。再經由群集分析所得到之分群結果與判別分析進行交叉比對，可得到整體預測分群正確率，表 5-22 訓練樣本判別率 95.8%及表 5-23 測試樣本判別率 94.0%。

表 5-22 93 年訓練樣本分類結果

群集觀察值個數 N=1827		預測各組成員					總合
		1	2	3	4	5	
個數	1	363	1	2	0	0	366
	2	0	126	1	3	8	138
	3	1	1	280	0	18	300
	4	10	0	0	306	2	318
	5	4	2	2	0	697	705
%	1	99.2	0.3	0.5	0.0	0.0	100.0
	2	0.0	91.3	0.7	2.2	5.8	100.0
	3	0.3	0.3	93.3	0.0	6.0	100.0
	4	3.1	0.0	0.0	96.2	0.6	100.0
	5	0.6	0.3	0.3	0.0	98.9	100.0

註：正確判別率 $95.8\%=(363+126+280+306+697)/1827$

表 5-23 93 年測試樣本分類結果

群集觀察值個數 N=913		預測各組成員					總合
		1	2	3	4	5	
個數	1	283	1	7	0	2	293
	2	0	81	0	7	0	88
	3	3	0	109	0	4	116
	4	3	1	0	251	5	260
	5	6	0	0	16	134	156
%	1	96.6	0.3	2.4	0.0	0.7	100.0
	2	0.0	92.0	0.0	8.0	0.0	100.0
	3	2.6	0.0	94.0	0.0	3.4	100.0
	4	1.2	0.4	0.0	96.5	1.9	100.0
	5	3.8	0.0	0.0	10.3	85.9	100.0

註：正確判別率 $94.0\%=(283+81+109+251+134)/913$

### 三、94 年肇事資料

將 94 年肇事資料的群集結果經由逐步判別分析所建立的判別函數及係數

及 Wilks' Lambda(A)值與卡方值檢定結果，如表 5-24 及表 5-25 所示：

表 5-24 94 年肇事資料函數檢定結果

函數檢定	Wilks' Lambda 值	卡方值	自由度	P 值	顯著性
1 到 5	0.000	24515.798	145	0.000	**
2 到 5	0.017	11240.358	112	0.000	**
3 到 5	0.185	4651.434	81	0.000	**
4 到 5	0.714	930.875	52	0.000	**
5	0.889	61.406	25	0.000	**

註：「\*\*」為 P 值 $<0.01$

表 5-25 94 年費雪線性判別函數係數

94 年	群集觀察值個數					
	1	2	3	4	5	6
(常數)	-20.046	-298.887	-261.847	-11.830	-44.254	-273.173
X <sub>a1</sub>	18.153	9.892	2.616	1.319	8.739	18.899
X <sub>a2</sub>	16.905	9.819	1.697	1.352	9.883	17.282
X <sub>a3</sub>	17.345	10.508	1.548	0.864	10.093	17.824
X <sub>a4</sub>	16.677	10.117	0.813	1.017	10.523	16.544
X <sub>a5</sub>	15.781	14.937	3.847	1.077	11.597	18.821
X <sub>a6</sub>	16.627	13.557	2.489	1.714	12.467	17.542
X <sub>a7</sub>	4.313	10.573	7.312	3.303	6.701	8.368
X <sub>a8</sub>	4.069	7.888	4.949	3.148	6.186	5.769
X <sub>c1</sub>	-1.517	14.804	-3.471	-0.716	14.371	-3.408
X <sub>c2</sub>	1.267	4.234	-1.982	1.344	6.345	-2.521
X <sub>d1</sub>	0.766	-1.379	1.746	0.664	-1.929	1.803
X <sub>d3</sub>	0.691	5.336	1.980	0.463	4.571	2.382
X <sub>d4</sub>	0.416	14.876	-0.061	0.738	14.441	0.856
X <sub>d5</sub>	1.272	38.136	2.142	0.510	39.405	2.557
X <sub>g1</sub>	3.642	7.027	4.571	3.839	5.541	4.490
X <sub>g5</sub>	4.846	10.131	10.227	4.637	4.887	10.109
X <sub>m4</sub>	0.202	3.167	-0.601	0.612	2.537	-0.988
X <sub>r4</sub>	4.180	3.665	1.831	4.208	6.549	1.765
X <sub>i3</sub>	1.431	-6.521	-0.003	1.023	-7.034	-0.282
X <sub>u2</sub>	12.965	14.701	14.007	12.953	11.506	13.500
X <sub>u3</sub>	11.080	16.061	11.930	11.933	9.496	10.766
X <sub>v1</sub>	4.098	1.711	0.576	2.992	2.989	0.640
X <sub>v2</sub>	2.130	0.796	3.206	2.514	5.187	3.013
X <sub>v3</sub>	3.042	-0.591	3.503	2.856	4.316	3.941
X <sub>v6</sub>	3.805	33.187	34.146	3.567	3.196	34.246
X <sub>v7</sub>	6.396	501.892	501.334	5.874	6.162	504.568
X <sub>v8</sub>	6.735	506.343	505.690	6.257	5.593	509.750
X <sub>v9</sub>	7.456	506.362	505.267	6.831	5.950	509.510
X <sub>v10</sub>	8.051	512.877	506.668	7.358	4.955	510.327

依據所得之判別函數進行分類時，將每一觀察值帶入上述六組函數，分類函數大者即屬於該群集。再經由群集分析所得到之分群結果與判別分析進行交叉比對，可得到整體預測分群正確率，表 5-26 訓練樣本判別率 97.9%及表 5-27 測試樣本判別率 96.4%。

表 5-26 94 年訓練樣本分類結果

群集觀察個數 N=1853		預測各組成員						總合
		1	2	3	4	5	6	
個數	1	407	0	0	0	5	0	412
	2	0	115	5	0	0	10	130
	3	0	0	300	1	0	0	301
	4	0	0	0	353	2	0	355
	5	5	0	0	11	311	0	327
	6	0	0	0	0	0	328	328
%	1	98.8	0.0	0.0	0.0	1.2	0.0	100.0
	2	0.0	88.5	3.8	0.0	0.0	7.7	100.0
	3	0.0	0.0	99.7	0.3	0.0	0.0	100.0
	4	0.0	0.0	0.0	99.4	0.6	0.0	100.0
	5	1.5	0.0	0.0	3.4	95.1	0.0	100.0
	6	0.0	0.0	0.0	0.0	0.0	100.0	100.0

註：正確判別率97.9%=(407+115+300+353+311+328)/1853

表 5-27 94 年測試樣本分類結果

群集觀察個數 N=926		預測各組成員						總合
		1	2	3	4	5	6	
個數	1	126	0	0	0	0	0	126
	2	0	82	4	0	0	3	89
	3	0	0	190	4	0	0	194
	4	0	0	0	207	1	0	208
	5	14	0	0	6	171	0	191
	6	1	0	0	0	0	117	118
%	1	100.0	0.0	0.0	0.0	0.0	0.0	100.0
	2	0.0	92.1	4.5	0.0	0.0	3.4	100.0
	3	0.0	0.0	97.9	2.1	0.0	0.0	100.0
	4	0.0	0.0	0.0	99.5	0.5	0.0	100.0
	5	7.3	0.0	0.0	3.1	89.5	0.0	100.0
	6	0.8	0.0	0.0	0.0	0.0	99.2	100.0

註：正確判別率96.4%=(126+82+190+207+171+117)/926

#### 四、95 年肇事資料

將 95 年肇事資料的群集結果經由逐步判別分析所建立的判別函數及係數

及 Wilks' Lambda(A)值與卡方值檢定結果，如表 5-28 及表 5-29 所示：

表 5-28 95 年肇事資料函數檢定結果

函數檢定	Wilks' Lambda 值	卡方值	自由度	P 值	顯著性
1 到 4	0.007	13700.260	92	0.000	**
2 到 4	0.062	7559.944	66	0.000	**
3 到 4	0.484	1976.750	42	0.000	**
4	0.914	243.838	20	0.000	**

註：「\*\*」為 P 值<0.01

表 5-29 95 年費雪線性判別函數係數

95 年	群集觀察值個數				
	1	2	3	4	5
(常數)	-23.526	-9.198	-20.381	-39.835	-2.872
$X_{a1}$	41.352	3.597	3.128	40.372	4.187
$X_{a2}$	40.846	3.432	2.357	39.763	4.028
$X_{a3}$	41.374	4.056	3.243	40.718	4.183
$X_{a4}$	40.796	3.756	3.712	41.058	4.134
$X_{a5}$	37.127	3.913	3.390	37.304	4.312
$X_{a6}$	20.840	3.695	2.089	22.850	4.129
$X_{a7}$	3.146	2.353	1.226	2.782	2.798
$X_{a8}$	0.686	1.277	2.490	1.589	0.542
$X_{c1}$	1.987	4.710	-0.003	1.469	0.033
$X_{c2}$	3.000	1.760	-0.110	2.007	0.969
$X_{d3}$	2.691	2.537	1.712	2.556	0.509
$X_{d4}$	2.391	3.331	0.761	1.239	0.465
$X_{d5}$	0.245	4.865	-2.042	-2.193	-0.368
$X_{g5}$	0.254	1.481	1.741	0.554	1.912
$X_{t3}$	-0.795	-4.052	-2.139	-1.866	0.703
$X_{u3}$	0.025	-0.085	-1.692	-2.793	0.510
$X_{v2}$	5.322	3.087	6.455	7.190	2.618
$X_{v3}$	4.422	4.373	7.437	7.612	2.502
$X_{v6}$	4.728	8.205	21.390	21.636	2.372
$X_{v7}$	5.759	11.583	38.165	40.348	1.724
$X_{v8}$	7.007	12.290	38.487	41.015	2.279
$X_{v9}$	6.538	11.661	38.363	40.700	2.045
$X_{v10}$	6.113	10.454	35.658	38.492	2.559

依據所得之判別函數進行分類時，將每一觀察值帶入上述五組函數，分類函數大者即屬於該群集。再經由群集分析所得到之分群結果與判別分析進行交叉比對，可得到整體預測分群正確率，表 5-30 訓練樣本判別率 94.8%及表 5-31 測試樣本判別率 93.2%。



表 5-30 95 年訓練樣本分類結果

群集觀察值個數 N=1825		預測各組成員					總合
		1	2	3	4	5	
個數	1	544	0	0	18	0	562
	2	0	122	52	0	8	182
	3	0	1	312	0	0	313
	4	2	0	2	431	0	435
	5	5	1	5	0	322	333
%	1	96.8	0.0	0.0	3.2	0.0	100.0
	2	0.0	67.0	28.6	0.0	4.4	100.0
	3	0.0	0.3	99.7	0.0	0.0	100.0
	4	0.5	0.0	0.5	99.1	0.0	100.0
	5	1.5	0.3	1.5	0.0	96.7	100.0

註：正確判別率  $94.8\% = (544 + 122 + 312 + 431 + 332) / 1825$

表 5-31 95 年測試樣本分類結果

群集觀察值個數 N=912		預測各組成員					總合
		1	2	3	4	5	
個數	1	81	1	0	1	5	88
	2	9	151	3	2	7	172
	3	0	0	173	0	2	175
	4	5	0	1	160	0	166
	5	16	5	5	0	285	311
%	1	92.0	1.1	0.0	1.1	5.7	100.0
	2	5.2	87.8	1.7	1.2	4.1	100.0
	3	0.0	0.0	98.9	0.0	1.1	100.0
	4	3.0	0.0	0.6	96.4	0.0	100.0
	5	5.1	1.6	1.6	0.0	91.6	100.0

註：正確判別率  $93.2\% = (81 + 151 + 173 + 160 + 285) / 912$

## 五、96 年肇事資料

將 96 年肇事資料的群集結果經由逐步判別分析所建立的判別函數及係數

及 *Wilks' Lambda*(*A*)值與卡方值檢定結果，如表 5-32 及表 5-33 所示：

表 5-32 96 年肇事資料函數檢定結果

函數檢定	Wilks' Lambda 值	卡方值	自由度	P 值	顯著性
1 到 6	0.001	15786.732	150	0.000	**
2 到 6	0.023	8838.872	120	0.000	**
3 到 6	0.130	4793.199	92	0.000	**
4 到 6	0.532	1482.601	66	0.000	**
5 到 6	0.784	569.941	42	0.000	**
6	0.959	99.095	20	0.000	**

註：「\*\*」為 P 值  $< 0.01$

表 5-33 96 年費雪線性判別函數係數

96年	群集觀察值個數						
	1	2	3	4	5	6	7
(常數)	-50.076	-53.396	-14.766	-46.046	-19.488	-4.631	-60.977
$X_{a1}$	18.275	18.974	23.051	-2.478	2.244	0.651	-2.327
$X_{a2}$	20.551	20.281	22.415	-0.073	3.450	1.283	0.843
$X_{a3}$	18.195	18.083	20.406	-1.356	6.269	1.041	1.599
$X_{a4}$	17.541	18.182	19.652	-0.400	6.846	0.891	3.894
$X_{a5}$	18.247	18.084	19.791	-0.264	7.290	1.309	4.716
$X_{a6}$	19.826	18.848	19.348	1.425	7.795	1.344	6.476
$X_{a7}$	9.792	9.071	7.745	1.653	4.708	3.164	5.618
$X_{c1}$	1.337	3.181	1.206	1.126	6.438	-0.305	9.895
$X_{d3}$	0.496	0.895	0.758	0.712	3.425	0.686	4.043
$X_{d4}$	3.042	3.438	0.975	2.283	7.173	0.612	7.794
$X_{d5}$	-2.539	-2.670	-3.387	0.054	17.412	-0.431	16.793
$X_{g1}$	0.786	1.098	1.681	0.422	2.322	1.643	1.976
$X_{m4}$	-1.163	-0.454	-1.988	0.714	1.554	0.255	2.886
$X_{p2}$	-0.438	7.035	0.518	-1.568	-21.955	-0.225	-26.726
$X_{r1}$	3.689	3.153	-0.164	3.214	-0.482	0.220	2.769
$X_{s1}$	6.554	0.221	2.856	3.437	1.925	3.115	1.569
$X_{t3}$	-4.059	-3.103	-0.837	-2.941	-4.954	-0.437	-4.755
$X_{v2}$	6.395	5.231	5.528	6.060	7.738	4.602	1.771
$X_{v3}$	5.624	4.867	5.111	5.547	7.624	4.447	1.060
$X_{v5}$	11.239	10.157	5.075	10.703	5.349	5.687	9.878
$X_{v6}$	19.462	18.420	5.369	19.501	5.691	5.988	17.708
$X_{v7}$	75.738	78.265	4.100	81.954	6.268	5.374	79.883
$X_{v8}$	77.397	86.176	5.732	85.984	8.142	6.315	85.699
$X_{v9}$	76.977	83.747	4.389	85.795	6.769	6.158	84.578
$X_{v10}$	77.549	85.604	4.734	86.599	5.226	6.377	89.804

