

國 立 交 通 大 學

運輸科技與管理學系

碩 士 論 文

比較 k-NN 模式與時變係數模式對高速公路

旅行時間預測之研究

A Study of Comparison of k-NN Model and Time-Varying  
Coefficient Model for Predicting Travel Time on Freeways

研 究 生：陳建旻

指 導 教 授：王晉元

中 華 民 國 九 十 八 年 六 月

# 比較 k-NN 模式與時變係數模式對高速公路旅行時間預測之研究

學生：陳建旻

指導教授：王晉元

國立交通大學運輸科技與管理學系碩士班

## 摘要

近年來政府大力推動智慧型運輸系統(Intelligent Transportation System, ITS)，而先進用路人資訊系統(Advanced Traveler Information System, ATIS)即為 ITS 中的一子系統。在 ATIS 中，為了要給予用路人準確的資訊，以作為路徑、運具選擇之依據，路徑旅行時間的預測是一項重要的課題。尤其在高速公路路網建置完成後，適當交通預測資訊之提供對用路人行為決策更顯得其重要，不僅可以作為駕駛者選擇適當之路徑與出發時間憑藉之依據，用路者亦可藉此選擇最短之旅行時間到達目的地，以真正發揮高速公路路網之整體績效。

本研究係針對國內高速公路為研究對象，利用探針車所蒐集到的即時交通資料，分別利用 k-NN 模式與時變係數模式預測未來旅行時間，並針對兩種模式的預測結果進行績效評估並作比較，以期能提供精準之旅行時間預測，提供用路人路徑選擇或是出發時間決策判斷之依據。

本研究以國道一號高速公路楊梅到泰山收費站作為實際測試對象，經由測試結果可得知，本研究所構建的 k-NN 與時變係數兩種旅行時間預測模式，都是屬於高精準預測，而 k-NN 模式又能夠得到比時變係數模式更好的預測結果。因此，從測試結果顯示出本研究的預測方法可實際應用在高速公路上，且可提供用路人精準的旅行時間預測。

**關鍵字：**k-NN 模式；時變係數模式；旅行時間預測；探針車

A Study of Comparison of k-NN Model and Time-Varying Coefficient Model  
for Predicting Travel Time on Freeways

Student : Jian-Min Chen

Advisor : Jin-Yuan Wang

Department of Transportation Technology and Management  
National Chio Tung University

**Abstract**

In recent years, the government is actively promoting Intelligent Transportation System (ITS), and Advanced Traveler Information System (ATIS) is a subsystem of ITS. Travel time prediction is a very important of ATIS. When drivers have to make a decision, it is more important for drivers to use suitable traffic information. Traffic information will allow drivers to select appropriate routes and departure time to avoid congestion and arrive in the destination with the shortest time.

In this study, the probe vehicles collect real-time traffic information, and use the k-NN model and Time-Varying Coefficients (TVC) model to predict the future travel time, respectively. Evaluation and Comparison of two models for forecasting the results, hope to provide accurate forecasts of travel time to travelers departure time or route choice decision-making judgements based on.

We use the 1st National Freeway Yangmei to Taishan Toll Station as the actual test object. The testing results show that k-NN model and TVC model are high precision prediction, and k-NN model predict better than TVC model. The prediction method can actually use on the highway, and can provide accurate prediction of travel time to drivers.

**Keyword:** k-NN model, TVC model, travel time prediction, Probe Vehicle

## 誌謝

兩年的研究所生活，很快的即將畫下句點，也即將脫離學生生活，準備踏入社會面對另一個嶄新的挑戰，在研究所期間，最感謝的人就是指導教授王晉元老師，其認真負責的教學態度，讓學生在學業或生活上都受益良多，每當我在學習上遇到瓶頸時，老師都會適時的給予鼓勵與幫助，讓我得以突破困境，面對接下來的挑戰，在此僅致上學生最誠摯的謝意與敬意，並感謝中華大學張靖老師與台北科技大學盧宗成老師能撥冗參與口試審查，提供一些論文不足處的修改意見，使本論文更臻完備。

很高興在研究室遇到那麼多優秀的學長，其中最感謝的學長彥佑大人，每當遇到問題時，都很有耐心的教導我並時常鼓勵我，讓我更有信心能順利完成論文，也因為有你的幫忙，我才能如此順利畢業，要感謝小松學長的教導，讓我開始接觸學習寫程式，感謝思文學長在課業學習上的幫助，以及好友芬傑與士勛在研究之餘一起看球賽、分享生活甘苦的日子真的很開心，老總不時的給予研究上的建議，並不厭其煩的仔細講解，聰儒、葉珮一起在課業上學習，以及可愛的學弟妹宇凡、佳儒、小惠、Daisy 的陪伴，讓我在研究所生活更多采多姿。

還要感謝的一直在背後默默支持我的爸媽，因為有你們的支持與照顧，讓我得以順利完成學業，兩位姐姐在我心情低落時給予打氣、加油，不斷地鼓勵我，以及嘉琳的支持讓我在學習上更具信心，表姊阿珍、表哥咪咪當我學習的好榜樣，讓我在學習路上知道如何向前邁進。

能完成這篇論文，要感謝的人實在是太多了，曾給過我幫助或鼓勵的人，我都不會忘記，非常感謝你們，也希望在此能與你們共同分享這份喜悅，謝謝大家！

陳建旻  
新竹交大  
2009/07/23

# 目 錄

中文摘要 .....	i
英文摘要 .....	ii
誌謝 .....	iii
目錄 .....	iv
表目錄 .....	vi
圖目錄 .....	viii
第一章 緒論 .....	1
1.1 研究動機 .....	1
1.2 研究目的 .....	2
1.3 研究範圍 .....	2
1.4 研究內容 .....	2
1.5 研究流程 .....	2
第二章 文獻回顧 .....	5
2.1 線性迴歸方法 .....	5
2.2 時間序列法 .....	6
2.3 類神經網路法 .....	7
2.4 k-NN 方法 .....	11
2.5 其他預測方法 .....	14
2.6 小結 .....	17
第三章 k-NN 模式構建 .....	18
3.1 k-NN 模式介紹 .....	18
3.1.1 k-NN 資料分群 .....	20
3.1.2 選擇距離量度 .....	20
3.1.3 k 值設定 .....	21
3.2 k-NN 模式構建 .....	22
3.3 k-NN 模式範例說明 .....	26
第四章 時變係數模式構建 .....	30
4.1 TVC 模式介紹 .....	30
4.1.1 TVC 資料分群 .....	32
4.1.2 迴歸基本統計假設 .....	32
4.1.3 檢定統計顯著性 .....	33
4.1.4 判定係數 $R^2$ .....	36
4.2 TVC 模式構建 .....	37

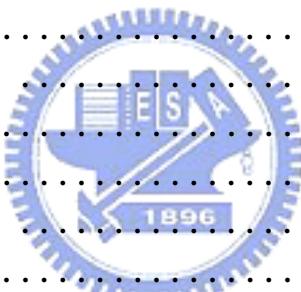
4.3 資料整理 .....	39
4.4 預測精確度之衡量 .....	42
<b>第五章 數值測試 .....</b>	<b>43</b>
5.1 模式構建與測試 .....	43
5.1.1 測試路段 .....	43
5.1.2 預測時段 .....	43
5.1.3 蒐集資料 .....	44
5.2 建立 k-NN 預測模式 .....	46
5.3 k-NN 預測結果 .....	47
5.4 建立 TVC 預測模式 .....	60
5.4.1 測試資料整理 .....	60
5.4.2 迴歸模式彙整 .....	81
5.5 TVC 預測結果 .....	82
5.6 績效評估 .....	90
<b>第六章 結論與建議 .....</b>	<b>92</b>
6.1 結論 .....	92
6.2 建議 .....	93
<b>參考文獻 .....</b>	<b>94</b>



## 表目錄

表 2.1 各種預測方法之比較 .....	16
表 3.1 各種資料庫之資料形式 .....	24
表 3.2 k-NN 距離量度資料表 .....	27
表 3.3 8:05 旅行時間資料表 .....	28
表 3.4 8:15 旅行時間資料表 .....	29
表 4.1 變異數分析表 .....	35
表 4.2 TVC 資料整理表 .....	40
表 4.3 MAPE 預測能力之等級 .....	42
表 5.1 各固定點編號名稱及位置 .....	43
表 5.2 旅行時間資料庫格式 .....	44
表 5.3 ETC 旅行時間資料庫格式 .....	45
表 5.4 門檻值對應 k 值(1) .....	48
表 5.5 k 值對應絕對誤差百分比(1) .....	49
表 5.6 k-NN 預測結果(1) .....	50
表 5.7 門檻值對應 k 值(2) .....	53
表 5.8 k 值對應絕對誤差百分比(2) .....	54
表 5.9 k-NN 預測結果(2) .....	56
表 5.10 上、下游固定點變數名稱對照 .....	61
表 5.11 k-NN 模式摘要(1) .....	62
表 5.12 變異數分析(1) .....	62
表 5.13 係數(1) .....	62
表 5.14 模式摘要(2) .....	63
表 5.15 變異數分析(2) .....	63
表 5.16 係數(2) .....	64
表 5.17 模式摘要(3) .....	64
表 5.18 變異數分析(3) .....	65
表 5.19 係數(3) .....	65
表 5.20 模式摘要(4) .....	66
表 5.21 變異數分析(4) .....	66
表 5.22 係數(4) .....	66
表 5.23 模式摘要(5) .....	67
表 5.24 變異數分析(5) .....	67
表 5.25 係數(5) .....	68
表 5.26 模式摘要(6) .....	68
表 5.27 變異數分析(6) .....	69

表 5.28 係數(6) .....	69
表 5.29 模式摘要(7) .....	70
表 5.30 變異數分析(7) .....	71
表 5.31 係數(7) .....	71
表 5.32 模式摘要(8) .....	72
表 5.33 變異數分析(8) .....	72
表 5.34 係數(8) .....	72
表 5.35 模式摘要(9) .....	73
表 5.36 變異數分析(9) .....	73
表 5.37 係數(9) .....	74
表 5.38 模式摘要(10) .....	74
表 5.39 變異數分析(10) .....	75
表 5.40 係數(10) .....	75
表 5.41 模式摘要(11) .....	76
表 5.42 變異數分析(11) .....	76
表 5.43 係數(11) .....	76
表 5.44 模式摘要(12) .....	77
表 5.45 變異數分析(12) .....	77
表 5.46 係數(12) .....	78
表 5.47 模式摘要(13) .....	78
表 5.48 變異數分析(13) .....	79
表 5.49 係數(13) .....	79
表 5.50 模式摘要(14) .....	80
表 5.51 變異數分析(14) .....	80
表 5.52 係數(14) .....	80
表 5.53 迴歸檢定彙整表 .....	81
表 5.54 TVC 預測結果(1) .....	83
表 5.55 TVC 預測結果(2) .....	86



## 圖目錄

圖 1.1 研究流程圖 .....	4
圖 3.1 k-NN 流程圖 .....	19
圖 3.2 k-NN 模式構建流程圖 .....	23
圖 3.3 k-NN 範例路段示意圖 .....	26
圖 4.1 TVC 流程圖 .....	31
圖 4.2 TVC 模式構建流程圖 .....	38
圖 4.3 TVC 範例路段示意圖 .....	39
圖 5.1 旅行時間資料庫 .....	45
圖 5.2 ETC 旅行時間資料庫 .....	46
圖 5.3 k-NN 預測結果(1) .....	51
圖 5.4 3/6 k-NN 預測結果 .....	51
圖 5.5 3/13 k-NN 預測結果 .....	52
圖 5.6 3/20 k-NN 預測結果 .....	52
圖 5.7 3/27 k-NN 預測結果 .....	53
圖 5.8 k-NN 預測結果(2) .....	57
圖 5.9 3/1 k-NN 預測結果 .....	57
圖 5.10 3/8 k-NN 預測結果 .....	58
圖 5.11 3/15 k-NN 預測結果 .....	58
圖 5.12 3/22 k-NN 預測結果 .....	59
圖 5.13 3/29 k-NN 預測結果 .....	59
圖 5.14 迴歸標準化殘差 P-P 圖(1) .....	61
圖 5.15 迴歸標準化殘差 P-P 圖(2) .....	70
圖 5.16 TVC 預測結果(1) .....	84
圖 5.17 3/6 TVC 預測結果 .....	84
圖 5.18 3/13 TVC 預測結果 .....	85
圖 5.19 3/20 TVC 預測結果 .....	85
圖 5.20 3/27 TVC 預測結果 .....	86
圖 5.21 TVC 預測結果(2) .....	87
圖 5.17 3/1 TVC 預測結果 .....	88
圖 5.18 3/8 TVC 預測結果 .....	88
圖 5.19 3/15 TVC 預測結果 .....	89
圖 5.20 3/22 TVC 預測結果 .....	89
圖 5.21 3/29 TVC 預測結果 .....	90

