

國立臺灣大學工學院土木工程學系

碩士論文

Department of Civil Engineering

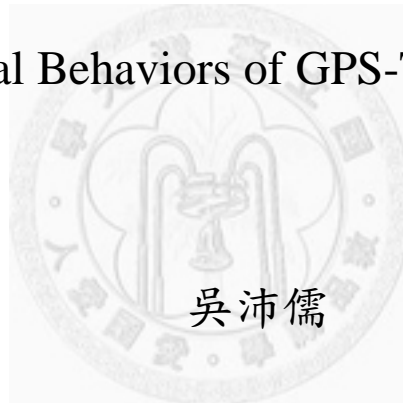
College of Engineering

National Taiwan University

Master Thesis

衛星派遣計程車駕駛人營業行為分析

Operational Behaviors of GPS-Taxi Driver



Wu, Pei-Ju

指導教授：張學孔 博士

Advisor: Chang, S.K. Jason, Ph.D.

中華民國 99 年 6 月

June 2010

誌謝

打開論文的檔案許久，卻遲遲無法在「誌謝」的標題下打上任何一個字，這一刻我明明盼望了很久，現在卻有種近鄉情怯的感覺。好像寫完致謝後，我的研究生涯就到此結束；好像寫完致謝後，我就不再是學生了；好像寫完致謝後，我就要與山林珊的大夥告別；好像寫完致謝後，我就必須邁向不一樣的人生。應該是開心的但帶有一些些的恐懼、擔心、煩惱以及不捨，那五味雜陳的感覺大概只有經歷過的人才能瞭解箇中滋味吧！

感謝張學孔老師總是給予我們許多的幫助，只要是跟研究有關的，老師總是二話不說的答應。除此之外，老師特殊的人生哲學 Appreciation 及 Variety (AV)，也讓我在談諧的言談中，獲益良多。感謝在口試時給予我很大讚賞與鼓勵的侯勝宗老師、洪鈞澤博士，讓我在研究的最後得到更多比建議更感人的收穫。感謝黃國平老師在口試時的指導，這讓我在逆境中學習成長，化壓力為力量。同時也要感謝台大所有教過我或沒教過我的老師，課業上的指導以及人格的養成，我想都是台大給我最大的禮物。

老師的忙碌也讓我們跟學長姐的關係特別密切，一路走來感謝英志、昱達與奇軒學長的教導以及協助，在我卡關的時候適時給予協助，也在我開始怠惰的時候給我適當的壓力，讓我能夠一步一步的往前進。更感謝一同進行 Taxi 研究的冠堯跟心怡，跟年紀相仿的你們相處及學習讓我更能敞開心房，面對不同的挑戰，也喜歡與你們一起談心的時刻。

除了課業之外，我很幸運的擁有一對很好的爸媽。感謝爸媽成為我強力的後盾，在我難過的時候給我安慰，在我迷惘的時候指引我方向，在我生氣

的時候當我的垃圾桶。雖然我不常回家，雖然我不習慣跟你們說感性的話，雖然我壓力很大的時候還會對你們發脾氣，但這血濃於水的親情是切不斷的，我愛你們。

在研究所這說長不長說短不短的兩年中，還好有山林珊的大家一同努力、玩樂、吵鬧，才能讓這苦悶的研生活增添色彩。這兩年中，也累積了好多好多的回憶。我想我以後一定會很懷念喜歡把成語隨意重組的芝旭、喝醉的時候喜歡吃餐具的小彩、小小隻卻愛說自己力氣很大的美夙、陰沉又愛超前進度的彭紀凱、喜歡小丸子跟園藝的志宜、愛消失又很有趣的忠賢、會讀書又會玩的中庭、萬事通卻心思纖細的張爸張立委、腦內存放星爺電影記錄片的黃君凱、深藏不露的電玩強手大師、能力很強卻愛叨叨念念的雅君、家裡在賣「馬信」的阿拔、喜歡吃大李的邱子揚、外表看似年長者卻熱愛搖滾的老頭、被用來當時間單位的雷德、常常惹雅君生氣的鄭鎧鋁、進來交通組後慢慢被同化的依葶、永遠像過動兒又能招來雨神的黃崇宇、總是散發出一股溫柔氣的小艾、默默偷交女朋友的李佳憲、來無影去無蹤的小美、對3C產品相當瞭解並且不浪費任何剩餘價值的心榆、熱心到幾乎忘記自己的大餅、NBA Live的紀錄保持者李宗晏、碩二下突然變得很開朗的珮群，還有陪我一起走過每個歡喜與悲傷、總是把我寵上天的寶貝熊熊，這一個個的懷念終將化成回憶被我珍藏在心中。

因為有這些永遠謝不完的人、永遠聊不完的感動，在我的碩士生涯中才能如此多采多姿。感謝每一個幫助過我的人，也感謝每一個傷過我的人。因為有你們，讓我的人生更加完整。

摘要

計程車駕駛人與乘客長期處於資訊不對稱的情況中，為改善此不利情勢，計程車衛星派遣的概念因應而生。對計程車產業而言，衛星派遣系統乃目前最佳的派遣方式，然而衛星派遣車隊卻遲遲無法達到永續經營之合理規模。推究其原因發現，現行派遣邏輯僅是單方面的考慮乘客所想，而忽略計程車駕駛人所望；換言之，衛星派遣車隊理應公平服務兩端的使用者，但現況卻完全以乘客需求為導向。然而，從過去的研究中，我們並不瞭解計程車駕駛人想要的是什麼。理論上藉由同時考慮計程車駕駛人與乘客的需要，方能發揮衛星派遣系統之最大效能，亦能吸引個人計程車駕駛人的加入，以期解決衛星派遣車隊規模數過小的問題，並創造更好的衛星派遣計程車營業環境。因此，本研究以計程車衛星派遣資料庫為基礎，進行計程車駕駛人行為的資料探勘(Data Mining)。為增加研究的廣度，本研究採用因素分析(Factor Analysis)對計程車駕駛人營業行為變數進行探討，並將派遣數等十二項營業行為影響變數減少為「營業狀況」、「空車率」及「營業習慣」等三個因素。接著利用集群分析(Cluster Analysis)對計程車駕駛人營業行為進行分群，透過兩階段集群分析的過程，將衛星派遣計程車駕駛人營業行為分為四群。此四群駕駛人營業行為的詳細內容除了分析原有營業行為變數外，並利用亂數表進行分群抽樣，分析駕駛人兩次載客間時間間隔等六項變數，同時將衛星派遣計程車駕駛人依其行為特性分為「智慧型」、「積極型」、「自由型」及「消極型」四群。最後根據各群計程車駕駛人的營業行為特性，分別提出派遣邏輯改善之具體建議。

關鍵字：計程車、駕駛行為、資料探勘、因素分析、集群分析、派遣邏輯

Abstract

Taxi drivers and passengers are normally involved in an asymmetric information situation. In order to improve this situation, the concept of automatic dispatching taxi was introduced. On the taxi industry, the GPS-based dispatching system is the best way to detach taxi nowadays; however, the fleet size of GPS taxi has not reached an economic scale for sustainable operation. One of the main reasons is that the dispatch logic only takes the passenger's expectation in consideration while leaving the taxi drivers' wanted out of consideration. However, from past researches, it seems lack of the taxi drivers' expectation. It is expected that taking the taxi drivers and passengers into account simultaneously, the GPS dispatch system shall have the maximum performance. Moreover, it will also attract individual taxi drivers to join the GPS taxi group and then create a better GPS taxi business environment. This study aims to use GPS-taxi dispatch database to analyze taxi driver operational behaviors by data mining approach. First,, this study uses factor analysis to explore key variables among the 12 operational variables while operating conditions, vacancy rate and operating habits are identified as the key variables. Then cluster analysis is then applied to classify the taxi drivers' operational behavior into four categories, namely intelligent type, positive type, free type and passive type. To find the drivers' business behavior of each category, this study uses the details of the original operational behavior variables and the random sampling analysis to analyze time interval between occupied and the other five variables. Finally, suggestions of improving dispatch logic for each type of drivers are proposed and evaluated based on analyses of their behaviors.

**Keywords : Taxi, Driving Behavior, Data Mining, Factor Analysis,
Cluster Analysis, Dispatch Logic**

目錄

誌謝.....	I
摘要.....	III
Abstract.....	IV
目錄.....	V
表目錄.....	VII
圖目錄.....	IX
第一章 緒論	1
1.1 研究背景與動機.....	1
1.2 研究目的.....	1
1.3 研究範圍與限制.....	2
1.4 研究流程.....	2
1.5 章節架構.....	4
第二章 文獻回顧	7
2.1 駕駛人行為相關文獻.....	7
2.1.1 私有運具駕駛人駕駛行為相關文獻.....	7
2.1.2 大眾運輸駕駛人駕駛行為相關文獻.....	12
2.1.3 計程車駕駛人駕駛行為相關文獻.....	14
2.2 資料探勘相關文獻.....	18
2.2.1 資料探勘於公共運輸之應用.....	19
2.2.2 資料探勘於其他運輸問題之應用.....	21
2.3 綜合評析.....	23
第三章 研究方法	27
3.1 資料探勘 (Data Mining).....	27
3.1.1 資料探勘基本概念.....	28
3.1.2 資料探勘應用流程.....	34
3.2 因素分析 (Factor Analysis, FA).....	36
3.2.1 因素分析基本概念.....	36
3.2.2 因素分析應用流程.....	40
3.3 集群分析 (Cluster Analysis, CA).....	45
3.3.1 集群分析基本概念.....	46
3.3.2 集群分析應用流程.....	56
3.4 方法論具體架構.....	59
第四章 派遣資料前置處理	61
4.1 分析流程.....	61
4.2 衛星派遣計程車營業資料選取及蒐集.....	64

4.3	衛星派遣計程車營業資料前置處理.....	64
4.4	衛星派遣計程車營業資料變數縮減.....	68
第五章	衛星派遣計程車駕駛人營業行為分析	73
5.1	衛星派遣計程車駕駛人月營業行為集群分析.....	73
5.1.1	集群分法與分析流程.....	73
5.1.2	階層式集群分析.....	74
5.1.3	非階層式集群分析.....	76
5.1.4	集群驗證.....	77
5.1.5	分析結果.....	77
5.2	衛星派遣計程車駕駛人日營業行為抽樣分析.....	82
5.2.1	抽樣方法與分析流程.....	82
5.2.2	T 檢定	83
5.2.3	分析結果.....	89
5.3	衛星派遣計程車駕駛人營業行為分群結果整理.....	91
第六章	派遣邏輯改善策略	93
6.1	智慧型駕駛人.....	93
6.2	積極型駕駛人.....	96
6.3	自由型駕駛人.....	99
6.4	消極型駕駛人.....	102
6.5	綜合評述.....	105
第七章	結論與建議	107
7.1	結論.....	107
7.1.1	研究成果.....	108
7.1.2	理論意涵.....	109
7.1.3	實務意涵.....	111
7.2	建議.....	111
參考文獻		115

表目錄

表 1-1：衛星車隊派遣邏輯.....	1
表 2-1：各群駕駛人營業行為特色表.....	18
表 2-2：駕駛人駕駛行為文獻整理表.....	23
表 2-3：資料探勘方法應用整理表.....	25
表 3-1：資料探勘之定義整理.....	28
表 3-2：資料探勘之定義整理（續）.....	29
表 3-3：資料探勘之定義整理（續）.....	30
表 3-4：KMO 檢定值與因素分析適合度關係.....	41
表 3-5：因素負荷量、解釋變異量及選取準則.....	42
表 3-6：直接轉軸法與斜交轉軸法綜合比較.....	44
表 4-1：車隊資料回傳頻率表.....	64
表 4-2：變數平均值、標準差及下限值表.....	66
表 4-3：衛星派遣資料庫變數之 KMO 檢定及 Bartlett 球形檢定結果.....	68
表 4-4：衛星派遣資料庫變數之總變異量表.....	69
表 4-5：衛星派遣資料庫變數之轉軸後的成份矩陣—最大變異數法.....	70
表 4-6：衛星派遣資料庫變數之因素命名表.....	71
表 5-1：階層集群分析法結果分析表.....	75
表 5-2：K-Means 集群分析結果.....	76
表 5-3：各分群因素與各集群之 ANOVA 檢定結果.....	77
表 5-4：四群駕駛人各觀察變數分析結果.....	79
表 5-5：四群駕駛人每日營業情形數據分析結果.....	81
表 5-6：各群駕駛人載客時間間距平均值及變異數整理表.....	83
表 5-7：各群駕駛人同區載客比例 T 檢定雙尾 P 值整理表.....	83
表 5-8：各群駕駛人載客距離間距平均值及變異數整理表.....	84
表 5-9：各群駕駛人載客距離間距 T 檢定雙尾 P 值整理表.....	84
表 5-10：各群駕駛人每日休息次數平均值及變異數整理表.....	85
表 5-11：各群駕駛人每日休息次數 T 檢定雙尾 P 值整理表.....	85
表 5-12：各群駕駛人同區載客比例平均值及變異數整理表.....	86
表 5-13：各群駕駛人同區載客比例 T 檢定雙尾 P 值整理表.....	86
表 5-14：各群駕駛人夜間時段營業比例平均值及變異數整理表.....	87
表 5-15：各群駕駛人夜間時段營業比例 T 檢定雙尾 P 值整理表.....	87
表 5-16：各群駕駛人凌晨時段營業比例平均值及變異數整理表.....	88
表 5-17：各群駕駛人凌晨時段營業比例 T 檢定雙尾 P 值整理表.....	88
表 5-18：計程車駕駛日營業行為分析整理表.....	90

表 5-19：各群駕駛人命名及特色整理表.....92
表 6-1：各群駕駛人派遣邏輯設計重點.....105



圖目錄

圖 1-1：研究流程圖.....	3
圖 2-1：駕駛人一般行為分類圖.....	9
圖 2-2：新加坡康福車隊科技採用模式.....	17
圖 2-3：台灣大車隊科技採用模式.....	17
圖 3-1：資料探勘目的架構圖.....	31
圖 3-2：知識探索應用流程.....	34
圖 3-3：探索性因素分析路徑圖.....	38
圖 3-4：驗證性因素分析路徑圖.....	39
圖 3-5：因素分析應用流程.....	40
圖 3-6：集群間距離圖.....	52
圖 3-7：集群分析應用流程.....	56
圖 3-8：資料分析模式.....	59
圖 4-1：分析流程.....	63
圖 4-2：KMO 檢定及平方和負荷量萃取值與資料刪除下限關係圖.....	67
圖 4-3：衛星派遣資料庫變數之因素陡坡圖.....	69
圖 5-1：集群方法分析流程.....	74
圖 5-2：係數變化率與集群數關係圖.....	75
圖 5-3：分群抽樣分析流程圖.....	82
圖 6-1：智慧型計程車駕駛人衛星派遣流程圖.....	95
圖 6-2：積極型計程車駕駛人衛星派遣流程圖.....	98
圖 6-3：自由型計程車駕駛人衛星派遣流程圖.....	101
圖 6-4：消極型計程車駕駛人衛星派遣流程圖.....	104

第一章 緒論

1.1 研究背景與動機

近年來由於全球油價上漲以及環保意識抬頭，民眾使用公共運輸的意願提高，而提供及門服務的計程車重要性也日益增加。然而計程車駕駛與乘客長期處於資訊不對稱的情況中。乘客在上車前無法得知計程車駕駛服務的優劣，並且也無法防止下次不再招到同一台計程車。另一方面，計程車駕駛對於需求發生的地點（乘客招車位置）也無法有效率的得知，再加以計程車市場供給過剩而導致空車繞行的情況嚴重，造成環境更大的負擔。

目前台北都會區計程車在營業的方式上，除了傳統的巡迴攬客外，尚有衛星派遣、無線電派遣、定點排班及駐行排班等方式。傳統的巡迴攬客雖為最方便且單純的營業模式，但空車巡迴不但增加道路流量，亦造成油料及時間的浪費，同時也增加了噪音及空氣污染等外部成本。此外，在 2000 年至 2002 年間蓬勃發展的無線電車隊，雖可舒緩計程車駕駛與乘客間資訊不對稱的情況，但無線電的廣播半徑範圍無法超過 40 公里，因此車隊僅能在固定區域內營業，減少了計程車的產能。再加上無法有效監控車輛的所在位置及計程車駕駛與總機人員的暗盤作業，導致無線電車隊時有分配不公平的現象。此外，駐行排班及定點排班的方式雖可舒緩巡迴攬客及無線電派遣的缺失，但仍分別有司機空等時間長及乘客搭乘不便等缺點。近年來推動的衛星派遣，除了可以摒除無線電派遣、駐行排班及定點排班的缺失外，亦可有效的減少巡迴攬客造成的污染及浪費；同時，計程車駕駛人與乘客間的資訊不對稱更能被有效的改善，因此衛星派遣將會是未來重要的營業方式之一。

然而，自 2001 年鼎華科技從新加坡康福車隊引進衛星派遣系統迄今，並無法有效的改善計程車產業。究其原因，一方面駕駛人對於新技術尚有疑慮，除了擔心自己的營業資料被監控外，對於衛星派遣技術能夠提昇他們多少收入也不甚樂觀，造成台灣的衛星派遣車隊規模尚嫌不足。規模不足的派遣車隊將會增加乘客的等候時間，進而降低乘客使用電話叫車的意願，導致衛星派遣的數量不足。如此一來，計程車駕駛人對於加入衛星派遣車隊之意願也會降低，造成惡性循環。根據鼎漢顧問公司研究顯示，衛星派遣車隊之車隊規模的損益平衡點為 6,000 台車。在大台北地區，目前最大的台灣大車隊之車隊規模在近期才達到 6,800 台，然而規模次之的大都會衛星車隊僅有 3,000 台車，更不用說其他衛星車隊，車輛數皆遠小於 6,000 台車。因此，如何吸引計程車駕駛人加入衛星派遣車隊為當務之急。另一方面，衛星派遣系統並未被有效的利用，衛星派遣系統除了可以被用在派遣機制上外，亦可提供駕駛寶貴的資訊。

長期以來，在計程車產業中計程車駕駛人總是被忽略的一群。觀察目前的派遣邏輯(見表 1-1)，除了空排點派遣外，皆是以乘客等候時間最小優先作為考量的依據。而事實上，本身空排點的設計亦是考慮乘客等候時間及需求點劃設。但是對衛星派遣車隊而言，它們向乘客收取費用，同時也向駕駛人收取隊費，理應將乘客與駕駛人放在同一個天平上去衡量，不應只關注乘客的需要而忽視駕駛人的期望。然而，計程車駕駛人需要的是什麼？過去對於計程車駕駛人行為的研究僅僅只有一篇，且該篇研究是以質性分析的方式進行，根據大量的訪談及問卷的發放，分析出計程車駕駛人的需要。但我們也知道，人們對自我的認知常與實際的行為不同。因此，本研究希望藉由實際的派遣資料庫，對於計程車駕駛行為進行數值面的分析。

表 1-1：衛星車隊派遣邏輯

派遣名稱	特色	派遣方式
即時派遣	<ul style="list-style-type: none"> ● 乘客預約時間在 30 分鐘內。 ● 目的地在大台北地區。 	接獲派遣指令時，透過衛星系統找出 <u>2 公里內</u> 、 <u>離乘客最近</u> 、 <u>等候派遣時間最久</u> 的空車。
市內預約派遣	<ul style="list-style-type: none"> ● 乘客預約時間超過 30 分鐘。 ● 目的地在大台北市區。 	開放競標，以 <u>最快按回覆鍵</u> 者獲得任務。
長途預約派遣	<ul style="list-style-type: none"> ● 乘客預約目的地超過大台北地區，單趟收入超過 1,000 元。 	所有車隊成員以 <u>輪流</u> 的方式來決定派遣順序。
定點排班	<ul style="list-style-type: none"> ● 除一般政府設立之排班點外，車隊於人潮密集處設立固定排班點。 ● 凡屬該車隊之隊員即可前往排班點進行排班。 	駕駛於排班點 <u>依序</u> 載客。
空排點派遣	<ul style="list-style-type: none"> ● 車隊設立若干空中排班區域，於排班區域內進行類似「郵局排隊」之排班。 ● 駕駛進入空中排班區域後可登記等候序號。 ● 車輛離開空中派遣區域時，序號自動取消。 	於空排區域內 <u>依序號</u> 派遣車輛。

資料來源：本研究彙整侯勝宗(2006)研究而得

1.2 研究目的

本研究將利用大都會衛星車隊計程車派遣資料庫為基礎，進行計程車駕駛行為的資料探勘，具體研究目的如下：

1. 分析衛星派遣計程車駕駛人的營業行為模式。
2. 針對衛星派遣計程車駕駛人的營業行為模式進行駕駛人分群。
3. 根據分析的結果，提出考慮計程車駕駛人的營業行為的派遣邏輯建議。

根據分析結果及派遣邏輯的設計，期能為改善派遣邏輯及車隊經營方式所參考，創造出衛星派遣車隊及計程車駕駛雙贏的局面，以提昇衛星派遣系統價值。

1.3 研究範圍與限制

本研究以大台北地區之衛星派遣計程車為主要分析對象，並以車隊規模較大之大都會衛星車隊之派遣資料庫為基礎，進行衛星派遣計程車駕駛人營業行為分析。大都會衛星車隊係於 2007 年 11 月正式成立並導入衛星派遣技術，2008 年 4 月正式開台商業營運。本研究囿於資料獲取與研究資源，僅選取大都會衛星車隊 2008 年 5 月的資料，藉此分析衛星派遣技術導入前期衛星派遣計程車駕駛人之營業行為。



1.4 研究流程

根據前述研究背景及動機，在確立研究目的及研究範圍後，首先進行相關文獻回顧及整理，文獻回顧的方向分為駕駛行為面、資料探勘及多變量分析方法論。派遣資料庫資料前處理(包含缺漏資料的處理及資料淨化)結束後，開始進行資料分析並對計程車駕駛進行分群，爾後針對各群駕駛行為分別設計較適合的派遣邏

輯，最後提出結論與建議，流程如圖 1-1

圖 1-1 所示。

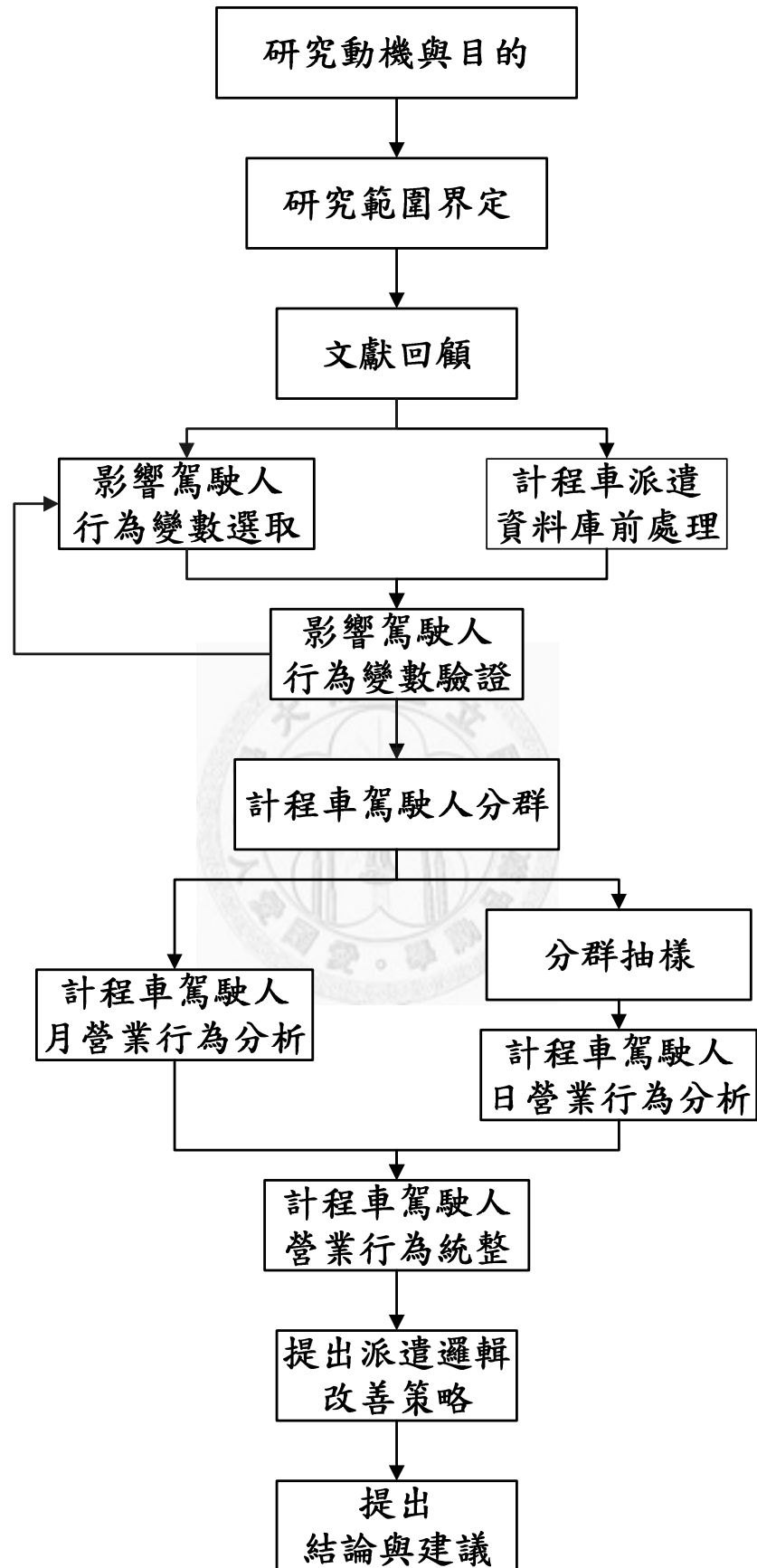


圖 1-1：研究流程圖

1.5 章節架構

本研究共分為七章，第一章為緒論；第二章為文獻回顧；第三章說明研究方法；第四章先進行派遣資料庫資料的前處理；第五章為研究結果討論及分析；第六章提出適合各群駕駛人之派遣邏輯；第七章則提出本研究之結論與建議。

第一章 緒論

說明研究背景、動機、目的、範圍、流程及章節架構。

第二章 文獻回顧

包含國內外駕駛行為的相關研究之比較以及資料探勘之定義、步驟、功能及應用。



第三章 研究方法

包含資料縮減及分析所需之方法，本研究主要採用因素分析中的主成分分析、階層集群分析法及 K-Means 集群分析法。

第四章 衛星派遣資料前置處理

包含缺漏資料之處理及淨化，亦同時利用 97 年營業情形調查之結果及試誤法設立分析門檻。爾後，利用主成分分析將各變項進行因素分析之縮減，並進行因素命名。

第五章 衛星派遣計程車駕駛人營業行為分析

利用各因素依序進行階層集群分析及 K-Means 集群分析。分群完成後，依據分群結果分別分析各群駕駛之月營業行為，並在各群隨機抽取若干個樣本進行駕駛每日行為分析，統整各群駕駛之月營業行為及每日營業行為後，依據各群駕駛的特色分別命名之。

第六章 派遣邏輯改善建議

針對前章分析出的各群駕駛特性，分別對各群駕駛人提出符合其需要之派遣邏輯改善建議。

第七章 結論與建議

針對研究結果提出具體結論與建議。





第二章 文獻回顧

本章節分為三大部分，首先回顧國內外與駕駛人行為相關之文獻，由於本研究的研究對象為計程車駕駛人，且私有運具駕駛人與公共運輸駕駛人之特性並不相同，故在回顧駕駛行為時將私有運具駕駛人、大眾運輸駕駛人及計程車駕駛人分開討論。接著進行資料探勘的文獻回顧，內容包含國內外在公共運輸領域使用資料探勘技術的文章及其他在運輸領域使用資料探勘的文獻。最後針對以上回顧項目進行文獻的整理、討論以及評析。

2.1 駕駛人行為相關文獻

國內外有關駕駛人行為的研究相當多，所用的分析方法也各不相同。本節將回顧國內外各類駕駛人行為分析的文獻，針對過去文獻的分析對象、分析項目及分析方法進行整理。另外，由於各類駕駛人之間皆有不同的特性，例如私有運具駕駛人及公共運輸駕駛人所考量的項目不盡相同，因此將分別討論私有運具駕駛人與公共運輸駕駛人的駕駛行為。最後針對以上回顧項目進行文獻的整理、討論以及評析。

2.1.1 私有運具駕駛人駕駛行為相關文獻

從駕駛目的上來看，私有運具駕駛人與公共運輸駕駛人差異很大，雖國內外對於駕駛人行為的研究行之有年，但研究對象多為私有運具駕駛人，對於公共運輸駕駛人行為的研究並不多。因此本節仍回顧私有運具駕駛人駕駛行為的相關文獻，希望從較為完整的私有運具駕駛人駕駛行為模式中，找出分析計程車駕駛人營業行為的基本概念、研究方法及影響較大的因素。以下將針對各類私有運具駕