依據所得之判別函數進行分類時，將每一觀察值帶入上述七組函數，分類函數大者即屬於該群集。再經由群集分析所得到之分群結果與判別分析進行交叉比對，可得到整體預測分群正確率，表 5-34 訓練樣本判別率 93.6%及表 5-35 測試樣本判別率 94.3%。

表 5-34 96 年訓練樣本分類結果

群集觀察值個數 N=1577		預測各組成員							總合
個數		1	2	3	4	5	6	7	
	1	165	1	23	0	0	0	0	189
	2	2	195	3	0	0	0	12	212
	3	0	0	435	0	41	0	0	476
	4	0	0	0	195	0	2	0	197
	5	0	0	1	0	162	10	0	173
	6	0	0	0	0	0	250	0	250
	7	0	0	0	6	0	0	74	80
%	1	87.3	0.5	12.2	0.0	0.0	0.0	0.0	100.0
	2	0.9	92.0	1.4	0.0	0.0	0.0	5.7	100.0
	3	0.0	0.0	91.4	0.0	8.6	0.0	0.0	100.0
	4	0.0	0.0	0.0	99.0	0.0	1.0	0.0	100.0
	5	0.0	0.0	0.6	0.0	93.6	5.8	0.0	100.0
	6	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0
	7	0.0	0.0	0.0	7.5	0.0	0.0	92.5	100.0

註：正確判別率96.3%=(165+195+435+195+162+250+74)/1577

表 5-35 96 年測試樣本分類結果

群集觀察值個數 N=788		預測各組成員							總合
個數		1	2	3	4	5	6	7	
	1	44	3	5	6	0	4	0	62
	2	0	38	0	0	0	0	0	38
	3	0	0	60	0	0	0	0	60
	4	0	9	0	158	0	1	0	168
	5	0	0	3	0	116	8	0	127
	6	0	0	0	0	0	267	0	267
	7	0	2	0	4	0	0	60	66
%	1	71.0	4.8	8.1	9.7	0.0	6.5	0.0	100.0
	2	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0
	3	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0
	4	0.0	5.4	0.0	94.0	0.0	0.6	0.0	100.0
	5	0.0	0.0	2.4	0.0	91.3	6.3	0.0	100.0
	6	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0
	7	0.0	3.0	0.0	6.1	0.0	0.0	90.9	100.0

註：正確判別率94.3%=(44+38+60+158+116+267+60)/788

## 六、97 年肇事資料

將 97 年肇事資料的群集結果經由逐步判別分析所建立的判別函數及係數

及 Wilks' Lambda(Λ)值與卡方值檢定結果，如表 5-36 及表 5-37 所示：

表 5-36 97 年肇事資料函數檢定結果

函數檢定	Wilks' Lambda 值	卡方值	自由度	P 值	顯著性
1 到 6	0.002	14607.390	144	0.000	**
2 到 6	0.034	7633.043	115	0.000	**
3 到 6	0.225	3369.034	88	0.000	**
4 到 6	0.572	1264.474	63	0.000	**
5 到 6	0.862	335.321	40	0.000	**
6	0.979	47.797	19	0.000	**

註：「\*\*」為 P 值 < 0.01

表 5-37 97 年費雪線性判別函數係數

97 年	群集觀察值個數						
	1	2	3	4	5	6	7
(常數)	-15.651	-6.501	-53.806	-62.346	-46.166	-25.072	-16.752
$X_{a1}$	3.499	0.615	1.055	28.775	3.720	32.521	25.204
$X_{a2}$	2.608	0.780	-0.283	28.069	2.793	31.793	25.285
$X_{a3}$	3.947	0.703	-1.444	26.108	1.186	31.415	25.505
$X_{a4}$	2.917	0.321	-0.259	26.954	2.408	29.738	25.327
$X_{a5}$	3.558	-0.058	-0.450	20.603	3.891	21.563	19.937
$X_{a6}$	2.215	-0.062	1.491	17.767	5.477	16.933	17.611
$X_{a7}$	2.872	1.089	2.180	5.868	3.293	6.284	6.982
$X_{b3}$	0.964	1.610	1.785	1.081	1.556	0.666	1.725
$X_{c1}$	1.699	-1.616	0.379	0.094	-1.767	1.570	-0.980
$X_{c2}$	4.960	1.482	2.230	1.549	-0.098	3.990	1.348
$X_{d4}$	2.270	0.068	0.464	-0.316	-0.095	1.899	-1.317
$X_{d5}$	4.919	-1.312	-1.268	-3.907	-1.418	-2.482	-2.851
$X_{g1}$	5.360	3.848	4.471	4.830	4.456	5.549	3.959
$X_{g5}$	3.464	4.550	3.138	2.895	4.801	2.964	4.645
$X_{r4}$	7.962	6.345	8.894	6.776	1.446	8.756	3.932
$X_{u3}$	-1.892	0.561	1.987	0.261	0.501	-2.362	-0.202
$X_{v1}$	4.683	4.323	5.495	5.316	4.060	3.551	6.109
$X_{v2}$	8.366	4.068	3.390	2.331	3.780	3.650	2.880
$X_{v3}$	8.821	4.149	4.048	4.571	4.653	6.342	4.823
$X_{v6}$	4.455	4.340	25.095	24.806	25.220	4.503	5.351
$X_{v7}$	4.041	3.379	79.870	75.931	79.042	-0.037	2.548
$X_{v8}$	5.063	3.742	93.044	90.581	88.127	3.066	5.811
$X_{v9}$	4.638	3.485	92.492	89.727	87.290	2.230	4.633
$X_{v10}$	1.718	3.429	92.300	90.009	86.990	1.319	5.242

依據所得之判別函數進行分類時，將每一觀察值帶入上述七組函數，分類函數大者即屬於該群集。再經由群集分析所得到之分群結果與判別分析進行交叉比對，可得到整體預測分群正確率，表 5-38 訓練樣本判別率 91.3%及表 5-39 測試樣本判別率 93.9%。

表 5-38 97 年訓練樣本分類結果

群集觀察值個數 N=1519		預測各組成員							總合
		1	2	3	4	5	6	7	
個數	1	71	9	0	0	0	24	0	104
	2	0	187	0	0	0	0	0	187
	3	0	2	138	0	2	0	0	142
	4	0	0	0	408	0	3	2	413
	5	0	4	3	24	110	0	0	141
	6	0	0	0	0	0	276	20	296
	7	0	0	0	11	0	28	197	236
%	1	68.3	8.7	0.0	0.0	0.0	23.1	0.0	100.0
	2	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0
	3	0.0	1.4	97.2	0.0	1.4	0.0	0.0	100.0
	4	0.0	0.0	0.0	98.8	0.0	0.7	0.5	100.0
	5	0.0	2.8	2.1	17.0	78.0	0.0	0.0	100.0
	6	0.0	0.0	0.0	0.0	0.0	93.2	6.8	100.0
	7	0.0	0.0	0.0	4.7	0.0	11.9	83.5	100.0

註：正確判別率  $91.3\% = (71+187+138+408+110+276+197)/1519$

表 5-39 97 年測試樣本分類結果

群集觀察值個數 N=759		預測各組成員							總合
		1	2	3	4	5	6	7	
個數	1	92	12	0	0	0	0	1	105
	2	1	198	0	0	0	0	0	199
	3	0	2	162	5	1	0	0	170
	4	0	0	0	33	0	1	0	34
	5	0	6	0	1	152	0	2	161
	6	0	0	0	0	0	5	1	6
	7	0	4	0	0	0	9	71	84
%	1	87.6	11.4	0.0	0.0	0.0	0.0	1.0	100.0
	2	0.5	99.5	0.0	0.0	0.0	0.0	0.0	100.0
	3	0.0	1.2	95.3	2.9	0.6	0.0	0.0	100.0
	4	0.0	0.0	0.0	97.1	0.0	2.9	0.0	100.0
	5	0.0	3.7	0.0	0.6	94.4	0.0	1.2	100.0
	6	0.0	0.0	0.0	0.0	0.0	83.3	16.7	100.0
	7	0.0	4.8	0.0	0.0	0.0	10.7	84.5	100.0

註：正確判別率  $93.9\% = (92+198+162+33+152+5+71)/759$

## 七、92-97 年肇事資料

將總肇事資料的群集結果經由逐步判別分析所建立的判別函數及係數

如，及 *Wilks' Lambda*( $\lambda$ )值與卡方值檢定結果，如表 5-40 及表 5-41 所示：

表 5-40 92-97 年肇事資料函數檢定結果

函數檢定	Wilks' Lambda 值	卡方值	自由度	P 值	顯著性
1 到 5	0.001	115944.031	175	0.000	**
2 到 5	0.033	52634.176	136	0.000	**
3 到 5	0.174	27048.903	99	0.000	**
4 到 5	0.743	4601.905	64	0.000	**
5	0.981	300.063	31	0.000	**

註：「\*\*」為 P 值  $< 0.01$

表 5-41 92-97 年費雪線性判別函數係數

總年度	群集觀察值個數					
	1	2	3	4	5	6
(常數)	-33.269	-17.602	-140.174	-27.526	-158.695	-145.324
$X_{a5}$	3.815	3.007	3.167	3.444	4.275	3.324
$X_{a6}$	4.273	3.164	3.100	3.668	4.898	3.506
$X_{a7}$	8.003	2.409	3.779	16.037	12.457	18.436
$X_{a8}$	8.304	2.061	2.591	18.852	11.578	19.869
$X_{a9}$	7.762	1.922	2.312	19.145	11.111	19.786
$X_{a10}$	6.892	1.519	2.676	18.910	11.002	20.551
$X_{a11}$	3.090	1.498	0.355	21.045	6.496	19.909
$X_{a12}$	2.942	1.576	0.866	21.404	6.655	20.502
$X_{c1}$	12.467	3.721	3.630	5.097	14.568	3.531
$X_{c2}$	4.341	1.862	0.061	2.602	3.292	0.284
$X_{c3}$	2.176	2.146	0.108	2.566	-0.023	0.371
$X_{d2}$	12.025	10.618	10.270	10.841	11.972	10.313
$X_{d3}$	11.568	7.819	7.286	8.359	11.110	7.228
$X_{d4}$	15.293	7.644	6.565	8.819	15.137	6.434
$X_{d5}$	21.420	4.772	5.418	4.874	22.621	3.699
$X_{g1}$	4.745	3.820	3.883	4.119	4.897	3.973
$X_{g5}$	6.235	6.583	8.283	6.053	7.863	8.056
$X_{h1}$	4.324	2.750	2.811	3.325	4.010	2.850
$X_{i4}$	2.378	2.924	2.353	2.796	2.339	2.512
$X_{l1}$	2.628	1.246	-0.033	-4.332	0.071	-5.222
$X_{m4}$	1.072	0.542	-0.575	0.598	0.607	-0.628
$X_{n2}$	0.034	2.343	1.228	2.297	-0.330	1.120
$X_{o2}$	2.630	1.556	2.627	1.492	-3.589	2.259
$X_{r4}$	6.405	4.918	2.940	5.572	4.098	3.248
$X_{t3}$	-0.619	3.943	1.729	3.696	-0.839	2.020
$X_{u2}$	12.509	13.148	13.530	12.995	15.060	13.749
$X_{u3}$	12.461	13.077	12.199	12.951	18.059	12.583
$X_{v1}$	3.455	3.820	2.057	3.495	3.060	1.969
$X_{v2}$	3.735	2.377	3.335	2.627	-1.229	3.525
$X_{v3}$	3.384	2.726	3.734	2.609	-2.151	3.541
$X_{v6}$	3.949	4.264	38.696	4.045	37.738	39.174
$X_{v7}$	5.022	4.460	248.157	4.591	243.026	244.072
$X_{v8}$	4.776	4.511	249.955	4.012	245.334	245.650
$X_{v9}$	5.755	5.725	250.885	5.492	246.362	246.527
$X_{v10}$	4.416	5.235	249.785	4.747	250.538	245.607

依據所得之判別函數進行分類時，將每一觀察值帶入上述六組函數，分類函數大者即屬於該群集。再經由群集分析所得到之分群結果與判別分析進行交叉比對，可得到整體預測分群正確率，表 5-42 訓練樣本判別率 94.3%及表 5-43 測試樣本判別率 96.7%。