# 第一章 緒論

## 1.1 研究動機

中山高速公路的興建，帶動台灣西部經濟走廊的基礎，自民國 67 年全線通車以來，車流量年年倍增，至今仍持續扮演台灣西部重要之旅次運輸角色。而隨著交通量急遽成長，其車流量遠超過原先所估計，使得中山高達到近飽和的狀態，交通壅塞的情形也變得比過去嚴重。為解決此問題，政府又興建了福爾摩沙高速公路(北二高)，在民國 93 已全線通車，加上近年持續興建的國道路線及東西向快速道路，著實形成高速公路路網，也達到促進地區均衡發展等目標。

然而隨著時間變化、道路狀況不同等因素，讓用路人對於高速公路旅行時間變化無法掌握，而用路人只能憑藉過往經驗來選擇欲行駛之道路，造成高速公路路網無法發揮預期之功效，某些路段嚴重壅塞。而隨著科技的日益進步，若能運用先進的運輸通訊設備，有效的預測未來旅行時間的變化，提供給用路人即時、準確的動態交通資訊，不但能協助用路人作出正確決策，節省行車時間，還可同時提升路網整體使用效率。

近年來政府大力推動智慧型運輸系統(Intelligent Transportation System, ITS)，而先進旅行者資訊系統(Advanced Traveler Information System, ATIS)即為 ITS 中的一子系統，主要將不同來源之交通資料作整合，提供用路人即時或預測的交通資訊，例如：交通擁塞路段、事件發生地點、旅行時間以及有關路況變化等資訊，此類動態交通訊息，讓用路人得以選擇適當行車路徑，避開交通壅塞情況，進而以最佳旅行時間完成旅次需要。所以，要如何正確預測旅行時間，提供用路人做出行前決策之依據，以降低在旅次發生時之不確定性，為 ATIS 應用中的一個關鍵課題。

目前全球定位系統(Global Positioning System, GPS)的發展已技術成熟，加上其定位精確度高、不受任何天氣影響，且全球覆蓋率高(高達 98%)等優點。因此，本研究選擇利用行駛在高速公路上，裝設有 GPS 設備的探針車，利用其所蒐集到的交通資訊，所推估出的旅行時間資料，使用 k-最鄰近模式(k-Nearest Neighbor Model, k-NN Model)和時變係數模式(Time-Varying Coefficient Model, TVC Model)，分別對高速公路旅行時間作預測，以比較兩種方法預測的準確性，以提供用路人準確的旅行時間預測。

若能有效提供用路人準確的即時路況資訊，進而對旅次之出發時間、

路徑選擇和旅行時間，皆能進行有效之決策判斷依據，讓用路人更機動掌握與調整行程，提昇運輸走廊之用路便利與智慧化，行駛於運輸走廊能獲得更精確、更廣泛之交通資訊，藉以導引用路人充分利用替代道路以避開壅塞路段，達到提升高速公路路網使用效率目的。

## 1.2 研究目的

本論文之研究目的，為利用高速公路上的偵測器位置作為固定點，將探針車所蒐集到的交通資訊，所推估出的兩兩固定點之間的旅行時間，利用 k-NN 和 TVC 兩種模式分別作旅行時間預測，並分析比較兩種預測模式準確度與績效，期望能提供用路人準確的高速公路旅行時間預測值作使用。

## 1.3 研究範圍

本研究以台灣地區國道一號高速公路收費站之間的旅行時間為主要研究範圍，並以楊梅收費站到泰山收費站為實際測試對象，蒐集此路段上的探針車旅行時間資料和高速公路電子收費系統(Electronic Toll Collection, ETC)旅行時間資料，利用探針車即時旅行時間資料預測未來的旅行時間，以 ETC 旅行時間當實際值檢驗預測結果。

## 1.4 研究內容

本研究主要探討內容分述如下：

1. 進行旅行時間預測模式之文獻蒐集與回顧
2. 構建 k-NN 旅行時間預測模式
3. 構建 TVC 旅行時間預測模式
3. 以實際資料進行模式驗證
4. 比較兩種預測模式的準確性
5. 綜合上述研究成果，最後提出本研究之結論與建議

## 1.5 研究流程

本研究流程如圖 1.1 研究流程圖所示，依照流程圖各步驟說明如下：

## **1. 確認研究問題**

本階段主要工作為分析研究目的，並了解外在限制與可用資源，以清楚描述研究問題。

## **2. 相關文獻回顧**

本階段主要工作為回顧國內外相關之實作研究，以分析過去研究優缺點，作為本研究模式構建之基礎。

## **3. 模式方法構建**

本階段主要工作為利用所蒐集到的探針車旅行時間資料，建立 k-NN 和 TVC 兩種旅行時間預測模式，以比較兩種模式在預測上的績效。

## **4. 實作應用與模式修正**

本階段主要工作為將旅行時間預測模式，應用在國號一號高速公路上，以驗證此模式是否適用於現實狀況，並由運作結果對模式做調整與修正。

## **5. 結論與建議**

對本研究之過程與結果作一結論並提出研究上的建議。

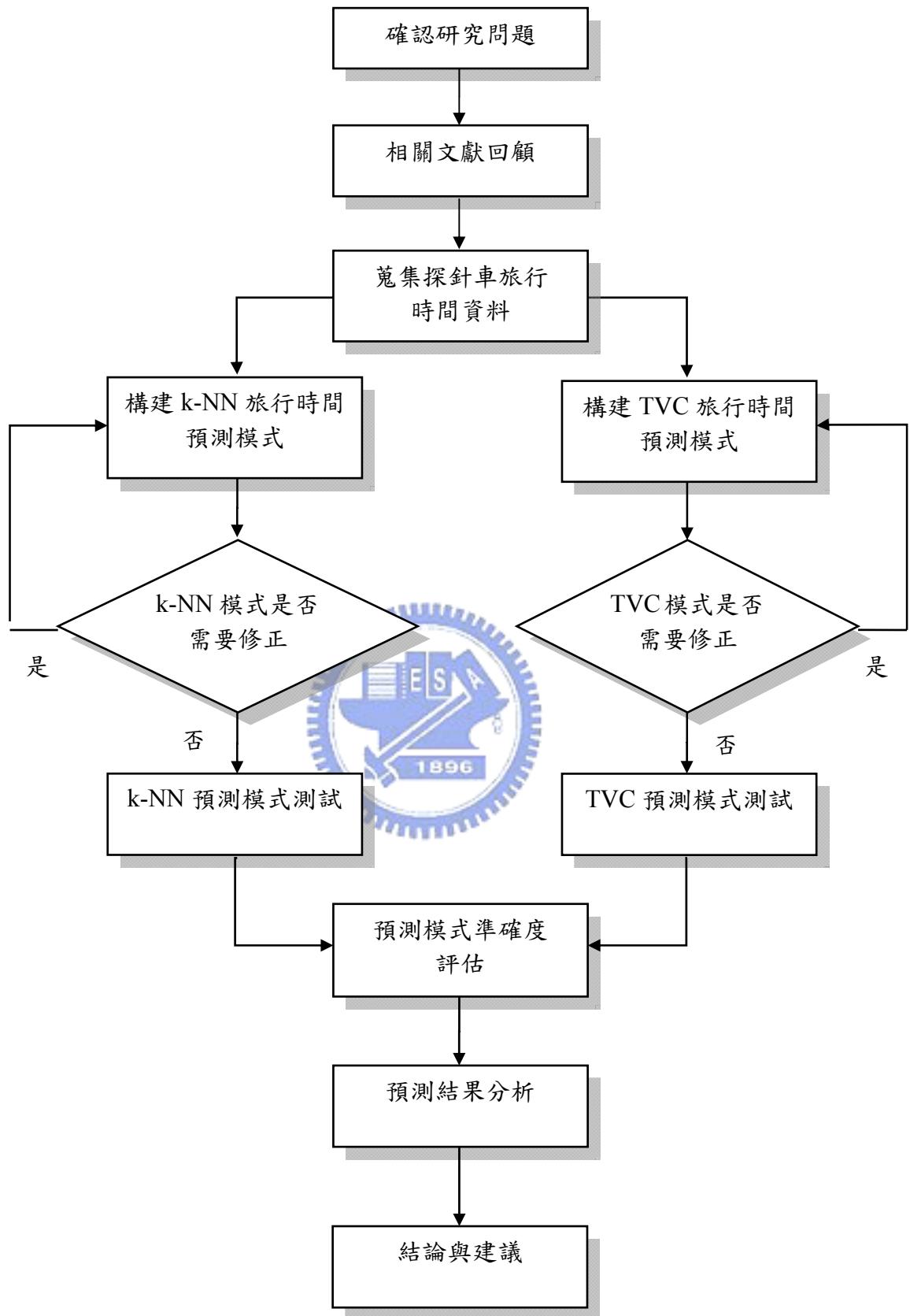


圖 1.1 研究流程圖

## 第二章 文獻回顧

旅行時間預測較常見的方法為利用即時或事後所偵測到的交通資料，用不同的方法來進行旅行時間預測，而旅行時間預測模式常用的方法有以下四種：線性迴歸模式 (Linear Regression Model)、時間序列模式 (Time Series Model)、類神經網路模式 (Neural Networks Model) 與 k-NN 模式 (k-Nearest Neighbor Model)，以下針對這幾種預測方法作一簡單介紹及國內外文獻回顧，並對各種方法進行優缺點比較分析。

### 2.1 線性迴歸方法

迴歸分析 (Regression Analysis) 是用來研究兩個或兩個以上變數間的關係，此方法的主要目的是要建立一個迴歸模式，然後根據一個或多個自變數 (Independent Variable) 之值，來預測依變數 (Dependent Variable) 或稱應變數 (Response Variable) 之值。因此，迴歸分析的主要目的是在建立變數與變數間之統計關係，然後利用此統計模式去做預測。迴歸分析可以分為簡單迴歸 (Simple Regression) 和複迴歸 (Multiple Regression)，簡單迴歸是用來探討一個應變數和一個自變數的關係，複迴歸是用來探討一個應變數和多個自變數的關係，整理表示式如下：

簡單迴歸表示式：

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

$\beta_0$  為常數， $\beta_1$  為迴歸係數， $\varepsilon$  為誤差

複迴歸表示式：

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

$\beta_0$  為常數， $\beta_1 \dots \beta_n$  為迴歸係數， $\varepsilon$  為誤差

Zhang [14] 認為目前的旅行時間與未來的旅行時間中存在著線性關係，因此利用此線性關係建立 TVC (Time-Varying Coefficient) 預測模式，利用目前高速公路車輛偵測器所收集的資料預測未來的旅行時間，並且利用平均絕對誤差百分比 (Mean Absolute Percentage Prediction Error, MAPPE) 來衡量此模式的預測結果。此模式適用於每個路段皆有佈設車輛偵測器的高速公路。

此研究的測試方法有二：

### (1)I-880

此測試資料來自一段在加州海沃(Hayward)長六英哩的 I-880 高速公路，在這段高速公路上每隔三分之一英哩就有一個雙迴圈車輛偵測器，因此全段總公有 35 個偵測器，而此測試範例是蒐集早上 5 點至早上 10 點北邊車道的偵測器資料。

### (2)I-405

此測試資料來自於加州的 I-405 公路，此研究蒐集早上 5 點至早上 10 點的北邊車道資料，研究規模較上一個測試範例大。

測試結果方面，在以 I-880 的資料測試下，依照時間的遠近誤差約在 5% 至 10%，而在以 I-405 的資料測試下，依照時間的遠近誤差約在 8% 至 13%，由於此模式的精確度會隨著預估時間的遠近，以及輸入資料量的多寡而有所不同，因此適用於短距離的公路上。

**Kwon [5]**利用資料庫內的流量以及佔有率資訊及探針車所得到的旅行時間資料來建立旅行時間預測模式。模式中將偵測到的流量、佔有率即當時的星期別及日期設定成一個集合  $X_i$ ，對探針車所得到的實際旅行時間值作線性迴歸模式，以此預測旅行時間，最後以 MAPPE 方式來檢定預測績效。