駛人的駕駛行為進行文獻回顧，並在最後整理出駕駛行為研究的研究方法及特色。

在進行駕駛人行為分類前，我們必須先瞭解如何描述駕駛行為？Fine 在 1965 年提出了一種描述方式，該研究利用臨界事件技巧(Critical Incident Technique)評估「駕駛人的判斷與決策」作為駕駛行為分群依據。駕駛行為的描述方式以「好的」或「壞的」來分。好的駕駛行為：駕駛人以注意的、有禮貌的、慎重、小心的駕駛行為，考慮人、車的限制，反映出安全、有秩序、迅速的交通流動。壞的駕駛行為：不適當的迴轉、未考慮其他用路者，做出錯誤的決定，妨礙交通的流動。另外，李東陽(1996)以速度、加減速、橫向移動面積等三個主要變數來分類駕駛行為，在決定駕駛行為之分類尺度則利用平均數及標準差、集群分析及判別分析、主觀決定等三種方式，研究結果發現以「主觀決定」作為分類尺度最適當。爾後利用四段駕駛環境的影片對受試者進行面談訪查，研究結果顯示複雜度愈高之駕駛環境，其安全性、舒適性愈低，操作困難度愈高。

瞭解了駕駛人駕駛行為描述方式後，一般駕駛人駕駛行為主要又分為哪些呢？根據龍天立等(1987)提出研究可知，一般駕駛行為特性可以圖 2-1 分類，由圖 2-1 可知一般駕駛行為可先初步分類為單獨行為及互動行為。單獨行為意指「駕駛者在沒有其他道路使用者顧慮下」的行為特性，此種情境能反應行為者許多期望的行為，如期望速度(Deaired Speed)等；互動行為則分為衝突行為、跟車行為及排隊行為，衝突行為意指同時接近衝突點的車輛間分享路權的行為，跟車行為指的是同一車到內車輛間依序追隨之行為，排隊行為則是指車輛間爭取服務先後次序的行為。有了前述背景知識後，該研究利用實地觀察、實地實驗、錄影分析集問卷調查四種方法對台北市駕駛人之態度與行為進行研究，研究結果顯示台北市駕駛人行為特性可歸納為「激進的駕駛行為」、「落後的駕駛文明」、「充斥的違規行為」、「橫行的計程車」、「爭先恐後的機車」、「執法的無力感」、「嚴重的認知

差距」、「沉默的大眾」等八項。其中「橫行的計程車」一項，意指計程車駕駛人的違規率有較高的傾向，其中以車道使用、行車間距、方向燈使用、臨時停車、變換車道等五類違規最常發生。在該研究中各車種駕駛人對個性的自我評定皆以溫和、中庸派佔最多數，且平均而言計程車駕駛人並不承認自己有較激進的駕駛行為，這個結果顯示駕駛人在面對問卷問項填答時，容易有隱惡揚善的情形發生。

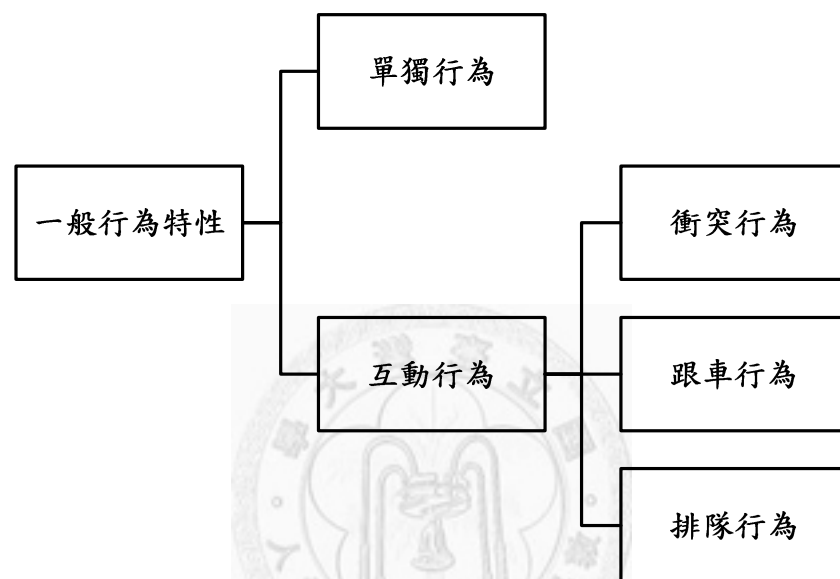


圖 2-1：駕駛人一般行為分類圖

資料來源：龍天立等(1987)

私有運具駕駛人駕駛行為之研究的分類方式有很多，在私有運具駕駛人駕駛行為被關注之初，研究多以駕駛人能力為分類依據。如程法泌(1979)探討汽車駕駛人之動作穩定能力、視野、深度知覺、手眼協調能力、瞬間提示反應能力、視覺辨別反應能力及時間速度預測能力與汽車駕駛行為間的關係，結果顯示汽車駕駛人之動作穩定能力、視野、由遠方至近方移動的深度知覺、視覺辨別反應能力、時間速度預測能力與汽車駕駛行為有顯著關係。

隨著時代的演進，私有運具駕駛人駕駛行為分類的方式日趨多元，分類的變數大致可分為駕駛人的筆事情形、駕駛人技能與經驗、駕駛人感知等三類。在駕

駛人的肇事情形方面，劉正華(1996)利用駕駛人行為問卷(Driver Behavior Questionnaire)與駕駛行為量表(Driving Behavior Inventory)，透過因素分析法、線性結構關聯模式與信度、效度分析設計適合國人特性的駕駛行為問卷，並以此進行駕駛人駕駛行為資料蒐集及分析。此外藉由集群分析與羅吉特回歸模式對駕駛者交通意外事故風險與駕駛者特性、駕駛行為進行關聯性分析，由此建立駕駛者肇事風險的評估方法。該研究結果顯示，駕駛行為中之侵略性及違規性失誤與交通意外事故風險有明顯的正向關係，而性別、年齡與駕車頻率亦是影響事故風險性的因子，該研究發現男性比女性有更高的肇事風險，而肇事概率會隨著年齡增加而下降，隨著駕車頻率增加而上升。此外，吊銷或吊扣駕照記錄、超車違規記錄與酒醉違規駕駛記錄亦與肇事風險有關，有以上違規記錄者其肇事風險亦較高。另外，陳國樑(2004)利用某公司的歷史肇事、保險理賠資料進行統計整理，同時對該公司的駕駛人進行問卷調查，調查的內容包含教育程度、服務年資、年齡分組等，並以一般統計分析、交叉分析、線性結構方程式分析該公司管理制度對駕駛人、駕駛行為(情緒反應)、與車輛管理制度的因果關係。研究結果顯示，公司管理制度與車輛管理制度、駕駛行為及駕駛人成顯著正向的直接影響關係，此外，駕駛人與駕駛行為之關係為顯著正向的直接影響關係。

在駕駛人技能與經驗方面，張新立等(2008)主要探討影響小客車駕駛人之駕駛技能與正向駕駛行為的因子，該研究利用駕駛技能調查表(Driver Skill Inventory, DSI)與駕駛人行為問卷(Driver Behavior Questionnaire, DBQ)分別量測駕駛技能與正向駕駛行為，經由因素分析將駕駛技能與正向駕駛行為各分為幾個構面，並以 Rasch 分析衡量駕駛人駕駛技能與正向駕駛行為的大小，藉以探討各因素之影響。研究結果顯示，性別、持有汽車駕照年數、平均每年駕駛里程、超車情形、高速公路駕駛速度對駕駛技能有顯著影響，而正向駕駛行為則受人格的影響為最大。此外，汪進財等(2009)提及在過去的研究中小客車及機車事故的研究中，年齡與性別為影響駕駛人行為的顯著因素，且大部分年輕人事故發生原因

並非由於經驗不足而是由於年輕人自身特性。該研究根據過去研究的結果設計年輕機車族群駕駛行為問卷，透過結構方程式與多群組分析進行研究。該研究結果顯示性別對年輕機車族群駕駛行為的影響最為顯著，遇到不如意的交通狀況下，男性駕駛人較容易產生高風險的駕駛行為。而年齡、駕駛經驗、違規經驗、事故經驗、職業等特性上亦具有顯著特性。

在駕駛人感知方面，王建仁(2003)利用風險感認理論，並參考過去道路使用者風險感認相關研究，建立國內機車使用者風險感認程度之量表，並以敘述統計、多變量分析、線性結構方程式，以及羅吉特回歸方程式等數量方法進行資料分析與統計檢定。研究結果顯示，性別、年齡、機車後照鏡之裝設、汽車駕照的取得與否、事故經歷、駕駛經驗、冒險性駕駛行為頻率、個人刺激尋求傾向及對於自身駕駛能力的樂觀態度等個人屬性，均與機車使用者對於整體機車駕駛與分項冒險性機車駕駛行為之風險態度有顯著關係。此外，國內機車使用者自認涉入交通事故之機率極低，但事實並非如此，此結果顯示不論是何種駕駛人的自我認知與事實皆存在一定程度的差異。除此之外，張新立(2009)利用試題反應理論，利用問卷收集的方式建立一「量測駕駛人憤怒駕駛傾向」之量表。進而透過Eysenck提出的人格特質量表，嘗試探索衝動性、冒險性與同理心對發生憤怒駕駛行為的影響。同時利用集群分析法與主成份分析法進行憤怒駕駛行為研究。研究結果顯示直接敵意的情境最容易激怒駕駛人，此外，性別、教育程度、年平均開車里程、開車年資、違規經驗、口頭衝突經驗、肢體衝突經驗、受傷事故經驗、財損事故經驗等特性對駕駛人憤怒駕駛傾向皆具有顯著影響。

私有運具駕駛人駕駛行為也常以模型建立的方式分析，這種方式多以分析某種外來訊息對駕駛人駕駛行為之影響為主，而模型建立的方式有很多種。Kuge等人(2000)利用隱藏式馬可夫模型(Hidden Markov Model, HMM)建立一個駕駛行為模型，試圖探討駕駛支援系統(Driver Assistance System)與駕駛行為間的關係。

研究結果顯示隱藏式馬可夫模型可有效的建立駕駛行為模型，並可利用車道與車輛間的關係即時偵測駕駛行為。另外，Peeta 等(2000)分析可變標誌(Variable message signs, VMS)對於駕駛人駕駛行為的影響，該研究利用敘述性偏好(Stated Preference)設計問卷，並藉由羅吉特模式建立一在可變號誌下的駕駛行為模式。研究結果顯示，可變號誌提供內容的詳細程度會顯著影響駕駛人改道的意願。其他重要因素包含駕駛人的社經地位、網路空間知識及對訊息的信賴程度。除此之外，胡守任等(2009)利用公路駕駛人問卷調查，藉由結構方程式分析平交道危險狀況與駕駛行為間之關聯性。研究結果顯示，對平交道法規認知不足與個人亦有違規屬性的駕駛人，在行經平交道前明顯有偏差的駕駛行為，同時與平交道前危險狀況有直接關聯。此外，有分心與違規傾向的駕駛人，其不良的飲酒習慣亦越高；然而，兩者與平交道前之偏差駕駛行為並無顯著的因果關係。

2.1.2 大眾運輸駕駛人駕駛行為相關文獻

上節回顧了私有運具駕駛人駕駛行為之相關文獻，本節將進行大眾運輸駕駛人駕駛行為相關文獻的回顧。由於大眾運輸駕駛人與私有運具駕駛人的目的不同，大眾運輸駕駛人需考慮收益的多寡，非單純顧慮本身之駕駛行為或者心情，故本研究將其單獨提出討論。然而，過去對於大眾運輸駕駛人之駕駛行為的研究並不多，又本研究認為影響大眾運輸駕駛人壓力感受的重要變數，亦將影響大眾運輸駕駛人駕駛行為，因此本節也納入大眾運輸駕駛人工作壓力相關研究的回顧。

在大眾運輸駕駛人駕駛工作壓力相關之研究方面，任維廉等(1993)利用深度訪談及問卷調查的方式對公車駕駛人進行工作壓力源與工作壓力反應的知覺研究，該研究之研究架構分為對工作壓力源知覺、干擾變數及工作壓力反應知覺三

部分。研究結果顯示，公車駕駛人知覺到主要的工作壓力源，依序為物理環境、獎懲制度、薪資制度、工作時間、心理環境、上司督導、同事相處。其他顯著的壓力源變數為年齡、年資及學歷。在過去研究的基礎上，任維廉等(1997)根據文獻回顧及公車公司管理人員與駕駛人訪談，推導出公車駕駛人工作壓力量表。公車駕駛人工作壓力的來源分別為工作時間、對工作上物理環境的感受、對工作上心理環境的感受、與同事關係、與站長關係、公司組織與結構、家庭與工作互動、生涯與成就、薪資獎金等九項；公車駕駛人工作壓力反應或結果分別為工作滿意、身體健康、心理健康、缺工、病歷、工作績效、工作傷害事故等七項。該研究結果顯示，公車駕駛人嚴重的工作壓力來源為休息時間太短、噪音與空氣污染嚴重、路上的交通狀況擁擠、底薪太低等，而造成嚴重的工作壓力反應為對公司解決衝突的方式、休假方式、福利待遇、工作保障與工作時間不滿意。除了工作壓力源、反應及壓力量表的設計外，張勝雄(1999)亦根據文獻回顧及業者、駕駛人訪談建構出公車駕駛人之薪資結構與工作壓力、工作滿足、服務品質之關係的概念模式及因果關係模式，藉此設計出問卷及量表進行問卷調查。研究結果顯示，在薪資結構方面，薪資滿足感、工作壓力、工作滿足及服務品質皆與之有顯著相關。在駕駛人薪資滿足感中，工作滿足與之有顯著正相關，而工作壓力則與之有顯著負相關，亦即駕駛人薪資滿足越低，駕駛人的工作滿足越低、工作壓力越高。

在大眾運輸駕駛人入與其營業行為相關之研究方面，分為駕駛人薪資以及激勵獎金兩個部份。劉建良(2000)利用深度訪談及問卷調查兩種方法進行公民營公車駕駛人薪資滿足感與工作壓力、工作滿足、服務品質的關係，研究結果顯示駕駛人人格特質為駕駛人服務品質的關鍵因素。在薪資滿足方面，公營公車駕駛人重視外部公平性，而民營公車駕駛人重視內部公平性。在工作壓力方面，公營公車駕駛人主要受工作本身因素影響，民營公車駕駛人則受薪資獎金影響最大。在工作滿足方面，公營公車駕駛人最重視工作關係，民營公車駕駛人則重視公司制

度。除此之外，曾瑋悅亦在 2002 年利用集群分析法、變異數分析法及線性結構模式分析公車薪資制度對駕駛人營業績效的影響。該研究除了將各家公車業者依據薪資制度分為高、中、低三種程度的激勵性獎酬薪資制度外，並探討各薪資制度別之公車駕駛人在薪資知覺壓力、薪資滿足感、營業績效上是否有差異，同時分析駕駛人在薪資知覺壓力、薪資滿足感、營業績效三者之關係。研究結果顯示，薪資知覺壓力以高激勵性獎酬薪資制度之駕駛人最高，而薪資滿足感以中激勵性獎酬薪資制度之駕駛人最高，營業績效則以中激勵性獎酬薪資制度之駕駛人最高。另外，因薪資制度的不同，駕駛人在薪資知覺壓力、薪資滿足感、營業績效上皆呈顯著差異。同時駕駛人之薪資知覺壓力會負向影響薪資滿足感，薪資滿足感會正向影響實施績效。除了薪資會影響大眾運輸駕駛人之營業行為外，績效獎金制度的建立對於大眾運輸駕駛人之營業行為亦會產生影響。任維廉等(1996)利用問卷調查的方式對公路客運業駕駛人進行績效獎金制度對駕駛人滿足感的有效程度影響程度，進而瞭解績效獎金制度對駕駛人工作動機的影響。研究結果顯示，績效獎金制度在工作量回饋、工作品質回饋、獎金與投入的公平性等變項和駕駛人滿足感皆為顯著正相關。對組織目標的認同感與對績效獎金制度的認同感是最重要的調節變項，認同感高低不同對績效獎金有效程度、滿足感、工作動機都會有顯著差異。該研究發現，若將獎金制度與薪資津貼合併觀察，金錢為駕駛人最重視的因素。

2.1.3 計程車駕駛人駕駛行為相關文獻

計程車在公共運輸系統中屬於副大眾運輸，相較於大眾運輸駕駛人，計程車駕駛人在營業限制上較大眾運輸駕駛人為少。計程車駕駛人沒有固定的營業路線及營業時間，其收入也較不固定。因此，計程車駕駛人的營業行為將會比大眾運輸駕駛人更為複雜。除此之外，衛星派遣系統的引進對計程車駕駛人營業行為也造成了影響。對衛星派遣車隊計程車駕駛人而言，如何利用衛星派遣技術增加其

收益將是一個很重要的課題。

由於計程車駕駛人之營業行為相當複雜，國內外對於計程車駕駛人之營業行為研究很少，然而，衛星派遣計程車駕駛人之營業行為該如何分析呢？侯勝宗(2006)利用田野調查與問卷調查兩種方式對台灣大車隊的計程車駕駛人進行駕駛行為分析，分析內容主要著重在科技意會的層次，亦即在計程車駕駛人的心裡，衛星派遣此項科技之於計程車駕駛人的意義為何？該研究首先對新加坡康福計程車(Comfort Taxi)車隊及台灣大車隊計程車駕駛人進行田野調查，藉此瞭解二地的計程車駕駛人如何採用衛星派遣科技，並歸納出科技採用者的科技心理擁有感將如何影響對科技的意會，與使用科技的在地知識類型，進而影響科技採用型態與採用頻次。其中科技採用頻次該研究採廣義的定義，亦即包含接獲衛星派遣之次數及因品牌之外溢效果而得之乘客數量。爾後透過對台灣大車隊中數百位計程車駕駛人駕駛行為的觀察、訪談及問卷調查，並利用科技心理擁有感、在地知識、科技意會量表的發展與線性結構方程式統計分析將各構面分別分為兩個類型，進而得知計程車駕駛的科技採用類型與採用績效。

透過前述的分析將新加坡康福車隊與台灣大車隊的計程車駕駛人分別分群，分群結果如圖 2-2 及圖 2-3 所示。由圖 2-2 可知，康福車隊的計程車駕駛人駕駛行為大致可分為聰明效率型、風險趨避型、勤奮冒險型及傳統保守型四群，在尋找乘客的方式來講仍以街道巡迴為主，而車機設定模式則以手動競標模式較多。另外，如圖 2-3 可知台灣大車隊計程車駕駛人駕駛行為則分為聰明效率型、風險趨避型、急功近利型、傳統保守型及關係經營型五群，這五群司機的駕駛行為特性與心理狀態分述如下：

a. 聰明效率型：

聰明效率型駕駛人大都較為年輕、對科技的接受度高，他們相信藉由科技可以帶來個人利益，且對組織的認同度高，有利他的歸屬感，願意介紹大

車隊的好處給其他司機，也不會害怕競爭。

b. 風險趨避型：

風險趨避型駕駛人的年紀較大、工作能力較差，所以科技使用頻率較低。雖安裝派遣科技系統，但卻較少依賴科技來進行營業行為。其加入台灣大車隊主要的原因是利用公司的專屬簽約排班點進行定點的排班，並享受台灣大車隊的品牌效益，在街道巡迴時較容易被叫車。

c. 急功近利型：

急功近利型駕駛人的開車行為相當積極，偏好在街上巡迴、同時等候衛星派遣機會。他們的組織歸屬感不高，對科技存在一種自私擁有的感受，且不願意介紹其他司機加入大車隊，深怕太多的車隊駕駛人會瓜分自己所能分配到的派遣數。在駕駛技能方面，急功近利型駕駛人熟悉大街小巷的地理知識，時常穿梭於小巷中以更迅速的接到更多的乘客。

d. 傳統保守型：

因為個人特質及地理知識較少的關係，傳統保守型駕駛人不敢利用衛星派遣的方式攬客。他們傾向於掌握人潮出現的時間與地點，進行小區域、特定時間的營業行為。其加入台灣大車隊的主要原因是利用台灣大車隊的品牌效益，增加他們路邊載客的機會。

e. 關係經營型：

關係經營型駕駛人對城市的地理知識並不豐富，喜好經營特定客群，以定點載客為主，較不喜歡街道巡迴攬客。他們通常學識較佳，部份熟悉外文，加入車隊前已有穩固的客群，其加入台灣大車隊的主要原因是希望藉由品牌的社會形象，提升他們的社會形象，同時增加乘客對駕駛人的尊重，並間接的擴大客源。

科技車機設定模式

		自動設定模式	手動競標模式
尋找乘客的實務作法	定點排班	型一：聰明效率型採用 科技應用在地知識 實用型科技意會 (13%)	型二：風險趨避型採用 機構基礎在地知識 悲觀型科技意會 (19%)
	街道巡迴	型三：勤奮冒險型採用 地理導向在地知識 功利型科技意會 (30%)	型四：傳統保守型採用 時間相關在地知識 控制型科技意會 (38%)

圖 2-2：新加坡康福車隊科技採用模式

資料來源：侯勝宗(2006)

		司機對科技的心理擁有感受	
		高	低
尋找乘客的實務作法	定點排班	型一：聰明效率型採用 科技應用在地知識 實用型科技意會 (11%)	型二：風險趨避型採用 機構基礎在地知識 保護型科技意會 (23%)
	街道巡迴	型三：急功近利型採用 地理導向在地知識 自利型科技意會 (26%)	型四：傳統保守型採用 時間相關在地知識 消極型科技意會 (31%)

型五：關係經營型採用
 顧客相關在地知識
 認同型科技意會
 (9%)

圖 2-3：台灣大車隊科技採用模式

資料來源：侯勝宗(2006)

除了前述的分群方法，侯勝宗等(2008)利用派遣數（媒合效率）及收入（生產力）為衛星派遣計程車駕駛人分群關鍵因素，將台灣大車隊計程車駕駛人分為理性、知性、感性、野性及鈍性五群。表 2-1 為理性、知性、感性及鈍性駕駛人的營業行為特色，由表 2-1 中可知衛星派遣使用率最高的為知性司機，經由知性司機自身經驗的累積在加上衛星派遣的利用，知性司機每月可創造十萬元的高薪。而收入最高的是感性司機，感性司機利用優良的服務創造高回客率以保障每日的載客數，衛星派遣系統之於感性司機猶如街道指引系統，彌補感性司機地理知識不足之缺憾。

表 2-1：各群駕駛人營業行為特色表

駕駛人分群	特色	派遣數 (件/月)	收入 (元/月)
理性	維持傳統定點排班及街道巡迴模式，同時也利用空中排班點進行派遣，在等候空排點派遣的時也在街道中巡迴攬客，每日上下班時間固定。	100~240	約四萬元
知性	擅長探勘黃金點，知道何時何地較易有計程車需求，並積極的接受派遣指令，更進一步研究出何處較易接獲派遣指令。	300 以上	約十萬元
感性	將衛星派遣科技融入服務中，利用衛星派遣科技建構自己的人際與顧客管理關係，提供高品質的載客服務以提高回客率，	50~100	約十四萬元至二十萬元
鈍性	車內髒亂、談吐失節，且不知如何利用派遣科技優勢來經營行車路線，遑論藉此創造更高收入。	90 以下	約二萬元

資料來源：本研究根據侯勝宗等(2008)之研究修正而得

2.2 資料探勘相關文獻

與其他常用的分析方法相比，資料探勘為一新興的技術。由於科技的進步及資料庫系統的普及，各領域儲存的資料量日益增加，若能有效的使用資料庫中的

資料，將能更真實的探索出問題的答案，於是資料探勘技術逐漸被應用在各種產業當中。在運輸的範疇中，也有許多資料庫系統的存在，舉凡智慧卡、衛星定位系統等都是運輸的資料庫系統。在這些資料庫系統中隱藏了許多有用的資訊，若不能將資料庫中龐大資料加以應用實屬可惜，因此研究者引入資料探勘技術對運輸資料庫系統進行分析。透過資料探勘技術，研究者可將大量的資料分析出有用的資訊，藉此解決更多問題。以下將針對利用資料探勘技術解決運輸上的問題進行文獻回顧，並將回顧的內容分為解決公共運輸問題及其他運輸問題兩部份。

2.2.1 資料探勘於公共運輸之應用

資料探勘技術被運用在許多領域中，在運輸領域也有很多的研究。本節將進行公共運輸領域資料探勘技術應用的文獻回顧，希望藉此增加對研究方法的瞭解。

在公共運輸的應用方面，以智慧卡資料庫之資料探勘較為多見。國內智慧卡資料庫的應用以台北悠遊卡系統資料庫之應用最為常見，洪浩凱(2002)利用台北悠遊卡系統資料庫，透過資料探勘的方式對搭乘捷運的顧客進行研究。由於對捷運業而言，如何針對顧客需求的逐步改變來改換其策略相當的重要，而智慧卡可大量儲存消費者的使用習性資訊以及其他各項特性正好能解決此需求。因此該研究以瑞士顧客滿意度指標模式(SWICS)及歐洲顧客滿意度指標模式(ECSI)為基礎，利用悠遊卡系統資料庫，分析從悠遊卡帶來的的特性對於顧客忠誠度的影響。另外，乘客搭乘紀錄的資料也被準確的紀錄在悠遊卡的資料庫中，因此羅帷元(2007)便利用資料探勘技術針對台北市悠遊卡的交易資料庫進行分析，針對悠遊卡交易資料庫的資料特性，將交易資料分為「班次內交易間的關係」及「相同卡號不同交易日間的關係」兩個面向探討，並將兩個面向所發展的方法結合處

理，找出公車使用者的起迄對，藉此產生公車旅次起迄分佈表。邱詩淳(2006)進一步結合台北市悠遊卡系統資料庫及公車衛星定位系統資料庫，透過悠遊卡系統資料庫及公車衛星定位系統資料庫的資料探勘，有效的改善了過去研究中欠缺完整且具代表性載客紀錄的缺點。該研究利用悠遊卡系統資料庫中之乘客實際交易資料及公車衛星定位系統內之公車到站時間資料，比對出乘客刷卡位置後，經由乘客搭乘紀錄的推估以建立乘客的 OD 矩陣，接著運用所得的乘客 OD 矩陣進行公車營業方式改善。

國外對於智慧卡資料庫資料探勘應用的例子也不少，如 Bagchi & White (2005)即利用智慧卡交易資料庫的資料探勘及問卷調查，分析在紹斯波特(Southport)、馬其賽特郡(Merseyside)及布拉福(Bradford)實施的敬老優惠政策效益。在分析的過程中，該研究利用智慧卡交易資料庫計算卡片交易率、旅次使用率並推斷每個車站的旅次。但由於公車僅在上車時刷卡，無法得知乘客下車位置，因此仍需透過問卷調查的方式進行旅次長度的計算。

除了智慧卡資料庫外，公路運輸仍有結合資料探勘技術與其他方法的例子。在 Yao(2007)的研究中想利用多元回歸模型預測公共運輸系統潛在的通勤旅次需求，其中預測變數的取得有兩種方式，一為需求指數法(Need Index Method)，另一個是利用資料探勘中的自組織神經網路(Self-Organizing Map, SOM)來尋找集群在高維向量空間中的預測變量。該研究利用美國喬治亞州(Georgia)首府亞特蘭大(Atlanta)在 2000 年的資料進行分析，分析結果顯示需求指數法適用於較高層次的潛在需求分析，而在較低層次的潛在需求分析則以資料探勘技術為佳。

在鐵道運輸的應用方面，有鐘志成等(2009)利用台鐵的票證紀錄檔為基礎，利用台鐵局之「平/假日之短/中/長程站間流量」彙整報表進行資料探勘，透過電腦模擬指派的方式重建旅客於系統內的搭乘行為，並藉此分析旅客於系統內的時

空分佈情形。此模式的發展將有助於決策者客觀檢視客運服務之使用狀況，作為列車班次之增刪、停靠站選擇或起迄點調整之參考。

在航空運輸的應用方面，亦不乏資料探勘技術的使用。林祥生等(2008)利用航空公司線上購票資料庫中的顧客資料及交易紀錄，透過資料探勘技術中的集群化、分類及關聯分析，建立一個新的旅客價值指標，協助航空業者明確分辨顧客對公司業績貢獻的高低，相較於目前業界多僅以哩程及頻率來作粗略判斷，更具客觀性。在航空運輸中的服務品質及顧客行為業可利用資料探勘來分析，Liou等(2010)利用資料探勘技術中的優勢關係粗糙集(Dominance-based Rough Set Approach, DRSA)分析航空公司的服務服務品質，藉由 DRSA 的應用除了可以刪除一些對服務品質影響不大的變數，並可藉此制定一組新的服務品質衡量標準。同年，Liou 等亦利用資料探勘中較新的 DRSA 技術來分析航空市場的顧客行為，藉由 DRSA 技術的導入，可以解決過去利用問卷調查分析顧客行為中的某些困境，過去在問卷設計中常因自然語言(Natural Language)影響受試者的回答，亦即每個問項的描述皆會影響受試者的答案，透過資料探勘技術可有效的解決這個問題。