表 5-42 92-97 年訓練樣本分類結果

群集觀察個數 N=10342		預測各組成員						總合
		1	2	3	4	5	6	
個數	1	1358	82	0	25	0	0	1465
	2	0	2030	0	82	0	0	2112
	3	0	0	1755	0	0	0	1755
	4	257	0	0	2082	0	0	2339
	5	6	0	36	0	666	46	754
	6	0	0	0	53	0	1864	1917
%	1	92.7	5.6	0.0	1.7	0.0	0.0	100.0
	2	0.0	96.1	0.0	3.9	0.0	0.0	100.0
	3	0.0	0.0	100.0	0.0	0.0	0.0	100.0
	4	11.0	0.0	0.0	89.0	0.0	0.0	100.0
	5	0.8	0.0	4.8	0.0	88.3	6.1	100.0
	6	0.0	0.0	0.0	2.8	0.0	97.2	100.0

註：正確判別率94.3%=(1358+2030+1755+2082+666+1864)/10342

表 5-43 92-97 年測試樣本分類結果

群集觀察個數 N=5172		預測各組成員						總合
		1	2	3	4	5	6	
個數	1	752	39	0	19	0	0	810
	2	5	1042	0	46	0	0	1093
	3	0	0	901	0	3	0	904
	4	11	0	0	1102	0	0	1113
	5	3	0	22	0	373	9	407
	6	0	0	0	14	1	830	845
%	1	92.8	4.8	0.0	2.3	0.0	0.0	100.0
	2	0.5	95.3	0.0	4.2	0.0	0.0	100.0
	3	0.0	0.0	99.7	0.0	0.3	0.0	100.0
	4	1.0	0.0	0.0	99.0	0.0	0.0	100.0
	5	0.7	0.0	5.4	0.0	91.6	2.2	100.0
	6	0.0	0.0	0.0	1.7	0.1	98.2	100.0

註：正確判別率96.7%=(752+1042+901+1102+373+830)/5172

經由上述判別分析之結果，從最佳分類的群集帶入判別模式，經由 *Wilks'* *Lambda*(*Λ*)檢定函數是否顯著及逐步判別分析建立費雪函數，最後再由訓練樣本及測試樣本得到整體分群的預測正確率，表 5-44 為各年度及 92-97 年判別分析之彙整，其分群數為 5~7，函數檢定皆為顯著分，分類正確率也都達 90%以上，代表其判別率佳。

表 5-44 判別模式彙整

年度		92	93	94	95	96	97	92-97
分群數		5	5	6	5	7	7	6
函數檢定		顯著						
分類正 確率(%)	訓練樣本	95.8	95.8	97.9	94.8	94.3	91.3	94.3
	測試樣本	93.5	94.0	96.4	93.2	94.3	93.9	96.7

## 5.8 判別模式綜合分析

根據群集分析法選取最佳群集數目，再套入判別模式的逐步判別分析，所選取出的變數下，哪個肇事變數在不同群集下所佔的比重為何？各肇事年度及總肇事年由哪些變數影響最鉅，其分析如下：

### 一、92 年肇事資料

經過逐步判別分析得到影響 92 年肇事資料的判別函數之預測變數共有 19 項，分別為變換車道或方向不當、未保持行車安全距離、酒醉(後)駕駛失控、違反號誌管制或指揮、違反特定標誌(線)標制、其他駕駛人因素、非駕駛人因素、一月到六月、速限大於 70 公里、國道、省道、無號誌、車與車、汽(機)車本身。

根據 92 年肇事資料判別分類函數與各肇事影響變數之關係，主要影響變數分別由肇事原因、發生月份、事故類型及型態。圖 5-10 及圖 5-11 為不同群集下，判別函數之肇事變數影響程度百分比及權重值。



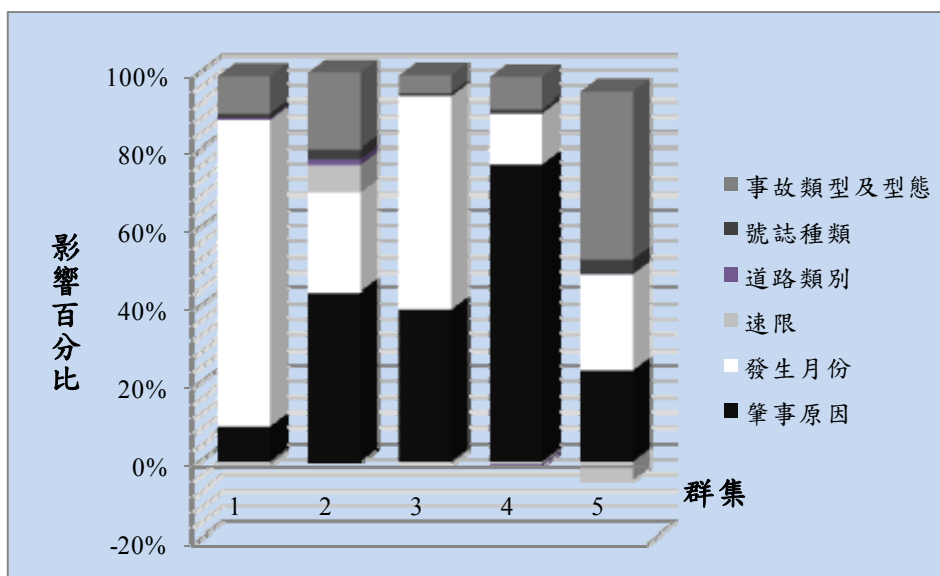


圖 5-10 92 年各群集肇事之變數所佔比例

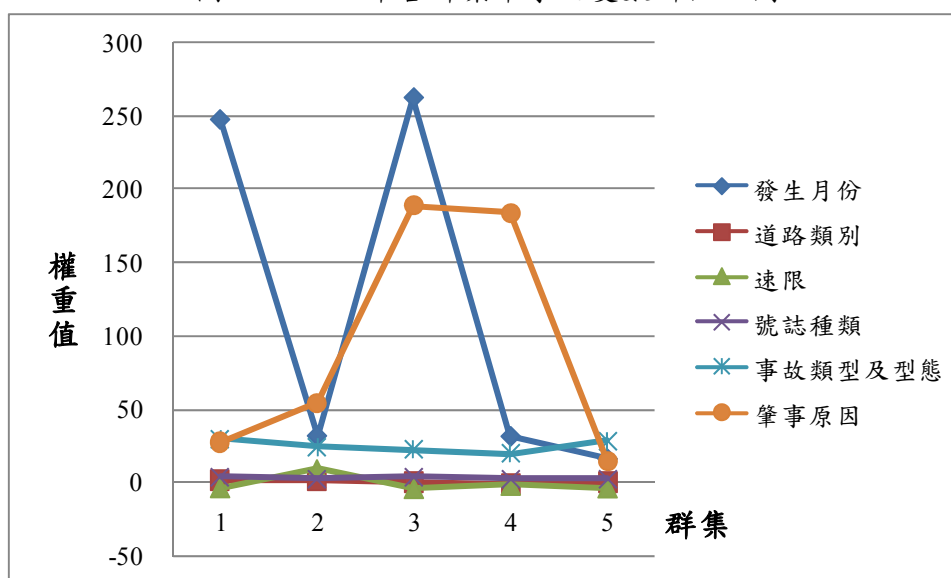


圖 5-11 92 年各群集肇事之變數權重值

## 二、93 年肇事資料

經過逐步判別分析得到影響 93 年肇事資料的判別函數之預測變數共有 24 項，分別為未依規定讓車、變換車道或方向不當、未保持行車安全距離、酒醉(後)駕駛失控、違反號誌管制或指揮、違反特定標誌(線)標制、其他駕駛人因素、非駕駛人因素、六到十二月、速限 60-70 公里及大於 70 公里、國道、省道、6-9 時、中央分隔島、無分向設施、無號誌、汽(機)車本身。

根據 93 年肇事資料判別分類函數與各肇事影響變數之關係，主要影響變數分別由肇事原因、發生月份、速限、道路類別。圖 5-12 及圖 5-13 為不同群

集下，判別函數之肇事變數影響程度百分比及權重值。

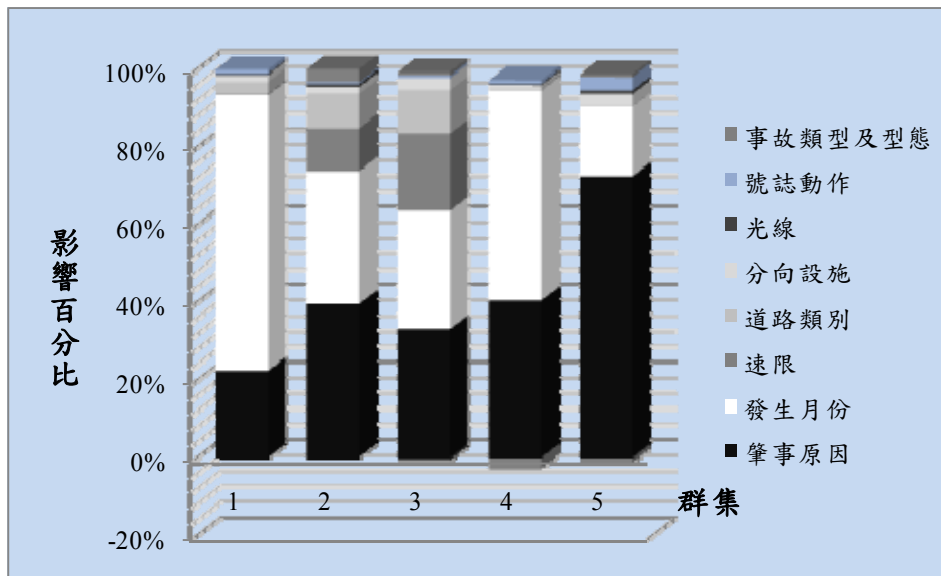


圖 5-12 93 年各群集肇事之變數所佔比例

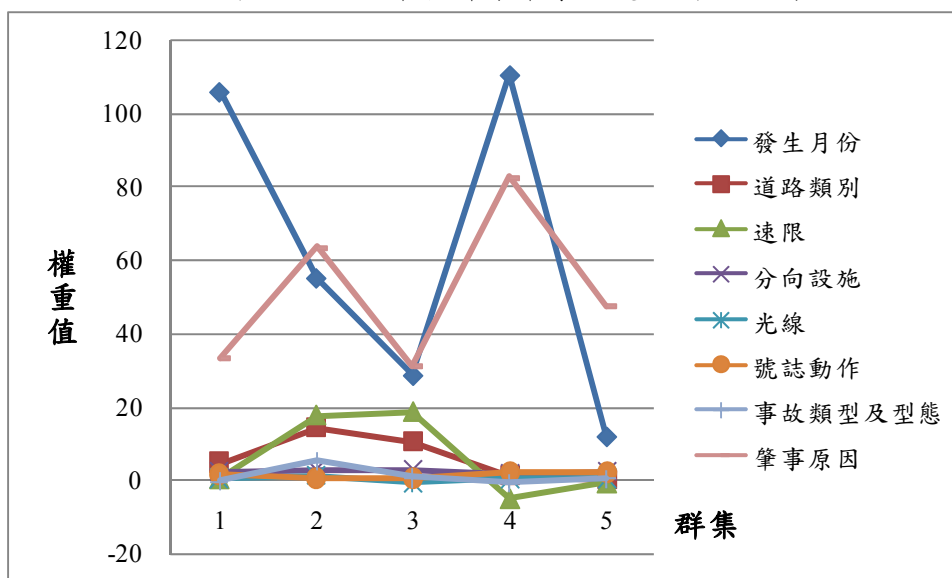


圖 5-13 93 年各群集肇事之變數權重值

### 三、94 年肇事資料

經過逐步判別分析得到影響 94 年肇事資料的判別函數之預測變數共有 29 項，分別為未依規定讓車、變換車道或方向不當、未保持行車安全距離、酒醉(後)駕駛失控、違反號誌管制或指揮、違反特定標誌(線)標制、其他駕駛人因素、非駕駛人因素、一到八月、速限小於 40 公里、速限 50-60 公里、速限 60-70 公里、速限大於 70 公里、國道、省道、中央分隔島、無分向設施、夜間無照明、無號誌、交流道、車與車、汽(機)車本身。

根據 94 年肇事資料判別分類函數與各肇事影響變數之關係，主要影響變數分別由肇事原因、發生月份、速限。圖 5-14 及圖 5-15 為不同群集下，判別函數之肇事變數影響程度百分比及權重值。