## 2.2 時間序列法

時間序列分析有許多模式，包括成長率模式、簡單移動平均模式 (simple moving average model)、指數平滑模式 (exponential smoothing model)、簡單線性迴歸模式 (以變數序列為因變數、時間為自變數)，以及 Box-Jenkins 模式等。其中，以 Box-Jenkins 模式之理論最為嚴謹，應用也最廣泛，其依變數序列之特性，又可分為自我迴歸模式(Autoregressive, AR)、移動平均模式(Moving Average, MA)、混合自我迴歸移動平均模式(ARMA)，以及整合自我迴歸移動平均模式(ARIMA)等四種，主要應用於分析歷史資料，並檢視資料本身之自我相關與偏自我相關等特性。

採用 ARIMA (Seasonal Autoregressive Integrated Moving Average) 所構建之預測模型，其構建過程分為四個階段，模型鑑定、參數估計、殘差檢定與模型選擇，透過此四步驟便可協助我們找到最佳的配適模型。其主要利用一連串的統計觀察值，依據時間的順序排列著，觀察的序列包含有實際過程(real process)與干擾(noise)兩部分，最重要的部分式如何削減外在環境所造成的干擾，因此自動迴歸與移動平均模式就這樣產生。基於這兩種

模型，任何不連續但固定兼具的序列則可以自動迴歸移動平均模式來表示，模式如下：

$$Y_t = a + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

$$(1 - \phi_1 B^1 - \phi_2 B^2 - \dots - \phi_p B^p) Y_t = a + (1 - \phi_1 B^1 - \phi_2 B^2 - \dots - \phi_p B^p) e_t$$

$$\phi_p(B)(1 - B)^d Y_t = a + \phi_q(B) e_t$$

其中， $Y_t$ 為一時間數列值； $a$ 為常數項； $p, d, q$  均為非負的整數，分別代表自我迴歸、差分階數及移動平均的參數個數； $B$  為後移運算子(Backward Shift Operator)，即  $B^d Y_t = Y_{t-d}$ ； $e_t$  為隨機干擾項，且為一白噪音(White Noise)。

**Smith [11]** 本研究討論統計上母數(parametric)與無母數(nonparametric)資料在單點短期車流量預測上的差異及比較。此研究在選擇母數模式的部份，以ARIMA作為預測模式，其主要是利用整合自我迴歸、移動平均兩種特性所構成的模式去處理時間序列，而無母數迴歸模式則是利用k-NN模式進行流量預測。

將以上兩種預測模式結果與naïve預測作比較，研究考量目前的流量可能會與先前的流量有關，但是隨著時間距離的增加，其影響程度愈小，所以嘗試將k-NN模式加入不同權重值進行預測值的計算。預測結果顯示，ARIMA預測結果最佳MAPE值約為8.8%，而k-NN模式的MAPE值約為9.4%，比naïve的MAPE值10%預測效果來得好，由於此研究預測所使用的歷史資料天數尚不足兩個月，因此研究結論提及若k-NN模式能利用更多歷史資料或其它的距離量度作資料篩選， k-NN模式或許還能提供更好的預測結果。

**Yang [13]**把時間序列模式應用在幹道的旅行時間預測研究上。透過實際採用裝有 GPS 系統的探針車於美國明尼蘇達州 194 號高速公路作實際研究。其方法是將蒐集所得資料視為時間序列，以 ARIMA 模式進行旅行時間預測，結果顯示此方法能夠有效預測短期內的旅行時間。

## 2.3 類神經網路法

人工神經網路是由許多的人工神經細胞(Artificial Neuron)所組成，人工

神經細胞又稱為類神經元、人工神經元或處理單元(Processing element)，每一個處理單元的輸出以扇狀送出，成為其他許多處理單元的輸入。處理單元其輸出值與輸入值的關係式，一般可用輸入值的加權乘積和函數來表示：

$$Y_j = f \left[ \sum_i W_{ij} X_i - \theta_j \right]$$

其中：

$Y_j$ =模仿生物神經元的模型輸出訊號。

$f$ =模仿生物神經元模型的轉換函數(Transfer Function)，是一個用以將從其他處理單元輸入的輸入值之加權乘積何轉換成處理單元輸出值的數學公式。

$W_{ij}$ =模仿生物神經元模型的神經節強度，又稱為連結加權值。

$X_i$ =模仿生物神經元模型的輸入訊號。

$\theta_j$ =模仿生物神經元模型的閾值。

介於處理單元間的訊號傳遞路徑稱為連結(Connection)。美一連結上有一個加權值 $W_{ij}$ ，用以表是第 i 處理單元對第 j 個處理單元的影響強度，如圖 所示。

人工神經網路的基本結構可分為三個層次：

1. **處理單元(Processing element, PE)**: 為人工神經網路組成的基本單位，處理單元的作用可以三個函數來說明：

(1) **集成函數(Summation Function)**: 將其它處理單元之輸出透過網路連結傳至之訊息加以綜合。

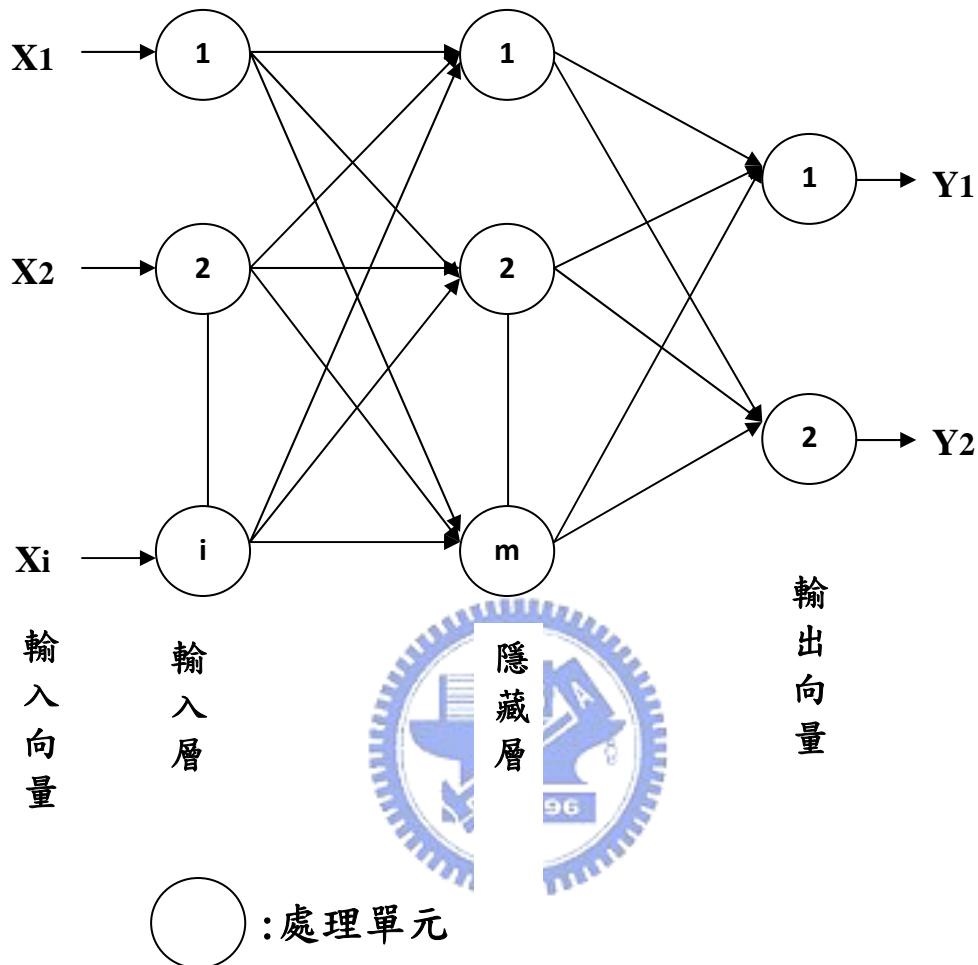
(2) **作用函數(Activity Function)**: 作用函數的目的是將即成函數值與處理單元目前的狀態加以綜合，但一般人工神經網路模式的作用函數採直接使用集成函數輸出，故其作用並不明顯。

(3) **轉換函數(Transfer Function)**: 轉換函數的目的是將作用函數的輸出值經轉換作用後成為處理單元的輸出，即處理單元即輸出值  $Y_j$ (作用函數輸出值)。而經由轉換函數的作用，使得人工神經元網路具有解決非線性問題的能力。

2. **層(Layer)**: 若干個具有相同作用的處理單元成為「層」，一般的人工神經

網路架構中主要分為三層。分別為輸入層、隱藏層與輸出層。

3. 網路(Network):若干個具有不同作用的「層」集合成為網路，而網路本身則具有學習與回憶的功能。



倒傳遞類神經網路架構圖

倒傳遞人工神經網路之學習演算過程一般包括下列幾項：

1. 設定網路參數、隨機設定出使權重  $\omega$  與閥值  $\theta$ 。
2. 計算能量函數(或稱誤差函數):學習演算過程旨在降低網路輸出結果與實際輸出值間之差距，而其間的差距一般以能量涵述表示其學習品質，其式如下：

$$E = (1/2) \sum (T_j - Y_j)^2$$

其中：

$E$ 為能量函數。

$T_j$ 為實際值。

$Y_j$ 為輸出層(j)之輸出結果。

3. 以訓練樣本(一般而言為歷史資料)訓練網路，並使學習過程之能量函數最小化。能量函數最小化的過程通常是以最陡坡將法(Gradient Steepest Descent Method)使能量函數最小化。而權重 $\omega$ 與閥值 $\theta$ 的改變量 $\Delta\omega$ 與 $\Delta\theta$ 等於：

$$\Delta\omega = -\eta \frac{\partial E}{\partial \omega}, \quad \Delta\theta = -\eta \frac{\partial E}{\partial \theta}$$

其中， $\eta$ 為學習效率，即控制每次加權值改變量的幅度，而學習效率之大小對網路的收斂有決定性。經由網路學習使能量函數最小化之過程，新的權重與閥值分別為 $\omega(\text{NEW}) = \omega + \Delta\omega$  與  $\theta(\text{NEW}) = \theta + \Delta\theta$ 。

4. 重複步驟三，直到收斂為止。

上述四個步驟為到傳遞人工神經網路主要的學習過程，藉由學習過程變可使網路修正原先隨機設定之權重，進而達到降低誤差的目的；而後以測試樣本(非學習樣本之歷史資料)測試網路之精確度。若網路輸出結果合乎要求，則可應用至待推案例中；若不合乎要求，則重複步驟三之學習過程，重新設定新的學習速率，反覆學習，值達合乎模式的要求為止。

李季森[16]探討國內高速公路駕駛人變換車道行為與變換車道時間，進行探討與推導相關公式，並進一步撰寫模擬程式，進而探討不同預測時間、流量、探針車混合比例與區段長度等相關參數之實驗組合，並利用探針車所收集之相關資料，透過倒傳遞類神經網路進行旅行時間之預測，以期提供精準之旅行時間預測。經由反覆的校估與測試結果可知，所構建旅行時間預測模式屬於「高精準預測」，可作為交通相關單位參考。

林士傑[17]利用運研所網站所提供的即時資訊、客運公車的GPS交通資訊，及中華顧問工程司之交通千里眼(E-traffic)所提供之即時交通播報資訊(如事件、施工、車輛偵測器等)，加上交通量調查報告與高速公路幾何資料等，運用倒傳遞類神經網路準確預測高速公路旅行時間。該研究以北部區域路段為研究範圍，構件預測模式，來提供用路人參考，以降低不確定性。

## 2.4 k-NN 方法

k-NN 方法基本的運算方式是先將歷史資料庫中的資料作分群，將輸入的現況資料與分群後的歷史資料庫中作比對，找尋  $k$  組與現況資料最相近的資料，利用這  $k$  組歷史資料的特性推估未來資料。所以在建立 k-NN 預測模式前，首先要做好分群的動作，通常依照星期(weekly)、季節(seasonal)、或是將每天(daily)視為一個 pattern，亦或是根據資料分析來找出明顯的特徵分群。k-NN 模式在進行資料比對時，以距離量度來判斷現況資料與歷史相近的程度，在資料選取時候，為了降低不同車道間所可能產生的偏差，模式也可以用權重不同的加權方式，對於不同車道對預測結果的影響作考慮，或是依照選取資料時間的相近程度，加上權重值作調整，增加比對資料的可靠性。