2.2.2 資料探勘於其他運輸問題之應用

除了公共運輸的領域外，在其他運輸領域也時常出現資料探勘技術的蹤跡。本節將對其他運輸領域使用資料探勘技術的相關文獻進行回顧，由此瞭解資料探勘技術在運輸領域中的應用項目。

在事故資料庫應用方面，吳冠宏等(2004)利用大量的車禍資料進行資料探勘，透過 K-Means 集群分析的方法對各類車禍進行分群的動作，由分群結果找出車禍事故與其發生地點特徵之關係，藉此於車禍事故具高發生率之處設置警示

標誌，以減少車禍事故的發生。另外，在肇事危險模式的建立上也有資料探勘技術導入的例子，黃湄清(2005)利用肇事資料分析事故之人為、車輛機械、天候環境、道路設計間的關聯性，透過資料探勘技術中的集群化技巧與判別分析，建立了一套適合台灣地區之肇事危險模式，並配合交通事故紀錄檔與易肇事目標之評定制，以增加肇事危險的預測能力提昇交通安全。在國外的例子中，也不乏事故資料庫資料探勘的應用，Chang(2006)利用資料挖掘探勘中之分類及回歸樹(Classification and Regression Tree, CART)，分析影響事故嚴重程度之主要原因與其影響程度，藉此改善一般統計方法面對大量事故資料時誤刪重要資訊的缺失，以提高運輸安全分析的精確度。

除了上述事故資料庫的應用外，資料探勘技術亦被應用於其他比較特別的議題當中。在最佳化問題之中，Wong(2008)利用資料探勘技術的導入提出一個相互配合以解決時制最佳化的方法，根據資料庫的數據利用演算法的迭代過程，該研究找出一種估計每日時制計劃表(Time-of-Day, TOD)的改善步驟，藉此解決 TOD 時間間隔不準確的缺點，以找出最佳的號誌時制。陳宜萍等(2009)利用資料探勘技術建立一路權取得輔助工具，該研究採用美國德州的路權管理資料庫，藉由資料探勘技術中的分類技術演算法建立一路權取得輔助工具，以瞭解影響延滯路權用地取得之因素，藉此加速路權取得之流程以減少交通工程興建之延滯。除此之外，人口普查資料之資料庫也可利用資料探勘技術將之有效運用，Clark(2009)利用英國政府 2001 年的人口普查資料，透過資料探勘的方式分析個人因素與家戶組成對汽車持有標準的影響，藉此瞭解未來道路基礎設施建設的方向。

2.3 綜合評析

上述駕駛行為的文獻回顧內容可整理如表 2-2，由表 2-2 可知，在公共運輸及計程車駕駛人的研究當中，顯示收入為影響其駕駛行為的重要變數，應特別注意之。此外，在計程車駕駛人駕駛行為的研究中，另將科技心理擁有感、在地知識、科技意會、科技採用頻次、派遣數納入討論，本研究將考慮採用該變數或其相應變數納入分析範圍中。

表 2-2：駕駛人駕駛行為文獻整理表

駕駛人類別	研究資料取得方法	顯著變數
私有機動運具	實地觀察、實地實驗、錄影分析、問卷調查、歷史資料統計、建立模型等，其中以問卷調查佔最多數。	性別、年齡、汽車駕照的取得與否、事故經歷、駕駛經驗、違規經驗、開車年資、職業為最常對駕駛行為有影響的變數。
大眾運輸	深度訪談、問卷調查、歷史資料，其中以深度訪談及問卷調查佔最多數。	年齡、年資、學歷、物理環境、獎懲制度、薪資制度、工作時間、心理環境、上司督導、同事相處為最常對駕駛行為有影響的變數。
計程車	田野調查及問卷調查。	科技心理擁有感、在地知識、科技意會、科技採用頻次、派遣數、收入為最常對駕駛行為有影響的變數。

在過去的研究中，不管是何種類別的駕駛人駕駛行為分析，研究資料的取得都是以問卷調查的方式佔最多數，亦即在駕駛行為分析的議題中，資料取得的來源並不多，需透過問卷調查的方式加強研究資料的廣度。然而，在龍天立等(1987)及王建仁(2003)的研究結果都曾提及受訪者容易有隱惡揚善的習慣，亦即利用問卷調查所得之結果不見得會與受訪者的實際行為相符。為改善此困境，本研究以計程車衛星派遣資料庫進行資料探勘，以實際的駕駛人營業數據分析計程車駕駛

人駕駛行為。透過實際數據的導入，還原駕駛人實際的駕駛行為。

第二部份資料探勘文獻回顧本研究整理如下頁表 2-3，由表 2-3 中得知，資料探勘技術在運輸議題中並不少見，且運用的層面相當廣泛，舉凡公共運輸裡的捷運、公車、鐵路、航空，其他部份如交通事故、肇事、車流、路權管理等，都曾經有導入資料探勘技術的例子。但由資料庫類別來看，有許多研究所使用的都是同一個資料庫，如在大眾運輸的議題中使用的資料庫幾乎都是台北悠遊卡資料庫，這也說明了事實上在同一個資料庫中，其隱含了相當多方面的資訊，足夠為多方研究所用。除此之外，同一個資料庫被使用的重複性高，也代表當某一資料庫被使用後，將增加其他研究者對該資料庫的關注，由此將由同一個資料庫發展出各類不同的研究。前述現象隱含了一個事實，截至目前為止尚有許多資料庫未被妥善利用，如本研究將使用的計程車衛星派遣資料庫即是一例，若能透過本研究引發其他研究者對計程車派遣資料庫的興趣與想法，對此資料庫進行其他議題之分析利用，將可發揮計程車派遣資料庫資料更大的效用，應為美事一樁。

綜合上述駕駛行為及資料探勘的文獻回顧，本研究擬利用計程車衛星派遣資料庫進行計程車駕駛人營業行為分析，此舉不但可以解決駕駛行為資料取得不易或不精確的困難，亦能將計程車衛星派遣系統的資料庫作妥善的運用，以發揮衛星派遣系統更大效能。

表 2-3：資料探勘方法應用整理表

學者	研究主題	研究對象	研究方法
洪浩凱 (2002)	探討捷運旅客忠誠度	台北悠遊卡系統 資料庫	典型相關分析結 果與結構方程式
Bagchi & White (2005)	探討敬老優惠政策效 益	紹斯波特、馬其賽 特郡及布拉福智 慧卡交易資料庫	配合問卷調查分 析
邱詩淳 (2006)	建立公車乘客的 OD 矩陣	台北悠遊卡系統 資料庫及台北公 車衛星定位系統 資料庫	集群分析及關聯 法則
羅帷元 (2007)	產生公車旅次起迄分 佈表	台北悠遊卡系統 資料庫	K-Means 演算法
Yao (2007)	預測公共運輸系統潛 在的通勤旅次需求	亞特蘭大公共運 輸系統資料庫	需求指數法及自 組織神經網路
林祥生等 (2008)	建立旅客價值指標	航空公司資料庫	集群分析、分類及 關聯分析
鐘志成等 (2009)	分析旅客於系統內的 時空分佈情形	台鐵票證紀錄檔	電腦模擬指派
Liou et al. (2010)	分析航空公司的服務 服務品質	航空公司資料庫	優勢關係粗糙集
Liou et al. (2010)	分析航空市場的顧客 行為	航空公司資料庫	優勢關係粗糙集
吳冠宏等 (2004)	分析車禍事故與其發 生地點特徵之關係	車禍資料	K-Means 集群分 析
黃湄清 (2005)	建立一適合台灣地區 之肇事危險模式	肇事資料	集群分析與判別 分析
Chang (2006)	分析影響事故嚴重程 度之原因與影響程度	事故資料	分類及回歸樹
Wong (2008)	建立最佳化的號誌時 制演算法	車流資料	演算法
陳宜萍等 (2009)	建立一路權取得輔助 工具	路權管理資料庫	分類演算法
Clark (2009)	分析個人因素與家戶 組成對汽車持有標準 的影響	人口普查資料	分類樹



第三章 研究方法

本研究主要採用資料探勘(Data Mining)的方式進行計程車駕駛人行為分析，跟隨資料探勘的分析方式進行資料選取、資料淨化、資料整合、資料選擇、資料轉換、資料探勘、特徵分析等步驟，藉以尋找出潛藏在計程車派遣資料庫中有關駕駛人行為的資訊。在進行資料探勘的過程中，由於本研究資料庫的資料量過於龐大，因此使用因素分析法(Factor Analysis, FA)進行變數縮減，以提高分析的效率及廣度。接著利用集群分析法(Cluster Analysis, CA)進行計程車駕駛人分群，將計程車駕駛人分成若干個集群，再分別回推各群司機的原始變數資料。最後分別對各群駕駛人進行隨機抽樣，利用一般統計檢定的方式找出各群計程車駕駛人的行為特性。本章將上述方法論進行深入探討後，提出本研究採用之資料分析模式。

3.1 資料探勘 (Data Mining)

資料探勘技術原是為了第二次世界大戰時人口普查及軍事用途而生，隨著資料庫系統的普及及資訊科技的進化，各領域儲存的資料日益增加，傳統的資料查詢方式已無法快速的應付如此龐大的資料量，因此資料探勘技術便廣泛的被應用於其他領域之中。

資料探勘技術的是在資料庫中，利用各種方法將過去的資料進行分析、歸納的方法。慣性及規律廣泛的存在於各個領域之中，利用資料探勘技術便可將資料中潛在的規律找出，做為決策時的參考依據進而改善現況。以下分別針對資料探勘技術的定義、目的及步驟一一說明。

3.1.1 資料探勘基本概念

資料探勘技術是知識探索(Knowledge Discovery in Database, KDD)的步驟之一，又名為資料發掘(Data Dredging)、資料考古學(Data Archaeology)、資料樣型分析(Data Pattern Analysis)及功能相依分析(Functional Dependency Analysis)等(駱至中等，2002)。透過合適的演算法之分析計算後，資料探勘技術可挖掘出潛藏在資料中的特徵(Pattern)、趨勢(Trend)、關聯性(Relationship)及模式(Model)，經過整理的過程萃取出我們所需的資訊。

資料探勘技術的定義在學界眾說紛紜，Thuraisingham(2000)提及資料探勘為從現有的資料中搜尋出有價值的隱藏事件，並且加以分析，使能從資料中獲取有更意義的資訊或是歸納出具結構化的型式，以作為決策時參考的依據。Frawley(1991)認為由資料庫中發掘出不明確的、先前未知的及潛在可能非常有用的資訊之過程即為資料探勘。Grupe&Owring(1995)認為資料探勘是從現存資料庫中，挖掘出新的及專家尚未發現的新事實及新關係。學界關於資料探勘之定義尚有很多，本研究將各學者對資料探勘之定義整理如表 3-1、表 3-2、表 3-3。

表 3-1：資料探勘之定義整理

學者	定義
Frawley (1991)	由資料庫中發掘出不明確的、先前未知的及潛在可能非常有用的資訊之過程即為資料探勘。
Grupe & Owring (1995)	資料探勘是從現存資料庫中，挖掘出新的及專家尚未發現的新事實及新關係。
Fayyad et al. (1996)	資料探勘是從大量資料中選取合適的資料，進行資料轉換、處理等工作，找出資料中有效的、潛在有用的、易於了解的特徵，再進行結果評析過程，亦即資料探勘只是知識探索的步驟之一。

參考資料：本研究彙整黃湄清(2005)、陳宜欣(2006)、王宜慶(2009)等研究而得

表 3-2：資料探勘之定義整理（續）

學者	定義	
Cabena et al. (1997)	資料探勘是將未知且有用的資訊從龐大的資料庫中取出的過程，除此之外並將萃取出來的有用資訊提供給上位者做為決策的參考。	
Berry & Linoff (1997)	資料探勘是為了發現有意義的模式或規則，以自動或半自動的方式，來勘查、分析大量資料所進行的流程。	
Kennedy et al. (1998)	資料探勘是一種結合資料視覺化(Data Visualization)、資料倉儲(Data Warehousing)、機器化學習(Machine Learning)及統計分析等多種方法的技術，利用這些方法可以從大量資料中，萃取出有規則的模式或有用的資訊。	
Peacock (1998)	最廣義	最廣義的資料探勘等同於知識探索(KDD)，它包含了以下六個步驟： 1. 內外部資料取得。 2. 資料的轉換、淨化及格式化。 3. 分析、確認及賦予資料意義。 4. 資料庫評析。 5. 建立與執行決策工具及系統，使資料探勘結果可提供給決策者使用。 6. 持續的修正模型與維護系統。
	廣義	廣義的資料探勘用於確定或測試探勘過程中所發現的關係。使用統計方法、建立假說或進行研究並確認關係以支持在狹義的資料探勘中發現的理論、模式。
	狹義	狹義的資料探勘指的是自動發現隱藏在資料庫中有趣但不明顯的樣式，可能會影響策略和組織目標。狹義的資料探勘圍繞在機器化學習的方法論上，主要強調發現的過程。
Thuraisingham (2000)	資料探勘是從現有的資料中搜尋出有價值的隱藏事件，並且加以分析，使能從資料中獲取有意義的資訊或是歸納出具結構化的型式，以作為決策時參考的依據。	
Han & Kamber (2001)	資料探勘是指由資料庫中的大量資料、資料倉儲或其他資訊儲存體中擷取出有價值之知識，亦即將資料轉換成知識的行為。	

參考資料：本研究彙整黃湄清(2005)、陳宜欣(2006)、王宜慶(2009)等研究而得

表 3-3：資料探勘之定義整理（續）

學者	定義
Roiger & Geatz (2003)	資料探勘是從整個資料庫中的資料利用一種或多種電腦技術以自動分析或擷取知識的過程，其目的為在龐大的資料中找出趨勢與特徵。
曾憲雄等人 (2005)	資料探勘可用以解釋資料庫的知識探索，亦即從大型的資料庫中萃取出一些有趣的知識、規則。
尹相志 (2007)	資料探勘是利用統計方法及機器學習的演算法，啟發性的從大量資料中找尋具有商業價值的趨勢及規則，作為自動化商業策略之應用。

參考資料：本研究彙整黃湄清(2005)、陳宜欣(2006)、王宜慶(2009)等研究而得

綜合各學者專家對資料探勘技術的定義，本研究將資料探勘技術的特色歸納為以下幾點：

1. 資料探勘是將未知且有用的資訊從龐大的資料庫中取出的過程，從現存大量的資料庫中選取合適的資料，以自動或半自動的方式進行資料轉換、處理等工作，挖掘出新的、不明確的、及潛在有用的新事實及新關係，再進行結果評析即完成資料探勘，它僅是知識探索的一部份。
2. 資料探勘的目的是藉由其分析方法，獲取有意義的資訊或是歸納出具結構化的型式，以作為決策時參考的依據。
3. 資料探勘是一種綜合的分析方法，它結合了資料視覺化(Data Visualization)、資料倉儲(Data Warehousing)、機器化學習(Machine Learning)及統計分析等多種方法的技術。
4. 資料探勘一種啟發性的方法，與演繹式的統計有根本上的不同。資料探勘除了傳統統計方法的假設，雖在過程中應用了統計的技巧，但資料探勘利用演算法自動搜尋有意義的規則，讓資料自然顯現出其意義所在。

5. 過去的研究中，傳統統計方法及透過問卷進行的市場調查一直頗為學術界質疑，因此僅能作為參考亦無法自動化的連結至商業行為中。資料探勘不但可確實的反映出研究對象的行為，亦可藉由一連串的自動化流程連結至實際情況中，增加其商業化應用及即時反映的效果。

由前述資料探勘定義可知，資料探勘是一種結合了許多種其他分析方法的技術。這些分析方法單獨使用時通常需被約束在某種限制下，經過各種分析方法的結合，大部分的問題都能被解決。針對不同的問題選擇各種不同的分析方法進行研究，這就是資料探勘的一個重要功能。圖 3-1 為資料探勘目的的架構圖，由圖 3-1 中可知資料探勘的目的分為預測性(Predictive)及描述性(Descriptive)兩類。預測性資料探勘又稱為監督式資料探勘(Directed Data Mining)，內容包含分類、推估及預測(Classification)、推估(Estimation)、預測(Prediction)；描述性資料探勘又稱為非監督式資料探勘(Undirected Data Mining)，內容包含集群化(Clustering)、視覺化(Visualization)、序列規則(Sequence Pattern)及關聯規則(Association Rule)。監督式與非監督式資料探勘的差別在於是否有特別觀察某個變數，非監督式資料探勘的目的是找出所有的變數中關聯性的有無，以下整理自黃湄清(2005)、郭信宏(2008)，分別針對各分析功能做詳細的介紹。

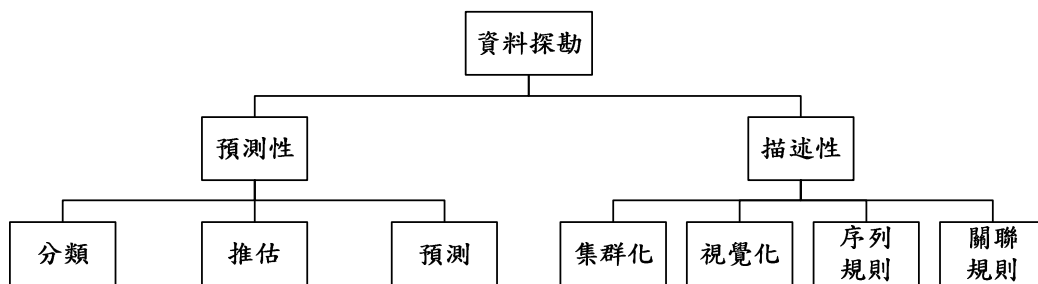


圖 3-1：資料探勘目的的架構圖

資料來源：本研究依 Dunham(2002)之研究整理而得

1. 分類

從資料庫中的歷史資料擷取已知的屬性及規則並加以分類，推導得存在資料中的特徵及模式，爾後檢視新資料的特性，並依據之將其歸類到預先定義好之分類中。常用的方法如決策樹(Decision Tree)、貝氏分類法(Bayes Classifier)、記憶基礎推論法(Memory Based Reasoning, MBR)、類神經網路(Artificial Neural Network)等。

2. 推估

推估與分類都是利用離散或連續的變數來推測資料潛在的特徵及模式，而其不同點在於，分類出來的結果是離散的，但推估所得的結果，則是連續性的數值。例如以一個人的月收入、機車的有無，推測其通勤花費稱為推估；若以相同的原始資料推測其運具選擇類型則稱為分類。常用的方法包含相關分析(Correlation Analysis)、回歸分析(Regression Analysis)、類神經網路方法等。

3. 預測

任何預測都只是推估的一部分，預測是根據某特定對象，觀察其過去行為或歷史資料建立模型，預測未來的狀態、數值以及趨勢。相關技術包括回歸分析、類神經網路、決策樹、時間序列分析(Time Series Analysis)、案例庫推理(Case-Based Reasoning)等。

4. 集群化

集群化又稱為資料切割，根據資料間的相似性，將資料分成若干群體，使組間差異最大、組內差異最小。當資料非常雜亂、或包含太多變數、太多

維度時，即可先進行集群化的動作將資料概廓的分類。集群化完成後，必須利用其他方法來瞭解集群的意義。常用的方法包含階層式集群分析(Hierarchical Cluster Analysis)、K-Means 集群分析(K-Means Cluster Analysis)等。集群化常與分類混淆，集群化和分類的不同點在於事前明確定義、規則的有無。分類是依照先前已有的類別將新資料逐一歸類，而集群化則是在不知道資料有何規則關聯性前即進行的分類行為。本研究先利用集群化的方式將計程車駕駛人進行分群，爾後利用統計方法找出各群駕駛人的特徵。

5. 視覺化

資料視覺化是一種強力的描述式資料探勘，透過資料視覺化的過程，研究者能很容易的從一個複雜的資料庫中解釋出一個狀態，透過良好的視覺化程序，亦能發現許多演算法無法判別的規則型態。

6. 序列規則

序列規則的目的在於了解不同時間點上各事件的關聯性。主要分為順序性與週期性兩種，順序性用於了解事件發生之時間先後關係，週期性用於瞭解某個時間區段內的變化，並觀察在其他相同時間區段內該事件是否亦會發生。常用的方法有 Apriori 演算法、FP-Growth 演算法、決策樹、類神經網路及時間序列分析。

7. 關聯規則

關聯法則的目的為瞭解龐大資料庫中某些資料項目彼此之間的關聯，這些關聯通常不是顯而易見的常識或規則，亦即找出在某一事件或資料中會同時出現的項目。典型的例子是判定超市中，哪些物品會一起被購買。常用的

分析方法如購物籃分析(Market Basket Analysis)、模糊集合(Rough Set)、連結分析(Link Analysis)等。同質分組的目的，是判定哪些事物會一起出現。

由上述介紹可知，許多方法可以達到多種的目的，如決策樹不但可以進行分類、預測，更是序列規則中不可或缺的方法之一。因此在進行資料探勘前，研究者需先瞭解各方法的優缺點並同時確定研究目的，以選擇最佳的資料探勘分析方法。

3.1.2 資料探勘應用流程

根據 Fayyad(1996)的定義得知，資料探勘僅為知識探索流程的其中一個步驟，完整的資料探勘或知識探索的流程分為資料選取(Data Selection)、資料處理(Data Processing)、資料轉換(Data Transformation)、資料探勘(Data Mining)、結果解釋與評估(Interpretation and Evaluation)五個步驟（如圖 3-2），以下分別介紹各步驟之詳細內容。

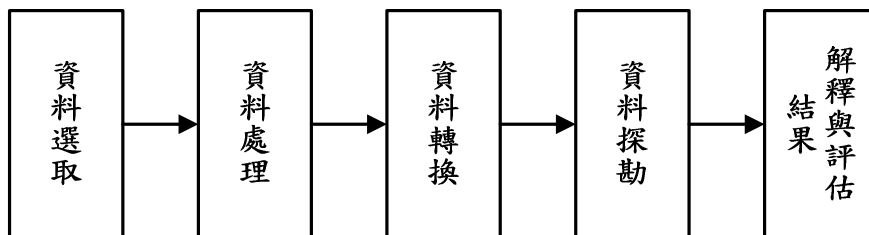


圖 3-2：知識探索應用流程

資料來源：Fayyad(1996)

1. 資料選取

知識探索的第一步，是根據知識探索的目的，從龐大的資料庫中選擇需要的資料做為分析基礎。然而資料選取並不需要使用特定資料庫中的所有資

料，從一個或多個資料庫中選取研究所需之資料庫進行資料探勘亦可。

2. 資料處理

資料處理又稱資料淨化(Data Cleaning)，它是知識探索中非常繁瑣卻相當重要的步驟。資料處理包含缺漏資料、異常資料的修正及重複資料、無效資料的刪除，以獲得正確、乾淨的資料。

3. 資料轉換

通常所需的資料類型與資料庫中資料的型式是不同的，因此在進行知識探索前需根據欲分析的問題，利用數學運算或其他方法，將目標變數由原始狀態轉換為可分析的形式，或將資料以不同維度(Domain)呈現，以突顯出目標變數的特徵。

4. 資料探勘

資料探勘方法有很多種如分類、推估、預測、集群化、關聯分析及序列規則等，利用前述適合的方法從轉換後的資料中，探索潛在的多種特徵、資訊及模式。

5. 結果解釋與評價

經過資料探勘分析後，透過文字或視覺化的圖形針對分析結果進行解釋，可將不同資訊探勘技術所獲得的知識進行整合，亦可經由專家的審查將不適宜的規則進行修正或刪除，最後綜合出結果納入系統中，並給予相應的評價。

根據以上知識探索的步驟，我們可以知道資料的選取及處理對知識探索結果影響很大。由於計程車派遣資料庫中的資料相當龐大，因此在資料探勘前本研究選擇因素分析法進行變數的縮減。另外，資料探勘方法的選擇也會影響知識探索結果之優劣，因此應根據研究目的選擇最適合的資料探勘方法。本研究希望對計程車駕駛人進行分群，因此選擇採用集群分析法將計程車駕駛人集群化。

3.2 因素分析 (Factor Analysis, FA)

因素分析起源於心理學，因為在心理學研究領域常遇到一些如智力、道德、操守等不能直接測量的因素，希望經由可測量的變數訂定出這些因素。為了解決這個問題，1904年 C.E. Spearman 在美國心理學雜誌(American Journal of Psychology)發表了一篇論文，這篇論文中將智力分為一般性因素(General Factor)及特殊性因素(Specific Factor)，藉由這個方法成功解決了人類智力衡量的問題，這就是因素分析的雛型。此方法剛開使用於解決心理學方面的問題，包含人格、動機核心理狀態的結構等，爾後逐漸被推廣至各個領域。

因素分析能將眾多的變數濃縮成為少數幾個有意義因素，且能保存住原有資料結構所提供的大部分資訊。利用因素分析可以有效的減少變數的數目，藉此減少在資料探勘過程中資料量過大帶來的負擔，以增加變數選取的廣度。以下分別針對因素分析的功能、模式、目的及步驟一一說明。

3.2.1 因素分析基本概念

因素分析是以少數幾個因素來解釋一群相互之間有關係存在的變數之數學模式，其主要功能包含潛在因素命名、潛在結構分析、因素相關分析及因素分數產生等。除此之外，因素分析對於研究測驗建構效度(Construct Validity)也相當適

合，藉由因素的發現可確立研究的結構成份。

黃俊英(2000)提出因素分析理論模式假設各觀察變數受各種因素影響，觀察值單位在某一變數上的反應是由二部分組成：共同因素(Common Factor)及獨特因素(Unique Factor)。共同因素又稱潛在因素(Latent Factor)，它代表各變數共有的部分，其個數會小於原始觀察變數；獨特因素代表各變數所獨有的部分，每個觀察變數只有一個。

因素分析是一種將 k 個變數 $X_i(i=1\sim k)$ 分別分解成少數幾個共同因素 $Y_j(j=1\sim n)$ 與獨特因素 e_i 的線性組合，由於每一個變數皆有一個獨特因素，故在 k 個變數的情形下，即有 k 個獨特因素，而共同因素的數目小於等於變數的數目，亦即 $n \leq k$ 。在此種情形下，因素分析的模式如下所示：

$$X_i = f_{i1}Y_1 + f_{i2}Y_2 + \dots + f_{ij}Y_j + e_i \quad (\text{式 3.1})$$

其中，

X_i 為第 i 個觀察變數， $i=1\sim k$ ，其數值由觀察而得

f_{ij} 為因素負荷量(Factor Loading)或組型負荷量(Pattern Loading)，亦即第 i 個觀察變數在第 j 個共同因素時的因素負荷（回歸係數）；

Y_j 為第 j 個共同因素， $j=1\sim n$ 無法直接觀察必須由因素分析產生；

e_i 為第 i 個觀察變數所對應的獨特因素。

上述因素分析模式，有下列幾個基本假設：

1. 共同因素 $Y_1\sim Y_k$ 間互相獨立且變異數為 1。
2. 獨特因素 $e_1\sim e_k$ 間互相獨立，且為平均數為 0 的常態分配。
3. 共同因素 $Y_1\sim Y_k$ 及獨特因素 $e_1\sim e_k$ 間皆獨立，亦即 $\text{Cov}(Y_j, e_i) = 0$ 。

根據因素分析的目的，可將其分為探索性因素分析(Exploratory Factor Analysis, EFA)及驗證性因素分析(Confirmatory Factor Analysis, CFA)兩種類型。前者是用以解釋一組變數的相關性，並藉此簡化原有變數；後者用於檢定變數間的假設關係或考驗理論架構，於線性結構關係的結構方程模式中較為常見(莊涵婷，2004)，以下分別說明之：

1. 探索性因素分析

探索性因素分析又稱解釋性因素分析，用於試探、描述、分類及分析一組變數的相關程度。通常研究者對其觀察變數所能表達哪些因素並不清楚，且沒有預先提出其觀察變數可測出幾個共同因素之研究假設，此時可利用探索性因素分析，藉由回歸方法分析變數間的關係同時找出共同因素，並計算因素分數以簡化原有變數。圖 3-3 為探索性因素分析路徑圖，由圖 3-3 中可知探索性因素分析的觀察變數 X_i 是由各個共同因素 Y_j 所組合而成。