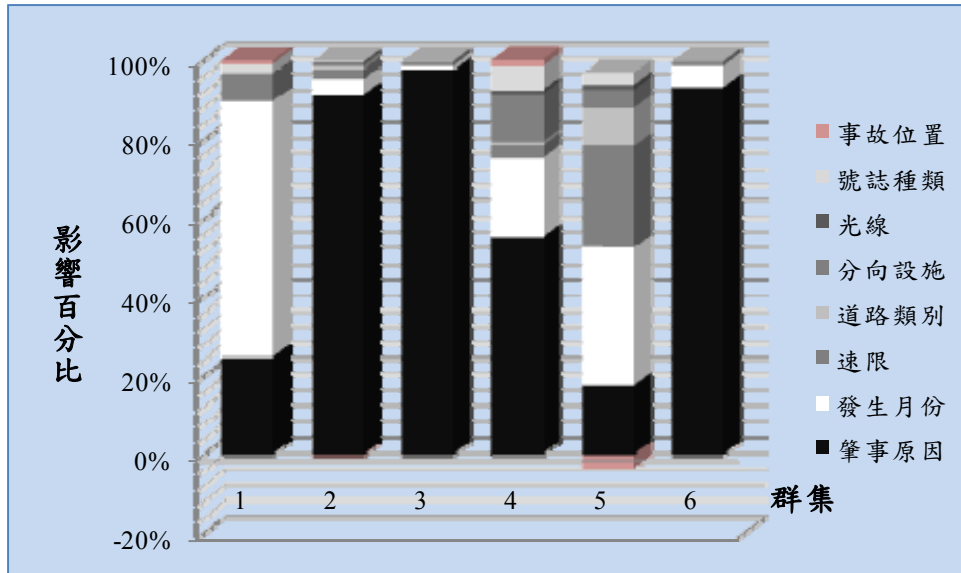


圖 5-14 94 各群集肇事之變數所佔比例

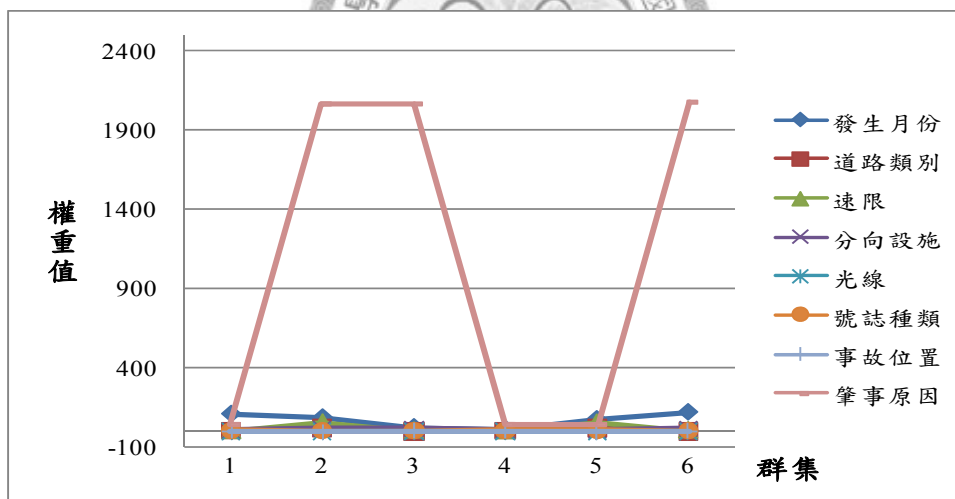


圖 5-15 94 年各群集肇事之變數權重值

#### 四、95 年肇事資料

經過逐步判別分析得到影響 95 年肇事資料的判別函數之預測變數共有 23 項，分別為變換車道或方向不當、未保持行車安全距離、酒醉(後)駕駛失控、違反號誌管制或指揮、違反特定標誌(線)標制、其他駕駛人因素、非駕駛人因素、一月到八月、速限 50-60 公里、速限 60-70 公里、速限大於 70 公里、國

道、省道、無分向設施、交流道、汽(機)車本身。

根據 95 年肇事資料判別分類函數與各肇事影響變數之關係，主要影響變數分別由肇事原因、發生月份、道路類別、速限。圖 5-16 及圖 5-17 為不同群集下，判別函數之肇事變數影響程度百分比及權重值。

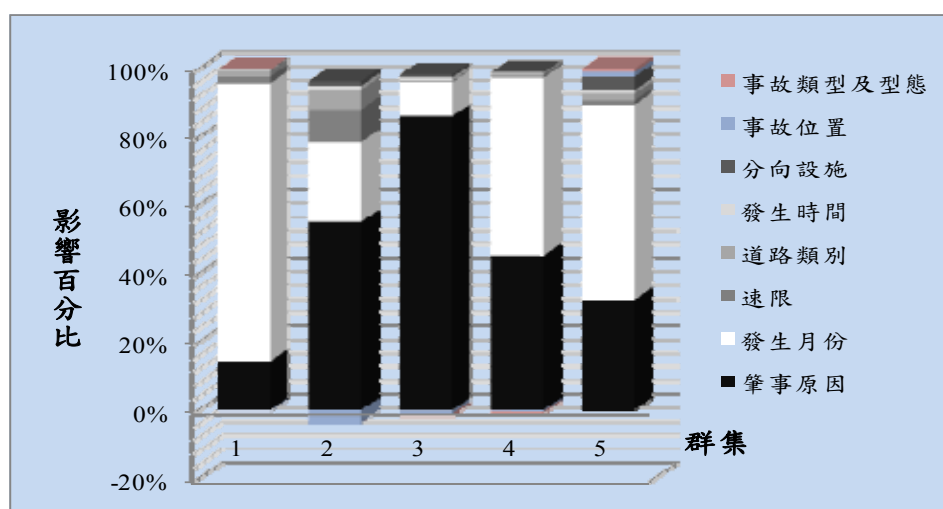


圖 5-16 95 年各群集肇事之變數所佔比例

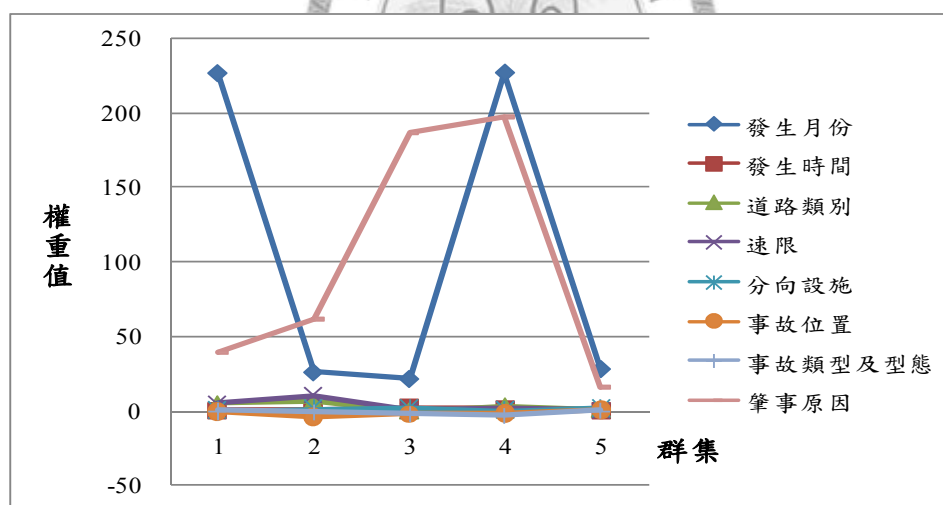


圖 5-17 95 年各群集肇事之變數權重值

## 五、96 年肇事資料

經過逐步判別分析得到影響 96 年肇事資料的判別函數之預測變數共有 25 項，分別為變換車道或方向不當、未保持行車安全距離、未注意車前狀態、酒醉(後)駕駛失控、違反號誌管制或指揮、違反特定標誌(線)標制、其他駕駛人因素、非駕駛人因素、一月到七月、速限 50-60 公里、速限 60-70 公里、速限大

於 70 公里、國道、中央分隔島、夜間無照明、有堆積物、行車管制號誌、號誌正常、交流道。

根據 96 年肇事資料判別分類函數與各肇事影響變數之關係，主要影響變數分別由肇事原因、發生月份、速限。圖 5-18 及圖 5-19 為不同群集下，判別函數之肇事變數影響程度百分比及權重值。

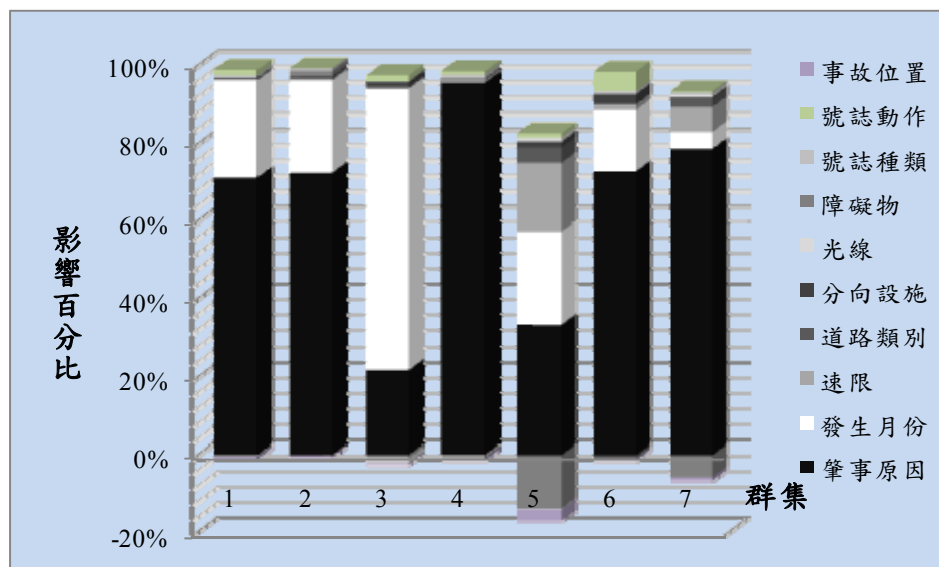


圖 5-18 96 年各群集肇事之變數所佔比例

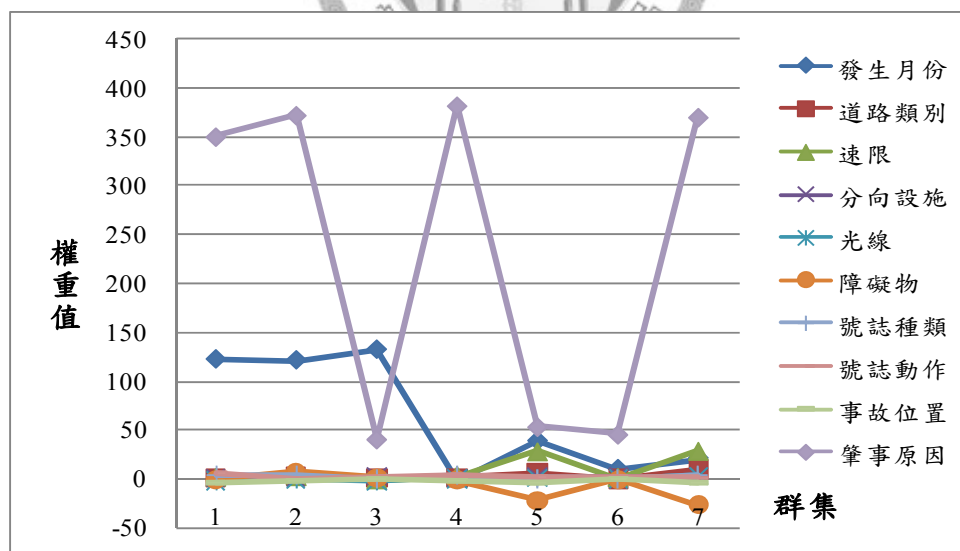


圖 5-19 96 年各群集肇事之變數權重值

## 六、97 年肇事資料

經過逐步判別分析得到影響 97 年肇事資料的判別函數之預測變數共有 24 項，分別為未依規定讓車、變換車道或方向不當、未保持行車安全距離、酒

醉(後)駕駛失控、違反號誌管制或指揮、違反特定標誌(線)標制、其他駕駛人因素、非駕駛人因素、一月到七月、速限 60-70 公里及大於 70 公里、國道、省道、6-9 時、中央分隔島、無分向設施、無號誌、汽(機)車本身。

根據 97 年肇事資料判別分類函數與各肇事影響變數之關係，主要影響變數分別由肇事原因、發生月份、分向設施、號誌種類。圖 5-20 及圖 5-21 為不同群集下，判別函數之肇事變數影響程度百分比及權重值。

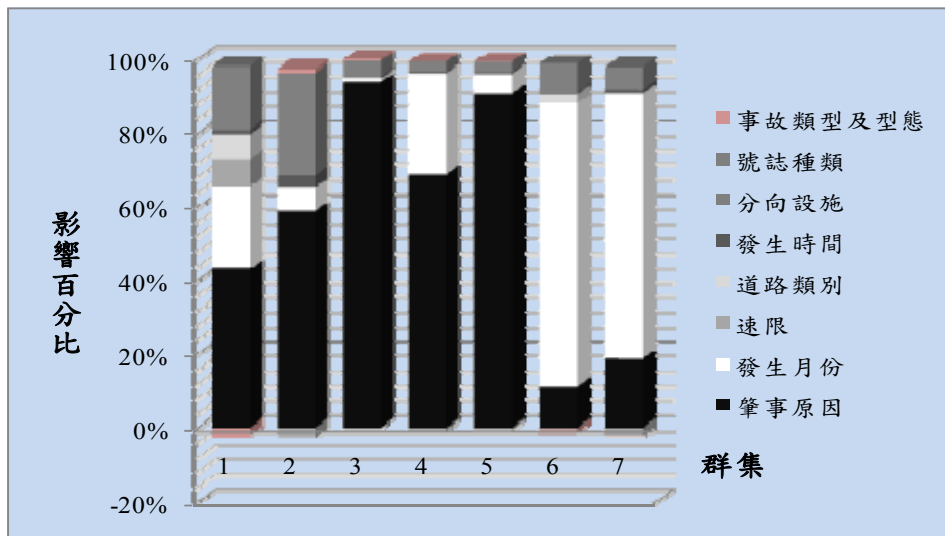


圖 5-20 97 年各群集肇事之變數所佔比例

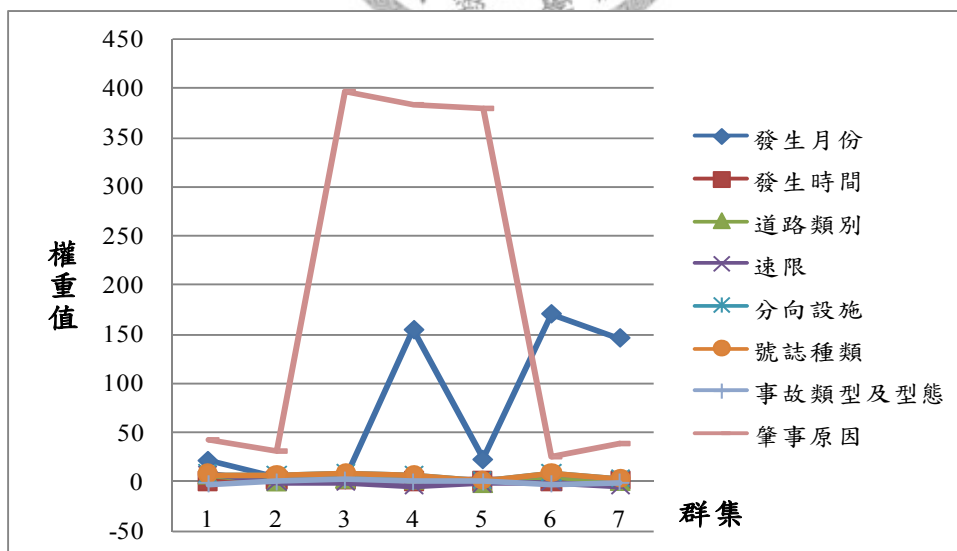


圖 5-21 97 年各群集肇事之變數權重值

## 七、92-97 年肇事資料

經過逐步判別分析得到影響總肇事資料的判別函數之預測變數共有 35

項，分別為未依規定讓車、變換車道或方向不當、未保持行車安全距離、酒醉(後)駕駛失控、違反號誌管制或指揮、違反特定標誌(線)標制、其他駕駛人因素、非駕駛人因素、五月到十二月、速限 40-50 公里、速限 50-60 公里、速限 60-70 公里、速限大於 70 公里、國道、省道、縣道、中央分隔島、無分向設施、禁止變換車道線(附標記)、快慢車道分隔線、暴雨、夜間無照明、水泥、油滑、無號誌、交流道、車與車、汽(機)車本身。