在尋找  $k$  筆資料過程中，如何設定  $k$  值是一項需要解決的問題，而多數的 k-NN 文獻中顯示，需測試不同的  $k$  值，來尋找使預測誤差值降至最低的解作為設定  $k$  值的最佳解，如此一來，可確保產生最小預測誤差，提供最佳的預測結果。

Altman[1]提出了 k-NN 法，該方法先將歷史資料區分群，再將輸入資料與分群後的歷史資料相比對，比對出該資料與那  $k$  組歷史資料較為類似，即可利用  $k$  組歷史資料推測未來資料。

Smith [9]研究對四種模式進行績效評估，分析比較以下四種交通流量的預測方法：歷史平均法、ARIMA 法、類神經網路法與 k-NN 模式，評估的方法為利用歷史的流量資料預測未來的流量再和實際值作比對，判斷哪一個方法的績效較差。結果發現，歷史平均法對於突發事件發生時，無法立即有效判斷造成不準確的預測，ARIMA 預測模式方面，對於遺失的資料無法作有效處理，且此方法較適合預測趨勢性與季節性等時間序列資料。而類神經網路需要繁複的訓練過程，當資料量大時，誤差也隨之變大，且其與 k-NN 比較其誤差值較大。

所以研究結果得到，雖然 k-NN 方法在搜尋相鄰近資料較複雜，但 k-NN 預測模式誤差值比其他方法來的小，因此利用 k-NN 模式針對流量進行預測是可行的。

Smith [11]將權重(Weight)的想法加入 k-NN 的模型，考慮目前的流量可能會與先前的交通流量有關係，但是隨著時間距離的增加，其影響的程度愈小，

所以利用權重的方法進行修正，將不同時間所造成影響納入考慮。

**Clark[4]**除了流量外，利用 k-NN 模式試著對其他可收集到的交通資訊與歷史資料進行比對，如速度和佔有率，找尋即時資訊和哪一天的歷史資訊相類似，而找尋的方式為歷史與即時資訊的平方差加總在距離量度內時，即判斷該分群的歷史資訊與即時資訊相類似。結果發現當同時利用流量、佔有率和速率這三項變數進行預測時，所得到的預測值相對於個別比對而言，誤差有下降的現象，所以愈多變數進行討論可以得到更精確的結果。判斷模式建立如下：

*tss: total sum of squares*

$$tss = \sum_{i=1}^L [(q_{ri} - q_{mi})/w_q] + \sum_{i=1}^L [(v_{ri} - v_{mi})/v_q] + \sum_{i=1}^L [(o_{ri} - o_{mi})/o_q]$$

$q$ : 流量  $v$ : 速率  $o$ : 佔有率

$L$ : 比對範圍的總偵測器數量。

$r_{ij}$ : 即時時間( $i$ )、偵測器( $j$ )之交通資訊。

$m_{ij}$ : 歷史時間( $i$ )、偵測器( $j$ )之交通資訊。

$w_q$ : 流量權重；  $v_q$ : 速率權重；  $o_q$ : 佔有率權重

**Robinson [8]**研究針對都市內的旅行時間進行預測，並提出了四點建立k-NN 模式時應先設定的條件，試著使用不同的距離量度來比對歷史資料，及找出k值最佳解，四項條件如下所示：

1. 決定每一個特徵包含哪幾種行為模式。
2. 設定距離量度(distance metric)。
3. 決定判斷分群的樣本數量。
4. 利用加權法減少 k-NN 模型的誤差。

**Lam[6]**利用每年的香港交通統計年報流量資料預測未來交通流量變化，本研究利用兩種母數模式 ARIMA 和類神經模式與兩種無母數模式 k-NN 和 Gaussian Maximum Likelihood(GML)作預測效果比較。研究結果顯示，無母數模式比母數模式預測誤差較小且效果較佳，k-NN 模式在短期流量預測效果又比 GML 較佳，而 GML 模式對於周期性穩定的交通流預測會比 k-NN 模式效果來得好，因此研究建議，若要進行短期的交通流量變化預測，利用 k-NN 預測模式較適合。

**Weng[12]**收集北京高速公路的歷史資料利用 k-NN 模式預測未來速度變化，將歷史資料依照交通流量特性作分群，選擇利用 Euclidean Distance 作為距離量度，並依照距離遠近加上權重值作調整，最後利用浮動車輛(floating car)所收集的資料驗證預測結果，其 EDP(Error Distributing Probability)超過 90%，是一項非常有效的預測方法。

**Smith[10]**研究利用即時流量資料，利用 k-NN 模式預測未來流量，模式建立前要先決定以下三項條件：

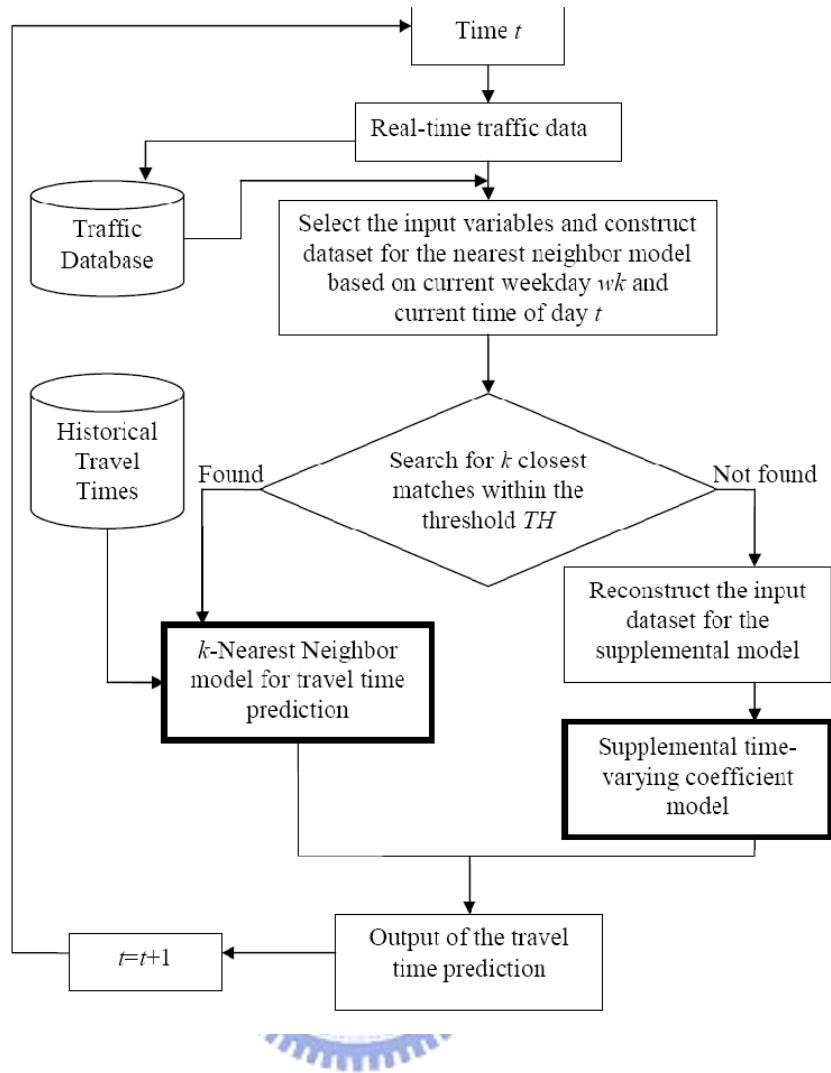
1. 狀態向量(State Vector)
2. 距離量度(Distance Metric)
3. 預測函數(Forecast Function)

利用 Euclidean Distance 作為距離量度，找出每筆歷史資料與即時資料的距離，選擇 k 筆最鄰近歷史資料作平均，即是研究所要的預測結果，最後，測試不同 k 值所得到預測誤差，選擇最小誤差的 k 值使預測精確度達到最高。

**Chang[2]**於美國馬里蘭(Maryland)在 25 英哩的快速道路間，使用 10 個偵測器與五個 LED 看版，將偵測所得資訊以旅行時間估計模型(Travel Time Estimation Module)來建立現有的歷史旅行時間資料庫，再透過此資料庫使用旅行預測模型(Travel Time Prediction Module)來預測旅行時間。

而事故偵測模型(Incident Detection Module)可將路段事故預先知會駕駛人，並將事故路段視為不可行，避免因為事故造成的擁塞使得預測結果偏差。遺失資料模型(Missing Data Module)可防止因為偵測器失靈所造成的資訊延滯與遺失，結合以上四種模型，可在各 LED 提供用路人通往下個交流道出口的估計旅行時間。

研究中將旅行時間預測模組分為 k-NN 模式(k-Nearest Neighbors Model)與增強時變係數模式(Enhanced Time-Varying Coefficient Model)兩大部分。k-NN 模式會根據目標路段的交通狀態(traffic pattern)和幾何特徵(geometric features)，在歷史資料庫中比對最接近目前交通狀態的資料，再拿來進行預測；若沒有找到 k 筆最接近目前交通狀態的歷史資料，則使用增強時變係數模式，找出每日與每週間旅行時間的關係來進行預測。整個模組運作的流程圖如下圖所示：



[資料來源：Chang, 2004]

## 2.5 其他預測方法

陳建名[19]以市區公車裝配 GPS 回傳的即時資料為主要來源，將公車旅行時間分為運行時間和停等時間，研究訂定一個門檻值，當公車在正常情況下行駛，使用歷史平均速度做預測，當實際速度與歷史平均速度差超過門檻值即判斷交通狀況出現異常，即利用前車資訊做預測。

研究將歷史資料庫依車速高低不同切割成數個不同交通時段，而在不同時段，研究以不同預測模式來預測公車於路口停等號誌之時間，研究結果顯示，在尖峰期間，受到交通管制影響，使得停等時間預估產生較大誤差，影響預測結果，而在路線較短的時候，也會產生較高的誤差。

吳佳峰[15]透過車輛歷史旅行資料預估車輛旅行時，模式採用裝設 GPS 之探針車輛所傳回之定位資訊推估車輛所在位置，從車輛速度變化觀察實際車流之交通情況。研究將車輛總旅行時間分段加總，將旅行時間分為車輛運行時間和車輛停等時間兩部分，當道路狀況出現了非重現性之壅塞時，以前車資訊調整預估車輛運行時間。

研究以實際國內客運業者車輛旅行資料對預估模式做實例之測試，從測試結果發現模式在未遭遇非重現性之壅塞時，預測旅行時間有著相當不錯之準度，而當遭遇非重現性壅塞時，模式亦能透過預估旅行時間之調整機制，控制在可接受之誤差範圍內。

Choi[3]利用GPS與數位化道路地圖(Digital Road Map)來計算路段之動態旅行時間。對於動態交通資訊而言，旅行時間預測為一項重要的要素，研究利用GPS與GIS技術來蒐集都市街道之動態路段旅行時間，並對於測量路段旅行時間的技術作比較。在都市幹道網路上，路段旅行時間為行駛時間(running time)與延滯時間(delay times)的總和，易受到號誌系統與交通環境所影響。研究利用流動車輛法(floating car method)，用以得到點至點的旅行時間，不過此方法須蒐集較多的交通資料，如各個車道、方向、時段的交通資料，最後利用實際收集到的交通路網資料來與本研究所計算出來的旅行時間值進行績效評估，證實預測結果準確。