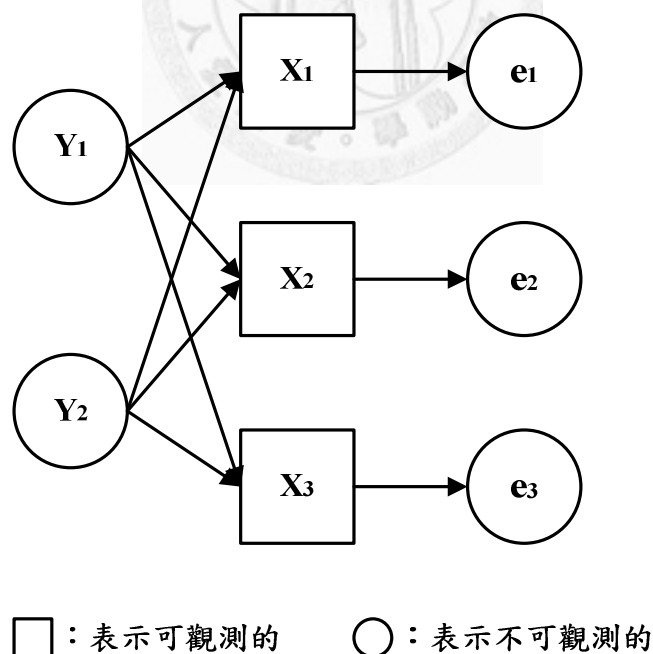


圖 3-3：探索性因素分析路徑圖

參考資料：羅惠瓊(2005)

2. 驗證性因素分析

驗證性因素分析又稱確認性因素分析，用於驗證理論架構與實際資料之相容性。在觀察變數 X_i 與所萃取之潛在共同因素 Y_j 間有一定理論架構時，利用驗證性因素分析可進行模式架構的確認。圖 3-4 為驗證性因素分析路徑圖，由圖 3-4 可知在驗證性因素分析中並非每個共同因素 Y_j 皆與觀察變數 X_i 間有連線，亦即在驗證性因素分析中觀察變數不一定是由每個共同因素所組合而成，有可能出現觀察變數與共同因素間無相關的情形。一般使用線性結構關係模式(Linear Structural Relations Model, LISREL)進行驗證性因素分析。

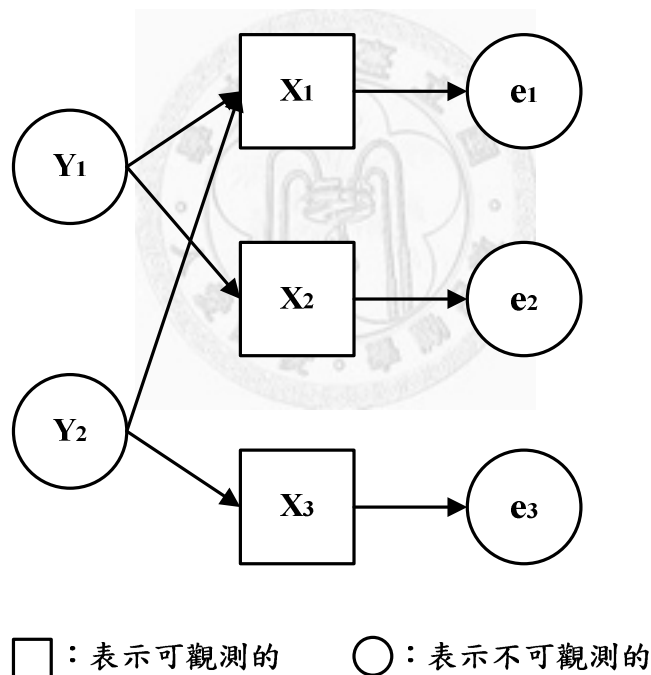


圖 3-4：驗證性因素分析路徑圖

參考資料：羅惠瓊(2005)

本研究自計程車衛星派遣資料庫中選取資料進行分析，在變數選取前無法得知其潛在結構性模式及關係，根據以上的介紹，可知本研究適合使用探索性因素分析進行變數縮減，因此以下將介紹探索性因素分析的應用流程。

3.2.2 因素分析應用流程

根據潘書麟(2006)，本研究將因素分析的流程分為五項(如圖 3-5)，由圖 3-5 可知因素分析的流程依序為變數共同性估計、因素萃取、因素數目決定、因素轉軸、因素解釋及命名，以下將分別介紹各步驟之詳細內容：

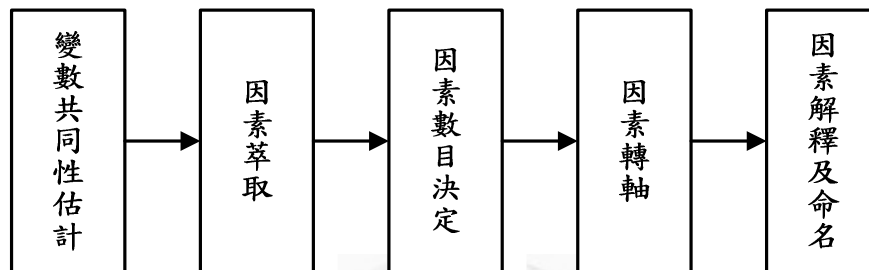


圖 3-5：因素分析應用流程

參考資料：潘書麟(2006)及本研究整理

1. 變數共同性估計

進行因素分析的變數間需具有一定程度的相關，一群相關太高或太低的變項，皆會造成執行因素分析的困難。變數間相關性太低難以抽取一組穩定的因素，不適合進行因素分析。然而，相關太高的變數，線性重合(Multicollinearity)的現象明顯，容易造成分析效度不足的問題，所獲得的因素結構價值不高。本研究利用 Kaiser-Meyer-Oklin 取樣適切量數(Kaiser-Meyer-Oklin Measure of Sampling Adequacy, KMO)檢定及 Bartlett 球形檢定(Bartlett's Test of Sphericity)進行變數共同性估計，以確立因素分析的效度。

- KMO 檢定的檢定值的意義為變數間共同因素的數量，它僅是一個比較值而非實際數量。KMO 檢定值越大，表示變數間的共同因素越多越適

合進行因素分析。一般而言，KMO 檢定值需高過 0.5 進行因素分析才有意義，下表 3-4 是 KMO 檢定值與因素分析適合度間的關係。

表 3-4：KMO 檢定值與因素分析適合度關係

KMO 檢定值	0.9 以上	0.8~0.9	0.7~0.8	0.6~0.7	0.5~0.6	0.5 以下
因素分析適合度	極適合	適合	尚可	勉強可	不適合	非常不適合

資料來源：羅惠瓊(2005)

- Bartlett 球形檢定可用來檢驗母體相關矩陣間是否有共同因素的存在，顯著的 Bartlett 球形檢定表示相關係數足以作為因素分析抽取因素之用。

2. 因素萃取

因素分析中萃取潛在因素的方法包含主成份分析法(Principal Component Analysis, PCA)、主軸法(Method of Principal Axes)及重心法(Centroid Method of Factoring)、最大概似法(Maximum Likelihood Analysis)等。過去研究中以主成份分析法及主軸法使用較為普遍，其中主成份分析法適用於簡化大量變項為較少數的成份時，以及做為因素分析的前置作業，與本研究研究目的較符合，因此本研究採用主成份分析法進行因素萃取。

進行因素萃取時因素負荷量是研究者所關切的主要項目之一，因素負荷量的觀念類似於回歸分析中自變數的權重，它反應了觀察變數對各共同因素的關聯強度。在進行因素萃取時，我們習慣於取用因素負荷量的平方值為估計依據，因素負荷量的平方值表示各共同因素可以解釋各變數的解釋變異量。亦即因素負荷量的平方值越大，萃取出因素越能反映出原變數的特性。因此根據各共同因素可解釋變異量的大小，可決定該項是否需納入共同因素

中。表 3-5 為因素負荷量、解釋變異量及選取準則，由表 3-5 中可知當因素負荷量為 0.71 時能解釋各變數 50% 的變異量，因素負荷量越高能解釋的變異量越高，但當因素負荷量為 0.32 時僅能解釋各變數 10% 的變異量，因素負荷量越低能解釋的變異量越低，應捨棄之。

表 3-5：因素負荷量、解釋變異量及選取準則

因素負荷量	解釋變異量	該項變數狀況
0.71	50%	相當理想(Excellent)
0.63	40%	非常好(Very Good)
0.55	30%	好(Good)
0.45	20%	普通(Fair)
0.32	10%	不好(Poor)

資料來源：Tabachnick and Fidell(2007)

3. 因素數目決定

共同因素萃取的原則為因素越少越好，而萃取出之因素能解釋各變數的變異數越大越好。一般而言，因素的個數決定於特徵值(Eigenvalue)的大小，特徵值表示某一因素可解釋的總變異量，特徵值越大亦即該因素的解釋力越強。以下列舉幾個常用的因素數目決定準則（陳菊萍，2003）：

a. Kaiser 準則：

學者 Kaiser 在 1960 年提出以特徵值等於 1 為因素決定的臨界值，由於每個變數的變異數為 1，因此將特徵值小於 1 的共同因素視為對總變異量貢獻度低並予以刪除，僅保留特徵值大於 1 的共同因素。此法也是目前電子計算機程式中最常用的一種。

b. 古特曼最強下限(Gutman's Strongest Low-Bound)準則：

所謂「古特曼最強下限」準則是以特徵值等於 0 為因素決定的臨界值，共同因素的特徵值為正予以保留，反之則刪除。這種準則是一種較為保守的標準，可以防止主要的共同因素被忽略。

c. 陡坡考驗法(Scree Test)：

陡坡考驗法是利用因素陡坡圖進行判斷，特徵值通常是由大至小依序出現，因素陡坡圖是以因素編號為橫軸、特徵值大小為縱軸畫成的一條自左上往右下驟降的圓 L 型負加速曲線圖。通常在出現幾個較大的特徵值後此曲線會急速下降，爾後下降情形變慢趨近於定值，此法選擇刪除右尾趨於定值之因素。然而，趨於定值的判斷過於主觀，需小心使用。

d. 經驗法則：

學者 Harman 在 1960 年提出有時在統計分析上發現有意義的共同因素，在實際應用上卻不甚重要，因此在相信統計分析結果時，必須同時考慮實際應用上的意義，才不會失去進行因素分析的意義。

本研究採用 Kaiser 準則，以特徵值大於 1 的因素為共同因素，並同時考慮取出的因素是否符合實際情況，以萃取出最精簡的因素。

4. 因素轉軸

完成上述步驟後，有時會發現因素分析初步的結果顯示某些變數同時在好幾個因素都有相當程度之負荷量，造成模式的解釋相當困難。為了將因素意義明朗化且利於進行因素之命名，因此在不改變共同性的前提下進行因素

轉軸。藉由因素轉軸法，因素負荷量可被調整為接近 1 或 0 的數值，亦即將因素負荷量大者變大、小者變小，使觀察變數與共同因素間的關係更明顯。轉軸後大部分的觀察變數只歸屬於一個或少數幾個因素，使得萃取出之共同因素與觀察變數間的關係更為明確。

常用的轉軸方法分成直交轉軸法(Orthogonal Rotation)及斜交轉軸法(Oblique Rotation)兩類。表 3-6 為直接轉軸法與斜交轉軸法之綜合比較，由表中可知斜交轉軸法較直交轉軸法貼近事實，然而在實際的情況中，研究者大多無法確定因素間的關係，再加上直交轉軸法的結構較為明確，大多數的研究仍採直交轉軸法進行分析。本研究將採用最普遍被運用的最大變異法進行因素轉軸，找出觀察變數與因素間的關係。

表 3-6：直接轉軸法與斜交轉軸法綜合比較

項目 \ 類型	直交轉軸法	斜交轉軸法
特色	在轉軸過程中，因素之間的軸線夾角為 90 度，亦即因素之間的相關為 0。	認為因素間有某種程度的關係，因此對各變數軸間不限制保持直交，容許因素間存在某種程度的共變。
常用方法	最大變異法(Varimax)、四次方最大值轉軸法(Quartimax)、相等最大值變異法(Equimax)。	直接斜交轉軸法(Direct Oblimin)、Promax 轉軸法。
優點	因素間的資訊不會重疊，對於因素結構的解釋較為容易，概念較為清晰。	精確的估計變項與因素關係並較貼近事實。
缺點	常與事實不符。	無法作不同研究間之比較。

資料來源：本研究依羅明旋(2009)之研究整理修增而得

5. 因素解釋及命名

各因素經轉軸後其意義將更明顯，利用各觀察變數與因素之關係，研究者可計算出各因素的數值，此值為各觀察變數經由加權得到的數值，其權重為該觀察變數在該因素中的因素負荷量。計算完成後，並可依據各因素內觀察變數的特性命名。

根據以上因素分析的步驟，我們可以順利找出對計程車駕駛人行為影響較大的數個因素，藉由這些因素便可進行計程車駕駛人集群化的動作。本研究採用集群分析進行計程車駕駛人分群，下節將詳細介紹集群分析法的基本概念及應用步驟。

3.3 集群分析 (Cluster Analysis, CA)

集群分析起源於分類學，在原始的分類學中，主要依據經驗和專業知識進行分類，並沒有定量分類的概念。隨著科學技術的發展，對分類精確度的要求也隨之增加，因此但憑經驗和專業知識進行的分類已無法滿足研究者的需求。有鑑於此，研究者將數學工具引入了分類學中，形成了數值分類學(Numerical Taxonomy)，爾後又加入多元分析技術的概念演變成今日的集群分析。

集群分析是將具有 n 個特性之觀察值，在沒有先驗知識的前提下，依其類似性將類似的觀察值歸於同一群的過程，所以同一集群中的觀察值相似性很高，而不同集群間的對象相異性很大（吳鴻春，2009）。集群分析應用於許多領域，在行銷學(Marketing)中，常依特性之不同，將市場加以分類，亦即所謂的市場區隔(Marketing Segmentation)。其他如數學，電腦科學，統計學，生物學和經濟學等，

都能發現集群分析方法的應用。本研究利用集群分析進行衛星派遣計程車駕駛人營業行為的分群，藉此進一步分析各群衛星派遣計程車駕駛人之行為特性。以下分別針對集群分析的目的、相似性衡量、集群建立方式及步驟一一說明。

3.3.1 集群分析基本概念

集群分析主要目的是利用客觀的計量方法分析資料彼此之間的相似程度，根據一組準則變數(Criterion Variables)將 n 個觀察值分為 k 個群別 (其中 $k \leq n$)，使得同一集群內的觀察值具有高度同質性(Homogeneity)，而不同集群間具有高度的異質性(Heterogeneity)，藉此找到集群結果並推論出隱含的特性及現象。由前述集群分析主要目的可知，在集群分析中判斷相似性為一重要課題，衡量資料間相似性(Inter-Object Similarity)的方法有很多種，根據黃俊英(2000)及林震岩(2007)大致可分為距離衡量(Distance-Type Measure)和配合衡量(Matching-Type Measure)及相關衡量(Correlation-Type Measure)三類。以下分別說明之：

1. 距離衡量：

距離衡量主要是以點與點間的距離來表示，點與點之間距離越短者表示相似性越高，亦即越傾向於同一集群。距離衡量的方法主要有明考斯基距離(Minkowski Distance)、城市街道距離(City-Block Distance)、歐基里得距離(Euclidean Distance)、馬氏距離(Mahalanobis Distance)四種。

a. 明考斯基距離：

明考斯基距離假設有 n 個觀察值，每個觀察值有 m 個屬性，則第 i 個觀察值與第 j 個觀察值間的距離為 d_{ij} ，則明考斯基距離為：

$$d_{ij} = \left(\sum_{k=1}^m |X_{ik} - X_{jk}|^r \right)^{\frac{1}{r}} \quad (\text{式 3.2})$$

其中， X_{ik} 與 X_{jk} 分別是在 m 維空間中， i 點與 j 點對變數 k 的投影。

明考斯基距離是一種較廣義的距離計算公式，很多的距離公式都是此公式的特例。

b. 歐基里得距離：

歐基里得距離是多維空間中兩點的幾何距離，假設有 n 個觀察值，每個觀察值有 m 個屬性，則第 i 個事物與第 j 個事物間的歐幾里德距離為：

$$d_{ij} = \sqrt{\sum_{k=1}^m (X_{ik} - X_{jk})^2} \quad (\text{式 3.3})$$

由式 3.3 可知，歐幾里得距離是明考斯基距離在 $r=2$ 時之特例，同時它也是在幾何空間中，我們計算距離最常使用的表示方式。另外，由於歐幾里得距離中有較麻煩的根號計算，因此在分析時常以歐幾里得距離平方計算，其公式如下：

$$d_{ij}^2 = \sum_{k=1}^m (X_{ik} - X_{jk})^2 \quad (\text{式 3.4})$$

c. 城市街道距離：

城市街道距離是各個變數差距之和，假設有 n 個觀察值，每個觀察值有 m 個屬性，則第 i 個事物與第 j 個事物間的城市街道距離為：

$$d_{ij} = \sum_{k=1}^m |X_{ik} - X_{jk}| \quad (\text{式 3.5})$$

由式 3.5 可知，城市街道距離是明考斯基距離在 $r=1$ 時之特例。在很多情況下，以城市街道距離做分群與歐幾里得距離做分群所得結果相似，但城市街道距離會將變數差距的效應減弱。在進行集群分析的情形下，由於需要處理大量的資料，判斷的依據為各觀察值間相對的距離，而非精確的距離，故一些電腦程式採用街道區距離，以節省電腦使用時間。然而，城市街道距離在使用上必須確認變數間沒有相關性，若非，則需使用馬氏距離。

d. 馬氏距離：

馬氏距離是歐幾里得距離平方的一種引申，假設有 n 個觀察值，每個觀察值有 m 個屬性，則第 i 個觀察值與第 j 個觀察值間的馬氏距離為：

$$d_{ij} = \frac{\sum_{k=1}^m X_{ik}}{\sqrt{\left(\sum_{k=1}^m X_{ik}^2\right)\left(\sum_{k=1}^m X_{jk}^2\right)}} \quad (\text{式 3.6})$$

馬氏距離加總組內變數矩陣以調整變數間的交互關係，特別適用於變數有相關性時。

2. 配合衡量：

配合衡量主要用在觀察值全部以類別變數(Nominal Variable)或虛變數來表示時，此時兩觀察值的相似性可用配合係數(Matching Coefficient)或相似比(Similarity Ratio)來衡量。

a. 配合係數：

由於類別變數無法用一般的數值方法進行分析，因此發展出配合係數的計算，配合係數的公式為：

$$S_{ij} = \frac{a+b}{m} \quad (\text{式 3.7})$$

$$a = \sum_{k=1}^m X_{ik} X_{jk} \quad (\text{式 3.8})$$

$$b = \sum_{k=1}^m (1 - X_{ik})(1 - X_{jk}) \quad (\text{式 3.9})$$

其中，a 為觀察值 i 和 j 皆有屬性數目，b 為觀察值 i 和 j 皆無的屬性數目，m 為屬性總數。

b. 相似比：

相似比為簡化配合係數而來，它僅考慮觀察值 i 和 j 皆有的屬性，不考慮觀察值 i 和 j 皆無的屬性，則相似比的公式為：

$$SR_{ij} = \frac{a}{m-b} \quad (\text{式 3.10})$$

3. 相關衡量：

相關衡量是先計算成對觀察值間的相關係數，爾後以各觀察值間相關係數的型態來判定觀察值間相似性。由於大多數集群分析的應用都強調數值的大小而非型態，因此相關衡量的實用性不高。

完成成對觀察值間的相似性衡量後，應考慮該如何進行觀察值歸納建立集

群。建立集群的方法分為階層式集群分析法(Hierarchical Clustering Method)與非階層式集群分析法(Nonhierarchical Clustering Method)兩大形式，結合兩種方法的集群分析則稱為兩階段集群分析法(Two Step Clustering Method)。階層式集群分析法是利用觀察值間某項量測的距離或相似性將觀察值連結，但事前並不知道分群的個數，通常可以樹狀圖(Dendrogram)表示。非階層式集群分析法則是於分析前依據其他研究或主觀認定事先決定分群數目，再進行集群化的過程，其中以K-Means 集群分析法最為常見。而兩階段集群分析法則是結合前述兩種方法，先利用階層式集群分析決定分群個數，接著以非階層式集群分析進行分群。以下將分別介紹階層式集群分析法、非階層式集群分析法及兩階段集群分析法。

1. 階層式集群分析法：

階層式集群分析法又稱系統聚類法，這種方法可將連續合併的過程以樹形圖表示，藉此找出最適合的分群數。階層式集群分析可分為凝聚法(Agglomerative Method)與分離法(Divisive Method)兩類，凝聚法是先將每個觀察值各自看成一類，而後根據相似性準則把相近的事物合併成為集群，直到所有的觀察值都併入同一集群為止。分離法與凝聚法的過程相反，它先把所有的觀察值視為同一集群，然後再依據相似性的準則把各事物劃分為較不相近的兩個集群，直到每個觀察值都自成一類為止。在一般的情況中，大多使用凝聚法進行集群分析，故以下將以凝聚法的介紹為主。

凝聚法進行的方式有很多種，各種方法主要的分歧點在於組間距離的計算方式，以下將介紹最常使用的單一連結法(Single Linkage Method)、完全連結法(Complete Linkage Method)、平均連結法(Average Linkage Method)、重心法(Centroid Method)、中位數法(Median Method)及華德法(Ward Method)六種。

a. 單一連結法：

單一連結法又稱最近法，其假設集群與集群間距離等於兩集群觀察值間最短的距離，其公式如下：

$$d_{A,B} = \underset{\substack{i \in A \\ j \in B}}{\text{Min}} d_{ij} \quad (\text{式 3.11})$$

由式 3.11 可知，A、B 兩群距離定義為 A 群內每一點到 B 群內每一點的距離最小值。單一連結法主要的缺點為「鏈接聚合」缺陷，容易造成大部分的觀察值皆被聚集在同一組，故最短距離法在研究上較少被使用。

b. 完全連結法：

完全連結法又稱最遠法，其分析方法一開始單一連結法相同，先將二個距離最短的點併為同個集群，但是其他點是否併入此集群中，是根據群內和群外的最長距離來決定其公式如下：

$$d_{A,B} = \underset{\substack{i \in A \\ j \in B}}{\text{Max}} d_{ij} \quad (\text{式 3.12})$$

由式 3.12 可知，A、B 兩群距離定義為 A 群內每一點到 B 群內每一點的距離最大值。

c. 平均連結法：

平均連結法一開始與前兩者相同，先將二個距離最短的點併為同個集群，但是其他點是否併入此集群中，是根據群內和群外各點的平均距離來決定，其公式如下：

$$d_{A,B} = \frac{\sum_{i \in A} \sum_{j \in B} d_{ij}}{n} \quad (\text{式 3.13})$$

其中，n 為全部距離的個數。

由式 3.13 可知，平均連結法把 A、B 兩群距離定義為 A 群內每一點到 B 群內每一點的距離平均，亦即把兩群間距定義為兩群中所有觀察值間距離的平均值，而非特殊點之間的距離。在應用上而言，平均連結法的集群效果較前兩者為佳應用較廣。

單一連結法、完全連結法及平均連結法在觀念上大同小異，主要的差別在於集群距離的計算方式，此三種方法集群距離計算方式可由圖 3-6 表示。

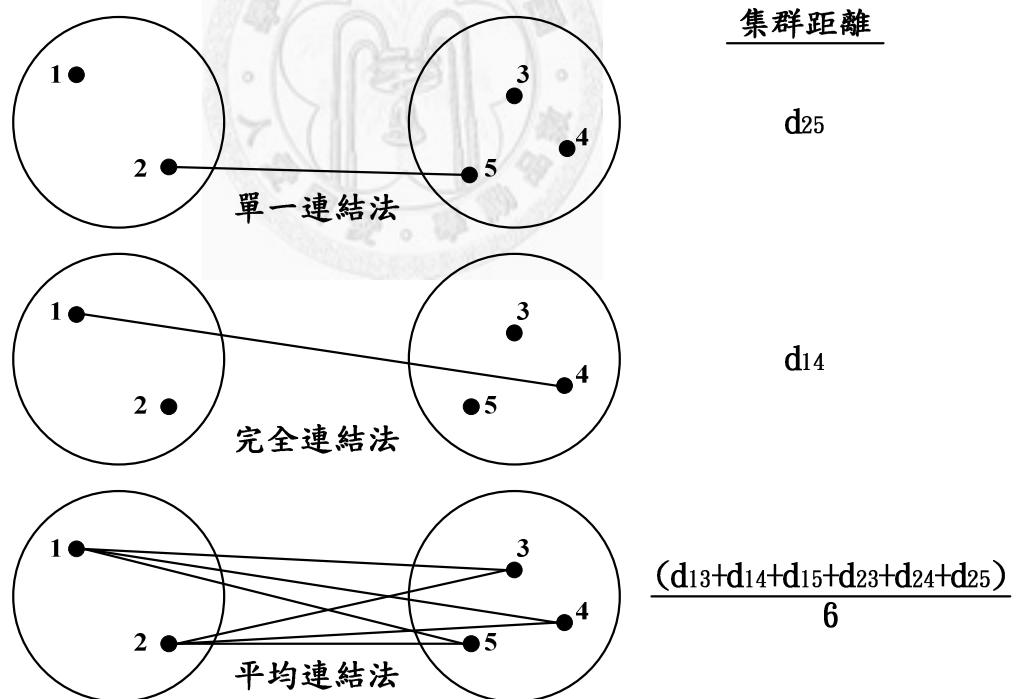


圖 3-6：集群間距離圖

資料來源：黃俊英(2001)

d. 中心法：

中心法亦稱形心法、重心法，每一群的中心點是該群中所有觀察值在各個變數上的平均值，中心法利用集群中心點決定是否需合併該二集群，除此之外每合併一次集群，都需要重新計算新集群的中心點。其公式如下：

$$d_{AB} = d(\overline{X}_A, \overline{X}_B) = \|\overline{X}_A - \overline{X}_B\|^2 \quad (\text{式 3.14})$$

由式 3.14 可知，A、B 兩群距離定義為 A 群中心點 \overline{X}_A 到 B 群中心點 \overline{X}_B 的距離。

e. 中位數法：

中位數法與中心法很類似，其差異僅在於中位數法把兩群之間的距離定義為兩群中位數之間的距離。其公式如下：

$$d_{AB} = d(\text{Med}(X_A), \text{Med}(X_B)) = \|\text{Med}(X_A) - \text{Med}(X_B)\|^2 \quad (\text{式 3.15})$$

由式 3.15 可知，A、B 兩群距離定義為 A 群中位數 $\text{Med}(X_A)$ 到 B 群中位數 $\text{Med}(X_B)$ 的距離。

f. 華德法：

華德法又稱最小變異數法(Minimum-Variance Method)，其概念為同一群內觀察值的變異數和應該較小，不同群之間觀察值的變異數和應該較大。其觀察值間距離衡量方式採用歐基里得距離。其公式如下：

$$d_{AB} = n_A \times \left\| \overline{X}_A - \overline{X} \right\|^2 + n_B \times \left\| \overline{X}_B - \overline{X} \right\|^2 \quad (\text{式 3.16})$$

由式 3.16 可知，A、B 兩群距離是以 A 群中心點 \overline{X}_A 到兩群合併中心點 \overline{X} 距離平方乘以 A 集群中觀察值的個數，與 B 中心點 \overline{X}_B 到合併中心點 \overline{X} 距離平方乘以 B 集群中觀察值個數之和。

在大部分的研究中，集群結果較佳的是華德法及平均連結法，而最差的是單一連結法，因此華德法與平均連結法為應用較廣泛的集群方法。

2. 非階層式集群分析法：

非階層式集群分析法又稱逐步聚類法，這種方法克服了階層式集群分析法在每一步的集群過程中都需要作很多計算並需儲存距離矩陣的缺點，有效的增加了可分析的樣本數，因此也適用於大樣本的集群分析。非階層式集群分析中最常使用的方法為 K-Means 集群分析法，K-means 集群分析與其他的集群分析法最大的差異在於：若觀察值的分群數為已知，則 K-Means 集群分析法可以將所有的觀察值分為指定的群數，另外相較於其他非階層式集群分析法 K-Means 集群分析法較不易受到不適當的屬性、離群值(Outlier)與不同的相似度之影響（蔡憲唐等，1987）。因此以下將利用 K-Means 集群分析法的步驟，介紹非階層式集群分析法的演算過程。K-Means 集群分析法的集群過程分為以下四個步驟：

- a. 指定要形成的集群數 k，並將觀察值分為 k 個集群，此時每個集群的觀察值成員是經過資料庫中任意選取產生的。
- b. 計算每個觀察值到各群中心點的歐基里德距離，並將每個觀察值歸入距中