根據總肇事資料判別分類函數與各肇事影響變數之關係，主要影響變數分別由肇事原因、發生月份、速限、事故類型及型態。圖 5-22 及圖 5-23 為不同群集下，判別函數之肇事變數影響程度百分比及權重值。

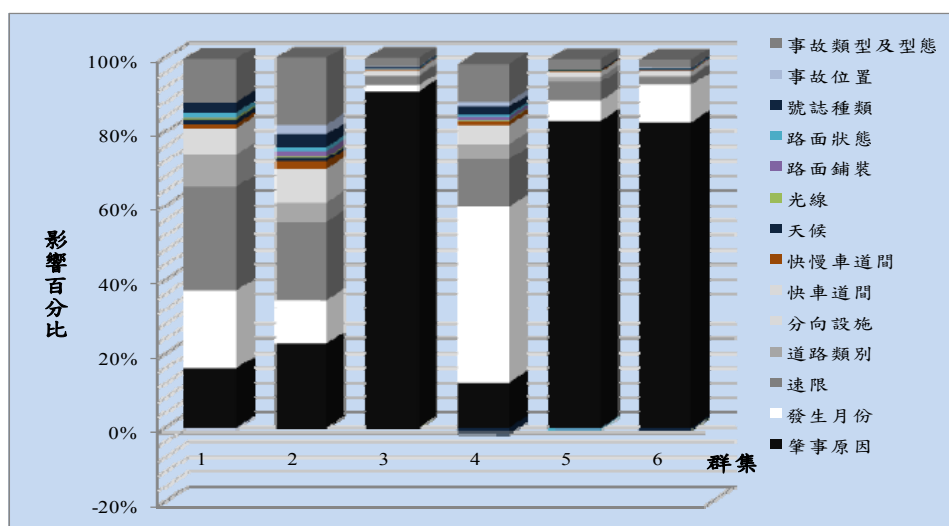


圖 5-22 92-97 年各群集肇事之變數所佔比例

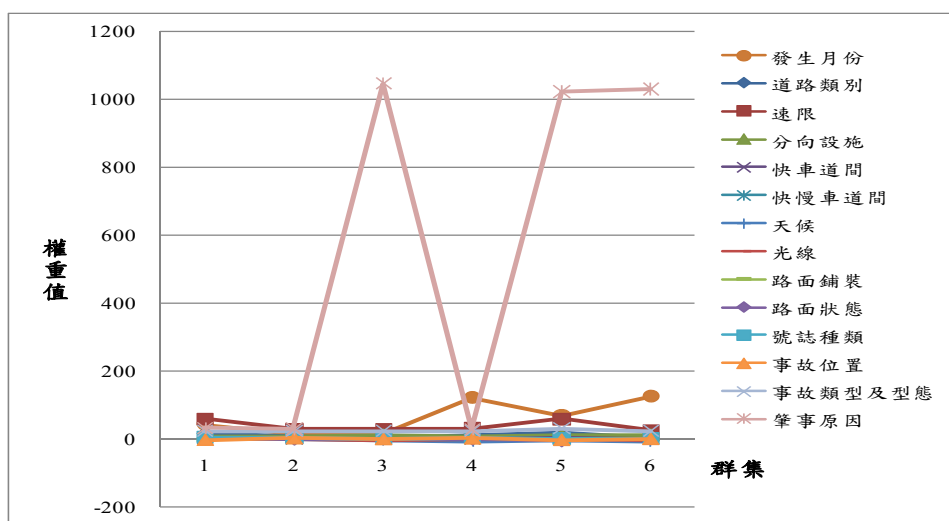


圖 5-23 92-97 年各群集肇事之變數權重值



由表 5-45 可知，在雪費函數加權值之比較中，發生月份、道路類別、速限、肇事原因在各年度皆會發生的肇事變數。又以肇事原因為影響各年度的主要變數且正相關；速限在 92 及 97 年呈微弱的負相關；事故位置在 94~96 年呈微弱的負相關；事故類型及型態在 95 及 97 年呈微弱的負相關；障礙物在只出現於 96 年且為負相關；其餘各年度出現不同的肇事變數皆為正相關。在 92-97 年中，從判別模式選取出快車道間、快慢車道間、天候、路面鋪裝、路面狀態在各年度所沒有的肇事變數，且呈微弱的正相關，其主要影響之變數為肇事原因。

表 5-45 費雪函數加權值比重

年度	92	93	94	95	96	97	92-97
發生月份	49.10%	47.34%	6.19%	49.37%	20.93%	26.90%	9.44%
發生時間				0.62%		0.49%	
道路類別	0.22%	4.65%	0.41%	1.50%	1.08%	0.77%	1.49%
速限	-0.22%	4.68%	1.89%	1.77%	2.90%	-0.28%	5.74%
分向設施		1.83%	1.06%	0.56%	0.47%	3.05%	1.63%
快車道間							0.48%
快慢車道間							0.36%
天候							-0.13%
光線		0.34%	0.07%		0.09%		0.04%
路面鋪裝							0.16%
障礙物					-2.04%		
路面狀態							0.17%
號誌種類	1.32%		0.32%		0.58%	2.28%	0.64%
號誌動作		1.08%			0.93%		
事故位置			-0.16%	-0.77%	-0.99%		0.24%
事故類型及型態	10.42%	0.87%		-0.38%		-0.06%	3.85%
肇事原因	39.16%	39.21%	90.23%	47.34%	76.06%	66.85%	75.90%

因為肇事原因在模式中的影響程度很高，且大客車發生肇事原因不外乎未依規定讓車、變換車道或方向不當、未保持行車安全距離、未注意車前狀態、酒醉(後)駕駛失控、違反號誌管制或指揮、違反特定標誌(線)標制、其他駕駛人因素(如超速失控、使用手持行動電話失控、違規停車或暫停不當、其他引起事故之違規或不當行為等)、非駕駛人因素(如煞車失靈、車輪脫落或輪胎爆裂、其他引起事故之故障、交通管制設施失靈或損毀等)。所以由表 5-46 可知，違反號誌管制或指揮、違反特定標誌(線)標制、其他駕駛人因素及非駕駛人因素都是影響各年度重要的預測變數約 16~23%；酒醉(後)駕駛失控 93 及 95 年約佔 11%、未依規定讓車在 93 年 12%。未依規定讓車只出現三年，分別為 93、94 及 97 年，但 94 及 97 年的影



響程度不大；未注意前車狀態只出現在 96 年。

表 5-46 肇事原因加權值之比重

	92年	93年	94年	95年	96年	97年	92-97年
未依規定讓車		12.20%	0.21%			2.59%	0.56%
變換車道或方向不當	3.91%	7.31%	0.27%	4.93%	2.31%	2.20%	0.45%
未保持行車安全距離	4.84%	5.80%	0.27%	5.26%	2.12%	2.89%	0.43%
未注意車前狀態					3.60%		
酒醉(後)駕駛失控	7.45%	10.20%	1.78%	11.65%	5.71%	7.25%	4.00%
違反號誌管制或指揮	21.83%	12.44%	24.16%	19.49%	20.55%	18.93%	23.43%
違反特定標誌(線)標制	19.10%	16.29%	24.38%	20.19%	22.03%	22.39%	23.58%
其他駕駛人因素	22.87%	16.69%	24.40%	19.84%	21.60%	22.00%	23.78%
非駕駛人因素	19.99%	19.07%	24.54%	18.63%	22.06%	21.73%	23.77%

1. 在 92 年的肇事因素裡，發生於一到六月、車與車及車本身的事故類型、無號誌、省道時為正相關，而速限高於 70 公里時為負相關。
2. 在 93 年的肇事因素裡，發生於六到十二月、國道及省道、速限高於 50 公里、中央分隔島、號誌動作正常、車本身的事故類型、夜間無照明時為正相關。
3. 在 94 年的肇事因素裡，發生於一到八月、速限高於 40 公里、無分向設施、國道及省道、夜間無照明、無號誌為正相關，而交流道時為負相關。
4. 在 95 年的肇事因素裡，發生於一到八月、速限高於 50 公里、省道、21-24 時、無分向設施為正相關，而事故位置在交流道及車本身的事故類型為負相關。
5. 在 96 年的肇事因素裡，發生於一到七月、速限高於 70 公里、省道、號誌正常、行車管制號誌、中央分隔島、夜間無照明為正相關，而有堆積物及事故位置在交流道為負相關。
6. 在 97 年的肇事因素裡，發生於一到七月、中央分隔島及無分向設施、無號誌、省道、6-9 時為正相關，而速限高於 60 公里及車本身的事故類型為負相關。
7. 在 92-97 年的肇事因素裡，發生於五到十二月、速限高於 40 公里、車與車及車本身的事故型態、中央分隔島及無分向設施、國道、省道、縣道、無號誌、禁止變換車道線、快慢車道分隔線、事故位置在交流道、路面為油滑及柏油、夜間無照明為正相關，而暴雨的天氣為負相關。

## 第六章 案例分析

前述之群集分析法的分群難以命名，因此本章加入了因子分析法，再套入群集及判別分析，找出最佳分類及判別率，並把各因子予以命名；再由死傷人數及因子數套入迴歸分析其各因子對死傷人數之顯著性。而先用 92-97 共六年間的大客車來分析，之後挑選出臺北市及南投縣分析比較。

### 6.1 臺灣地區

大客車臺灣地區的樣本數共有 15514 筆，利用主成份法進行因素分析之參數估計，並配合直交轉軸之最大變異法進行轉軸，而共通因子數目的取決方面，選取標準將配合 Kaiser 準則(特徵值大於 1 者)及陡坡考驗法，決定初步選取的共通因子數目，最後套入判別分析，選擇最佳的共通因子數及分群方式。其初始特徵值及因素陡坡圖，而因子所主要代表之變數，將於轉軸後之結果說明如表 6-1 所示。

表 6-1 臺灣地區解說總變異量表

成份	初始特徵值		
	總和	變異數的百分比	累積百分比
1	4.021	21.161	21.161
2	1.907	10.036	31.197
3	1.551	8.162	39.359
4	1.433	7.544	46.903
5	1.267	6.668	53.571
6	1.192	6.276	59.847
7	1.043	5.490	65.337
8	0.999	5.262	70.599
9	0.975	5.132	75.731
10	0.914	4.813	80.544
11	0.834	4.388	84.932
12	0.805	4.237	89.169
13	0.527	2.774	91.943
14	0.492	2.588	94.531
15	0.400	2.107	96.638
16	0.230	1.211	97.849
17	0.187	0.986	98.835
18	0.169	0.890	99.725
19	0.052	0.275	100.000

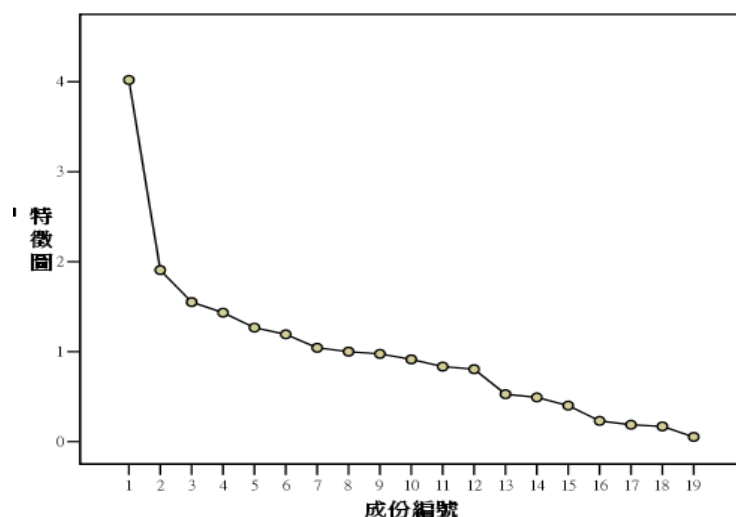


圖 6-1 臺灣地區因素陡坡圖

由上列之解說總變異量表可看出，第 1-7 個成份之特徵值都大於 1，而累積至第 7 個成份所解釋的變異約有 65%。透過 Kaiser 準則(特徵值大於 1)，因子數目在 7 個以內皆屬合理。另外，對應之特徵值繪製因素陡坡圖，圖中特徵值大於 1 的因子從第 3 個開始平坦，透過陡坡圖考驗法得知，因子數目為 3-7 時，為較適當之因子數目。因此，分別取出 3-7 個因子，然後分成 3-10 群集數，套入判別分析，比較其判別正確率；判別正確率高者即為選定的因子及分群數目。如表 6-2 所示：