綜合以上相關文獻，彙整表格如下表2.1所式：

表 2.1 各種預測方法之比較

預測方法	理論內容	優點	缺點
線性迴歸	利用一個或多個自變數來預測應變數，其中自變數與應變數皆為線性關係，利用所獲得之樣本資料去估計模型中參數的計量分析方法。	1. 根據統計理論基礎，解釋變數與因變數之關係，較有說服力。 2. 有同趨勢之規律性時，根據大量樣本個數，即可計算出線性迴歸方程式。	1. 系數固定，故對外在因素的改變，缺乏反映之彈性。 2. 不適用於少量樣本之場合。
時間序列	發掘時間序列變數現在與過去的關係，預測此變數未來的趨勢值，時間相隔越短之兩觀測值，其相關性越大，此方法基本上不採用其他的變數，只採用過去的資料來構建預測模式。	1. 對於週期性、季節及循環性之趨勢易於掌握。 2. 純粹以變數歷史數據作為預測基礎，資料收集容易。	1. 模式選擇需高度技巧與經驗。 2. 缺乏統計理論基礎，造成模式解釋不易。
類神經網路	模擬人類腦神經組織，以歷史或模擬資料作為訓練樣本，利用輸入、輸出、隱藏層等各種不同方式連結，透過訓練的方式，讓類神經網路反覆學習，直到對每個輸入都能正確對應到所需要的輸出。	1. 能解決較複雜、非線性關係的問題。 2. 事前無須任何假設輸入與輸出變數之間的關係。 3. 應用範圍相當廣泛，舉凡生物、醫學、運輸…等皆有所應用。	1. 模式需經過足夠之樣本進行訓練始能使用。 2. 容易產生過度訓練或訓練不足。 4. 最佳隱藏層數目及神經元數目決定無規則可循。
k-NN	是一種利用歷史資料特性推估未來資料的方法。先將資料作分群，透過現況資料與歷史資料進行比對，找出k筆最近的歷史資料預測未來資料。	1. 當歷史資料有遺漏或錯誤而沒有辦法得到完整資料時，能適時利用其他歷史資料作彌補。 2. 資料經過分群後，大幅減少搜尋時間，加快預測速度。	1. 搜尋相鄰近資料過程較複雜。 2. 必須找出k值最佳解。

資料來源：本研究整理

## 2.6 小結

經由本章節對於各種預測方法之介紹後，考量各種預測方法之適用性，作為本研究旅行時間預測模式之構建基礎。從上述文獻回顧中可發現線性迴歸方法是一種可以的方式是一種概念較簡單，但又不失預測準確度的方法，可運用在預測旅行時間週期性的變化上。在以往的線性迴歸預測模式中，多採用一個迴歸模式作預測，當時間變化或外在因素改變時，因迴歸模式缺少彈性變化，無法即時反應旅行時間變化情況，容易產生較大預測誤差，因此本研究採用的 TVC 模式，能夠改善這項缺點，其作法為先將歷史資料做分群，以此建立各分群的迴歸模式，並隨時更新參數，當收到一筆即時資料時，可立即判斷該筆資料屬於哪一分群，代入該分群的迴歸模式輸出預測結果，如此可即時反應各個時段旅行時間的變化，因此本研究選擇 TVC 模式作為旅行時間預測模式之一。

而由於考量路段上會碰到偵測資料遺失的情況，造成旅行時間預測誤差變大，甚至造成無法預測的情況發生，經由以上的文獻回顧可得知，k-NN 預測模式在資料遺失處理上，能適時利用其他歷史資料作彌補，不會因資料缺漏而有無法預測的情況發生，以及在國外研究文獻上，也證實k-NN模式的預測準確度高，因此本研究選擇使用k-NN作為另一預測模式，並與TVC 模式作預測比較，評估兩種方法應用在高速公路旅行時間預測的準確性，以提供用路人更準確的預測資訊。

## 第三章 k-NN 模式構建

本研究發展之旅行時間預測模式，為採用裝設 GPS 定位之探針車輛所回傳之定位資訊，所推估出的旅行時間作為構建預測模式的資料來源，分別利用 k-NN 模式和 TVC 模式進行旅行時間預測，並比較兩種預測模式的準確度與績效，以提供用路人有效的預測資訊。本章將先針對 k-NN 預測模式進行構建，首先介紹 k-NN 模式建立流程，依照建立模式流程各步驟進行探討，以構建適合本研究預測模式完整架構，待確認模式流程架構後，即依照流程各步驟開始進行。

### 3.1 k-NN 模式介紹

建立旅行時間預測模式的目的在於，從即時的交通資訊預測未來的旅行時間為何，本研究採用 k-NN 方法進行模式構建，其基本的運算方法為先將歷史資料作分群，當收到一筆即時資料時，可立即判斷即時資料屬於哪一分群，在分群的歷史資料庫中利用距離量度比對即時資料與歷史資料，如此先進行資料分群動作，可減少資料比對的時間。

在比對完歷史資料後，可得知各筆歷史資料的距離量度，從各距離量度中尋找  $k$  筆與即時資料最鄰近的歷史資料，表示這  $k$  筆歷史資料與即時資料相似，將這  $k$  筆歷史資料加入權重值作調整，則此加權後的平均值即是本研究所取得的預測結果。因此要建立一個完整的 k-NN 模式，必須先確認各個步驟的作法，本研究將依照 k-NN 流程圖的每個步驟進行逐步探討，以建構完整 k-NN 旅行時間預測模式。k-NN 流程如圖 3.1 所示：

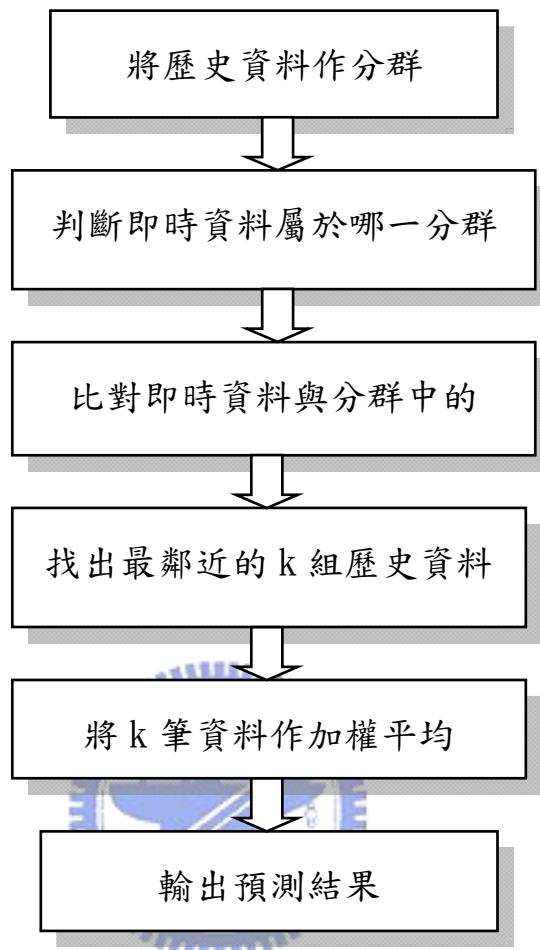


圖 3.1 k-NN 流程圖

### 3.1.1 k-NN 資料分群

在建立預測模式前，首先要做好分群的動作，當接收到一筆即時資料，能夠利用較短時間在龐大的資料庫中搜尋需要的歷史資料，減少比對時間，迅速將預測結果提供給用路人。因此在建立 k-NN 預測模式前，做好歷史資料分群是首要工作，研究文獻中有依照星期別、季節、或是將每天視為一個分群，或是根據資料分析來找出明顯的特徵分群。若是按照每天的資料作分群，可再將當天的交通狀況區分出尖峰與離峰交通狀態，或是將每周的交通狀況區分出假日與非假日的交通趨勢，也可以將每季的交通狀況則區分出不同季節的交通變化，如夏季的交通行為與冬季的交通行為就會有所不同。

### 3.1.2 距離量度(Distance Metric)

在 k-NN 模式中，距離量度的主要功能為比對現況資料與歷史資料相近程度的一種判斷工具。當距離量度小於門檻值，則可判斷此筆現況與歷史資料相類似，相反的，當距離量度大於某門檻值，則判斷此筆現況資料與歷史資料差異較大。距離量度較常見的主要有下列三種，分述如下：

#### (一) 歐幾里得距離(Euclidean distance)

設有  $n$  個事物，每個事物有  $m$  個屬性，則第  $i$  個事物與第  $j$  個事物間的歐幾里得距離如式 3-1 所示：

$$d_{ij} = \left[ \sum_{p=1}^m (x_{ip} - x_{jp})^2 \right]^{1/2} \quad (3-1)$$

#### (二) 馬式距離 (Mahalanobis distance, $D^2$ )

馬式距離  $D^2$  可用矩陣如式 3-2 來表示：

$$D_{ij}^2 = (X_i - X_j)' C_W^{-1} (X_i - X_j) \quad (3-2)$$

上式中， $C_W$  為聯合組內共變數矩陣(pooled within-group covariance matrix)，馬式距離  $D^2$  是歐幾里得距離平方的一種引申：

$$d_{ij}^2 = (X_i - X_j)' (X_i - X_j) \quad (3-3)$$

$$\begin{aligned}
&= (\chi_{i1} - \chi_{j1}, \chi_{i2} - \chi_{j2}, \chi_{i3} - \chi_{j3}, \dots, \chi_{im} - \chi_{jm}) \begin{bmatrix} \chi_{i1} - \chi_{j1} \\ \chi_{i2} - \chi_{j2} \\ \vdots \\ \chi_{im} - \chi_{jm} \end{bmatrix} \\
&= \sum_{p=1}^m (\chi_{ip} - \chi_{jp})^2
\end{aligned}$$

意即，在求  $d_{ij}^2$  時， $C_W^{-1} = C_W = I$  (單元矩陣)。換言之，當各軸的單位長度相等，且各軸互為垂直時，歐幾里得距離平方與馬式距離  $D^2$  相同。

### (三) 街道區距離 (city block distance)

設有  $n$  個事物，每個事物有  $m$  個屬性，則第  $i$  個事物與第  $j$  個事物間的街道區距離可以如式 3-4 衡量：

$$d_{ij} = \sum_{p=1}^m |\chi_{ip} - \chi_{jp}| \quad (3-4)$$

#### 3.1.3 k 值設定

$k$  值的選定，是決定預測誤差大小的關鍵步驟，若是選取太大或太小的  $k$  值，容易產生較大的預測誤差，導致預測結果產生偏差，因此一般研究在  $k$  值設定這一步驟時，通常都是測試選取不同  $k$  值，找出各個  $k$  值所得應到的誤差結果，由此取得使誤差最小的  $k$  值，以此設定為  $k$  值最佳解，如此可確保預測誤差達到最小。

本研究  $k$  值的設定，主要先設定不同的門檻值來進行  $k$  值選取，當比對的距離量度小於此門檻值，即判斷該筆資料與即時資料相近，依照各門檻值內所得到的不同  $k$  值，找出該時段預測誤差最小的  $k$  值，以此作為該時段的最佳  $k$  值解，由於每五分鐘即預測一次， $k$  值也隨著每五分鐘變化一次，以找出適合該時段的最佳  $k$  值解。

### 3.2 k-NN 模式構建

本研究構建之一旅行時間預測模式，為使用 k-NN 方法依照上述 k-NN 流程步驟，構建高速公路旅行時間預測模式。首先，在資料分群的步驟，依照星期別作區分，與即時資料進行比對，比對時，將同時段、兩兩固定點之間的即時旅行時間資料與歷史旅行時間資料互相作比對，找出 k 筆最鄰近的歷史旅行時間資料，將這 k 筆旅行時間資料依照距離量度的大小作加權平均，以此輸出旅行時間預測值。