心最近的集群以調整分群。

- c. 重新計算每群觀察值資料向量的中心點。

$$m_i = \frac{\sum_{p \in C_i} p}{|C_i|} \quad (\text{式 3.17})$$

其中， $|C_i|$ 表示 C_i 類別的集群觀察值數目。

- d. 重複步驟 b、c，直到各集群沒有觀察值可以再調整為止。

3. 兩階段集群分析法：

階層式集群分析法在每一步的集群過程中都需要作很多計算並需儲存距離矩陣，並且受離群值、相似衡量和不合適的分群變數的影響很大。非階層式集群分析法雖可彌補前述缺失，但非階層式集群分析法的集群結果對初始分群數非常敏感，因此將兩種方法結合起來的兩階段集群分析法因應而生。

兩階段集群分析法先利用華德法或其他階層式集群分析法決定集群個數，同時檢查是否有離群值並刪除之，接著對剩下的觀察值重新進行分群，以階層式集群分析法得到各個集群的中心點，做為非階層式集群分析法的初始分群中心。爾後以 K-Means 集群分析法進行集群化，移動各群組內的個體，直到各集群沒有觀察值可以再調整為止。

本研究希望利用計程車衛星派遣資料庫中駕駛人之營業資料對計程車駕駛人進行分群，進而分析計程車駕駛人之行為特性。因此採用兩階段集群分析法進行計程車駕駛人分群，以下將介紹兩階段集群分析法之應用流程。

3.3.2 集群分析應用流程

根據黃俊英(2001)之分析，本研究將集群分析的流程分為六項（見圖 3-7），由圖 3-7 可知及集群分析的流程依序為研究問題確定、變數選擇、相似性衡量、集群方法選擇、集群數決定、集群結果解釋，以下將分別介紹各步驟之詳細內容：

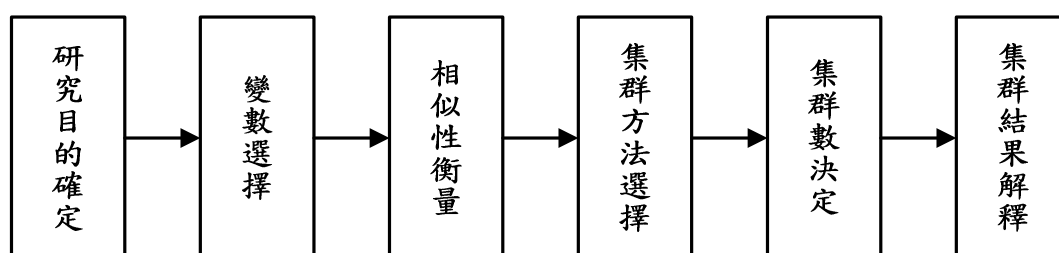


圖 3-7：集群分析應用流程

參考資料：本研究依黃俊英(2001)之研究修增而得

1. 研究目的確定：

首先要確定集群分析的主要目的是探索性或是驗證性。傳統的集群分析常用於探索性的用途，亦即將觀察值客觀的分群，使得每群群內元素對集群變數的同質性高。

2. 變數選擇：

無論研究的目的是探索性或驗證性，集群分析的結果都受到所選變數的限制。由於集群分析技術無法區分變數與議題間的相關性，加入與研究目的不相關的變數選將會增加產生離群值的機會，對於集群分析的結果會有重大的影響。因此變數的選擇必須兼顧理論、觀念和實務的考量，屏除盡可能加入更多變數的錯誤觀念。

3. 相似性衡量：

相似性衡量亦即量測觀察值間的相似性做為分群的基礎，集群分析將觀察值間相似性高的集合在一起，以形成數個集群。衡量觀察值間相似程度的方法大致可分為距離衡量、配合衡量及相關衡量三種，前節已詳細介紹在此不再贅述。本研究選擇距離衡量中的最常被使用的歐基里得距離做為相似性衡量的方式。

4. 集群方法選擇：

將成對事物間的相似性加以衡量後，接著就可利用集群方法將各觀察值歸入集群中。集群的方法可分成階層式集群分析法、非階層式集群分析法及兩階段集群分析法三類。此三類方法前節已詳細介紹在此不再贅述，惟階層式集群分析法只能得到各種不同集群數的集群結果，而非階層式集群分析法需自行決定分群數目且對分群數目的敏感度相當高，因此本研究採兩階段集群分析法進行分析。首先選擇階層式集群分析法中的華德法進行集群數目的決定，再根據決定的集群數目進行非階層式集群分析法中的 K-Means 集群分析，找出適當的分群集群。藉由兩階段集群分析的方式，不但可彌補非階層式集群分析法對集群數目敏感的缺點，亦可進一步檢定每個分群變數對集群產生的效果，以對每個集群進行描述性分析。

5. 集群數決定：

集群數目的決定，是集群分析中一項重要決策。理論上而言，集群數最好在二至四群間，因為當超過五群時，集群結果的解釋與命名將相當困難。但目前為止，集群數目的決定並沒有一套可依循的標準。惟可知集群數目越多，集群內觀察值的同質性越高，但越不易找出資料的結構；集群數目越少，

雖越能看出資料的簡單結構，但集群內觀察值的同質性則越低。因此，決定集群數目端看研究者的取捨。

集群數目的決定雖無一定的標準，但通常可以利用以列幾種方法進行集群數目的選擇（蕭文龍，2007）：

a. 理論值或實務上的考量：

有些議題在進行集群分析前已有確定的集群數，或者是在實際使用時已給定一個集群大小區間，此時便可直接利用理論值或者評估集群大小區間內的幾個集群數目來決定該次研究的集群數目。

b. 集群數距離激增：

執行階層式或非階層式集群分析的程序時，會發現集群步驟突然增加很多，此時代表群體間有較大的差異，可藉此決定集群數目。

本研究採用華德法中的階層凝聚係數來決定集群數目，當凝聚係數驟增時，表示要將集群從 n 群凝聚為 $n-1$ 群時的困難度急速增加，因此選擇將觀察值分為 n 群。

6. 集群結果解釋：

由集群分析找出集群後，應找出足以代表該集群的名稱。一般而言，研究者常以集群之重心(Centroid)來描述該集群，亦即使用集群內的各觀察值在各變數上的平均值描述。此外，亦可以集群的變異情形來描述之，如以集群內各點間的平均距離或各點與重心間之平均距離描述之。

3.4 方法論具體架構

根據上述方法論探討之後，本研究依循知識探索的步驟進行衛星派遣計程車駕駛人行為分析。由於本研究所採用的資料庫資料龐大，因此在主要分析前利用因素分析法進行變數的縮減，接著利用資料探勘進行衛星派遣計程車駕駛人行為分析。資料探勘使用的方法有很多，根據本研究之研究目的，選定集群分析法為本研究資料探勘的主要方法。利用集群分析法將衛星派遣計程車駕駛人分群，藉此找出衛星派遣計程車駕駛人之營業行為特性。本研究資料分析模式如圖 3-8 所示。

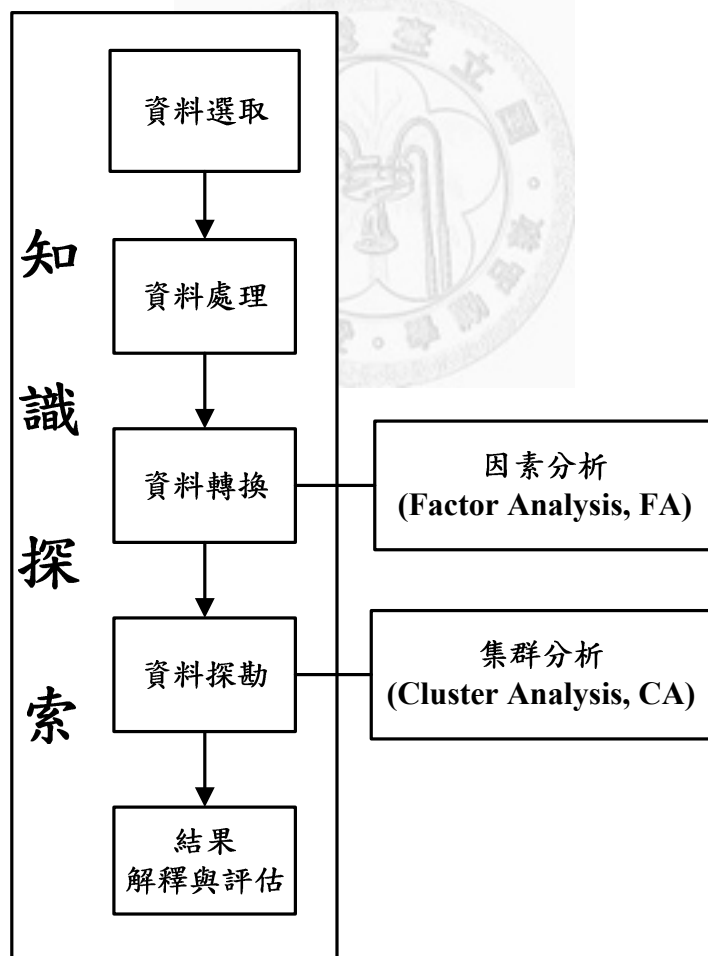


圖 3-8：資料分析模式



第四章 派遣資料前置處理

本章根據前述研究目的及相關研究探討，設計並說明衛星派遣計程車營業資料的蒐集及分析流程，並針對衛星派遣計程車營業資料進行前置處理，以利後續研究分析。

4.1 分析流程

本研究分析流程如圖 4-1 所示，包含資料選取、資料前置處理、資料變數縮減、月資料集群分析、日資料抽樣分析、集群結果整理、派遣邏輯改善建議，內容分別敘述如下：

1. 資料選取：

在資料探勘的過程中，首先需選定要探勘及分析之資料。然而資料選取並不需要使用特定資料庫中的所有資料，從一個或多個資料庫中選取研究所需之資料庫進行資料探勘亦可。因此，分析時應依據所需資料項目進行篩選，惟使用不同資料庫之資料時，應注意不同資料庫間資料的格式及單位等問題。

2. 資料前置處理

進行資料探勘時，資料的正確性與完整性是相當重要的。因此在分析前，需先將已選取之資料進行資料淨化，資料淨化包含缺漏資料的處理、重複資料及無效資料的刪除等，確認資料之正確及完整。爾後選取影響駕駛人營業行為之變數，並將資料轉換為分析軟體所能讀取之格式，以利研究的進行。

3. 資料變數縮減

由於本研究所選取之資料量過於龐大，故在進行分析前需先進行變數的縮減。本研究採用的變數縮減方式為因素分析法中之主成份分析法，利用主成份分析法將原始變數縮減為較具代表性的幾個因素，以利後續分析。

4. 月資料集群分析

由於單次載客資料過於龐大，本研究將資料分析分成兩階段。首先利用衛星派遣計程車月資料進行集群分析，爾後進行日資料抽樣分析。根據研究目的，本研究採用集群分析中的階層集群分析法(Hierarchical Clustering Methods)及 K-Means 集群分析法(K-means Clustering Methods)從派遣資料庫中挖掘駕駛人營業特性，並藉由集群分析的結果將衛星派遣計程車駕駛人進行分群。

5. 日資料抽樣分析

經過前述衛星派遣計程車駕駛人分群後，根據分群的結果，分別利用亂數表對各群駕駛人進行隨機抽樣。隨後分別對各群駕駛人樣本進行統計分析，藉此找出更多各群衛星派遣計程車駕駛人特徵，以利後續派遣邏輯的設計。

6. 營業行為分群結果整理

經過集群分析及抽樣統計分析後，可以得到最適之駕駛人分群，根據駕駛人分群統整各群駕駛人之營業行為特徵，此駕駛人分群特徵為後續派遣邏輯設計之依據。

7. 派遣邏輯改善建議

根據前述駕駛人分群特徵，分別提出考慮各群駕駛人特徵及需要之派遣邏輯改善建議，並粗略設計派遣邏輯。

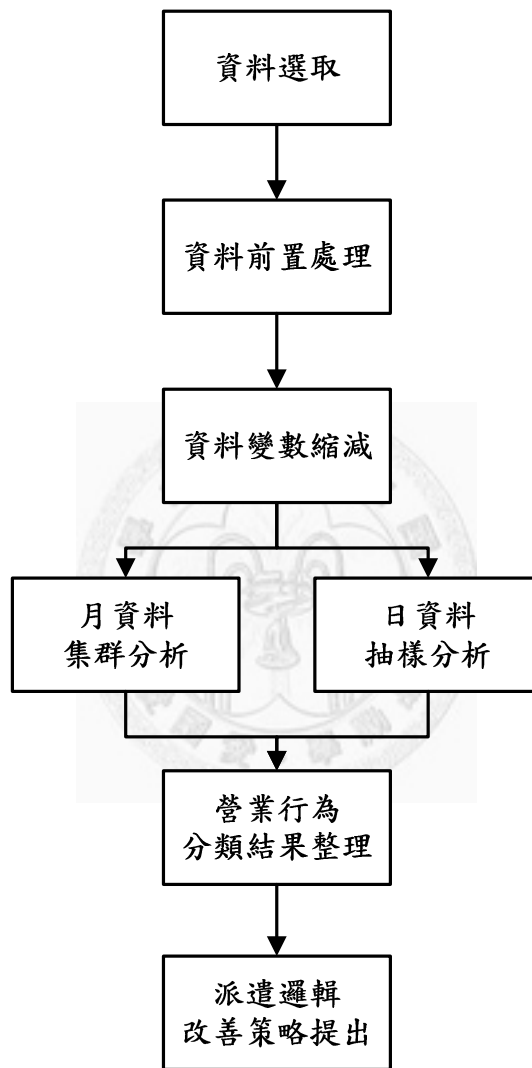


圖 4-1：分析流程

4.2 衛星派遣計程車營業資料選取及蒐集

本研究以大台北地區之衛星派遣計程車為主要分析對象，在大台北地區的衛星派遣車隊中，以台灣大車隊及大都會衛星車隊的規模最大，能取得之資料數較充足，故自該二車隊挑選分析之對象。由於本研究目的為改善計程車駕駛人之工作環境，故視駕駛人收入為重要資料，然而派遣資料中並不包含此項目，因此須以車輛之軌跡點及時間來推斷其收入。下表 4-1 為該二車隊之行車資料回傳頻率，其中大都會衛星車隊在有載客時 30 秒即回傳一次，對於推斷駕駛收入較不易有誤差，故選擇大都會衛星車隊之派遣資料庫進行分析。在時間範圍上，擷取 2008 年 5 月的資料進行分析。

表 4-1：車隊資料回傳頻率表

車隊名稱	車輛狀態	
	無載客	有載客
台灣大車隊	5 分鐘	3 分鐘
大都會衛星車隊	2 分鐘	30 秒

4.3 衛星派遣計程車營業資料前置處理

資料的前置處理為進行資料探勘前一個很重要的過程，不同領域的資料探勘所需之前置處理技術各不相同。本研究之資料前置處理步驟包含變數選取、資料雜訊清除、無效資料刪除、資料屬性及格式轉換、資料庫構建，詳細內容敘述如下：

1. 變數選取：

本研究根據 97 年營業情形調查之調查項目及侯勝宗(2006)的研究內容，同時根據研究者之觀察選取載客數、載客時程、載客里程、營業時程、營業里程、營業天數、總收入、平均日收入、距離空車率、時間空車率、平均速度及單次載客收入等十二項變數進行分析。十二項變數中，除收入資料外皆為派遣資料庫中包含之項目，駕駛人收入資料為本研究由派遣資料庫所提供之軌跡點推估而得。

2. 資料雜訊清除：

將各項資料中含有亂碼、有部份缺漏或是明顯不合理之資料予以刪除，並且檢核字元之正確性，同時刪除分析時不需要的資料。

3. 無效資料刪除：

由於計程車產業是一個相對自由的工作，駕駛人選擇辛勤工作或暫時休假都是不被約束的，因此在清除資料雜訊後，某些變數仍有不合理或不具代表性之處。為此，本研究根據九十七年度台北地區營業情形調查及試誤法進行資料刪除。本研究認為營業數據過高為駕駛人選擇辛勤工作以獲取更高的收入，而營業數據過低則會使資料不具代表性，因此選擇設立下限而不設立上限的方式進行資料刪除。

本研究首先根據九十七年度台北地區營業情形調查進行資料刪除，判斷的變數包含每日載客次數、每日載客里程、每日載客時程、每日營業里程、每日營業時程、每日營業收入。假設母體為常態分佈，根據中央極限定理 (Central Limit Theorem)，約有 95% 的數值介於 $\mu \pm 2\sigma$ 之間，因此設立平均值

減去 2 個標準差為下限值。表 4-2 為各變數所能接受的下限值，表 4-2 中某些變數之 $\mu-2\sigma$ 數值為負值以 0 計。由表 4-2 可知，雖對 6 個變數進行下限值的設立，其中僅有每日營業里程及每日營業時程兩個變數的下限值不為 0，分別以此 2 變數進行資料的刪除。

表 4-2：變數平均值、標準差及下限值表

變數 \ 項目	平均值	標準差	下限值
每日載客次數 (次)	11.06	5.71	0
每日載客里程 (公里)	54.61	30.17	0
每日載客時程 (小時)	2.38	1.26	0
每日營業里程 (公里)	128.87	62.73	3.41
每日營業時程 (小時)	12.17	5.58	1.01
每日營業收入 (元)	1,758.56	900.28	0

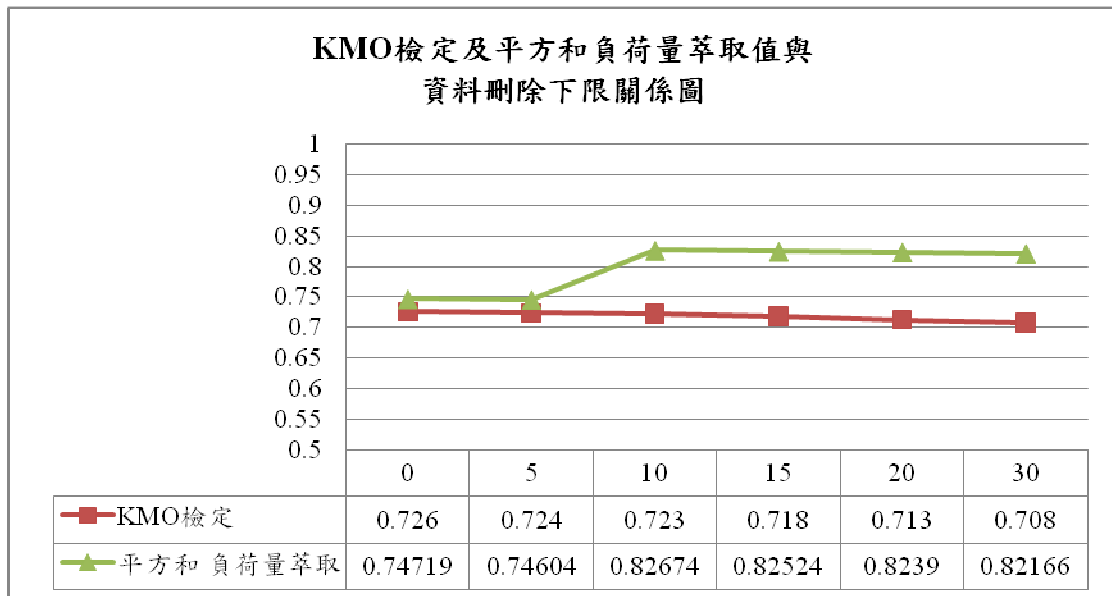
註：小於 0 之下限值以 0 計。

資料來源：本研究根據張學孔等(2008) 調查結果整理而得。

完成上述資料刪減後，觀察發現仍有些許該月份營業時數較短的資料，而九十七年營業情形調查並未調查駕駛人每月營業時數的資料，因此本研究以主成份分析當中所得之 KMO 檢定值及平方和負荷量萃取值為標準進行試誤，藉此找出最適合的每月最小營業時數。KMO 檢定的檢定值的意義為變數間共同因素的數量，此值最少需大於 0.5 方通過檢定，其值越高則顯示此次分析顯著性越高。平方和負荷量萃取值的意義則為萃取出成份能夠代表原變數群的程度，此值越高隱含萃取出成份越能代表原變數群。因此，本研究找出 KMO 檢定值及平方和負荷量皆高時之值設為下限值。圖 4-2 為 KMO

檢定及平方和負荷量萃取值與資料刪除下限關係，由圖可知其 KMO 檢定值皆落在通過檢定的區域且數值變化不大，而平方和負荷量萃取值在營業時數為 10 小時時驟增，因此選擇 10 小時為每月營業時數的下限值。

圖 4-2：KMO 檢定及平方和負荷量萃取值與資料刪除下限關係圖



4. 資料屬性及格式轉換：

為了順利進行分析，需將原始派遣資料庫資料屬性欄位在合理的情況下進行資料屬性定義轉換。原始整理過後的派遣資料庫為 Access 檔案，依照需要將資料庫檔案轉換為 Excel 及 SPSS 格式。

5. 資料庫構建：

將原始派遣資料經合併、轉換、篩選與整理完成後，依據研究目的建立本研究所需之資料庫，以進行後續分析。

4.4 衛星派遣計程車營業資料變數縮減

本研究採用因素分析法進行變數縮減，因素分析的第一步驟是檢定資料庫變數是否適合進行因素分析，一群相關太高或太低的變數都不適合進行因素分析，因此首先需對資料庫變數進行 KMO 檢定及 Bartlett 球形檢定。表 4-3 為本研究資料庫變數 KMO 檢定及 Bartlett 球形檢定的定結果，由表可知本研究資料庫變數的 KMO 檢定值為 0.723，此值高於 0.5 通過 KMO 檢定，顯示本研究資料庫變數有足夠的共同因素數量。Bartlett 球形檢定近似卡方值為 18161.94、自由度 66、顯著性 $p < 0.001$ ，顯示母體相關矩陣間有共同因素的存在。透過 KMO 檢定及 Bartlett 球形檢定，證實本研究採用之資料庫變數適合進行因素分析。

表 4-3：衛星派遣資料庫變數之 KMO 檢定及 Bartlett 球形檢定結果

項目	數值	
KMO 取樣適切性量數	0.723	
Bartlett 球形檢定	近似卡方分配	18161.94
	自由度	66
	顯著性	0

確認資料庫變數適合進行因素分析後開始進行因素萃取，本研究採用因素分析法中的主成份分析法進行因素萃取，並以相關矩陣為分析工具，利用 Kaiser 準則取相關矩陣特徵值大於 1 的因素為主要共同因素，輔以陡坡考驗法進行驗證，找出最適的因素個數。由表 4-4 可知在 12 個資料庫變數中特徵值大於 1 的變數有 3 個，累積的總解釋變異量為 82.67%，根據因素陡坡圖（圖 4-3）的趨勢，發現在成份四以後已趨於平坦，故本研究採用三個因素進行分析。

表 4-4：衛星派遣資料庫變數之總變異量表

成份	初始特徵值			平方和負荷量萃取			轉軸平方和負荷量		
	總和	變異數的%	累積%	總和	變異數的%	累積%	總和	變異數的%	累積%
1	6.853	57.106	57.106	6.853	57.106	57.106	6.696	55.803	55.803
2	1.944	16.198	73.304	1.944	16.198	73.304	2.036	16.969	72.771
3	1.124	9.370	82.674	1.124	9.370	82.674	1.188	9.903	82.674
4	0.916	7.635	90.308						
5	0.498	4.147	94.455						
6	0.402	3.354	97.809						
7	0.099	0.825	98.634						
8	0.067	0.562	99.196						
9	0.057	0.478	99.674						
10	0.022	0.184	99.858						
11	0.017	0.140	99.998						
12	0.000	0.002	100.000						

萃取法：主成份分析。

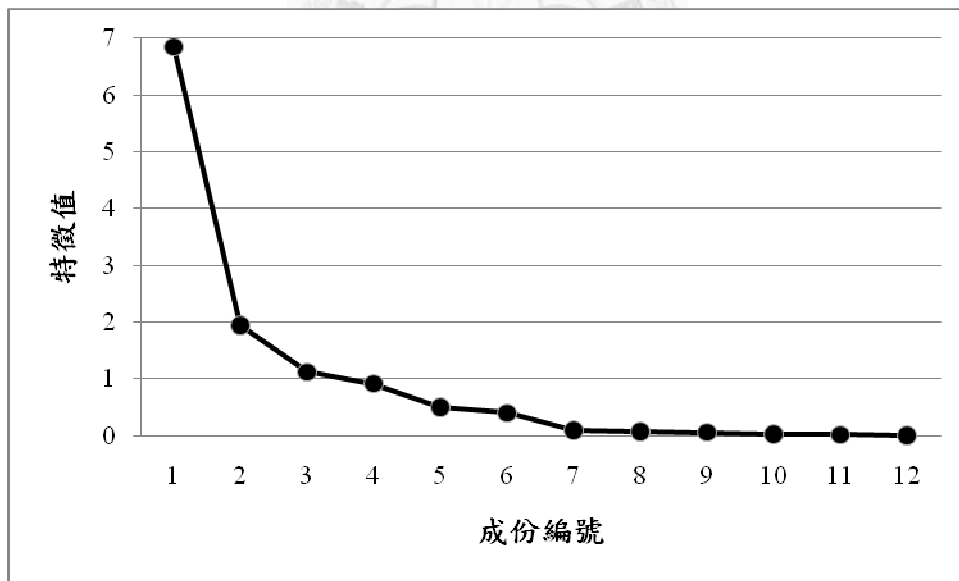


圖 4-3：衛星派遣資料庫變數之因素陡坡圖

決定因素數目後，再利用最大變異數法進行因素轉軸，使得萃取出之共同因素與資料庫變數間的關係更為明確，並以因素負荷量大於 0.5 為標準，選擇適合表達該因素的資料庫變數。表 4-5 為利用最大變異數法轉軸後的成份矩陣，由表 4-5 可知「成份一」包含載客數、載客時程、載客里程、營業時程、營業里程、營業天數、總收入及平均日收入八項，「成份二」包含距離空車率、時間空車率兩項，「成份三」包含平均速度、單次載客收入兩項。其中各資料庫變數之因素負荷量，即為其因素縮減時的權重。

表 4-5：衛星派遣資料庫變數之轉軸後的成份矩陣—最大變異數法

	成份		
	1	2	3
載客數	0.962	-0.186	-0.102
載客時程	0.944	-0.155	-0.074
載客里程	0.966	-0.193	0.086
營業時程	0.898	0.246	0.137
營業里程	0.931	0.046	0.062
時間空車率	0.040	0.897	0.175
距離空車率	-0.190	0.809	-0.112
總收入	0.973	-0.203	0.024
平均速度	0.096	0.261	0.727
營業天數	0.862	0.128	0.016
單次載客收入	-0.057	-0.170	0.755
平均日收入	0.731	-0.473	-0.003

萃取方法：主成份分析。

旋轉方法：含 Kaiser 常態化的 Varimax 法。

最後根據三個成份所包含的資料庫變數，分別對三個成份命名，賦予其實用上的意義。命名結果如表 4-6 所示，本研究分別將三個成份依據其特性命名為營業狀況、空車率及營業習慣。至此，本研究已由原本的十二項變數萃取出三項，下章則依據此三項成份進行計程車駕駛人分群並整理出各群計程車駕駛人之行為特性。

表 4-6：衛星派遣資料庫變數之因素命名表

成份	變數 個數	評估項目
營業狀況	8	載客數、載客時程、載客里程、營業時程、營業里程、營業天數、總收入、平均日收入
空車率	2	距離空車率、時間空車率
營業習慣	2	平均速度、單次載客收入





第五章 衛星派遣計程車駕駛人營業行為分析

透過前章的計程車衛星派遣資料前置處理，本研究已得到較正確且精簡的分群因素。本將將透過集群分析法對計程車駕駛營業行為進行分群，此處所用之分群因素為計程車駕駛人月營業行為之總和資料。爾後根據分群結果對計程車駕駛人進行分群抽樣，藉此分析各群計程車駕駛人每日營業行為特性。最後將統整前述計程車駕駛人月營業行為及日營業行為，以作為集群命名的依據。

5.1 衛星派遣計程車駕駛人月營業行為集群分析

本節將闡述計程車駕駛人營業行為分群的過程，本研究利用前章因素分析而得之分群因素，透過集群分析法的應用，將計程車駕駛人分為若干個集群，並回推各群駕駛人之營業行為特性，藉此分析計程車駕駛人之營業行為，並作為下節計程車駕駛人分群抽樣的依據。

5.1.1 集群分法與分析流程

根據第三章所述，集群分析方法分為階層式集群分析、非階層式集群分析及綜合兩者的兩階段集群分析。本研究採用兩階段集群分析的概念，先利用階層式集群分析中的華德法進行分析，並以凝聚係數驟增為衡量標準，由此找出適合的集群數目。爾後輸入決定的集群數目進行非階層式集群分析，在非階層式集群分析的部份，本研究使用 K-Means 集群分析法進行分群。分群完成後，對各因素進行單因子變異數分析(Analysis of Variance, ANOVA)，以確認因素的顯著性，並刪除不顯著之因素，同時重新進行前述集群分析過程，直到所有被考慮的分群因