表 6-2 臺灣地區各因子下分群數之判別正確率

分群數目	3 因子	4 因子	5 因子	6 因子	7 因子
3	44.1%	68.2%	50.2%	54.9%	47.0%
4	68.3%	52.1%	35.1%	50.3%	48.5%
5	70.2%	58.6%	31.6%	48.4%	54.9%
6	76.3%	62.9%	42.2%	54.4%	47.2%
7	57.9%	55.0%	43.3%	50.9%	46.8%
8	69.9%	49.5%	42.0%	47.6%	55.2%
9	60.0%	53.4%	52.0%	44.9%	49.4%
10	72.9%	68.7%	44.8%	53.5%	55.3%

由上表可知，在 3 因子時，整體判別正確率都高於其他因子，所以選取 3 因子，而又從 3 因子中分 6 群時的判別正確率最高，因此選定 3 因子分 6 群。

由於選定 3 因子分 6 群後，表 6-3 為 3 因子轉軸後因子負荷矩陣，如下所示：

表 6-3 臺灣地區 3 因子轉軸後因子負荷矩陣

肇事變數	因子構面一	因子構面二	因子構面三	共通性
號誌動作	0.906			0.063
號誌種類	0.893			0.050
道路型態	0.824			0.868
事故位置	0.673			0.115
障礙物				0.678
速限		-0.783		0.655
道路類別		0.780		0.774
快車道間		0.756		0.014
分向設施		0.633		0.505
視距				0.618
路面鋪裝				0.004
路面缺陷				0.014
天候			0.927	0.854
路面狀態			0.920	0.033
光線				0.099
發生時間				0.805
發生地區				0.836
快慢車道間				0.490
事故類型及型態				0.003
特徵值	4.021	1.907	1.551	
變異數的百分比	21.161	10.036	8.162	
累積百分比	21.161	31.197	39.359	
<b>KMO 為 0.690</b>				

註：本表僅列出因子負荷值大於 0.5 者

由表 6-3 知，KMO 大於 0.5 是可以作因素分析，所以因子構面一主要包含的變數為號誌動作、號誌種類、道路型態、事故位置，其特徵值為 4.201，所能解釋總變異的百分比為 21.161；因子構面二主要包含的變數為速限、道路類別、快車道間、分向設施，其特徵值為 1.907，所能解釋總變異的百分比為 10.036；因子構

面三主要包含的變數為天候、路面狀態，其特徵值為 1.551，所能解釋總變異的百分比為 8.162。分別把因子構面一至三命名為行車環境、道路特性、環境特性。

表 6-4 為群集分析分 6 群後的最後群集中心點，各分群之樣本數、死亡人數、受傷人數，如下所示：

表 6-4 臺灣地區 3 因子分 6 群之構面

	群集 1	群集 2	群集 3	群集 4	群集 5	群集 6
行車環境	0.6857	0.8478	-0.4258	-1.3300	0.5084	-1.3474
道路特性	0.6793	0.9514	-0.2881	0.6176	-1.3118	-0.6051
環境特性	0.4735	-2.1610	-2.3006	0.4038	0.3036	0.2592
樣本數	5406	929	1349	2518	3549	1763
死亡人數	324	62	37	111	87	54
受傷人數	6719	1205	1166	2897	1355	2022

臺灣地區大客車肇事案件由 3 因子並分成 6 群之群集分析的結構，構建群集之事故嚴重程度迴歸預測模式，了解影響事故嚴重程度的因子，藉以對發生大客車之肇事特性有進一步的認知。

死亡人數和傷亡人數為因變數，行車環境、道路特性、環境特性為自變數，套入迴歸模式，其結果如表 6-5 及表 6-6 所示：

表 6-5 臺灣地區死亡人數之迴歸模式

	平方和	自由度	平均平方	F 值	P 值	顯著性
迴歸	5.525	3	1.842	20.849	0.000	**
殘差	1370.106	15510	0.088			
總和	1375.631	15513				
	β 之估計值			t 值	P 值	顯著性
(常數)	0.044			18.234	0.000	**
行車環境	0.009			3.588	0.000	**
道路特性	0.017			7.039	0.000	**
環境特性	-0.001			-0.356	0.722	—

註：「\*」為  $0.01 \leq P \text{ 值} \leq 0.05$ ，「\*\*」為  $P \text{ 值} < 0.01$ 。

迴歸模式中，死亡人數在環境特性之因子為不顯著，而環境特性之肇事變數為天候及路面狀態。代表大客車肇事發生時，天候及路面狀態不是影響肇事之原

因，而行車環境及道路特性為影響肇事之因子。

表 6-6 臺灣地區受傷人數之迴歸模式

	平方和	自由度	平均平方	F 值	P 值	顯著性
迴歸	1868.876	3	622.959	234.530	0.000	**
殘差	41197.673	15510	2.656			
總和	43066.549	15513				
	$\beta$ 之估計值			t 值	P 值	顯著性
(常數)	0.990			75.685	0.000	**
行車環境	-0.075			-5.745	0.000	**
道路特性	0.339			25.895	0.000	**
環境特性	0.002			0.115	0.909	—

註：「\*」為  $0.01 \leq P \text{ 值} \leq 0.05$ ，「\*\*」為  $P \text{ 值} < 0.01$ 。

迴歸模式中，受傷人數在環境特性之因子為不顯著，而環境特性之肇事變數為天候及路面狀態。代表大客車肇事發生時，天候及路面狀態不是影響肇事之原因，而行車環境及道路特性為影響肇事之原因。

由傷亡人數之迴歸模式得知，環境特性之變數：天候及路面狀態不為影響大客車肇事發生的原因；行車環境：號誌動作、號誌種類、道路型態、事故位置及道路特性：速限、道路類別、快車道間、分向設施為影響大客車肇事之因素。

## 6.2 臺北市

臺北市大客車肇事案件共有 2567 筆，利用主成份法進行因素分析之參數估計，並配合直交轉軸之最大變異法進行轉軸，而共通因子數目的取決方面，選取標準將配合 Kaiser 準則(特徵值大於 1 者)及陡坡考驗法，決定初步選取的共通因子數目，最後套入判別分析，選擇最佳的共通因子數及分群方式。其初始特徵值及因素陡坡圖，而因子所主要代表之變數，將於轉軸後之結果說明如表 6-7 所示。

表 6-7 臺北市解說總變異量表

成份	初始特徵值		
	總和	變異數的百分比	累積百分比
1	3.039	16.886	16.886
2	1.979	10.995	27.881
3	1.817	10.096	37.977
4	1.500	8.331	46.308
5	1.315	7.307	53.615
6	1.205	6.692	60.307
7	1.113	6.182	66.489
8	1.024	5.688	72.177
9	0.976	5.422	77.599
10	0.963	5.351	82.949
11	0.813	4.515	87.464
12	0.735	4.081	91.545
13	0.458	2.543	94.088
14	0.350	1.942	96.030
15	0.284	1.580	97.610
16	0.229	1.274	98.884
17	0.176	0.980	99.864
18	0.024	0.136	100.000

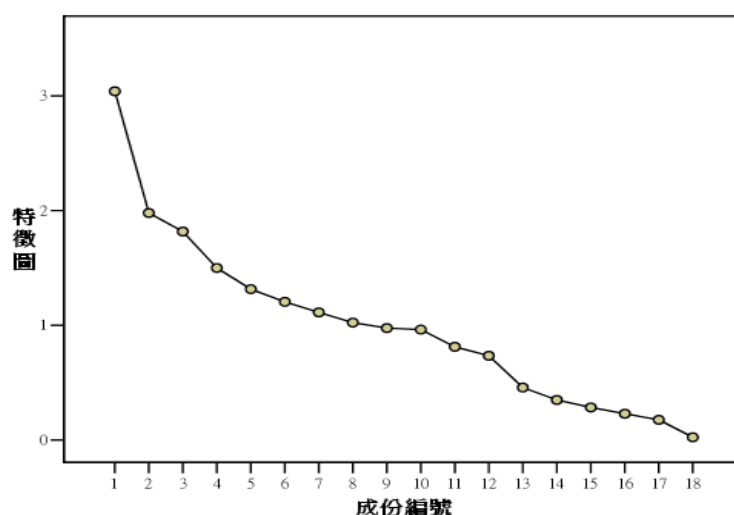


圖 6-2 臺北市因素陡坡圖

由上列之解說總變異量表可看出，第 1-8 個成份之特徵值都大於 1，而累積至第 8 個成份所解釋的變異約有 72%。透過 Kaiser 準則(特徵值大於 1)，因子數目在 8 個以內皆屬合理。另外，對應之特徵值繪製因素陡坡圖，圖中特徵值大於 1 的因子從第 4 個開始平坦，透過陡坡圖考驗法得知，因子數目為 4-8 時，為較適當之因子數目。因此，分別取出 4-8 個因子，然後分成 3-10 群集數，套入判別分析，比較其判別正確率；判別正確率高者即為選定的因子及分群數目。如表 6-8 所示：

表 6-8 臺北市在各因子下分群數之判別正確率

分群數目	4 因子	5 因子	6 因子	7 因子	8 因子
3	40.0%	45.8%	50.5%	55.3%	53.4%
4	29.3%	57.1%	63.8%	62.1%	53.7%
5	55.2%	62.0%	50.3%	71.8%	62.1%
6	61.7%	46.1%	66.3%	63.0%	63.3%
7	54.4%	46.2%	50.6%	75.8%	63.0%
8	52.9%	45.1%	64.1%	55.2%	62.7%
9	56.4%	53.4%	47.5%	60.6%	47.5%
10	66.5%	53.1%	50.6%	54.1%	62.8%

由上表可知，在 7 因子時，整體判別正確率都高於其他因子，所以選取 7 因子，而又從 7 因子中分 7 群時的判別正確率最高，因此選定 7 因子分 7 群。



由於選定 7 因子分 7 群後，表 6-9 為 7 因子轉軸後因子負荷矩陣，如下所示：

表 6-9 臺北市 7 因子轉軸後因子負荷矩陣

肇事 變數	因子 構面一	因子 構面二	因子 構面三	因子 構面四	因子 構面五	因子 構面六	因子 構面七	共通性
號誌動作	0.933							0.828
號誌種類	0.925							0.906
道路型態	0.823							0.824
事故位置	0.632							0.833
路面狀態		0.951						0.753
天候		0.947						0.784
道路類別			-0.869					0.277
速限			0.835					0.677
事故類型								0.776
發生時間				0.909				0.774
光線				0.902				0.135
快慢車道間					0.822			0.908
快車道間					0.783			0.602
障礙物						0.773		0.337
路面缺陷						0.509		0.903
視距						0.500		0.910
路面鋪裝								0.686
分向設施							0.800	0.056
特徵值	3.039	1.979	1.817	1.500	1.315	1.205	1.113	
變異數%	16.886	10.995	10.096	8.331	7.307	6.692	6.182	
累積%	16.886	27.881	37.977	46.308	53.615	60.307	66.489	

KMO 為 0.582

註：本表僅列出因子負荷值大於 0.5 者

由表 6-9 知，KMO 大於 0.5 是可以作因素分析，所以因子構面一主要包含的變數為號誌動作、號誌種類、道路型態、事故位置，其特徵值為 3.309，所能解釋總變異的百分比為 16.886；因子構面二主要包含的變數為路面狀態、天候，其特徵值為 1.979，所能解釋總變異的百分比為 10.995；因子構面三主要包含的變數為道路類別、速限，其特徵值為 1.817，所能解釋總變異的百分比為 10.096；因子構面四主要包含的變數為發生時間、光線，其特徵值為 1.500，所能解釋總變異的百分比為 8.311；因子構面五主要包含的變數為快慢車道間、快車道間，其特徵值為

1.315，所能解釋總變異的百分比為 7.307；因子構面六主要包含的變數為障礙物、路面缺陷、視距，其特徵值為 1.205，所能解釋總變異的百分比為 6.692；因子構面七主要包含的變數為分向設施，其特徵值為 1.113，所能解釋總變異的百分比為 6.182。分別把因子構面一至七命名為行車環境、環境特性、道路特性、時間特性、車道特性、道路障礙、分向設施。

表 6-10 為群集分析分 7 群後的最後群集中心點，各分群之樣本數、死亡人數、受傷人數，如下所示：

表 6-10 臺北市 7 因子分 7 群之構面

	群集 1	群集 2	群集 3	群集 4	群集 5	群集 6	群集 7
行車環境	-0.2119	0.2414	-0.0157	-0.8228	-0.0969	0.7942	0.0217
環境特性	0.0436	-0.0198	-0.0034	1.0380	0.0243	-0.1632	-0.1041
道路特性	-0.4460	-0.5708	-0.1814	1.6149	-0.0650	3.7637	-0.0214
時間特性	0.0103	0.2003	0.0102	-0.6075	-0.0870	0.0373	-0.1813
車道特性	-2.1221	0.1388	0.2604	1.7627	0.2906	-0.1293	-0.6176
道路障礙	0.1675	-3.0643	0.1936	-4.7939	0.3032	0.0440	-8.0699
分向設施	-0.2072	0.7792	-0.4278	-0.3928	1.5882	-0.2237	-0.4410
樣本數	259	109	1609	8	438	129	15
死亡人數	12	6	42	1	10	1	2
受傷人數	354	136	2220	8	536	34	21