研究將針對各時段的 k 值進行分析，找出 k 值最佳解。以下將相關預測步驟進行說明，本研究 k-NN 模式構建之相關流程可表示如圖 3.2 所示。



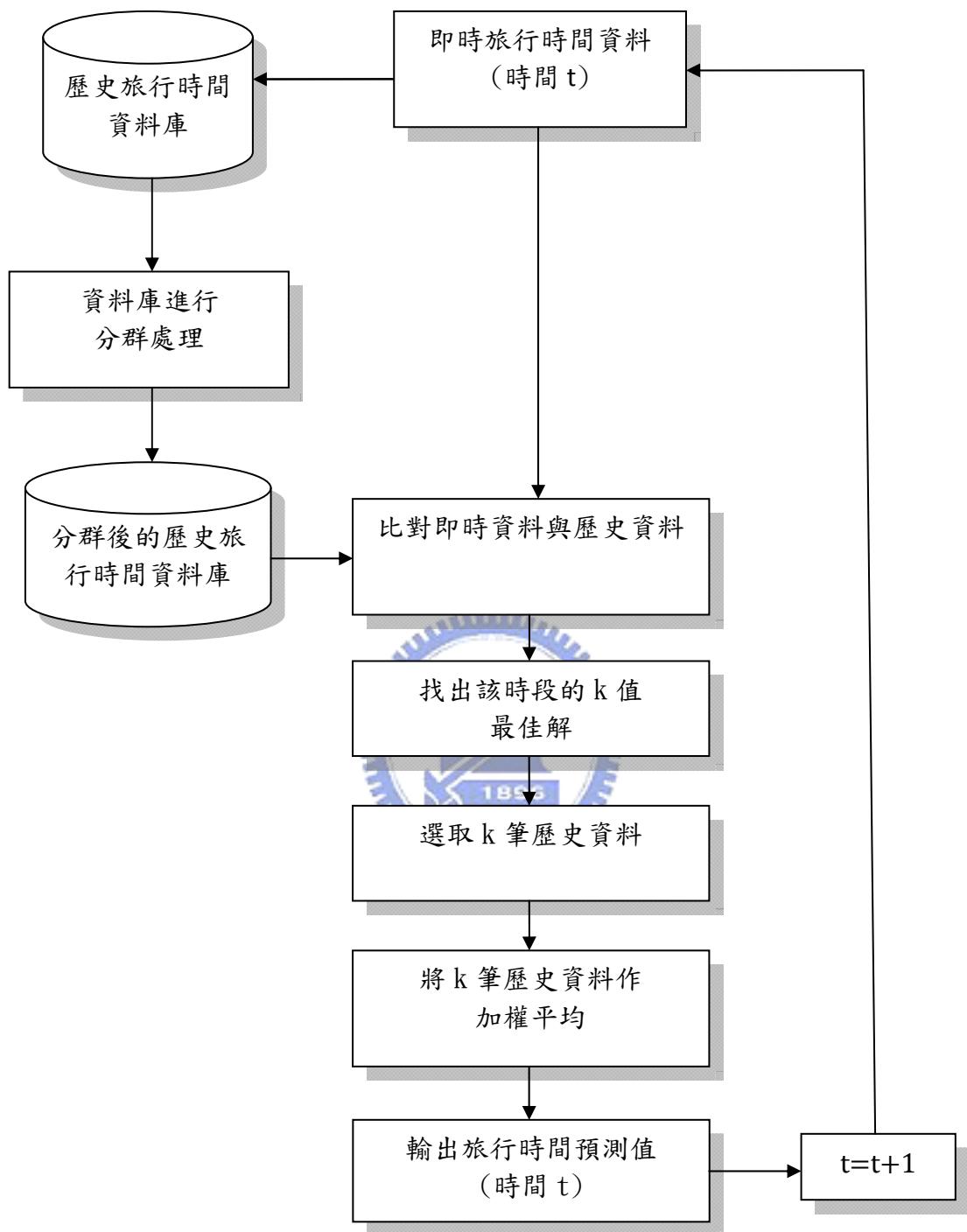


圖 3.2 k-NN 模式構建流程圖

而本研究 k-NN 模式構建流程運作主要可分為下列五個步驟，說明如下：

### Step 1. 交通資料蒐集

本研究利用行駛在國道一號高速公路上裝設有 GPS 的大客車探針車蒐集完整交通資訊，利用路段上偵測器位置作為固定點位置，推估出路段上兩兩固定點之間的旅行時間資料作為預測資料來源，檢驗預測結果時，以 ETC 所蒐集兩收費站之間的旅行時間當作實際資料與預測結果作比對，以評估預測績效。由於本研究主要預測小客車的旅行時間，因此在 ETC 資料方面，將使用小客車的旅行時間作使用。各種資料庫之資料形式，整理如表 3.1 所示：

表 3.1 各種資料庫之資料形式

資料庫 資料格式	主要資料內容	資料形式
<b>GPS 探針車 資料庫</b>	車輛編號  經度座標  緯度座標  日期、時間  車速	ex. 車輛 656-FC  東西經(度、分)  南北緯(度、分)  年/月/日 時:分:秒  公里/小時
<b>歷史旅行時間 資料庫</b>	上、下游固定點編號  日期、時間  上、下游固定點間旅行時間  行車方向	S:南 ; N:北  年/月/日 時:分:秒  秒  S:南 ; N:北
<b>ETC 旅行時間 資料庫</b>	日期、時間  ETC 起、迄點代碼  平均旅行時間  行車方向  車種代碼	年/月/日 時:分:秒  22002:楊梅收費站 14003:泰山收費站  分鐘  S:南 ; N:北  31:小客車，41:大客車

資料來源：本研究整理

## **Step 2. 將歷史資料作分群**

建構 k-NN 模式時需先將歷史資料作分群，當收到一筆即時資料時能夠立即判斷資料屬於哪一分群，再比對即時資料與群中的歷史資料，如此可減少比對時間，增加預測速度。本研究 k-NN 模式將依照不同的星期別將資料作分群，將一周七天資料分成七群，若遇到特殊節日，須獨立出來考慮。

## **Step 3. 比對即時資料與歷史資料**

利用距離量度比對即時與歷史資料，以得到各筆歷史資料與即時資料的差異程度。本研究的距離量度採用歐幾里德距離(Euclidean Distance)算法，比對與即時資料同星期別、前後半小時的歷史資料，將兩兩固定點間的即時旅行時間資料與歷史旅行時間資料，依照歐幾里得距離方法作平方差加總，找出該分群資料庫中各筆歷史資料與即時資料的距離量度。

## **Step 4. 選取 k 筆資料**

依照距離量度的大小排序，選取 k 筆最小距離量度的歷史資料，而 k 值設定方式，必須要找到 k 值最佳解，也就是讓預測誤差最小的 k 值解，作為設定 k 值的條件。



## **Step 5. 利用加權法減少 k-NN 模式誤差**

由於現況的旅行時間會與先前的旅行時間相關，但是隨著時間距離的增加其影響程度越小，所以我們可以根據歷史資料距離即時資料的時間差距或是距離量度的大小，將這 k 筆資料加入權重值做調整，以減少旅行時間預測誤差。

## **Step 6. 輸出旅行時間預測值**

將預測的旅行時間輸出，並將此旅行時間預測值存入歷史資料庫中，提供給下一個時段的旅行時間預測模式作使用。

### 3.3 k-NN 模式範例說明

以下舉一範例作說明：要預測 2008/3/6(四)早上 8 點從 A 收費站出發到 B 收費站間的旅行時間，其中有五個固定點  $x_1$ 、 $x_2$ 、 $x_3$ 、 $x_4$ 、 $x_5$ ，在早上 8 點一組收到兩點間即時旅行時間資料。

範例假設如下：

1. 歷史旅行時間資料庫內只有 2008 年 1、2 月份的資料。

2. 只搜尋同時段旅行時間資料。

在 2008/3/6(四)早上八點鐘收到即時旅行時間資料，若只搜尋同時段旅行時間資料，也就是在比對歷史資料的時候，只比對 2008 年 1、2 月份同星期別(禮拜四)早上八點鐘的資料。

3. 權重值皆設定為 1。

4.  $k$  值設定為 3。

之後使用實際資料時會重新作設定。

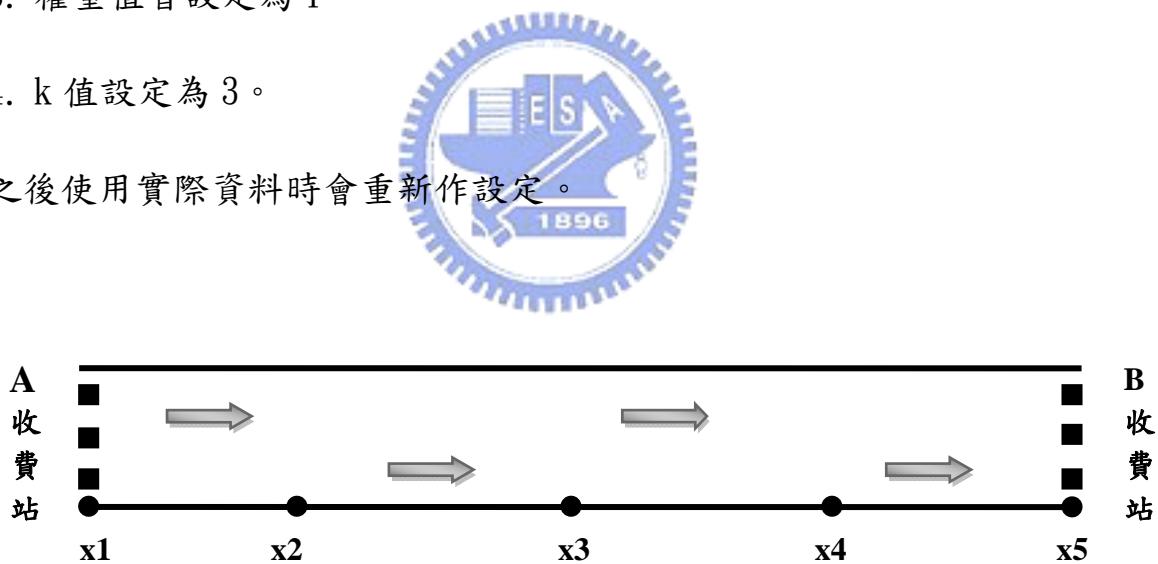


圖 3.3 k-NN 範例路段示意圖

本範例將分成三個步驟來進行，說明如下：

## Step1. 比對即時資料與歷史資料

因為已將1、2月份歷史資料庫依照同星期別作好分群，因此，在比對資料時只比對同時段的歷史資料，所以總共有1/3(四)、1/10(四)、1/17(四)、1/24(四)、1/31(四)、2/7(四)、2/14(四)、2/21(四)、2/28(四)這九天，比對各偵測器間八點鐘的歷史旅行時間資料。

## Step2. 找出各筆歷史資料的距離量度

各筆歷史資料的距離量度如表3.2距離量度資料表所示。

表3.2 k-NN距離量度資料表

時間	x1~x2 旅行時間	x2~x3 旅行時間	x3~x4 旅行時間	x4~x5 旅行時間	距離量度
1/3 AM 8:00	200	250	290	360	181.9341
1/10 AM 8:00	260	390	500	280	227.1563
1/17 AM 8:00	250	440	280	330	147.3092
1/24 AM 8:00	300	460	390	360	130.384
1/31 AM 8:00	280	300	500	420	187.3499
2/7 AM 8:00	400	360	250	380	110.4536
2/14 AM 8:00	350	500	400	290	198.2423
2/21 AM 8:00	290	470	390	300	171.4643
2/28 AM 8:00	330	360	330	410	181.9341
3/6 AM 8:00	280	370	390	350	

時間單位:秒

各筆歷史資料與即時資料的距離量度算法如下所示：

$$1/3 : \sqrt{(280 - 200)^2 + (370 - 250)^2 + (390 - 290)^2 + (350 - 260)^2} \\ = 181.9341$$

$$1/10 : \sqrt{(280 - 260)^2 + (370 - 390)^2 + (390 - 500)^2 + (350 - 280)^2} \\ = 227.1563$$

⋮

$$2/28 : \sqrt{(280 - 330)^2 + (370 - 360)^2 + (390 - 330)^2 + (350 - 410)^2} \\ = 181.9341$$

### Step3. 選擇 k 筆資料

因為 k 值設定為 3，所以從距離量度中篩選最小的三筆，分別為 1/17、1/24、2/7 這三天。

可得：

$$\underline{VD1 \sim VD2 \text{ 旅行時間為: } (250+300+400)/3=316.67(\text{秒}) \doteq 5.28(\text{分鐘})}.$$

由於每五分鐘可收到一筆即時資料，而  $x_1 \sim x_2$  的旅行時間已超過五分鐘，因此  $x_2 \sim x_3$  的旅行時間必須搜尋下一時段(8:05)的旅行時間資料如表 3.3 所示。

表 3.3 8:05 旅行時間資料表

時間	$x_1 \sim x_2$ 旅行時間 (秒)	$x_2 \sim x_3$ 旅行時間 (秒)	$x_3 \sim x_4$ 旅行時間 (秒)	$x_4 \sim x_5$ 旅行時間 (秒)
1/17 AM 8:05	270	280	300	310
1/24 AM 8:05	320	300	360	370
2/7 AM 8:05	430	290	330	350

從表 3.3 可得：

$$\underline{x_2 \sim x_3 \text{ 旅行時間為: } (280+300+290)/3=290(\text{秒}) \doteq 4.83(\text{分鐘})}.$$

因為未超過五分鐘，因此  $x_3 \sim x_4$  的資料還是用 8:05 分的資料來計算：

$$\underline{x_3 \sim x_4 \text{ 旅行時間為: } (300+360+330)/3=330 (\text{秒}) = 5.5(\text{分鐘})}.$$