素皆顯著時方確立分群結果，最後對分群結果進行分析，詳細流程如圖 5-1 所示。此外，本研究在進行分析前已先將分群因素標準化，藉此平衡各因素的數值大小，避免因其原始數值大小影響分群效果。

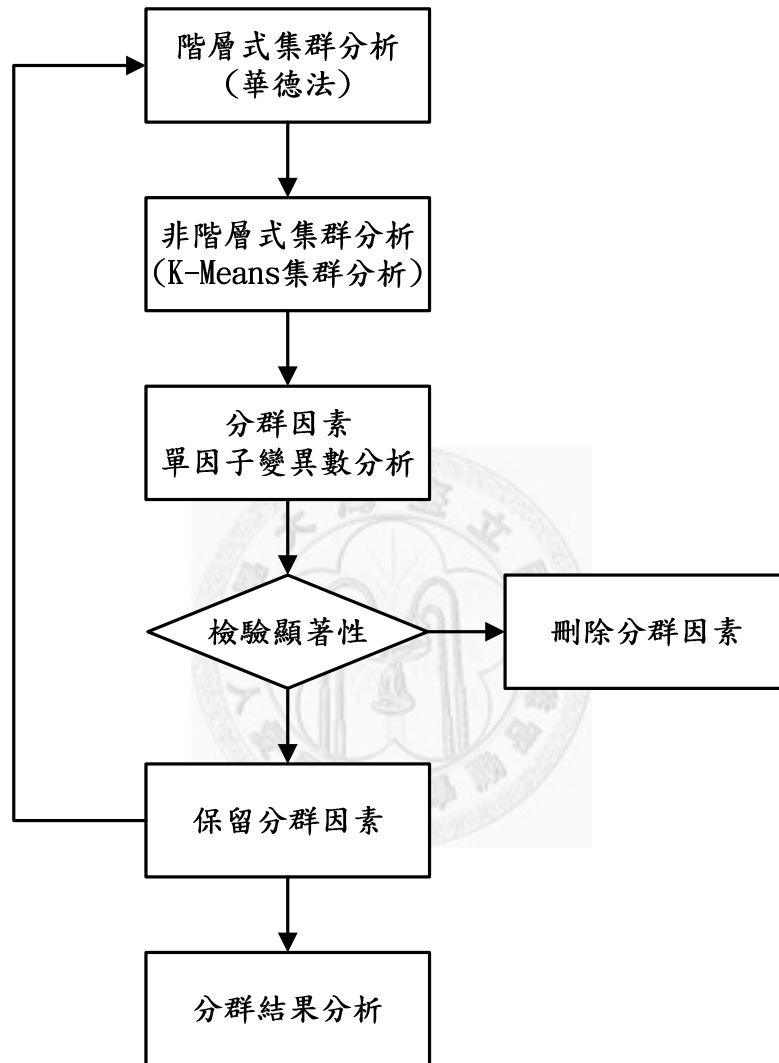


圖 5-1：集群方法分析流程

5.1.2 階層式集群分析

本研究採用階層式集群分析中最常被使用的華德法進行分群數目的決定，華德法屬於凝聚法的一種，亦即在其分析之初將各觀察值分別視為一個群體，接著

利用觀察值間距一步一步的縮減集群數，直到整個樣本皆為同一個集群為止。然而，在凝聚集群的過程中，會發現在某些步驟時凝聚的難度大幅增加，此現象將表現在凝聚係數上。當凝聚係數增量大幅增加時，顯示要從 n 群縮減為 $n-1$ 群時困難度驟增，亦即此時被凝聚的集群距離其併入的集群較遠，故應選擇 n 群為最佳集群數。

本研究的階層式集群分析結果如圖 5-2 及表 5-1 所示，由表 5-1 可知當集群數從四群變為三群時凝聚係數增量驟增。此結果亦可從圖 5-2 看出，在圖 5-2 中可發現集群數由四群降為三群時該段斜率明顯大於其它兩段。亦即，若要將群集數由四群降為三群時，凝聚群集的困難度將會快速增加，故選擇四群為最佳的群集數。

圖 5-2：係數變化率與集群數關係圖

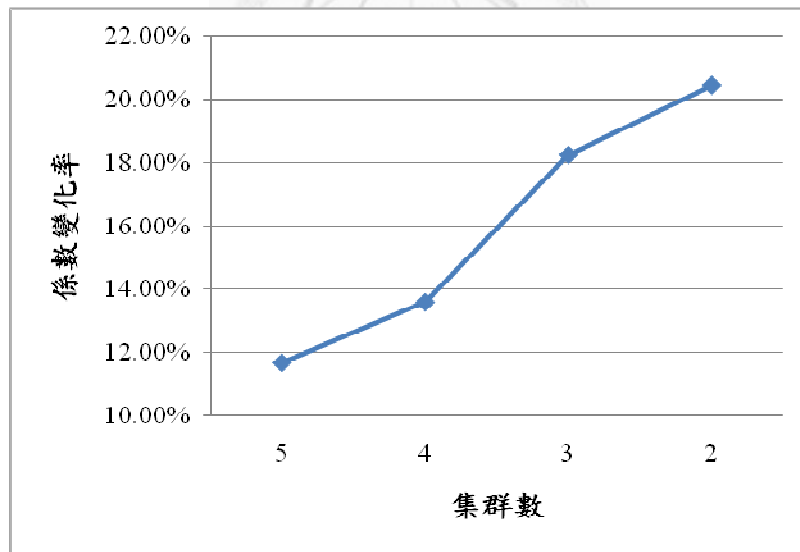


表 5-1：階層式集群分析法結果分析表

集群數	係數	係數變化	係數變化率
6	911.769	--	--
5	1032.334	120.565	11.68%
4	1194.826	162.492	13.60%
3	1461.808	266.982	18.26%
2	1837.982	376.174	20.47%

5.1.3 非階層式集群分析

根據前述階層式集群分析法的結果，得知本研究利用之計程車派遣資料庫資料的最佳分群數為四群，利用此結果帶入非階層式群集分析的 K-Means 集群分析中，以計算分群結果，K-Means 集群分析之集群過程如下：

1. 將觀察值分割成四個原始群落，此時每個集群的觀察值成員是經過資料庫中任意選取產生的。
2. 利用歐基里得距離計算每群觀察值到各群中心的距離，接著將觀察值分派至最近的群落以調整分群。
3. 重新計算每群觀察值資料向量的中心點。
4. 重複步驟 2、3，直到各集群沒有觀察值可以再調整為止。
5. 最後集群在各群落中分群因素的中心點位置。

根據前述步驟，本研究將計程車駕駛人營業行為分為四個集群，分群結果如表 5-2 所示，各集群內容將經過集群驗證後再回推整理。

表 5-2：K-Means 集群分析結果

項次		個數
集群	1	252
	2	169
	3	181
	4	181
有效值		783
遺漏值		0

5.1.4 集群驗證

完成集群化的動作後，研究者需確認分群因素在不同集群的集合中是否存在顯著差異。一般而言，為確認此項目需計算各集群在各分群因素上的平均值，對平均值進行比較分析，若集群數為二群進行 T 檢定分析，若集群數為三群以上則進行單因子變異數分析。本研究的集群數為四群，因此利用單因子變異數分析進行集群的驗證。其統計假設如下：

H_0 ：該分群因素在各集群有顯著差異

H_1 ：該分群因素在各集群無顯著差異

若分群因素的大小與各集群有關即留下，無關則刪除之，並回到階層式集群分析階段建立新的集群。本研究之單因子變異數分析結果如表 5-3 所示，由表 5-3 可知，本研究所採用之三個分群因素之 P 檢定值皆小於 0.001，亦即在 99.9% 的信賴水準下，各分群因素在各集群有顯著差異，顯示此集群分析結果有效。

表 5-3：各分群因素與各集群之 ANOVA 檢定結果

因素名稱	集群		誤差		F 檢定	P 檢定	顯著性
	平均平方和	自由度	平均平方和	自由度			
營業狀況	123.479	3	0.38	779	326.711	0	顯著
空車率	85.353	3	0.36	779	236.348	0	顯著
營業習慣	145.653	3	0.42	779	349.826	0	顯著

5.1.5 分析結果

經由前述集群分析的過程，本研究已得到衛星派遣計程車駕駛人之營業行為分群，利用分群的結果回推可得到營業狀況、空車率及營業習慣三個分群因素的範圍。然而，分群因素是由許多影響衛星派遣計程車駕駛人營業行為的觀察變數

所組成，單以分群因素數值表達的衛星派遣計程車駕駛人營業行為不易瞭解。因此本研究將利用分群結果回推至原始的觀察變數，以利瞭解各群駕駛人的實際營業情形。

表 5-4 為四群駕駛人各觀察變數數值之統計整理表，表中灰底為該項觀察變數數值最高的集群、粗體為該項觀察變數數值最低的集群。以下將分述營業情形、空車率及營業情況三個分群因素中觀察變數所顯示出的衛星派遣計程車駕駛人營業行為。



表 5-4：四群駕駛人各觀察變數分析結果

觀察變數		第一群			第二群			第三群			第四群		
		最小值	最大值	平均數	最小值	最大值	平均數	最小值	最大值	平均數	最小值	最大值	平均數
營業情形	載客數	187	745	383	3	578	249	6	345	145	2	330	135
	載客時程(hr)	52.8	178.8	84.7	0.2	124.0	48.4	1.0	90.0	32.9	0.3	68.2	29.0
	載客里程(km)	1,099.6	3,693.4	1,888.5	4.4	4,854.0	1,361.6	20.0	1,712.6	701.2	7.2	1,526.8	657.1
	營業時程(hr)	220.9	698.0	421.9	10.2	668.9	366.5	10.9	392.5	151.3	10.2	585.1	228.5
	營業里程(km)	2,564.9	14,833.3	4,390.9	64.0	10,183.0	3,280.4	76.5	3,252.7	1,544.4	68.5	6,187.7	2,054.8
	營業天數	18	31	29	3	31	25	2	31	16	1	31	18
	總收入	35,265	116,775	60,988	265	119,560	41,479	805	56,490	22,964	250	49,590	21,254
平均日收入	1,252	4,381	2,131	88	3,856	1,614	268	3,326	1,471	125	3,034	1,151	
空車率	時間空車率	59.9	88.4	77.6	67.7	93.4	82.5	56.0	87.3	73.4	73.7	96.6	85.1
	距離空車率	42.3	72.9	56.4	40.0	78.7	58.2	38.9	67.5	54.2	50.4	90.3	67.0
營業習慣	平均速度(kph)	17.4	27.0	22.0	24.4	38.3	28.0	15.1	26.4	21.5	16.1	29.5	22.6
	單次載客收入	136	246	159	88	338	167	115	482	163	97	412	160

1. 營業情形：

由表 5-4 中可知，營業情形中的觀察變數以第一群駕駛人為最高、第二群駕駛人次之，顯示該兩群駕駛人對於營業行為較為積極打拼。然而，營業情形的最小值並未出現在同一群當中，載客數、載客時程、載客里程、總收入及平均日收入的最小值出現在第四群中，但營業時程、營業里程及營業天數的最小值則出現在第三群中，這個結果顯示雖然第三、四群的駕駛人都是相對較不積極的，但第四群駕駛人的營業時程、里程與天數顯然無法有效的增加其載客的時程及里程，導致其為收入最低的一群。反觀第三群駕駛人，雖然其營業態度相當不積極，但相較於第四群駕駛人其為較有營業技巧的一群，第三群駕駛人可以用較短的時間就得到與第二群駕駛人相去不遠的日收入。

2. 空車率：

空車率這項分群因素更能說明第三群與第四群駕駛人在營業情形中的差異，由表 5-4 中可知時間空車率及距離空車率最大的皆為第四群駕駛人，而時間空車率及距離空車率最小的均為第三群駕駛人，這點再次佐證了第三群駕駛人為各群駕駛人當中相對有營業技巧的一群。第三群駕駛人營業的時間、距離及天數，均為較有效的巡迴。

3. 營業習慣：

在營運習慣方面，平均速度及單次載客收入最高點皆落在第二群駕駛人，平均速度的最低點落在第三群駕駛人，但其與第一、四群駕駛人之平均速度相去不遠，因此可視為無差異。反觀第二群駕駛人在平均速度一項較其他駕駛人高出一個階層，本研究推測第二群駕駛人在營業時可能有開快車之習慣，抑或是偏好行駛於車輛較少且路況較好之處，此點特性隱含其可能較常穿梭於巷弄間或者有夜間營業的習慣。在單次載客收入一項最低點落在第

一群駕駛人，但四群駕駛人此項數據差異僅在十元以內，因此將此項觀察變數視為無差異。

分析完原始的十二項觀察變數後，本研究認為這些觀察變數尚有其他分析的可能，因此本研究針對營業情形進行進一步的分析，分析的項目分為每日營業情形及淨收入。每日營業情形由載客數、營業里程、營業時程、載客里程及載客時程分別除以其營業天數而得，而淨收入由總收入減去總成本而得。其中總成本的估算，本研究利用台北地區 97 年營運情形調查所提供的燃油成本乘以營業里程而得，使用燃油成本而不用車公里成本的原因為車公里成本的估計是由政府定定的十二項成本計算而來，但在 97 年營運調查中提及十二項成本有許多重複計算與不甚合理的部份，因此本研究採用單純的燃油成本估算總成本以得到淨收入的數值。

表 5-5 為四群駕駛人每日營業情形數據分析整理表，表中灰底為該項觀察變數數值最高的集群、粗體為該項觀察變數數值最低的集群。由

表 5-5 可看出每日營業情形數據仍是以第一群駕駛人為最高、第二群駕駛人次之，在營業里程、時程及燃油成本中是以第三群駕駛人最低，每日載客數、載客里程、載客時程及淨收入則是以第四群駕駛人為最低。觀察在淨收入的部份，僅剩下第一群駕駛人收入較高，第三及第四群駕駛人每月淨收入已不足二萬元，第四群駕駛人甚至已逼近一萬元。

表 5-5：四群駕駛人每日營業情形數據分析結果

觀察變數 集群	每日 載客數	每日 載客里程 (km)	每日 載客時程 (hr)	每日 營業里程 (km)	每日 營業時程 (hr)	燃油 成本	淨 收入
第一群	13	66.0	3.0	153.2	14.7	16,642	44,347
第二群	10	52.9	1.9	127.5	14.2	12,433	29,047
第三群	9	45.0	2.1	98.8	9.8	5,853	17,110
第四群	7	35.8	1.6	110.1	12.6	7,788	13,466

5.2 衛星派遣計程車駕駛人日營業行為抽樣分析

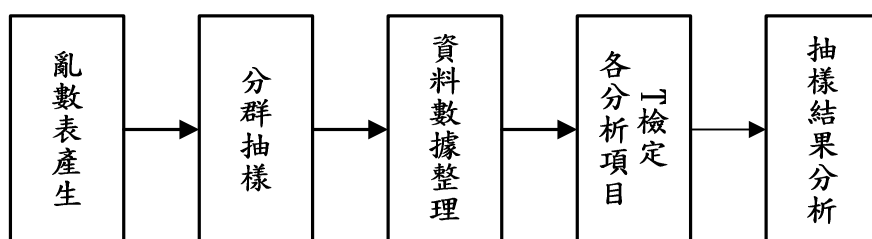
完成衛星派遣計程車駕駛人月營業行為分析後，本研究仍對某些問題感到好奇，如各群駕駛人在兩次載客中間會有什麼特性？然而資料庫資料量過大而無法對全部的駕駛人進行分析，因此本節將利用分群抽樣的方式，對衛星派遣計程車駕駛人日營業行為進行分析。

5.2.1 抽樣方法與分析流程

本研究利用 Excel 系統自動產生的亂數表進行隨機抽樣分析，針對各群駕駛人分別抽出 30 個樣本進行每日營業行為之詳細分析，分析的內容包含載客時間間距、載客距離間距、每日休息次數及同區載客比例。透過數據整理後，分別對各群駕駛人之各分析項目變數進行 T 檢定，以找出各群駕駛人與各觀察變數是否有顯著關係，藉此瞭解在各分析項目是否有其特徵存在。詳細流程如圖 5-3

圖 5-3 所示。

圖 5-3：分群抽樣分析流程圖



5.2.2 T 檢定

分群抽樣完成後，本研究將資料庫資料整理為適合分析的形式，爾後分別針對載客時間間距、載客距離間距、每日休息次數及同區載客比例進行 T 檢定，觀察各群駕駛人在各分析項目中是否有顯著特徵，以下將分別說明之。

1. 載客時間間距：

載客時間間距意指計程車駕駛人在兩次載客間間隔多少時間，利用此數據可看出駕駛人下客後花多少時間可攬得下一組乘客，並與攬客技巧有關。表 5-6 為各群駕駛人載客時間間距平均值及變異數整理表。由表 5-6 中可知，載客時間間距最高者為第一群駕駛人，最低者為第四群駕駛人。

表 5-6：各群駕駛人載客時間間距平均值及變異數整理表

	第一群	第二群	第三群	第四群
平均值	974.9	935.65	948.45	852.45
變異數	14,913.46	12,546.34	48,467.73	16,277.78

為瞭解計程車駕駛人載客時間間距是否有顯著差異，因此對各群駕駛人之載客時間間距進行 T 檢定，表 5-7 為各群駕駛人載客時間間距 T 檢定之 P 值整理表，由表 5-7 中可看出第四群駕駛人與其他群間的 P 值皆小於 0.1，亦即第四群駕駛人的載客時間間距顯著低於其他群。此結果顯示若機會允許，第四群駕駛人習慣於在短時間內找到下一組乘客。

表 5-7：各群駕駛人同區載客比例 T 檢定雙尾 P 值整理表

	第一群	第二群	第三群	第四群
第一群	--	0.296165	0.641859	0.002539
第二群	0.296165	--	0.818422	0.025453
第三群	0.641859	0.818422	--	0.085434
第四群	0.002539	0.025453	0.085434	--

2. 載客距離間距：

載客距離間距意指計程車駕駛人在兩次上客地點間距離差距多少，利用此數據可看出駕駛人下客後在多遠的距離可攬得下一組乘客，此數據與駕駛人習慣有關，利用此數據可看出駕駛人喜歡在下客處附近巡迴攬客，或者習慣於在其熟悉處巡迴攬客。表 5-8 為各群駕駛人載客距離間距平均值及變異數整理表。

表 5-8：各群駕駛人載客距離間距平均值及變異數整理表

	第一群	第二群	第三群	第四群
平均值	1,721.513	2,099.914	1,834.063	1,652.702
變異數	83,172.17	93,848.69	273,408	83,481.71

為瞭解計程車駕駛人載客距離間距是否有顯著差異，因此對各群駕駛人之載客距離間距進行 T 檢定，表 5-9 為各群駕駛人載客距離間距 T 檢定之 P 值整理表，由表 5-9 中可看出第二群駕駛人與第一、四群間的 P 值皆遠小於 0.05，而第二群駕駛人與第三群駕駛人之 P 值亦有小於 0.06，亦即第二群駕駛人的載客距離間距顯著高於其他群。此結果顯示第二群駕駛人習慣於在完成載客後，往較遠處巡迴攬客，而不會在下客處的鄰近地區繞行。

表 5-9：各群駕駛人載客距離間距 T 檢定雙尾 P 值整理表

	第一群	第二群	第三群	第四群
第一群	--	0.000264	0.40595	0.455602
第二群	0.000264	--	0.058812	2.9E-05
第三群	0.40595	0.058812	--	0.184695
第四群	0.455602	2.9E-05	0.184695	--

3. 每日休息次數：

每日休息次數意指計程車駕駛人在每日營業時程中斷了幾次，利用此數據可看出在乘客較少的時間駕駛人是否會選擇暫時休息，此數據不但與駕駛人巡迴習慣有關，亦有與駕駛人之積極程度相關。表 5-10 為各群駕駛人每日休息次數平均值及變異數整理表。由表 5-10 可看出第一群計程車駕駛人每日休息次數最高，最低則落在第三群駕駛人一項中。

表 5-10：各群駕駛人每日休息次數平均值及變異數整理表

	第一群	第二群	第三群	第四群
平均值	1.890709	0.949976	0.784951	1.180196
變異數	0.488731	0.300434	0.230591	0.374019

為瞭解駕駛人每日休息次數是否有顯著差異，因此對各群駕駛人之每日休息次數進行 T 檢定，表 5-11 為各群駕駛人每日休息次數 T 檢定之 P 值整理表，由表 5-11 中可看出第一群駕駛人與其他群間的 P 值皆遠小於 0.01，亦即第一群駕駛人的每日休息次數顯著高於其他群。此結果顯示第一群駕駛人習慣在計程車需求較少時暫時休息，不會無限巡迴。

表 5-11：各群駕駛人每日休息次數 T 檢定雙尾 P 值整理表

	第一群	第二群	第三群	第四群
第一群	--	3.36E-05	1.43E-06	0.001536
第二群	3.36E-05	--	0.317752	0.21762
第三群	1.43E-06	0.317752	--	0.029079
第四群	0.001536	0.21762	0.029079	--

4. 同區載客比例：

同區載客比例意指計程車駕駛人在下客處載到下一個乘客的比例有多少，利用此數據可看出駕駛人下客是否願意在下客處巡迴攬客。表 5-12 為各群駕駛人載客時間間距平均值及變異數整理表。由表 5-12 可看出第一群計程車駕駛人同區載客比例最高，而第四群駕駛人同區載客比例最低。

表 5-12：各群駕駛人同區載客比例平均值及變異數整理表

	第一群	第二群	第三群	第四群
平均值	0.41964	0.387108	0.403036	0.35565
變異數	0.007984	0.003672	0.007134	0.002746

為瞭解計程車駕駛人同區載客比例是否有顯著差異，因此對各群駕駛人之同區載客比例進行 T 檢定，表 5-13 為各群駕駛人同區載客比例 T 檢定之 P 值整理表，由表 5-13 中可看出第四群駕駛人與第一、三群間的 P 值皆小於 0.05，而第四群駕駛人與第二群駕駛人之 P 值亦有小於 0.1，亦即第四群駕駛人的同區載客比例顯著小於其他群。此結果顯示第四群駕駛人習慣於在完成載客後，往他處巡迴而不會留在下客處。

表 5-13：各群駕駛人同區載客比例 T 檢定雙尾 P 值整理表

	第一群	第二群	第三群	第四群
第一群	--	0.186972	0.549484	0.009555
第二群	0.186972	--	0.497844	0.087362
第三群	0.549484	0.497844	--	0.040789
第四群	0.009555	0.087362	0.040789	--

5. 夜間時段營業比例：

夜間時段營業比例意指計程車駕駛人夜間時段載客次數佔總載客次數的比例，利用此數據可分析各群駕駛人對夜間時段營業的偏好程度。本研究根據 97 年度台北地區計程車營運情形調查所提出的時間分段，將夜間營業時段設定為 21:00~24:00。表 5-14 為各群駕駛人夜間時段營業比例平均值及變異數整理表。由表 5-14 可看出第二群計程車駕駛人之夜間時段營業比例最高、第四群次之，最低則落在第三群駕駛人中。

表 5-14：各群駕駛人夜間時段營業比例平均值及變異數整理表

	第一群	第二群	第三群	第四群
平均值	0.079056	0.151512	0.069374	0.132558
變異數	0.004591	0.005468	0.004705	0.013384

為瞭解駕駛人夜間時段營業比例是否有顯著差異，因此對各群駕駛人之夜間時段營業比例進行 T 檢定，表 5-15 為各群駕駛人夜間時段營業比例 T 檢定之 P 值整理表，由表 5-15 中可看出第二群駕駛人與第一群及第三群駕駛人之 P 值皆遠小於 0.01，而第四群駕駛人與第一群及第三群駕駛人之 P 值皆小於 0.1，亦即第二群及第四群駕駛人的夜間時段營業比例顯著高於其他群。此結果顯示第二群及第四群駕駛人相較於其他駕駛人偏好在夜間時段營業載客。

表 5-15：各群駕駛人夜間時段營業比例 T 檢定雙尾 P 值整理表

	第一群	第二群	第三群	第四群
第一群	--	0.00255	0.655893	0.08411
第二群	0.00255	--	0.000804	0.541386
第三群	0.655893	0.000804	--	0.043873
第四群	0.08411	0.541386	0.043873	--

6. 凌晨時段營業比例：

凌晨時段營業比例意指計程車駕駛人凌晨時段載客次數佔總載客次數的比例，利用此數據可分析各群駕駛人對凌晨時段營業的偏好程度。本研究根據 97 年度台北地區計程車營運情形調查所提出的時間分段，將凌晨營業時段設定為 00:00~04:00。表 5-16 為各群駕駛人凌晨時段營業比例平均值及變異數整理表。由表 5-16 可看出第二群計程車駕駛人之凌晨營業比例最高，而凌晨營業比例最低則落在第三群駕駛人中。

表 5-16：各群駕駛人凌晨時段營業比例平均值及變異數整理表

	第一群	第二群	第三群	第四群
平均值	0.11771	0.451943	0.110714	0.17773
變異數	0.017597	0.033861	0.023595	0.043813

為瞭解駕駛人凌晨時段營業比例是否有顯著差異，因此對各群駕駛人之凌晨時段營業比例進行 T 檢定，表 5-17 為各群駕駛人凌晨時段營業比例 T 檢定之 P 值整理表，由表 5-17 中可看出第二群駕駛人與其他群間的 P 值皆遠小於 0.01，亦即第二群駕駛人的凌晨時段營業比例顯著高於其他群。此結果顯示第二群駕駛人相較於其他駕駛人偏好在凌晨時段營業載客。

表 5-17：各群駕駛人凌晨時段營業比例 T 檢定雙尾 P 值整理表

	第一群	第二群	第三群	第四群
第一群	--	1.3E-07	0.878321	0.286833
第二群	1.3E-07	--	2.01E-07	8.84E-05
第三群	0.878321	2.01E-07	--	0.256177
第四群	0.286833	8.84E-05	0.256177	--

5.2.3 分析結果

本研究將前述檢定結果與解釋整理如表 5-18，由表 5-18 中可知第一群駕駛人習慣於計程車需求較低時就暫時休息，相較於其他群駕駛人第四群駕駛人偏好在夜間時段營業，而第二群駕駛人則習慣在夜間時段及凌晨時段營業，另外，第四群駕駛人習慣於在完成載客後往遠處巡迴。然而觀察第四群駕駛人在載客距離間距與同區載客比例的表現似乎有違常理，一般而言同區載客比例高載客距離間距應較小，但該兩項分析項目在第四群駕駛人的表現卻是同區載客比例與載客距離間距皆小。在觀察第四群駕駛人之載客時間間距也為最小，這似乎也與同區載客比例有某種程度上的衝突。事實上，此現象隱含了第四群駕駛人某種營業行為特性。

本研究在進行之載客時間間距資料處理時給定一個假設，此假設為載客時間間距超過 1 小時者視為暫時休息，未超過 1 小時者才放入載客時間間距計算。而載客距離間距則是以計算時間間距之資料來計算的，亦即載客時間、距離間距兩者之取用資料是相同的。然而同區載客比例是以所有抽樣的資料進行分析，亦即所有被選用的資料皆被用來計算此項目。因此載客時間、距離間距與同區載客比例間的選用資料不同，同區載客比例選用的資料較能代表整體的狀況。

若將此兩組分析項目綜合陳述，本研究發現第四群駕駛人載密度差異很大，在尖峰時段與其他群駕駛人相比，其能在較短時間內找到下一組乘客。然而在旅次需求不高的時候，第四群駕駛人傾向於向各處漫無目的的巡迴攬客，造成其表現在短時間項目的載客時間、距離間距為最短，但若脫離尖峰時間，其巡迴時間及距離即會變長，導致其同區載客比例較低，距離空車率及時間空車率皆較高的現象。

表 5-18：計程車駕駛日營業行為分析整理表

分析項目	排序	結果	解釋
載客時間 間距	1>3>2>4	第四群在 $\alpha=0.1$ 時顯著小於第一、二、三群	第四群駕駛人習慣於在短時間內就找到下一個乘客
載客距離 間距	2>3>1>4	第二群在 $\alpha=0.01$ 時顯著大於第一、四群，在 $\alpha=0.06$ 時顯著大於第三群	第二群駕駛人完成載客後傾向於往遠處巡迴
每日 休息次數	1>4>2>3	第一群在 $\alpha=0.01$ 時顯著大於第二、三、四群	第一群駕駛人習慣於旅次需求較低時就暫時休息，不會無限巡迴
同區載客 比例	1>3>2>4	第四群在 $\alpha=0.1$ 時顯著小於第一、二、三群	第四群駕駛人甲地下客甲地巡迴的比例最低
夜間時段 營業比例	2>4>1>3	第二群駕駛人在 $\alpha=0.01$ 時顯著大於第一、三群，而第四群駕駛人在 $\alpha=0.05$ 時顯著大於第三群，在 $\alpha=0.1$ 時顯著大於第一群	第二群及第四群駕駛人相較於其他駕駛人偏好在夜間時段營業載客
凌晨時段 營業比例	2>4>1>3	第二群駕駛人在 $\alpha=0.01$ 時顯著大於第一、三、四群	第二群駕駛人相較於其他駕駛人偏好在凌晨時段營業載客