臺北市 6 年大客車肇事案件由 7 因子並分成 7 群之群集分析的結構，構建群集之事故嚴重程度迴歸預測模式，了解影響各事故嚴重程度的因子，藉以對發生大客車之肇事特性有進一步的認知。

死亡人數和傷亡人數為因變數，行車環境、環境特性、道路特性、時間特性、車道特性、道路障礙、分向設施為自變數，套入迴歸模式，其結果如表 6-11 及表 6-12 所示：

表 6-11 臺北市死亡人數之迴歸模式

	平方和	自由度	平均平方	F 值	P 值	顯著性
迴歸	0.700	7	0.100	1.952	0.058	—
殘差	131.167	2559	0.051			
總和	131.867	2566				
	$\beta$ 之估計值			t 值	P 值	顯著性
(常數)	0.029			6.451	0.000	**
行車環境	0.001			0.251	0.802	—
環境特性	-0.002			-0.380	0.704	—
道路特性	-0.009			-1.975	0.048	*
時間特性	0.010			2.242	0.025	*
車道特性	0.001			0.214	0.830	—
道路障礙	-0.009			-2.031	0.042	*
分向設施	-0.003			-0.596	0.551	—

註：「\*」為  $0.01 \leq P \text{ 值} \leq 0.05$ ，「\*\*」為  $P \text{ 值} < 0.01$ 。

迴歸模式中，死亡人數整體的預測模式為不顯著。在行車環境、環境特性、車道特性、分向設施之因子為不顯著，其變數有號誌動作、號誌種類、道路型態、事故位置、路面狀態、天候、快慢車道間、快車道間、分向設施；道路特性、時間特性、道路障礙之因子為顯著，其變數有道路類別、速限、發生時間、光線、障礙物、路面缺陷、視距。

表 6-12 臺北市受傷人數之迴歸模式

	平方和	自由度	平均平方	F 值	P 值	顯著性
迴歸	217.625	7	31.089	12.435	0.000	**
殘差	6397.898	2559	2.500			
總和	6615.522	2566				
	$\beta$ 之估計值			t 值	P 值	顯著性
(常數)	1.289			41.305	0.000	**
行車環境	0.003			0.101	0.920	—
環境特性	0.025			0.785	0.432	—
道路特性	-0.280			-8.975	0.000	**
時間特性	-0.025			-0.801	0.423	—
車道特性	0.026			0.845	0.398	—
道路障礙	-0.001			-0.021	0.983	—
分向設施	-0.066			-2.123	0.034	*

註：「\*」為  $0.01 \leq P \text{ 值} \leq 0.05$ ，「\*\*」為  $P \text{ 值} < 0.01$ 。

迴歸模式中，受傷人數在行車環境、環境特性、時間特性、車道特性、道路障礙之因子為不顯著，而道路特性及分向設施為顯著之因子。代表大客車發生肇事會受傷之影響變數為道路類別、速限及分向設施。

由傷亡人數之迴歸模式得知，道路特性對死亡和受傷皆影響，其變數為道路類別及速限，代表在臺北市的道路裡大客車都是因為速度而發生肇事。行車環境、環境特性及車道環境對傷亡皆沒有顯著的影響，其變數為號誌動作、號誌種類、道路型態、事故位置、路面狀態、天候、快慢車道間、快車道間。

### 6.3 南投縣

南投縣大客車肇事案件共有 418 筆，利用主成份法進行因素分析之參數估計，並配合直交轉軸之最大變異法進行轉軸，而共通因子數目的取決方面，選取標準將配合 Kaiser 準則(特徵值大於 1 者)及陡坡考驗法，決定初步選取的共通因子數目，最後套入判別分析，選擇最佳的共通因子數及分群方式。其初始特徵值及因素陡坡圖，而因子所主要代表之變數，將於轉軸後之結果說明如表 6-13 所示。

表 6-13 南投縣解說總變異量表

成份	初始特徵值		
	總和	變異數的百分比	累積百分比
1	3.254	18.080	18.080
2	2.141	11.892	29.972
3	1.742	9.680	39.652
4	1.574	8.747	48.399
5	1.390	7.723	56.121
6	1.277	7.095	63.216
7	1.007	5.593	68.809
8	0.976	5.422	74.232
9	0.931	5.171	79.402
10	0.779	4.330	83.733
11	0.622	3.453	87.186
12	0.526	2.924	90.110
13	0.490	2.725	92.834
14	0.414	2.299	95.134
15	0.383	2.130	97.263
16	0.217	1.207	98.470
17	0.181	1.003	99.474
18	0.095	0.526	100.000

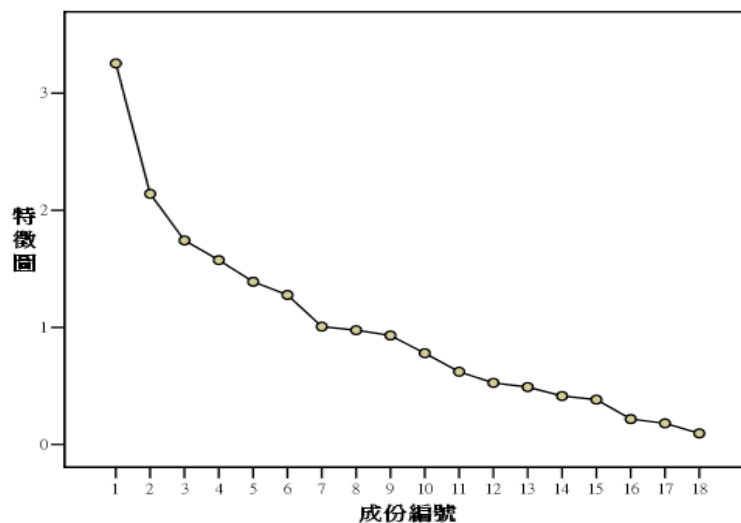


圖 6-3 南投縣因素陡坡圖

由上列之解說總變異量表可看出，第 1-7 個成份之特徵值都大於 1，而累積至第 7 個成份所解釋的變異約有 69%。透過 Kaiser 準則(特徵值大於 1)，因子數目在 7 個以內皆屬合理。另外，對應之特徵值繪製因素陡坡圖，圖中特徵值大於 1 的因子從第 3 個開始平坦，透過陡坡圖考驗法得知，因子數目為 3-7 時，為較適當之因子數目。因此，分別取出 3-7 個因子，然後分成 3-10 群集數，套入判別分析，比較其判別正確率；判別正確率高者即為選定的因子及分群數目。如表 6-14 所示：

表 6-14 南投縣在各因子下分群數之判別正確率

分群數目	3 因子	4 因子	5 因子	6 因子	7 因子
3	56.7%	51.7%	46.7%	44.7%	53.6%
4	65.1%	45.9%	43.8%	42.6%	51.9%
5	78.8%	49.3%	44.7%	43.3%	42.3%
6	57.7%	56.2%	43.5%	54.5%	51.0%
7	57.2%	45.9%	30.1%	43.8%	46.2%
8	64.1%	41.9%	42.3%	43.3%	43.5%
9	66.7%	50.7%	47.4%	40.2%	34.4%
10	64.4%	55.5%	35.4%	35.9%	47.4%

由上表可知，在 3 因子時，整體判別正確率都高於其他因子，所以選取 3 因子，而又從 3 因子中分 5 群時的判別正確率最高，因此選定 3 因子分 5 群。

由於選定 3 因子分 5 群後，表 6-15 為 3 因子轉軸後因子負荷矩陣，如下所示：

表 6-15 南投縣 3 因子轉軸後因子負荷矩陣

肇事變數	因子構面一	因子構面二	因子構面三	共通性
號誌動作	0.889			0.025
道路型態	0.858			0.784
號誌種類	0.858			0.097
事故位置	0.744			0.290
視距				0.471
路面鋪裝				0.765
事故類型及型態				0.150
速限		-0.682		0.510
快車道間		0.655		0.465
分向設施		0.637		0.223
道路類別		0.526		0.054
快慢車道間		0.505		0.821
路面缺陷				0.172
路面狀態			0.856	0.190
天候			0.815	0.742
障礙物				0.794
光線				0.568
發生時間				0.016
特徵值	3.254	2.141	1.742	
變異數的百分比	18.080	11.892	9.680	
累積百分比	18.080	29.972	39.652	
KMO 為 0.622				

註：本表僅列出因子負荷值大於 0.5 者

由表 6-15 知，KMO 大於 0.5 是可以作因素分析，所以因子構面一主要包含的變數為號誌動作、道路型態、號誌種類、事故位置，其特徵值為 3.254，所能解釋總變異的百分比為 18.080；因子構面二主要包含的變數為速限、快車道間、分向設施、道路類別、快慢車道間，其特徵值為 2.141，所能解釋總變異的百分比為 11.892；因子構面三主要包含的變數為路面狀態、天候，其特徵值為 1.742，所能

解釋總變異的百分比為 9.680。分別把因子構面一至三命名為行車環境、道路特性、環境特性。

表 6-16 為群集分析分 7 群後的最後群集中心點，各分群之樣本數、死亡人數、受傷人數，如下所示：

表 6-16 南投縣 3 因子分 5 群之構面

	群集 1	群集 2	群集 3	群集 4	群集 5
行車環境	0.6601	0.5048	0.5809	-1.2726	0.2059
道路特性	0.7865	-0.8268	2.2606	0.1670	-0.8503
環境特性	0.2657	0.6897	-2.8009	0.2021	-2.1039
樣本數	129	107	6	121	55
死亡人數	21	4	0	5	2
受傷人數	275	146	13	140	149

南投縣 6 年大客車肇事案件由 3 因子並分成 5 群之群集分析的結構，構建群集之事故嚴重程度迴歸預測模式，了解影響各事故嚴重程度的因子，藉以對發生大客車之肇事特性有進一步的認知。

死亡人數和傷亡人數為因變數，行車環境、道路特性、環境特性為自變數，套入迴歸模式，其結果如表 6-17 及表 6-18 所示：

表 6-17 南投縣死亡人數之迴歸模式

	平方和	自由度	平均平方	F 值	P 值	顯著性
迴歸	0.852	3	0.284	1.818	0.143	—
殘差	64.698	414	0.156			
總和	65.550	417				
	β 之估計值			t 值	P 值	顯著性
(常數)	0.077			3.959	0.000	**
行車環境	0.034			1.739	0.083	—
道路特性	0.023			1.167	0.244	—
環境特性	0.020			1.034	0.302	—

註：「\*」為  $0.01 \leq P \text{ 值} \leq 0.05$ ，「\*\*」為  $P \text{ 值} < 0.01$ 。

迴歸模式中，死亡人數整體預測模式為不顯著。在行車環境、道路特性、環境特性之因子皆為不顯著，代表大客車發生肇事變數對死亡沒有很明顯的原因。



表 6-18 南投縣受傷人數之迴歸模式

	平方和	自由度	平均平方	F 值	P 值	顯著性
迴歸	127.786	3	42.595	4.024	0.008	**
殘差	4382.666	414	10.586			
總和	4510.452	417				
	$\beta$ 之估計值			t 值	P 值	顯著性
(常數)	1.730			10.869	0.000	**
行車環境	0.335			2.103	0.036	*
道路特性	0.004			0.023	0.982	—
環境特性	-0.441			-2.766	0.006	**

註：「\*」為  $0.01 \leq P \text{ 值} \leq 0.05$ ，「\*\*」為  $P \text{ 值} < 0.01$ 。

迴歸模式中，受傷人數在道路特性之因子為不顯著，而道路特性之變數為速限、快車道間、分向設施、道路類別、快慢車道間。代表大客車發生肇事時，速限、快車道間、分向設施、道路類別、快慢車道間為是影響肇事之原因，而行車環境及環境特性為影響肇事之原因。

而由傷亡人數之迴歸模式得知，死亡沒有顯著的因子影響，而受傷為行車環境及環境特性所影響，其影響大客車肇事變數有號誌動作、道路型態、號誌種類、事故位置、路面狀態、天候。

## 6.4 案例比較

由臺灣地區整體來作因子分析、群集分析、判別分析與迴歸分析的整體預測模式，之後用一樣的方法再挑選出比較偏都市型態的臺北市及偏山區型態的南投縣作比較，其結果彙整如表 6-19 所示：

表 6-19 案例分析之彙整

		臺灣地區	臺北市	南投縣
KMO 檢定		0.690	0.582	0.622
因子數目		3	7	3
因子命名		行車環境 道路特性 環境特性	行車環境、環境特性、 道路特性、時間特性、 車道特性、道路障礙、 分向設施	行車環境 道路特性 環境特性
分群數目		6	7	5
判別正確率		76.3%	75.8%	78.8%
死亡	整體迴歸	顯著	不顯著	
受傷	預測模式		顯著	
死亡	顯著因子	行車環境 道路特性	道路特性、時間特性、 道路障礙	—
受傷			道路特性、分向設施	行車環境、環境特性
死亡	不顯著因子	環境特性	行車環境、環境特性、 車道特性、分向設施	行車環境、道路特 性、環境特性
受傷			行車環境、環境特性、 車道特性、時間特性、 道路障礙	道路特性

KMO 檢定都達 0.5 以上，代表適合作因子分析。而因子數目臺灣地區和南投縣為 3 因子，臺北市為 7 因子，本研究推測可能全臺發生大客車肇事的型態和南投縣類似，因此分成同樣的因子數目；臺北市因為為都市地區，所以肇事發生的因素比較複雜，因此分成 7 個因子來解釋。另外，從文獻回顧的遊覽車肇事的案例來看，全臺會發生事故的地區大都為山區，因此全臺會發生事故的整個情況就和南投縣類似，而南投縣為山區環境的地區。其分群數目為 5-7 群，判別正確率也皆達到 75%，代表此分類效果良好。