由於  $VD2 \sim VD3$  旅行時間 4.83(分鐘)加上  $x_3 \sim x_4$  旅行時間 5.5(分鐘)總共為 10.33 分鐘，已超過十分鐘，因此  $x_4 \sim x_5$  的旅行時間，需要蒐尋下兩個時段(8:15)的旅行時間資料如表 3.4 所示。

表 3.4 8:15 旅行時間資料表

時間	x1~x2 旅行時間 (秒)	x2~x3 旅行時間 (秒)	x3~x4 旅行時間 (秒)	x4~x5 旅行時間 (秒)
1/17 AM 8:15	280	270	320	320
1/24 AM 8:15	320	320	310	350
2/7 AM 8:15	330	280	300	330

從表 3.4 可得：

$$\underline{x4 \sim x5 \text{ 旅行時間為: } (320+350+330)/3=333.33 \text{ (秒) } \doteq 5.56 \text{ (分鐘)。}}$$

#### Step4. 輸出旅行時間預測值

整理以上的旅行時間資料結果可得到：

$$x1 \sim x2 \text{ 旅行時間為: } (250+300+400)/3=316.67 \text{ (秒) } \doteq 5.28 \text{ (分鐘)。}$$

$$x2 \sim x3 \text{ 旅行時間為: } (280+300+290)/3=290 \text{ (秒) } \doteq 4.83 \text{ (分鐘)。}$$

$$x3 \sim x4 \text{ 旅行時間為: } (300+360+330)/3=330 \text{ (秒) } = 5.5 \text{ (分鐘)。}$$

$$x4 \sim x5 \text{ 旅行時間為: } (320+350+330)/3=333.33 \text{ (秒) } \doteq 5.56 \text{ (分鐘)。}$$

因此，預測 3/6 早上 8 點鐘 A 收費站到 B 收費站旅行時間為：

$$\underline{(5.28 + 4.83 + 5.5 + 5.56) = 21.17 \text{ (分鐘)。}}$$

## 第四章 時變係數模式

在上一章已介紹完 k-NN 旅行時間預測模式，本章將介紹本研究所使用的另一項旅行時間預測模式-時變係數模式(Time-Varying Coefficient Model, TVC Model)，其基本概念為多元線性迴歸模式(Multiple Linear Regression Model)為基礎，本章將從 TVC 基本概念作介紹，經由構建流程每一步驟深入作探討，以建構完整 TVC 旅行時間預測模式。

### 4.1 TVC 模式介紹

上述提到的多元線性迴歸模式，又稱為複迴歸模式(Multiple Regression Model)，其主要在自變數的部分，為探討兩個以上的自變數與因變數之間的關係，利用最小平方法(Method of Least Squares) 校估各個迴歸係數，建立多元線性迴歸方程式，當得到自變數  $x_i$  時，即可預測依變數  $y$ 。其模式型態如式 4-1 所示：

$$y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon_i \quad (4-1)$$

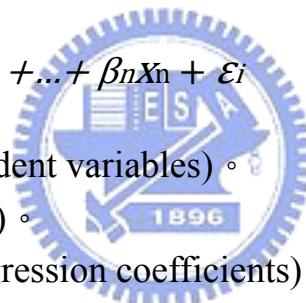
$y$  為依變數 (dependent variables)。

$\alpha$  為常數 (constant)。

$\beta_i$  為迴歸係數 (regression coefficients)。

$x_i$  為自變數 (independent variables)。

$\varepsilon_i$ : 誤差值 (error)。



TVC 模式也就是利用複迴歸模式為基礎，依照時間變化係數跟著作改變。而本研究利用蒐集資料的時間差異，所對應不同的迴歸係數，以此建立不同時間點所形成的迴歸模式，TVC 流程圖整理成圖 4.1 所示。因此，建構 TVC 模式前可先利用分群方式將資料作處理，將整理好的分群歷史資料利用統計軟體 SPSS 進行線性迴歸模式分析，依此建立各個分群的線性迴歸模式，當輸入一筆即時資料，立即可判斷即時資料屬於哪一分群，將此即時資料代入相同分群的迴歸模式，即可輸出所要的預測結果。

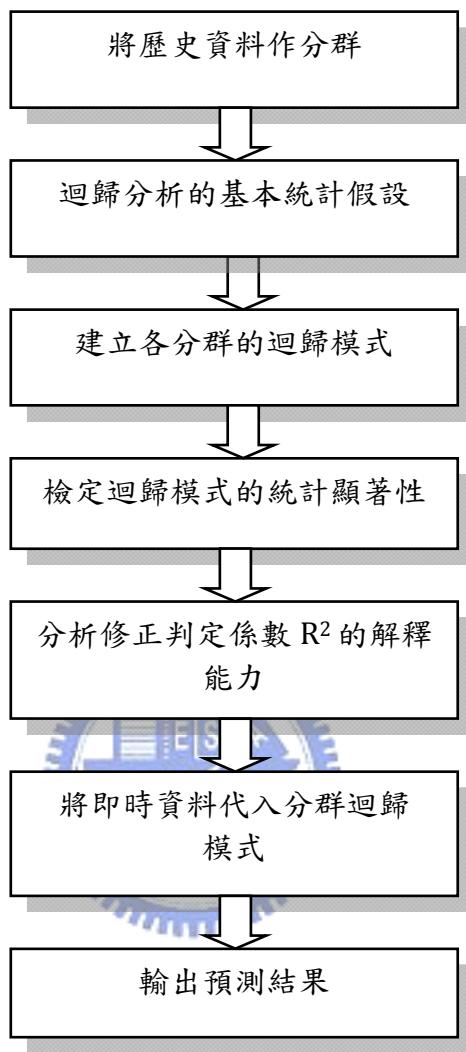


圖 4.1 TVC 流程圖

以下將會依照TVC模式流程圖每個步驟作一詳細說明，以建構完整TVC旅行時間預測模式。

#### 4.1.1 TVC資料分群

在上一章構建k-NN模式中也有提到資料分群步驟，在本研究k-NN模式與TVC模式分群的目的及方法並不完全相同。在k-NN模式中，分群的目的主要是能減少資料比對的時間，增加預測速度，因此依照不同的星期別來作分群。而TVC模式分群的目的是為了找出旅行時間的周期性變化，以此構建各分群的迴歸模式，讓迴歸預測模式能夠隨著時間作改變。

因此，本研究為了找出旅行時間週期性的變化，將依照不同星期別以及將每天的交通狀況區分為尖峰、離峰時段作分群，尖峰時段為星期一到星期五早上七點到九點和下午五點到七點，其餘皆屬離峰時段，試著利用此分群方式，構建本研究的TVC旅行時間預測模式。

#### 4.1.2 迴歸基本統計假設

在使用迴歸分析前，必須要確認資料是否符合迴歸分析的基本統計假設，否則，當資料違反迴歸分析的基本統計假設時，會導致統計推論偏誤的發生。

迴歸分析的基本統計假設有下列四項：

➤ 線性關係

依變數和自變數之間的關係必須是線性，也就是說，依變數與自變數存在著相當固定比率的關係，若是發現依變數與自變數呈現非線性關係時，可以透過轉換(transform)成線性關係，再進行迴歸分析。

➤ 常態性(normality)

若是資料呈現常態分配 (normal distribution)，則誤差項也會呈現同樣的分配，當樣本數夠大時，檢查的方式是使用簡單的 Histogram (直方圖)，若是樣本數較小時，檢查的方式是使用Normal Probability Plot (常態機率圖)。

### ➤ 誤差項的獨立性

自變數的誤差項，相互之間應該是獨立的，也就是誤差項與誤差項之間沒有相互關係，否則，在估計迴歸參數時，會降低統計的檢定力，我們可以藉由殘差(Residuals)的圖形分析來檢查，尤其是與時間序列和事件相關的資料，特別需要注意去處理。

### ➤ 誤差項的變異數相等(Homoscedasticity)

自變數的誤差項除了需要呈現常態性分配外，其變量數也需要相等，變量數的不相等(Heteroscedasticity)會導致自變數無法有效的估計應變數，例如：殘差分佈分析時，所呈現的三角形分佈和鑽石分佈，在SPSS軟體中，我們可以使用Levene test，來測試變異數的一致性，當變異數的不相等發生時，我們可以透過轉換(transform)成變異數的相等後，再進行迴歸分析。

## 4.1.3 檢定統計顯著性

複線性迴歸分析除了求取參數係數值外，還必須確定所得到的參數係數值是否有效，再求解迴歸方程式前，線性迴歸模型只是一種假設，因此還要進行統計測試，以確保所得之參數在統計上式有意義價值，以下介紹迴歸分析所需進行知統計測試，原則上區分成兩個部分，整體迴歸模式建立之測試和個別參數係數值顯著性之測試，整體的檢定滿足顯著性後，才有意義進一步進行個別參數之檢定，如下：

### (一) 整體迴歸模式建立之測試

在迴歸分析中，欲知n個自變數對應變數是否有影響，則必須對模式建立一統計檢驗，迴歸模式的顯著性檢定，一般都使用F test (檢定)，F檢定將所有自變數計算進來，看應變數 Y 和所有自變數 X<sub>n</sub> 是否有統計的顯著性。

F 檢定的虛無假設(Null hypothesis)如下：

$$\left\{ \begin{array}{l} H_0 : \beta_1 = \beta_2 = \dots = \beta_n = 0 \\ H_1 : \text{Not all } \beta_i = 0 \quad (i = 1, 2, \dots, n) \end{array} \right. \quad (4-2)$$

F 值的計算公式如4-3所示：

$$F = \frac{\frac{SSR}{k}}{\frac{SSE}{(N-k-1)}} = MSR/MSE \quad (4-3)$$

- SSR:迴歸平方和(Sum of Squares Regression, SSR)，是Y的預測值和Y的均值之間差異平方的總和，計算公式如式4-4所示：

$$SSR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 \quad (4-4)$$

$\hat{Y}_i$ :  $Y_i$ 的預測值； $\bar{Y}$ : Y的平均值

- SSE:誤差平方和(Sum of Squares Error, SSE)，是Y的實際值和Y在迴歸上的預測值之間差異平方的總和，計算公式如下：

$$SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (4-5)$$

- 總平方和(Sum of Squares Corrected Total, SST) ，計算公式如下：

$$SST = \sum_{i=1}^n (Y_i - \bar{Y})^2 = S_{YY} \quad (4-6)$$

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 \quad (4-7)$$

$$SST = SSE + SSR \quad (4-8)$$

- MSR: Mean Square Regression
- MSE: Mean Square Error

變異數分析(ANOVA)如表4.1所示，其中k表共有k個待判定的迴歸係數，N共有多少個觀測量。

表4.1 變異數分析表

來源	平方和	自由度	平均平方和	F檢定
迴歸	SSR	k	MSR	MSR/MSE
殘差	SSE	N-k-1	MSE	
總和	SST	N-1		

算出 F 值後，由 F 分佈表所查到之  $F_{(\alpha,k,N-k-1)}$  值作比較， $\alpha$  為選定之顯著水準。若根據統計量計算所得的值  $F > F_{(\alpha,k,N-k-1)}$ ，則我們在顯著水準  $\alpha$  下，拒絕  $H_0$ ，認為線性迴歸方程式是顯著的，表示此組係數項  $\beta$  對 Y 是有貢獻的；反之，認為線性迴歸方程式不顯著，不管提供之 F 值為多少，對 Y 的值都沒有貢獻，表示所建立之線性迴歸方程有問題。

## (二) 個別參數值顯著性之測試

若係數項  $\beta$  對 Y 是有貢獻的，但並非代表係數項  $\beta$  中之美一個係數都有貢獻未確定哪些係數是否對 Y 有貢獻，則必須針對參數值之顯著性作一測試如下：

$$\left\{ \begin{array}{l} H_0 : \beta_i = 0 \\ H_1 : \beta_i \neq 0 \end{array} \right. \quad (4-9)$$



利用 t 統計量進行檢定：

$$t = \frac{\hat{\beta}_i}{S(\hat{\beta}_i)} \sim t(n-k-1) \quad (4-10)$$

由此方法可算得每個參數的 t 統計量，再根據選定的顯著水準即可判定是否接受  $H_0$ ，若拒絕  $H_0$ ，表示此估計參數是顯著的，自變數  $X_i$  對應變數 Y 有顯著影響，在迴歸模式中有其存在的必要。若結果是接受  $H_0$ ，表示此參數不顯著，此參數在迴歸模式中無其存在之必要，一般一次 t 檢驗後，只剔除一個變量，這個變量是所有不顯著變量中 t 值最小，然後重新建立迴歸模式，在進行檢驗，直到所建立的迴歸模式即自變量都是顯著時為止。