5.3 衛星派遣計程車駕駛人營業行為分群結果整理

根據衛星派遣計程車駕駛人月及日營業行為的分析結果，本研究將各群駕駛人分別命名，第一群駕駛人命名為智慧型駕駛人，第二群駕駛人命名為積極型駕駛人，第三群駕駛人命名為自由型駕駛人，第四群駕駛人命名為消極型駕駛人，其分別的特性如表 5-19 所示。由表 5-19 可知智慧型駕駛人為四群駕駛人中月收入最佳的一群，他們不但營業態度積極且洞悉應該休息的時間，利用旅次需求較少時暫時暫停巡迴攬客，以減少油耗的花費，不管在總收入或淨收入上面，智慧型駕駛人都有很好的表現。積極型駕駛人偏好在夜間及凌晨時段營業，而營業時平均速度較快，其營業技巧不佳的缺點由營業時程增加來彌補，因此積極型駕駛人之月收入尚不算太差。自由型駕駛人與消極型駕駛人的營運態度都不夠積極，但自由型駕駛人屬於較聰明的一群，他們瞭解該如何以最少的營業時程載到最多的客人，但他們的不積極也反應在月收入上，其月收入僅優於消極型駕駛人。消極型駕駛人偏好在夜間時段營業，而其營業態度不積極且營業技巧不佳，僅能靠尖峰時段旅次需求大增的效果來彌補其缺失，但在離峰時段消極型駕駛人僅會在各區無限巡迴碰運氣，故收入為四群駕駛人中最差的一群。

表 5-19：各群駕駛人命名及特色整理表

駕駛人類型	樣本數	特色
智慧型	252	營業狀況變數皆最高想是其 <u>營業態度相當積極勤奮</u> ，卻也不會白白浪費時間巡迴攬客，傾向於 <u>客人較少時就暫時休息</u> 。再加上空車率次低營業效率高，本群月收入為最高的 <u>61,000</u> 元左右。
積極型	169	偏好於 <u>夜間時段及凌晨時段營業</u> ， <u>平均速度較其他群駕駛人為快</u> ，完成載客後傾向於 <u>往較遠處巡迴攬客</u> ，非在下客處繞行。 <u>營業及載客時程距離皆長</u> ，但 <u>營業效率卻不佳</u> ，然而由於其勤奮工作，月收入仍在第二名的位置，約為 <u>41,000</u> 元。
自由型	181	營業狀況變數皆為次低顯示其 <u>營業態度也屬於不積極</u> 的，每日營業里程、時程為最低，但每日載客里程、時程卻非最低，且其空車率最低，顯示其為 <u>營業效率最高</u> 的一群。他們對於 <u>招客的技巧很有一套但個性上較不積極</u> ，因此月收入也落在倒數第二名的 <u>23,000</u> 元左右。
消極型	181	相較於其他群駕駛人其偏好於 <u>夜間時段營業</u> ，營業狀況變數皆為最低顯示其 <u>營業態度較不積極</u> ，營業及載客時程里程皆低。每日營業里程、時程非最低，但每日載客里程、時程卻為最低，再加上空車率為全部最高的一群，顯示其 <u>營業效率相當低</u> 。在 <u>尖峰時間習慣在短時間內就找到下一個乘客</u> ，但在 <u>離峰的情形下巡迴攬客技巧不佳</u> ，習慣於 <u>漫無目的的往遠處巡迴攬客</u> 。對於攬客屬於 <u>土法煉鋼型</u> 卻又不積極，月收入為最低的 <u>21,000</u> 元左右。

根據以上分析結果，下章將針對各群駕駛人提出派遣邏輯改善建議，並概略設計出新的派遣邏輯，作為未來改善之參考。

第六章 派遣邏輯改善策略

經由前述章節的分析後，本研究將衛星派遣計程車駕駛人依其特性分為積極型、消極型、自由型及智慧型四群，本節依各群駕駛人之特性分別提出派遣邏輯改善建議，並概略設計考慮駕駛人特性之派遣邏輯。

在派遣邏輯的設計方面，本研究引入空中排班點的概念進行考慮駕駛人營業行為特性之派遣邏輯設計，在此設計中需對各空中排班區域進行乘客需求的預測，本研究採用侯勝宗(2009)提出之需求預測模型進行預測。

6.1 智慧型駕駛人

智慧型駕駛人的特色為營業態度相當積極，但不會白白浪費時間巡迴攬客，傾向於客人較少時就暫時休息，然而其休息時間不長，很快又再次進行載客，營業效率相對較高，而對於下客後巡迴的區域沒有特別偏好。根據以上特性，本研究設計適合智慧型駕駛人之派遣邏輯如圖 6-1，流程分別說明如下。

流程 1：訊號的發出分為主動式及被動式，主動式為當駕駛完成載客後車機自動向派遣中心發出訊號；被動式為駕駛主動向派遣中心發出繞行區域建議需求。

流程 2：當派遣中心接獲計程車駕駛需求後，首先需對車輛的位置進行定位，以進行接下來的步驟。

流程 3：確定車輛即時位置後，系統由地圖上找出與該位置相鄰的系統劃分區域。

流程 4：分別預測由車輛即時位置到達流程 3 找出之區域所需的車程時間，車程預測時以一般的駕駛速率進行預測。

流程 5：由需求模型預測該區域即時的需求情形，同時取出該區域當時的空中排班點排班情形，以推斷其排班順位。

流程 6：根據流程 5 所得之資料，預測在該區域等候空中排班所需的時間。例如：由需求預測得知該區域 10 分鐘有一需求產生，而該車的空排順位為 2，則可推得其需在該區等候 20 分鐘方可順利載客。

流程 7：加總流程 4 及流程 6 分別得到的車程時間及空排等候時間，可得出該車到達每一區域的總等候時間。

流程 8：選擇總等候時間最短者為此次運算的最佳派遣區域。

流程 9：判斷由流程 8 所選出的最佳派遣區域的總等候時間是否小於 15 分鐘，若非，則進行至流程 10；若非，則進行至流程 11。其中總等候時間是以智慧型駕駛人載客時間間距平均值決定，本研究認為若系統給出的巡迴指引較駕駛人本身所能達到之更差則不應接受此結果。

流程 10：由於智慧型駕駛人習慣於在旅次需求較少時暫時休息，故經由前述流程計算過後若無法得到合宜的派遣區域，建議智慧型駕駛人暫時休息。

流程 11：由派遣中心發出訊號通知駕駛何處為最佳繞行區域，以提供駕駛更好的繞行方向建議。

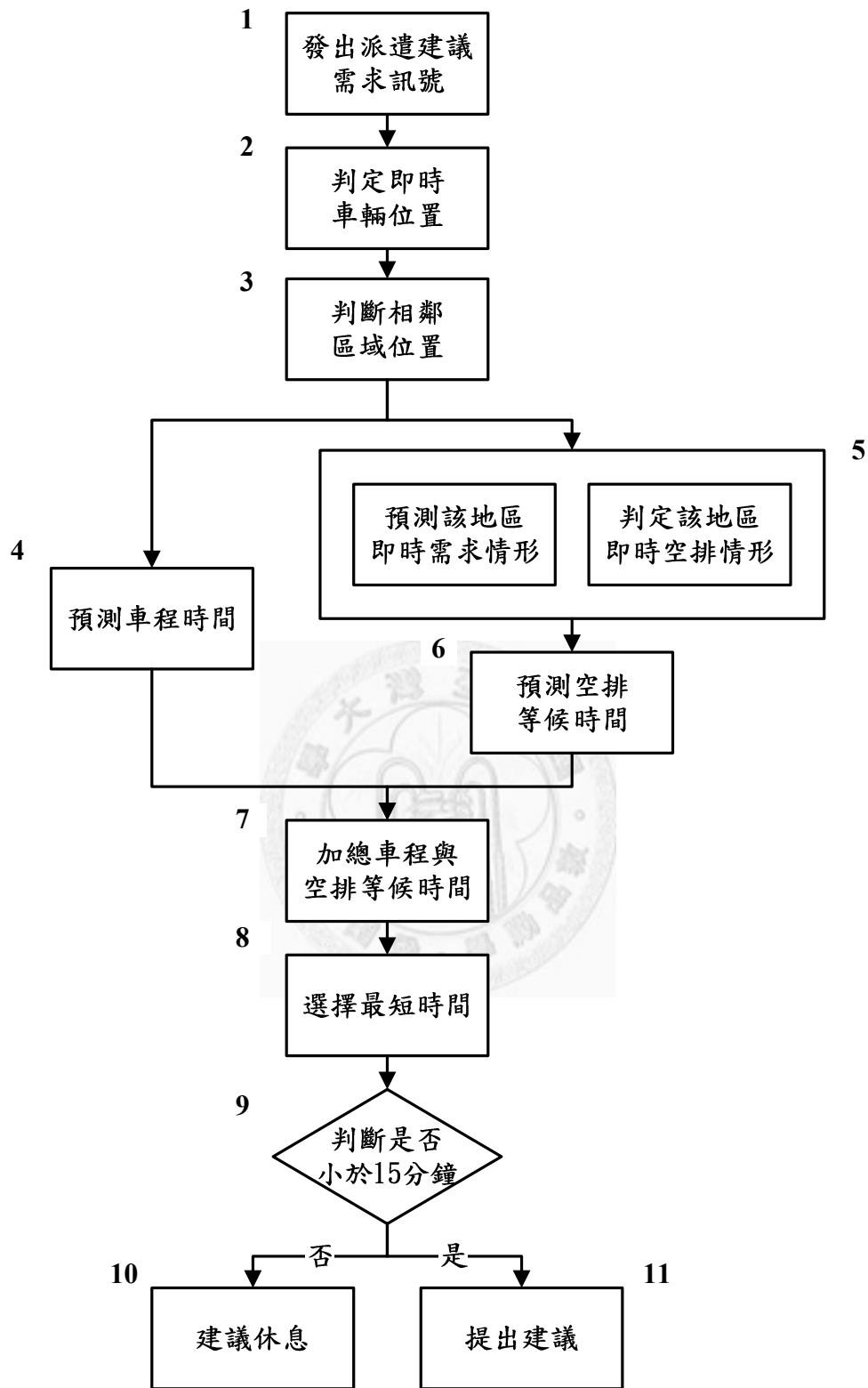


圖 6-1：智慧型計程車駕駛人衛星派遣流程圖

6.2 積極型駕駛人

積極型駕駛人最大的特色為平均駕駛速度較其他駕駛人為快，且完成載客後傾向於往較遠處巡迴攬客，非在下客處繞行。對於營業及載客較為積極，對於無客時暫時休息較不在意。根據以上特性，本研究設計適合積極型駕駛人之派遣邏輯如圖 6-2，各流程分別說明如下。

流程 1：訊號的發出分為主動式及被動式，主動式為當駕駛完成載客後車機自動向派遣中心發出訊號；被動式為駕駛主動向派遣中心發出繞行區域建議需求。

流程 2：當派遣中心接獲計程車駕駛需求後，首先需對車輛的位置進行定位，以進行接下來的步驟。

流程 3：確定車輛即時位置後，系統由地圖上找出與該位置相鄰的系統劃分區域。

流程 4：分別預測由車輛即時位置到達前一流程找出之區域所需的車程時間，若為同一區則車程時間為 0，進行車程預測時需將積極型駕駛人駕駛速率較快考慮進去。

流程 5：由需求模型預測該區域即時的需求情形，同時取出該區域當時的空中排班點排班情形，以推斷其排班順位。

流程 6：根據流程 5 所得之資料，預測在該區域等候空中排班所需的時間。例如：由需求預測得知該區域 10 分鐘有一需求產生，而該車的空排順位為 2，則可推得其需在該區等候 20 分鐘方可順利載客。

流程 7：加總流程 4 及流程 6 分別得到的車程時間及空排等候時間，可得出該車

到達每一區域的總等候時間。

流程 8：選擇總等候時間最短者為此次運算的最佳派遣區域。

流程 9：判斷由流程 8 所選出的最佳派遣區域的總等候時間是否小於 15 分鐘，若非，則進行至流程 10；若是，則進行至流程 11。其中總等候時間是以積極型駕駛人載客時間間距平均值決定，本研究認為若系統給出的巡迴指引較駕駛人本身所能達到之更差則不應接受此結果。

流程 10：由鄰近區域無法得到總等候時間小於 15 分鐘之派遣區域，故擴大試算範圍，找出較相鄰區域更遠一些的区域後，再回到流程 4 及流程 5 進行分析。

流程 11：由派遣中心發出訊號通知駕駛何處為最佳繞行區域，以提供駕駛更好的繞行方向建議。



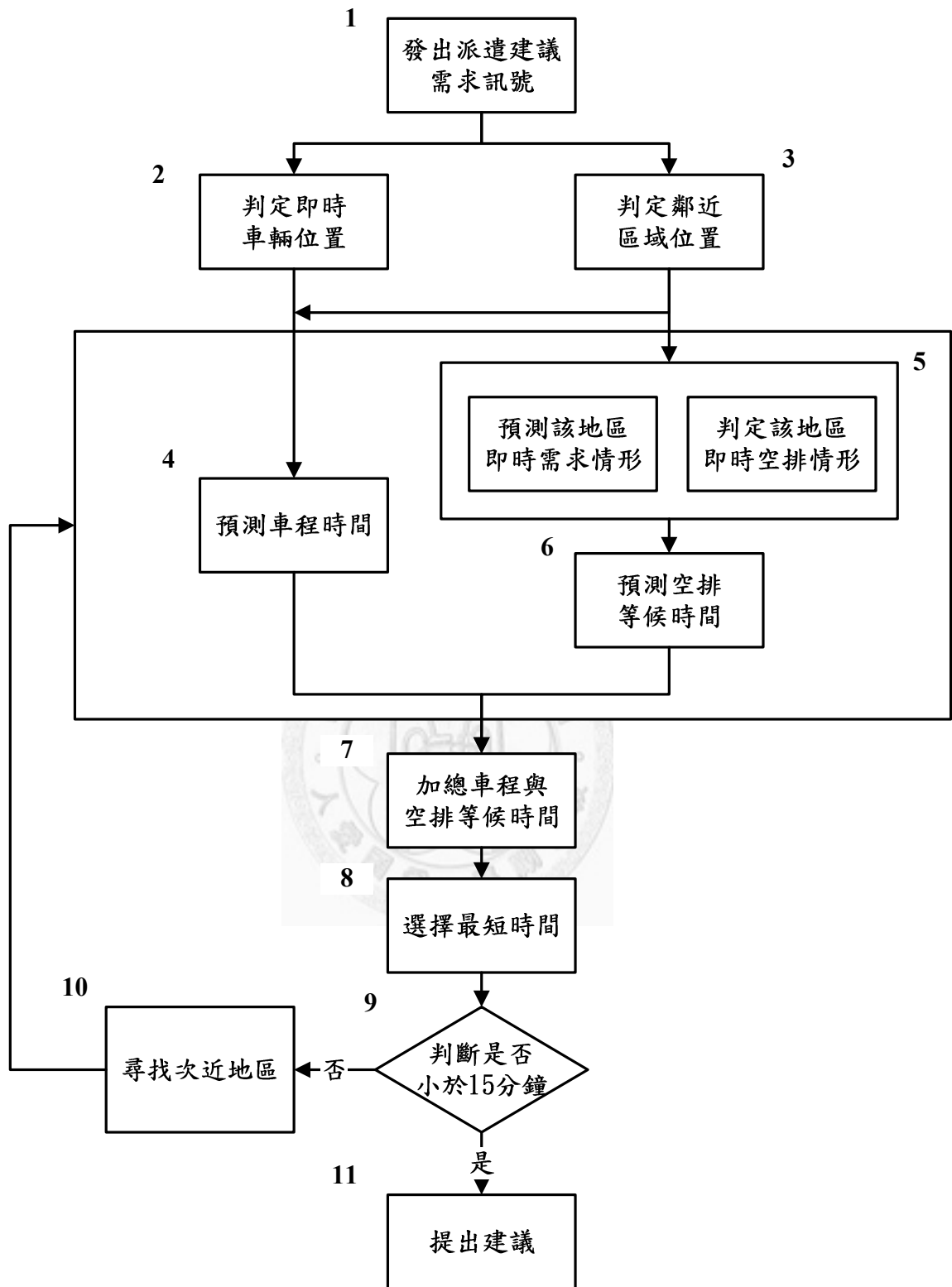


圖 6-2：積極型計程車駕駛人衛星派遣流程圖

6.3 自由型駕駛人

自由型駕駛人的特色為營運態度較不積極，但招客技巧佳、營運效率高，完成載客後並沒有特別傾向於往遠處或在原區巡迴攬客，對於無客時暫時休息也較不在意。根據以上特性，本研究設計適合自由型駕駛人之派遣邏輯如圖 6-3 所示，流程分別說明如下。

流程 1：訊號的發出分為主動式及被動式，主動式為當駕駛完成載客後車機自動向派遣中心發出訊號；被動式為駕駛主動向派遣中心發出繞行區域建議需求。

流程 2：當派遣中心接獲計程車駕駛需求後，首先需對車輛的位置進行定位，以進行接下來的步驟。

流程 3：確定車輛即時位置後，系統由地圖上找出與該位置相鄰的系統劃分區域。

流程 4：分別預測由車輛即時位置到達流程 3 找出之區域所需的車程時間，車程預測時以一般的駕駛速率進行預測。

流程 5：由需求模型預測該區域即時的需求情形，同時取出該區域當時的空中排班點排班情形，以推斷其排班順位。

流程 6：根據流程 5 所得之資料，預測在該區域等候空中排班所需的時間。例如：由需求預測得知該區域 10 分鐘有一需求產生，而該車的空排順位為 2，則可推得其需在該區等候 20 分鐘方可順利載客。

流程 7：加總流程 4 及流程 6 分別得到的車程時間及空排等候時間，可得出該車到達每一區域的總等候時間。

流程 8：選擇總等候時間最短者為此次運算的最佳派遣區域。

流程 9：判斷由流程 8 所選出的最佳派遣區域的總等候時間是否小於 15 分鐘，若非，則進行至流程 10；若是，則進行至流程 11。其中總等候時間是以自由型駕駛人載客時間間距平均值決定，本研究認為若系統給出的巡迴指引較駕駛人本身所能達到之更差則不應接受此結果。

流程 10：由鄰近區域無法得到總等候時間小於 15 分鐘之派遣區域，故擴大試算範圍，找出較相鄰區域更遠一些的区域後，再回到流程 4 及流程 5 進行分析。

流程 11：由派遣中心發出訊號通知駕駛何處為最佳繞行區域，以提供駕駛更好的繞行方向建議。



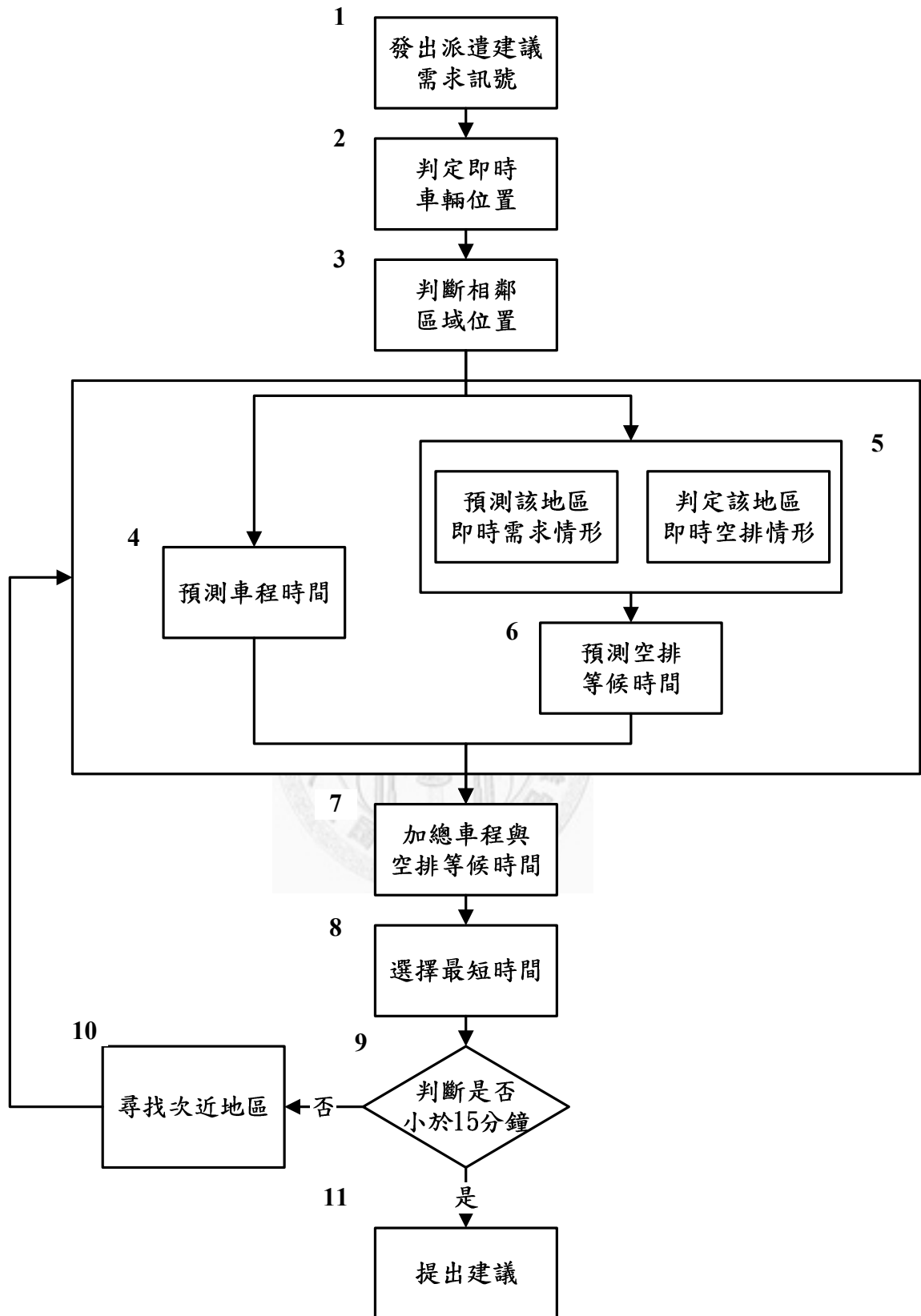


圖 6-3：自由型計程車駕駛人衛星派遣流程圖

6.4 消極型駕駛人

消極型駕駛人的特色為同區載客比例最低，亦即其習慣於不在下客處巡迴攬客。對於載客屬於較不積極的一群，且兩次載客時間間距較長，營運效率較低。根據以上特性，本研究設計適合消極型駕駛人之派遣邏輯如圖 6-4，流程分別說明如下。

流程 1：訊號的發出分為主動式及被動式，主動式為當駕駛完成載客後車機自動向派遣中心發出訊號；被動式為駕駛主動向派遣中心發出繞行區域建議需求。

流程 2：當派遣中心接獲計程車駕駛需求後，首先需對車輛的位置進行定位，以進行接下來的步驟。

流程 3：由於消極型駕駛人的駕駛特性在尖離峰時間差異很大，尖峰時段傾向於在短時間及距離內找到乘客，離峰時段較不在乎往遠處巡迴。因此加入判斷該時段屬於尖峰或離峰的步驟，離峰時接續步驟 4、尖峰時直接進行步驟 5 及步驟 6。

流程 4：確定車輛即時位置後，系統由地圖上找出與該位置相鄰的系統劃分區域。

流程 5：分別預測由車輛即時位置到達前一流程找出之區域所需的車程時間，若為同一區則車程時間為 0，進行車程預測以一般駕駛人平均速度決定。

流程 6：由需求模型預測該區域即時的需求情形，同時取出該區域當時的空中排班點排班情形，以推斷其排班順位。

流程 7：根據流程 6 所得之資料，預測在該區域等候空中排班所需的時間。例如：由需求預測得知該區域 10 分鐘有一需求產生，而該車的空排順位為 2，則可推

得其需在該區等候 20 分鐘方可順利載客。

流程 8：加總流程 5 及流程 7 分別得到的車程時間及空排等候時間，可得出該車到達每一區域的總等候時間。

流程 9：選擇總等候時間最短者為此次運算的最佳派遣區域。

流程 10：判斷由流程 9 所選出的最佳派遣區域的總等候時間是否小於判斷的時間，若非，則進行至流程 9；若是，則進行至流程 10。尖峰時之總等候時間應小於 12 分鐘，離峰時則可在 20 分鐘內。其中總等候時間是以消極型駕駛人載客時間間距平均值決定，本研究認為若系統給出的巡迴指引較駕駛人本身所能達到之更差則不應接受此結果。

流程 11：由鄰近區域無法得到總等候時間小於判斷時間之派遣區域，故擴大試算範圍，找出較相鄰區域更遠一些的区域後，再回到流程 5 及流程 6 進行分析。

流程 12：由派遣中心發出訊號通知駕駛何處為最佳繞行區域，以提供駕駛更好的繞行方向建議。

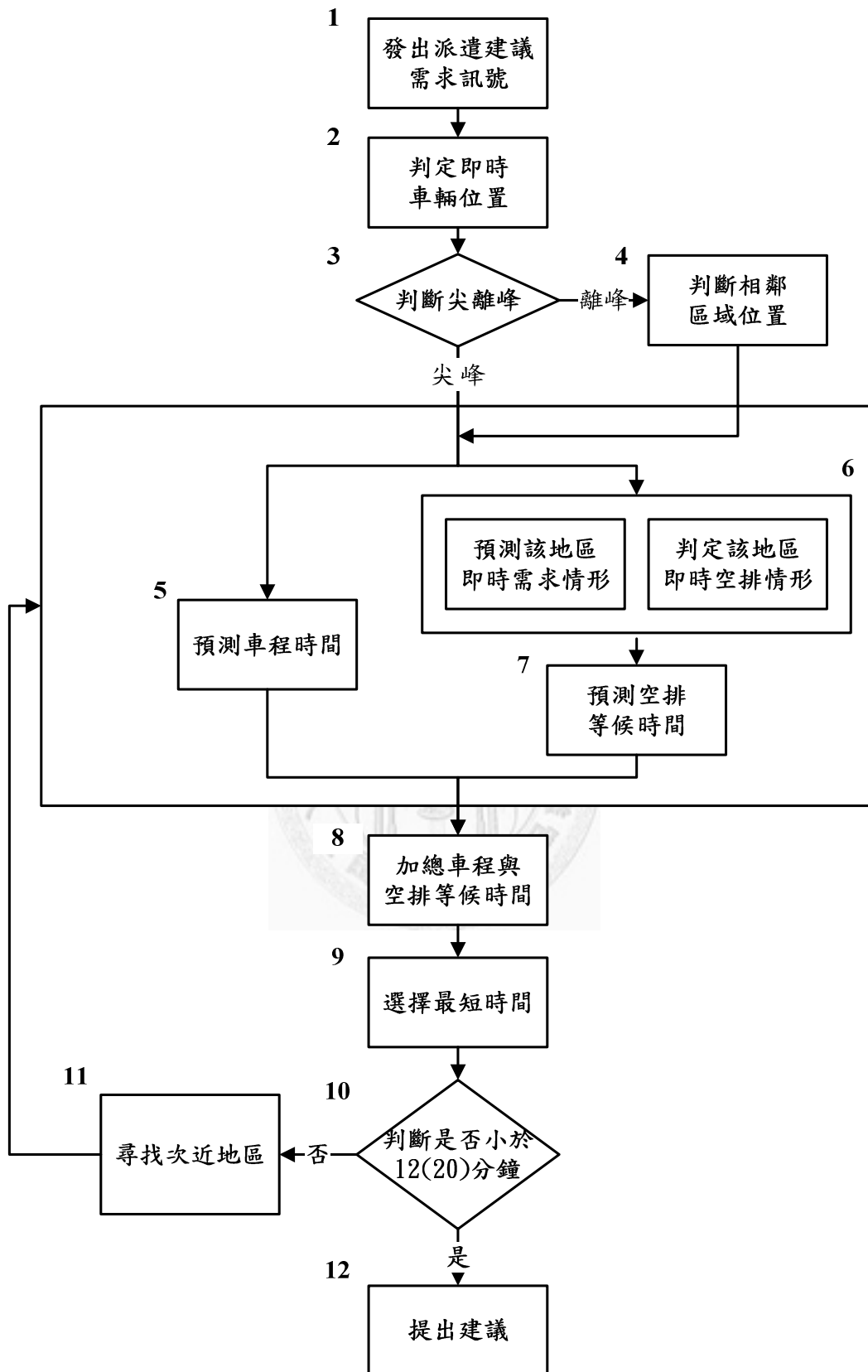


圖 6-4：消極型計程車駕駛人衛星派遣流程圖

6.5 綜合評述

綜合以上派遣邏輯改善建議，本節將整理前述各群駕駛人派遣邏輯設計重點，並整理如表 6-1。由表 6-1 中可知本研究考慮各群駕駛人較特別的營業行為特性作為派遣邏輯設計之依據，透過不同的派遣邏輯，可望讓各群駕駛人對派遣邏輯較有信心，增加衛星派遣計程車駕駛人對該衛星派遣車隊之認同感，藉此吸引游離的個人計程車駕駛人加入衛星派遣車隊，加速衛星派遣車隊達到損益平衡點，進而促使良性循環的發生。