從事故嚴重程度迴歸預測模式來看：臺灣地區整體的死亡及受傷人數的迴歸模型皆為顯著；而臺北市及南投縣的死亡人數迴歸模型為不顯著，受傷人數的迴歸模型為顯著。大客車發生事故幾乎都為受傷居多，會發生死亡都為嚴重的肇事，其案件數不多，因此從迴歸的預測模式看臺北市及南投縣二個地區，死亡人數就沒有那麼明顯；整個臺灣地區的死亡人數會顯著，代表著案例分析的二個地區不顯著，但不代表其他地區不顯著。

由個別因子作的嚴重程度迴歸預測模式來看：由於全臺和南投縣因子數、分類的肇事變數及情境較為類似，因此試其比較其死亡和受傷人數的因子顯著性。臺灣地區死亡及受傷的顯著因子為行車環境及道路特性；南投縣死亡並沒有顯著之因子，受傷顯著因子為行車環境及環境特性。臺灣地區及南投縣受傷時的行車環境都為相同，不同的因子為而臺灣地區的道路特性及南投縣的環境特性各為顯著。代表著臺灣地區發生的大客車事故可能由速限、道路類別、快車道間及分向設施的四個肇事變數，南投縣由天候及路面狀態的二個肇事變數。臺北市死亡的顯著因子為道路特性、時間特性及道路障礙，受傷的顯著因子為道路特性及分向設施。皆影響臺北市死亡及受傷的大客車事故由道路類別和速限的二個肇事變數；另外，影響死亡的肇事變數還有發生時間、光線、障礙物、路面缺陷及視距；另外，影響受傷的肇事變數還有分向設施。

## 第七章 結論與建議

過去交通肇事的分析大多使用羅吉特迴歸、負二項迴歸、卜瓦松迴歸等方法，而本研究採用資料探勘於大客車交通事故的分析上，利用群集分析、判別分析及因子分析的方法找出影響大客車影響變數。本章依據前面各章所得之結論作歸納整理，並提出後續研究建議。

### 7.1 結論

#### 一、資料分析

本研究使用群集分析法的 K-Means 法做分析，各六年間及 92-97 年的群集數目為 5~7 群。由分群結果與肇事變數做卡方檢定，其顯著的變數個數為 19~21 個，不顯著變數為路面缺陷、路面鋪裝、障礙物、視距、路面狀態，依各年度的不同所帶入的保留及剔除之變數有也有不同。

依各年度及 92-97 年最佳分群結果套入判別分析中，其函數 Wilks' Lambda( $\Lambda$ )檢定皆為顯著，而分類正確率分成訓練樣本和測試樣本，皆達 90%，代表其分類效果良好。

經由逐步判別分析中，萃取出的變數有發生月份、發生時間、道路類別、速限、分向設施、光線、路面鋪裝、障礙物、號誌種類、號誌動作、事故位置、事故類型及型態及肇事原因；總年度多了快車道間、快慢車道間、天候、路面鋪裝、路面狀態等變數。

經由判別模式的加權值可知，主要影響交通事故的變數為肇事原因影響比例最高，而肇事原因中以變換車道或方向不當、未保持行車安全距離、酒醉(後)駕駛失、違反號誌管制或指揮、違反特定標誌(線)標制、其他駕駛人因素、非駕駛人因素等因子左右事故的發生。

## 二、案例分析

分析臺灣地區、臺北市及南投縣的大客車肇事案例，其 KMO 檢定皆大於 0.5 代表可以做因子分析。臺灣地區及南投地區的因子數目皆為 3，臺北市的因子數目為 5，本研究推測可能全臺發生大客車肇事的型態和南投縣類似，因此分成同樣的因子數目；臺北市因為為都市地區，所以肇事發生的因素比較複雜，因此分成 7 個因子來解釋。其分群數目為 5-7 群，判別正確率也皆達到 75%，代表此分類效果良好。

從事故嚴重程度迴歸預測模式來看：臺灣地區整體的死亡及受傷人數的迴歸模型皆為顯著；而臺北市及南投縣的死亡人數迴歸模型為不顯著，受傷人數的迴歸模型為顯著。由個別因子作的嚴重程度迴歸預測模式來看：臺灣地區死亡及受傷顯著的因子皆為行車環境及道路特性；臺北市的死亡顯著因子為道路特性、時間特性及道路障礙，受傷的顯著因子為道路特性及分向設施；南投縣死亡的因子皆不顯著，受傷的顯著因子為行車環境及環境特性。

## 7.2 建議

- 一、 由於本研究的資料只有針對大客車的「道路交通事故調查表」，其肇事真正的原因再進一步探討與研究，並加入車子本身因素、交通因素、道路幾何因素等相關方面考量，或許將來有助於交通安全的改善。另外，也未考量曝光量的問題，後續可以研究如何把交通事故調查表的資料及曝光量作結合。
- 二、 資料探勘有許多種方法，目前在交通事故之分析研究上不多，未來建議可以使用決策樹、類神經網路及基因演算作分類及預測影響死亡及受傷的肇事因子，用關聯分析探討交通事故中的變數之間是否有很高的關聯程度等方法作比較。
- 三 資料探勘之軟體，本研究僅用 spss 軟體操作，未來可用 SAS、IBM、SQL

Server 等另外專門為資料探勘設計之軟體，但取得上不易且軟體金額過於昂貴，因為資料探勘軟體都是專門授權給相關研究單位使用。

四、經由內政部警政署提供的資料中發現，資料中有許多錯誤及登錄方式尚不完整，其原因可能為記錄員警之專業不足或紀錄上的缺失，未來應加強交通警察與記錄人員之訓練，盡可能降低肇事黑數，以求未來在交通安全領域上有更完整的研究。

五、本研究對於所蒐集之肇事資料皆透過編號方式形成變數資料，而編號乃是一種權宜措施，將屬性資料轉換成有特殊涵義的編號，並藉以進行後續的模式推導，所以編號的方式可能有未盡合理或過於主觀之結果，建議後續研究者可以思考更符合邏輯之編號方式或利用其他的方法論探討肇事變數。

六、事故嚴重程度預測模式部分，本研究採用迴歸模式，建議後續研究可選取其他適合的模式進行校估預估或比較分析，以求取更準確的結果。

七、事先的預防勝於事後的補救，本研究是先以肇事資料來分析其肇事發生的原因，從事後的資料得知主要影響大客車的肇事因子，提出長短期的建議希望能夠減少大客車的交通事故。

#### 1. 短期

- 現有的交通設施及設計有缺失加以改善，包括道路的維修、道路的品質、標誌標線的正確性、修復的即時性。
- 強化傳統的標誌及標線的設置及引進國外成效良好的標線(如山形標線、立體視覺減速標線)。
- 環境因素及事故發生之位置，應該要設置注意、警告標語及強化號誌。

#### 2. 長期

- 公路總局監理單位、業者、駕駛及民眾都應常教育及宣導，如公路總局監理單位確保大客車相關定期檢定(改造車體、使用再生胎等)及車

齡問題、業者駕駛每天檢查車輛的行車紀錄器等。

● 引進新有的 ITS 系統於大客車增加其安全性：

- 自動車況偵測顯示：胎壓及胎溫警示系統、煞車溫度警示系統、先進車輛控制及安全系統。
- 先進車輛控制及安全系統：駕駛疲勞警示系統、適應性頭指系統、前方障礙物碰撞預防輔助系統、側邊障礙物警告系統、車道偏離輔助系統、緊急煞車預先警告系統、適應性行駛控制及煞車控制、自動事故回報系統。
- 緊急救援系統。



## 參考文獻

1. 內政部警政署，<http://www.npa.gov.tw/NPAGip/wSite/mp>
2. W. Frawley, G. Piatetsky-Shapiro and C. Matheus, “Knowledge Discovery in Databases: An Overview”, AI Magazine, Fall 1992, pp.213-228, 1991.
3. F. H.Grupe, M. M. Owrang, “Data Base Mining” Information Systems Management, Vol.12, No.4, pp.26-31, 1995.
4. U.M. Fayyad, D. Haussler and P. Stolorz, “Mining scientific data”, Communications of ACM, Vol.39, pp.55-57, 1996.
5. U. M. Fayyad, G. Piatetsky-Shapiro and C. Matheus, “The KDD process for extracting useful knowledge from volumes of data”, Communication of the ACM, Vol.39, No.11, pp.27-34, 1992.
6. M. J. A. Berry and G. S. Linoff, Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, Wiley Publishing, Inc, 2000.
7. J. Han and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann Publisher, Inc, 2001.
8. R.J. Roiger, M.W. Geatz, “Data mining : a tutorial-based primer” Addison Wesley, 2003.
9. 曾憲雄、蔡秀滿、蘇東興、曾秋蓉、王慶堯，「資料探勘」，旗標出版社，民國 94 年。
10. 尹相志，「SQL Server 2005 Data Mining 資料採礦」，悅知文化，民國 96 年。
11. U. M. Fayyad, “Data mining and knowledge discovery: making sense out of data,”IEEE expert, Vol. 11, No.5, pp.20-25, 1996.
12. 謝邦昌、鄭宇庭、蘇志雄、郭良芬，「Microsoft SQL Server 在資料採礦與商業智慧之應用」，中華民國採礦協會，民國 96 年。
13. T. Raymond & H. Jiawei, “Efficient and Effective Clustering Methods for Spatial Data Mining”, Proceedings of the 20th VLDB Conference Santiago, Chile, 1994.
14. L.K. Kao & C.C. Chiu, “Mining the customer credit by using the neural network model with classification and regression tree approach”, IFSA World Congress and 20th NAFIPS International conference 2001 Joint 9th , Vol.2, pp.923-928, 2001.
15. Q. Wang, Y. Wu, J. Xiao, G.F. Pan, “The applied research based on decision tree of data mining in third-party logistics”, International Conference on Automation and Logistics, China, 2007.
16. F.M. Eduardo, M.H. Dulce, F.R. Andres, “Mining Road Accidents”, MICAI LNAI 2313, pp.516-525, 2002.
17. R. Marukatat, “Structure-Based Rule Selection Framework for Association Rule Mining of Traffic Accident Data”, Springer-Verlag Berlin Heidelberg, 2007.



18. 吳冠宏、吳信宏、郭廣清，「應用分群技術於交通事故資料分析」，品質學報，Vol.13，No.3，頁305-311，民國95年。
19. 陳雯龍，「交通事故調查作業規範之研究」，中央警察大學交通管理研究所碩士論文，民國九十一年。
20. 交通部運輸研究所，「易肇事地點改善作業技術參考手冊」，民國92年。
21. 交通部運輸研究所「易肇事地點改善作業手冊之教育訓練計劃」，民國92年。
22. A. S., Al-Ghamdi, Analysis of traffic accidents at urban intersections in Riyadh, Accident Analysis and Prevention, Vol.35, No.5, pp.717-724, 2003.
23. E. Amoros, J. L. Martin, B. Laumon, "Comparison of road crashes incidence and severity between some French counties", Accident Analysis and Prevention, Vol.35, No.4, pp.537-547, 2003.
24. A.H. Lee, M.R. Stevenson, K. Wang, K.K.W. Yau, "Modeling young driver motor vehicle crashes: data with extra zeros", Accident Analysis and Prevention Vol.34, No.4, pp.515-521, 2002.
25. 楊思瑜，「小型車事故嚴重程度預測以桃竹苗地區為範圍」，逢甲大學交通工程與管理學研究所碩士論文，民國92年。
26. 邱裕鈞、王福聖、紀竣傑、郭曜彰、謝侃君，「兩車碰撞事故之鑑定決策支援系統」，道路交通安全與執法國際研討會，頁341-353，民國93年。
27. 蕭文龍，「多變量分析入最佳入門實用書」，基峰資訊，民國96年。
28. 陳順宇，「多變量分析」，華泰書局，民國94年。
29. 黃俊英，「多變量分析」，中國經濟企業，民國87年。
30. 林岩震，「多變量分析：SPSS的操作與應用」，智勝文化，民國95年。
31. 張云濤、龔玲，「資料探勘原理與技術」，五南圖書，民國96年。
32. 丁一賢、陳牧言，「資料探勘」，滄海書局，民國94年。
33. 謝邦昌教授統計電子學校，<http://www.stat.fju.edu.tw/teachonline/benchang/>。
34. 謝邦昌、鄭宇庭、蘇志雄、郭良芬，「EXCEL在資料採礦應用」，中華民國採礦協會，民國96年。
35. 陳正昌、程炳林、陳新豐、劉子鍵，「多變量分析方法-統計軟體應用」，五南圖書，民國96年。
36. 高啟涵，「運用資料採礦探討數位式行車紀錄器於公路客運駕駛員異常操作行為管理之研究」，淡江大學運輸管理學系運輸科學碩士班碩士論文，民國95年。
37. 易緯工程顧問，「大客車禁行路段及行駛應特別注意路段改善計劃(期中報告書)」，民國97年。

## 簡歷



基本資料			
姓名	王宜慶		
出生日期	1982.10.21		
性別	男		
出生地	臺灣臺北		
求學經歷			
	學校名稱	科系	就學期間
研究所	臺灣大學	土木所交通工程組	96.09－98.06
大學	淡江大學	運輸管理系	93.09－96.06
五專	致理技術學院	企業管理科	88.09－93.06
國中	恆毅中學	—	84.09－87.06
國小	光華國小	—	78.09－84.06