表 6-1：各群駕駛人派遣邏輯設計重點

駕駛人類型	派遣邏輯設計重點
智慧型	<ul style="list-style-type: none"> ➤ 對下客後巡迴之地點沒有特殊偏好，故在第一輪即考慮原區巡迴及鄰近區域巡迴之可能性。 ➤ 習慣於在無客時暫時休息，故不再次計算更遠之區域，若原區與鄰近區域皆無符合適合派遣的情形即建議休息。
積極型	<ul style="list-style-type: none"> ➤ 平均速度較快，移動時間較短。 ➤ 下客後習慣於往遠處巡迴，因此在第一輪先不將原區巡迴考慮在內。
自由型	<ul style="list-style-type: none"> ➤ 對下客後巡迴之地點沒有特殊偏好，故在第一輪即考慮原區巡迴及鄰近區域巡迴之可能性。
消極型	<ul style="list-style-type: none"> ➤ 尖峰時段習慣在短時間內就找到下一個乘客，因此依其需求縮短判斷時間。 ➤ 離峰習慣往遠處巡迴攬客，因此在第一輪先不將原區巡迴考慮在內。



第七章 結論與建議

台北地區大眾運輸路網已逐漸完成，計程車的重要性日益增加，然而計程車駕駛人與乘客長期處於資訊不對稱的情況中，為改善此情勢而產生之衛星派遣系統亦無法發揮其最大效用，導致台北地區計程車空車率居高不下，造成產能閒置及環境污染。而派遣車隊規模無法擴大有許多因素，但根本的因素為計程車駕駛人對加入衛星派遣車隊所能增加的收益存疑，因此對衛星派遣系統興致缺缺。回顧目前的派遣邏輯，發現其僅考慮乘客的需要忽略駕駛人的特性，造成衛星派遣車隊的駕駛人僅能選擇配合或者不利用衛星派遣系統，這個結果對於對衛星派遣系統不甚瞭解之個人計程車駕駛人有相當大之推力。因此本研究以大都會車隊之衛星派遣資料庫為基礎，針對衛星派遣計程車駕駛人之營業行為特性進行資料探勘。首先透過「因素分析」中的主成份分析將衛星派遣資料庫資料之觀察變數縮減為營業情形、空車率及營業習慣三個因素，接著透過「兩階段集群分析」的過程找出衛星派遣計程車營業行為特性之分群，並以此結果回推原始觀察變數之特性，同時配合分群抽樣分析的進行，找出各群衛星派遣計程車駕駛人之營業行為特性，最後根據各群衛星派遣計程車駕駛人之營業行為特性分別為智慧型駕駛人、積極型駕駛人、自由型駕駛人與消極型駕駛人四群。並針對四群駕駛人提出派遣邏輯改善建議。以下依據研究心得分別針對衛星派遣計程車駕駛人之營業行為特性及未來研究課題提出結論與建議。

7.1 結論

本研究將結論分為研究成果、理論意涵及實務意涵三個部份，本節首先整理本研究之具體成果，並分別論述本研究在理論及實務方面的意涵。

7.1.1 研究成果

1. 本研究透過因素分析及集群分析的過程將衛星派遣計程車駕駛人分為四群，依各群駕駛人之營業行為將衛星派遣計程車駕駛人分群為智慧型、積極型、自由型及消極型四群。在營業行為因素以智慧型為最高，而營業行為因素中的營業時程、里程及天數以自由型駕駛人為最低，載客數、載客里程、時程、總收入及平均日收入方面以消極型駕駛人為最低。在空車率因素方面，自由型駕駛人之時間、距離空車率皆為最低，而消極型駕駛人的空車率因素皆為最高。在營業習慣因素中，單次載客收入並沒有顯著差異，而平均速度方面以積極型駕駛人為最高，其他類型駕駛人並無顯著差異。
2. 每日營業情形數據仍是以智慧型駕駛人為最高、積極型駕駛人次之，在營業里程、時程及燃油成本中是以自由型駕駛人最低，每日載客數、載客里程、載客時程及淨收入則是以消極型駕駛人為最低。在淨收入的部份，僅剩下智慧型駕駛人收入較高，消極型及自由型駕駛人每月淨收入已不足二萬元，消極型駕駛人甚至已逼近一萬元。
3. 智慧型駕駛人之營業狀況變數皆最高想是其營業態度相當積極勤奮，但智慧型駕駛人的營業技巧較積極型駕駛人為佳，他們不會白白浪費時間巡迴攬客，在旅次需求較少時傾向於暫時休息，減少油耗的花費。因而其空車率低、營業效率高，本群月收入為最高的 61,000 元左右。
4. 積極型駕駛人偏好於夜間及凌晨時段營業，其營業時之平均速度較快，完成載客後傾向於往較遠處巡迴攬客，非在下客處繞行。營業及載客時間距離皆長，但營業效率卻不佳，然而由於其勤奮工作，月收入仍在第二名的位置，約為 41,000 元。

5. 自由型駕駛人之營業狀況變數皆為次低顯示其營業態度也屬於不積極的，每日營業里程、時程為最低，但每日載客里程、時程卻非最低，且其空車率最低，顯示其為營業效率最高的一群。他們對於招客的技巧很有一套但個性上較不積極，因此月收入也落在倒數第二名的 23,000 元左右。
6. 消極型駕駛人偏好於夜間時段營業，其營業狀況變數皆為最低顯示其營業態度較不積極，營業及載客時間距離皆低。每日營業里程、時程非最低，但每日載客里程、時程卻為最低，再加上空車率為全部最高的一群，顯示其營業效率相當低。在尖峰時間習慣在短時間內就找到下一個乘客，但在離峰的情形下巡迴攬客技巧不佳，習慣於漫無目的的往遠處巡迴攬客。對於攬客屬於土法煉鋼型卻又不積極，月收入為最低的 21,000 元左右。

7.1.2 理論意涵

1. 過去對於計程車駕駛人營業行為的相關研究主要是以派遣數及駕駛人收入做為駕駛人分群依據，本研究利用由載客數、載客里程、載客時程、營業時程、營業里程、營業天數、總收入及平均日收入組成的「營業行為」，包含時間及距離空車率的「空車率」，以及由平均速度及單次載客收入組成的「營業習慣」三個主要因素進行衛星派遣計程車駕駛人分群。透過分群因素的增加，能有效的增加研究的廣度，對於衛星派遣計程車駕駛人營業行為的分析應更為全面。
2. 在侯勝宗(2006)的研究中採用的分群因素為「科技心理擁有感」、「在地知識」、「科技意會」及「科技採用頻次」，這些分群因素主要代表衛星派遣計程車駕駛人之心理感受及駕駛技能；而本研究所採用的分群因素為「營業行為」、「空車率」及「營業習慣」，這些分群因素主要代表衛星派遣計程車駕駛

人之實際營業狀況。因此，在研究範圍中，本研究將衛星派遣計程車駕駛人行為分析的資料來源由駕駛人之自我認知推進至實際營業情形。透過實際營業情形變數的導入，可排除駕駛人回答問題時隱惡揚善的現象，藉此得到較貼近駕駛人真實營業行為之分群因素。

3. 衛星派遣資料庫中蘊含許多車隊成員的資訊，若能妥善運用可望解決許多衛星派遣車隊面臨的困境，進而改善目前的計程車產業。再者，對衛星派遣車隊而言，衛星派遣資料庫為固有的資產，車隊無須另外付費即可取得，過去未能有效的應用實屬可惜。因此，本研究引入資料探勘技術的概念，藉以分析衛星派遣計程車駕駛人營業行為。透過衛星派遣資料庫的使用，可將衛星派遣計程車駕駛人營業行為之研究由質性分析的角度導向數值面的分析。同時，藉由資料探勘技術的引入，本研究開啟了衛星派遣資料庫新的分析概念及使用方式。未來可望利用資料探勘技術，解決更多計程車產業所面臨之難題。
4. 根據 95 年度及 97 年度台北地區營運情形調查可知，95 年度及 97 年度台北地區之時間空車率分別為 72.3% 及 80.0%。本研究採用衛星派遣技術導入初期之大都會衛星車隊資料進行分析，分析的結果顯示自由型駕駛人為空車率最低的一群，其時間空車率為 73.4%，其與普查之資料相去不遠。然而衛星派遣技術導入較早之台灣大車隊在 95 年度的時間空車率即僅有 43%，明顯較普查之資料為小。這個現象顯示駕駛人對於新技術的導入需要一段適應的時期，在科技導入的初期（本研究所採用之大都會衛星車隊資料），衛星派遣車隊駕駛人對新科技心存疑慮且不知如何善用之，故對於空車率的減少有限。然而，當科技的導入日趨成熟（台灣大車隊 95 年度之資料），衛星派遣車隊駕駛人懂得如何利用新技術提升其營業效率，這個現象使得其空車率大幅減少，營業效率較其他未使用衛星派遣技術之計程車高。

7.1.3 實務意涵

1. 由前述理論意涵可知，計程車駕駛人在初接觸衛星派遣技術時有一段摸索期。在這段時期，計程車駕駛人無法善加利用衛星派遣技術。因此，可在計程車駕駛人加入車隊時給予計程車駕駛人職前訓練，藉以縮短計程車駕駛人對衛星派遣技術之摸索期，加速衛星派遣技術優點之發揮。
2. 本研究提出一套「類空中排班」的派遣邏輯，除改善原有空中排班邏輯上未考慮計程車駕駛人需要之缺點外，亦根據各群駕駛人營業行為的差異，分別設計符合駕駛人需求之派遣邏輯。智慧型駕駛人的派遣邏輯設計重點為：無巡迴地點偏好及習慣於在無客時暫時休息；積極型駕駛人的派遣邏輯設計重點為：習慣於遠處巡迴及平均速度較快；自由型駕駛人的派遣邏輯設計重點為：無巡迴地點偏好；消極型駕駛人的派遣邏輯設計重點為：尖離峰時段的營業行為差異。
3. 對人力資源管理層面而言，教育訓練會影響衛星派遣計程車駕駛人的營業狀況。然而，衛星派遣計程車駕駛人的情況較為特殊，統一的教育訓練無法一一滿足各群駕駛人之需要。因此，車隊應針對各群駕駛人之特性，設計出不同的教育訓練目標及方式。加強各群駕駛人好的營業行為，同時改善其較差的營業行為。藉由教育分流的概念，分別提供各群駕駛人所需之營業觀念及技巧，以改善各群駕駛人之營業狀況。

7.2 建議

1. 由本研究之衛星派遣計程車駕駛人營業行為分析可發現，各群駕駛人的月收

入差異懸殊，消極型駕駛人之淨月收入僅有一萬三千元左右，而智慧型駕駛人之淨月收入有四萬四千元左右，其中超過三倍的收入起因於營運情況以及營運技巧。建議應對消極型駕駛人進行宣導與教育訓練，並給予其營運技巧的正確觀念，不但可增加消極型駕駛人之收入，更能進而減少無謂的巡迴繞行造成之環境污染。

2. 本研究僅提出現有衛星派遣計程車駕駛人之營業行為分群及分析，若因派遣邏輯的改善促使更多計程車駕駛的加入，勢必會遇到新加入之駕駛人如何歸類的問題。建議可利用判別分析設計一套歸納流程，讓新加入的駕駛人亦可迅速歸納入原派遣邏輯分群。
3. 本研究因資料取得的問題僅選用大都會衛星派遣車隊之衛星派遣應用前期資料庫進行分析，未來若能取得規模擴大後以及其他車隊之衛星派遣資料庫，建議可利用本研究之方法，進一步比較不同車隊規模及不同公司之計程車駕駛人營業情形、行為以及績效等議題。
4. 排班計程車駕駛人為目前計程車產業中為數不少的一群，而且有逐步增加趨勢，本研究並未特別將之提出討論，建議未來可透過載客起點之衛星定位，區分出排班計程車駕駛人，並分析排班計程車駕駛人之營業行為特性。
5. 本研究僅以計程車駕駛人營業行為特性進行分析，然而在眾多的研究中顯示駕駛人之個人特性（如性別、年齡等）對其駕駛行為之影響亦相當顯著。故建議可利用車隊隊編連結本研究之成果及駕駛人個人資料，並設計駕駛人心理狀態問卷，以進一步分析駕駛人個人特性、心理狀況與營業行為間之關係。
6. 本研究對象大都會衛星車隊之週末營業額比例明顯高於其他的衛星派遣車隊，其成因眾說紛紜，至今尚未獲得一明確之合理解釋，建議未來可針對此

現象進行深入探討。

7. 本研究僅討論計程車駕駛人之營業行為，亦即僅分析計程車市場之供給面。然而，市場的需求的改變亦會影響供給面的狀況。故建議未來的研究可加入計程車市場需求面的探討，以較完整以及貼近現實的方式，分析計程車產業運作之狀況。
8. 本研究提出之派遣邏輯改善策略僅針對計程車駕駛人，而派遣邏輯的設計僅為概念式的提出一套考慮計程車駕駛人營業行為特性之流程，建議未來衛星派遣計程車隊應參考本研究之設計，進一步加入乘客需求的觀念，設計一套考慮駕駛及乘客之派遣邏輯。





參考文獻

1. 尹相志(2007)，SQL Server 2005 Data Mining 資料採礦，悅知文化。
2. 王建仁(2003)，「台灣地區機車使用者風險感認與駕駛行為關聯之研究」，國立交通大學運輸科技與管理研究所碩士論文。
3. 王偉(1986)，「台北市汽車駕駛行為特性分析」，國立台灣大學土木工程學研究所碩士論文。
4. 任維廉、徐聖智、張傳琳(1997)，「公車駕駛人工作壓力量表設計與測試：以大有巴士、福和客運為例」，運輸學刊，第十卷，第一期，pp.91-118。
5. 任維廉、黃清信(1993)，「台北都會區公車駕駛員工作壓力源之研究」，中華民國運輸學會第八屆論文研討會論文集，pp.465-472。
6. 任維廉、鄭亮傑(1996)，「公路客運公司員工對績效獎金制度態度之研究—以新竹客運為例」，中華民國運輸學會第十一屆學術研討會論文集，pp.685-697。
7. 吳冠宏、吳信宏、郭廣洋(2004)，「應用資料挖掘於交通事故資料分析」，中華民國品質學會第40屆年會高雄市分會第30屆年會暨第10屆全國品質管理研討會論文集，pp.72-81。
8. 吳鴻春(2009)，「國民中小學教科書數位學習系統多準則評估模式之建構」，開南大學資訊管理學碩士論文。
9. 呂錦山、方菟萍(2001)，「高雄港貨櫃市場行銷策略之研究」，運輸計畫季刊，第三十卷，第二期，pp.439-480。
10. 呂錦山、陳李逸(2003)，「定期貨櫃航商電子商務服務屬性需求之探討」，運輸計畫季刊，第三十二卷，第一期，pp.177-193。
11. 李東陽(1996)，「駕駛環境複雜度之衡量方法及特性分析」，淡江大學交通管理研究所碩士論文。
12. 汪進財、鐘易詩、蕭力文(2009)，「年輕機車族群高風險駕駛行為異質性剖析」，中華民國運輸學會98年學術研討會論文集，pp.877-900。

13. 阮士峰(2004)，「資料探勘在股市序列型樣的應用」，國立東華大學資訊工程研究所碩士在職專班論文。
14. 林清山(1980)，多變量分析統計法，台灣東華書局股份有限公司。
15. 林祥生、劉益豪(2008)，「應用資料採礦探討國際線航空旅客之線上購票行為」，運輸計畫季刊，第三十七卷，第二期，pp.197-236。
16. 林震岩(2007)，多變量分析—SPSS 的操作與應用，智勝文化事業有限公司，再版。
17. 邱詩淳(2006)，「運用悠遊卡及資料探勘求解公車營運改善方案」，中華大學運輸科技與物流管理研究所碩士論文。
18. 侯勝宗(2006)，「科技心理擁有感、在地知識與科技採用：科技意會觀點」，國立政治大學科技管理研究所博士論文。
19. 侯勝宗、蕭瑞麟(2008)，科技意會，台灣培生教育出版股份有限公司。
20. 侯勝宗、洪鈞澤等(2009)，「從使用者觀點探討服務派遣技術商業化之研究」，經濟部商業司委託逢甲大學及季鈞管理顧問有限公司辦理。
21. 施念青(2002)，「公路運輸部門能源消費與 SO_x、NO_x、CO₂ 排放特性分析」，國立成功大學環境工程研究所碩士論文。
22. 洪浩凱(2002)，「台北悠遊卡使用者對於台北大眾捷運公司顧客忠誠度之研究」，南華大學旅遊事業管理研究所碩士論文。
23. 胡守任、李志綱、鄭珮琪(2009)，「鐵路平交道前公路駕駛人行為分析」，中華民國運輸學會 98 年學術研討會論文集，pp.1733-1759。
24. 張明智(2005)，「以空間資料搜尋支援資料探勘之研究」，國立成功大學測量及空間學系碩士論文。
25. 張勝雄(1999)，「大眾運輸業之薪資結構與工作壓力、工作滿足、服務品質之關係探討—台北市公民營公車駕駛員之比較研究」，行政院國家科學委員會專題研究計畫成果報告。

26. 張媛婷(2008),「都市土地使用管制對土地價格之影響—階層線性模式之應用」,國立成功大學都市計畫研究所碩士論文。
27. 張新立、廖哲聖(2009),「憤怒駕駛傾向之量測與人格特質及性別角色之影響」,中華民國運輸學會 98 年學術研討會論文集, pp.901-931。
28. 張新立、蔡維唐(2008),「影響小客車駕駛人駕駛技能與行為因素研究」,中華民國運輸學會 97 年學術論文國際研討會。
29. 張蓓蒂、李宸邦(2006),「行銷創新、品牌形象與銷售績效關聯性之研究—以多層次傳銷事業個案公司為例」,第十一屆直銷學術研討會論文集, pp.77-101。
30. 張學孔、吳奇軒、陳育生(2009),「計程車產業政策關鍵因素分析」,運輸計畫季刊,第三十八期,第二卷, pp.173~200。
31. 張學孔、周文生(2008),「九十七年度台北地區計程車營運情形調查期末報告」,台北市政府交通局與台北縣政府交通局共同委託中華民國運輸學會辦理, pp.14-98。
32. 莊涵婷(2004),「通勤旅程中非工作活動產生模式之研究」,中原大學土木工程研究所碩士論文。
33. 陳宜欣(2006),「以資料探勘技術探討顧客忠誠方案—以某信用卡發卡銀行為例」,國立中正大學行銷管理研究所碩士論文。
34. 陳宜萍、何庭武、曾思敏、周建成(2009),「應用資料探勘決策工具協助路權取得流程之研究」,中華民國運輸學會 98 年學術論文研討會論文集, pp.2393-2403。
35. 郭信宏(2008),「應用資料探勘技術於面板檢測實證研究」,國立中央大學工業管理研究所碩士論文。
36. 陳國樑(2004),「公路客運公司管理制度與駕駛員行為特性關係之研究」,逢甲大學交通工程與管理學系碩士論文。
37. 陳國樑(2004),「公路客運公司管理制度與駕駛員行為特性關係之研究」,逢甲大學交通工程與管理學系碩士論文。

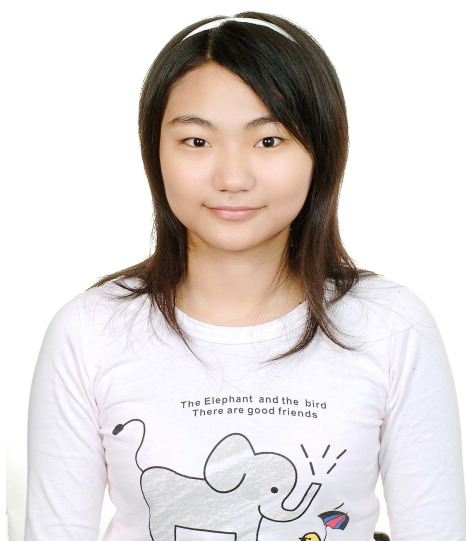
38. 陳菊萍(2003)，「地區推動永續發展行為意向之研究—以大鵬灣周邊鄉鎮為例」，國立中山大學公共事務管理研究所碩士論文。
39. 陳順宇(2005)，多變量分析，華泰書局，四版。
40. 曾國雄、鄧振源(1986)，多變量分析(一)—理論應用篇，松岡電腦圖書資料有限公司。
41. 曾瑋悅(2002)，「薪資制度對公車駕駛人壓力、滿足感與績效之影響—以台北市聯營公車為例」，淡江大學運輸管理研究所碩士論文。
42. 曾憲雄、蔡秀滿、蘇東興、曾秋蓉、王慶堯(2005)，資料探勘，旗標出版社。
43. 程法泌(1979)，「汽車駕駛行為之研究與預測 (I)」，國立政治大學學報，第39期。
44. 黃俊英(2000)，多變量分析，中國經濟企業研究所，華泰文化事業公司總經銷，第七版。
45. 黃致諳(2004)，「資料探勘技術用於結合客戶價值與客戶消費行為的直銷行銷之研究」，國立台灣科技大學資訊管理研究所碩士論文。
46. 黃國平、賴博階(2008)，「衛星派遣計程車之乘客選擇意象比較」，運輸學刊，第二十卷，第二期，pp.119-145。
47. 黃湄清(2005)，「利用資料探勘技術於台灣地區肇事危險判別之研究」，國立中央大學土木工程研究所碩士論文。
48. 楊浩二(1984)，多變量統計方法，華泰書局，初版。
49. 劉正華(1996)，「駕駛行為之風險評估研究」，東海大學統計學系碩士論文。
50. 劉建良(2000)，「公車駕駛員之薪資滿足感與工作壓力、工作滿足、服務品質之關係探討」，淡江大學運輸管理學系運輸科學研究所碩士論文。
51. 潘書麟(2006)，「台灣物流業營運效率之比較分析」，國立中央大學土木工程學系碩士論文。
52. 蔡憲唐、周建新(1987)，「物以類聚—集群分析」，科學月刊雜誌社，第二零九期。

53. 蔡鎮安(2009),「集群分析」,中國醫藥大學生統E報,第三十六期。
54. 駱至中、林錦昌(2002),「以民意調查資料的智慧型分析看資料探勘於政治學研究之運用」,國家政策研究基金會第二屆政治與資訊學術研討會國政研究報告。
55. 龍天立(1987),「台北市車輛駕駛行為特性之研究」,市政建設專題研究報告第159輯,台北市政府研究發展考核委員會委託台灣大學土木工程學系辦理。
56. 羅明旋(2009),「貓空地區觀光意象對遊客選擇行為影響之路徑分析」,國立政治大學地政學系私立中國地政研究所碩士論文。
57. 羅惟元(2007),「以悠遊卡交易資料探索公車路線之旅客起迄」,淡江大學運輸管理研究所碩士論文。
58. 鐘志成、張恩輔(2009),「以票證紀錄資料模擬台鐵旅客時空分佈特性」,中華民國運輸學會98年學術研討會論文集,pp.1053~1071。
59. Cabena, P., Hadjinian, P. O., Stadler, R., Verhees and J., and Zanasi, A. (1997), *Discovering Data Mining from Concept to Implementation*, New Jersey: Prentice-Hall, Inc., pp.41-59.
60. Kennedy, R. L., Lee, Y., Van, R. B., Reed, C. D. and Lippman, R. P. (1998), *Solving Data Mining Problems through Pattern Recognition*, Prentice Hall.
61. Liou, J. H., Yen, L. and Tzeng, G. H. (2010), “Using Decision Rules to Achieve Mass Customization of Airline Services”, *European Journal of Operational Research*, Vol. 205, No. 3, pp.680–686.
62. Liou, J. H., Yen, L. and Tzeng, G. H. (2010), “A Dominance-Based Rough Set Approach to Customer Behavior in the Airline Market”, *Information Sciences*, Vol. 180, No. 11, pp.2230–2238.
63. Clark, S. D. (2009),”The Determinants of Car Ownership in England and Wales from Anonymous 2001 Census Data”, *Transportation Research C*, Vol. 17, No. 5, pp.526–540.

64. Wong, Y. K. and Woon, W. L. (2008), “An Iterative Approach to Enhanced Traffic Signal Optimization”, *Expert Systems with Applications*, Vol. 34, No. 4, pp.2885–2890.
65. Tabachnick, B. G. and Fidell, L. S. (2007), *Using Multivariate Statistics*, Allyn and Bacon, 4th ed.
66. Yao, X. (2007), “Where are Public Transit Needed – Examining Potential Demand for Public Transit for Commuting Trips”, *Computers, Environment and Urban Systems*, Vol. 31, No.5, pp.535-550.
67. Chang, L. Y. and Wang, H. W. (2006), “Analysis of Traffic Injury Severity: An Application of Non-Parametric Classification Tree Techniques”, *Accident Analysis and Prevention*, Vol. 38, No. 5, pp.1019–1027.
68. Bagchi, M. and White, P. R. (2005), “The Potential of Public Transport Smart Card Data”, *Transport Policy*, Vol. 12, No. 5, pp.464-474.
69. Roiger, R.J. and Geatz, M.W. (2003), *Data Mining: A Tutorial-Based Primer*, Addison Wesley.
70. Dunham, M.H. (2002), *Data Mining, Introductory and Advanced Topics*, Prentice Hall.
71. Han, J. and Kamber, M. (2001), *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers.
72. Han, J. and Kamber M. (2001), *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publisher, Inc.
73. Thuraisingham, B. (2000), “A Primer for Understanding and Applying Data Mining”, *IT Professional*, Vol. 2, No. 1, pp.28-31.
74. Berry, M.J.A. and Linoff, G.S. (2000), *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*, Wiley Publishing, Inc.
75. Kuge, N., Yamamura, T., Shimoyama, O. and Liu A. (2000), “A Driver Behavior

- Recognition Method Based on a Driver Model Framework”, Society of Automotive Engineers, Vol. 109, No. 6, pp.469-476.
76. Peeta, S., Ramos, J. L. and Pasupathy, R. (2000),”Content of Variable Message Signs and On-Line Driver Behavior”, Transportation Research Record, Vol. 1725, No. 970, pp.102-108.
77. Peacock, P.R. (1998), “Data Mining in Marketing: Part1,” Marketing Management, Vol. 6, No. 4, pp.8-18.
78. Fayyad, U. M., Haussler, D. and Stolorz, P. (1996), “Mining Scientific Data”, Communications of ACM, Vol. 39, pp.55-57.
79. Fayyad, U. M. (1996), “Data Mining and Knowledge Discovery: Making Sense out of Data”, IEEE Expert, Vol. 11, No. 5, pp.20-25.
80. Grupe, F. H. and Owrang, M. M. (1995), “Data Base Mining”, Information System Management, Vol. 12, No. 4, pp.26-31.
81. Frawley, W., Piatetsky-Shapiro, G. and Matheus, C. (1992), “Knowledge Discovery in Database: An Overview”, AI Magazine, pp.213-218.
82. Fayyad, U. M., Piatetsky-Shapiro, G. and Matheus, C. (1992), “The KDD Process for Extracting Useful Knowledge from Volumes OD Data”, Communications of ACM, Vol. 39, No. 11, pp.27-34.
83. Taschdjian, E. (1986), “Typology and Factor Analysis”, Nanzan Institute for Religion and Culture, Folklore Studies, Vol. 2, pp.185-200.
84. Fine, J. L. (1965), “The Development of a Criterion for Driver Behavior”, The Safety Research and Education Project, Teachers College, Columbia University.
85. Herman, H. H. (1960), Modern Factor Analysis, The University of Chicago Press.
86. Kaiser, H. F. (1960).” The Application of Electronic Computers to Factor Analysis”, Educational and Psychological Measurement, Vol. 10, pp.141-151.

簡歷



基本資料	
姓名	吳沛儒
出生日期	74.06.13
性別	女
出生地	台灣高雄

學歷			
	學校名稱	院系科別	修業年限
研究所	台灣大學	土木工程學系 交通工程組	97.09-99.06
大學	台灣大學	土木工程學系	92.09-97.06
高中	聖功女中	--	89.09-92.06
國中	聖功女中	--	86.09-89.06
國小	十全國小	--	80.09-86.06