## 4.1.4 判定係數 $R^2$

### (一) 判定係數

迴歸估計最理想的情況是所有的觀測點皆落在估計的迴歸線上，此時誤差測度的部份為零，即  $SSE=0$ ， $Y$  之變異情形可完全由  $X$  的線性函數解釋之，也就是  $Y$  之總變異等於迴歸變異， $SST=SSR$ 。另一種情形與上述完全相反，迴歸估計造成較大的誤差，而最差的狀況是所有變異皆為誤差變異，即  $SST=SSE$ ， $SSR=0$ 。

由上面的敘述，我們得到了依值限迴歸模型的指標係數，稱為判定係數(Coefficient of Determination)，用以衡量線性關係之強度，以符號  $R^2$  表示，定義如下：

$$R^2 = \frac{SSR}{SST} \quad , 0 \leq R^2 \leq 1 \quad (4-11)$$

$R^2$  之值亦表示  $Y$  的總變異中可藉由迴歸線解式的比例，其值越接近 1，表示迴歸線的解釋能力越強，反之，其值越接近 0，則表示此組資料配式此直線關係並不適當。其分子部份是因引進自變數  $X$  進入迴歸式後，應變數  $Y$  的總變異減少的百分比，當  $R^2$  越大，表示  $X$  解釋  $Y$  變異之能力越強。

### (二) 修正判定係數

在多元迴歸模式中，一般皆認為  $R^2$  愈大愈好。但事實上，判定係數  $R^2$  愈大，並不一定表示該模式愈好，因為在  $R^2$  公式中，自變數每多放一個，則其誤差平方和必定會降低，所不同的只是降低幅度大小問題而已，因此每增加一個自變數時必定會增加。因此當研究放入一些毫無意義之自變數時，必定可使  $R^2$  增加，而誤認該變數或該模式是有意義的，因此要修正上述之缺失，就必須考慮以修正後或調整後判定係數(Adjusted Determination Coefficient)，來檢驗模式是否良好，而調整判定係數之公式定義為：

$$\begin{aligned} R_a^2 &= 1 - \frac{SSE(N-1)}{SST(N-k-1)} \\ &= 1 - (1-R^2) \left[ \frac{N-1}{N-k-1} \right] \end{aligned} \quad (4-12)$$

## 4.2 TVC 模式構建

構建TVC旅行時間預測模式前，須先將歷史資料作分群，本研究的TVC模式將依照不同的星期別及尖峰、離峰時段將資料作分群，並建立各個分群的迴歸模式，當收到一筆即時旅行時間資料時，判斷該筆資料屬於哪一分群，將資料代入符合該分群的TVC模式，即可輸出旅行時間預測值。以下將相關預測步驟進行詳細說明，其構建模式之相關流程可表示如圖 4.2 TVC 模式構建流程圖所示。



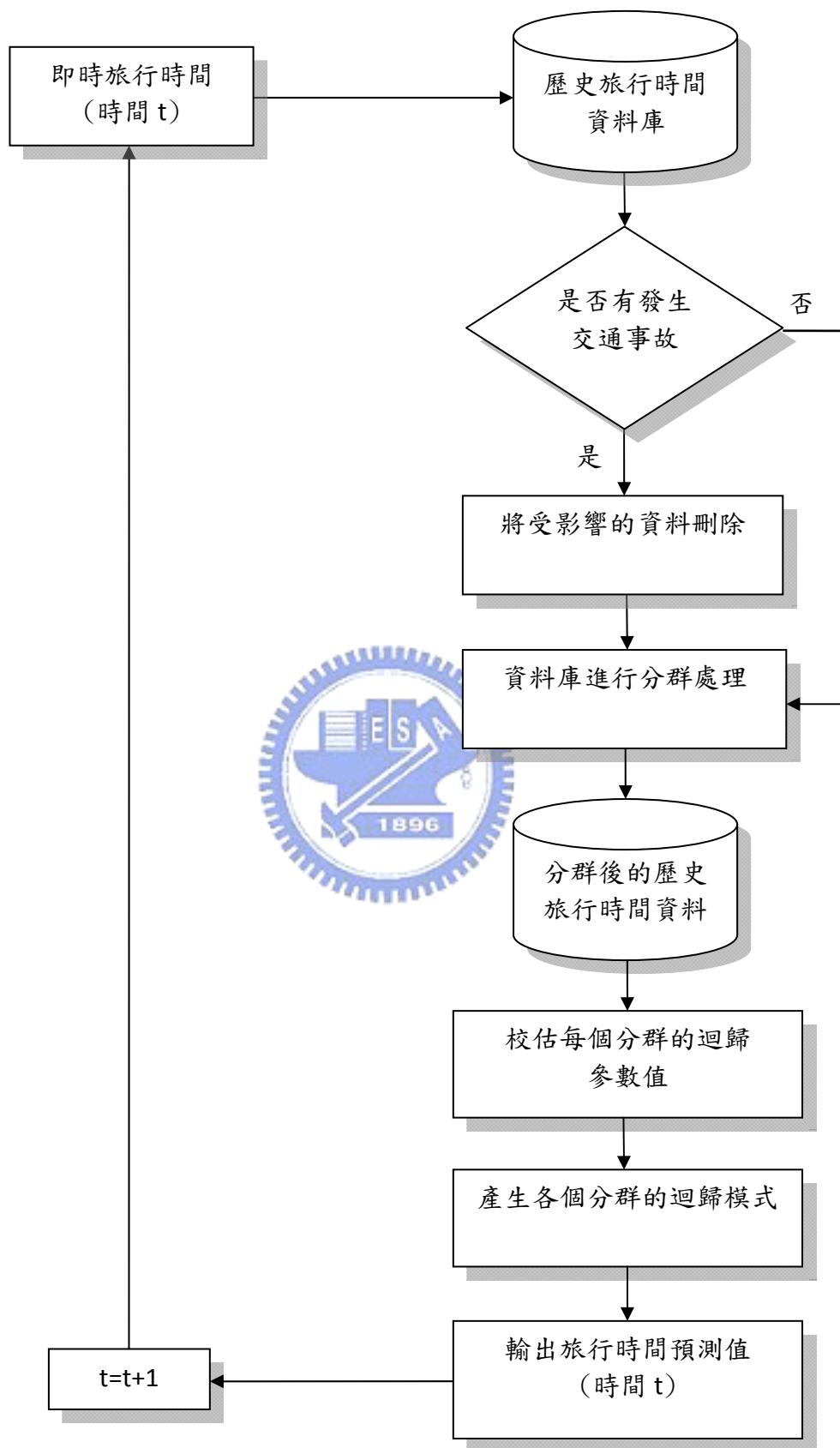


圖 4.2 TVC 模式構建流程圖

## 4.3 TVC 模式資料整理

### (一) 確認路段上固定點的數量

在 TVC 模式中，本研究利用 ETC 旅行時間當作應變數 Y，兩固定點之間的旅行時間當作自變數  $X_i$ ，由於本研究利用偵測器的位置作為固定點，因此要先確認所要預測的路段上共有多少偵測器來決定固定點的數量，才可得知需要設多少自變數，由於我們要預測的路段是泰山到楊梅南下路段的旅行時間，假設這路段上共有五個固定點  $d_1$ 、 $d_2$ 、 $d_3$ 、 $d_4$ 、 $d_5$  如圖 4.3 所示，每五分鐘可蒐集到一組資料，共有四筆同時段、兩兩固定點之間的旅行時間資料( $d_1 \sim d_2$ 、 $d_2 \sim d_3$ 、 $d_3 \sim d_4$ 、 $d_4 \sim d_5$ )，因此可將這四筆旅行時間資料設為自變數  $X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ ，同時段從泰山收費站出發到楊梅收費站的 ETC 旅行時間資料設為應變數 Y。

### (二) 將資料作整理

假設歷史資料庫中有 2008 年 1、2 月份各禮拜四的資料，我們要預測 3 月份各禮拜四早上 8:00 的旅行時間，首先依照分群方式先整理 1、2 月份個禮拜四尖峰時段 7:00~9:00 的資料，共有 1/3、1/10、1/17、1/24、1/31、2/7、2/14、2/21、2/28 這九天，將 ETC 旅行時間資料設為應變數 Y，兩點間的旅行時間資料為自變數  $X_i$ ，可將資料整理成如表 4.2 所示。當資料整理完成後，即可匯入本研究所使用的統計軟體 SPSS 進行線性迴歸模式分析，以建立完整線性迴歸預測模式。

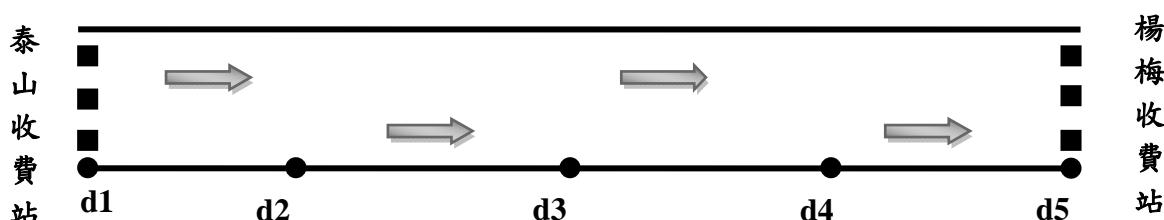


圖 4.3 TVC 範例路段示意圖

表 4.2 TVC 資料整理表

日期、時間	資料型態				
	Y	X1	X2	X3	X4
1/3(四) AM 7:00	1/3 早上 7:00 從泰山出發到楊梅 ETC 旅行時間	1/3 7:00 d1 到 d2 旅行時間	1/3 7:00 d2 到 d3 旅行時間	1/3 7:00 d3 到 d4 旅行時間	1/3 7:00 d4 到 d5 旅行時間
1/3(四) AM 7:05	1/3 早上 7:05 從泰山出發到楊梅 ETC 旅行時間	1/3 7:05 d1 到 d2 旅行時間	1/3 7:05 d2 到 d3 旅行時間	1/3 7:05 d3 到 d4 旅行時間	1/3 7:05 d4 到 d5 旅行時間
1/3(四) AM 7:10	1/3 早上 7:10 從泰山出發到楊梅 ETC 旅行時間	1/3 7:10 d1 到 d2 旅行時間	1/3 7:10 d2 到 d3 旅行時間	1/3 7:10 d3 到 d4 旅行時間	1/3 7:10 d4 到 d5 旅行時間

依照 TVC 模式構建流程圖所示，其詳細流程運作主要可分為五個步驟，以下將詳細作說明：

### Step1. 交通資料蒐集

本模式利用行駛在國道一號高速公路上楊梅到泰山之間的探針車所蒐集的交通資料，所推估的兩兩固定點之間的旅行時間資料當作資料來源，以此作為迴歸模式的自變數  $X_i$ ，ETC 旅行時間資料當作迴歸模式的應變數 Y，因此需要蒐集固定點之間旅行時間資料與 ETC 旅行時間資料。

### Step2. 將歷史資料作分群

本模式依照不同的星期別及尖峰、離峰時間，將資料作分群處理，其中如果遇到特殊節日，須獨立出來考慮。而當歷史資料發現該時段有發生交通事故或道路施工等特殊狀況，會造成車速減緩影響正常行車時間，必須將受到影響的資料予以刪除，若不予以刪除，將造成迴歸模式預測時產生誤差，誤認該時段為塞車情況，產生不正確的預測結果。

### Step3. 校估迴歸模式參數值

本模式利用統計軟體 SPSS 進行迴歸模式分析，採用最小平方法校估各分群的迴歸模式係數，並依照係數的顯著程度作調整，若有係數呈現不顯著性，則予以刪除，再重新作回歸模式分析，直到所有係數都為顯著為止。

#### Step4. 產生各個分群的時變係數模式

校估後的迴歸模式係數，代入迴歸式中建立完整迴歸式，並建立各個分群的迴歸模式，提供旅行時間預測使用。

#### Step5. 輸出旅行時間預測值

將收到的同時段、兩兩偵測器間的旅行時間資料帶入該分群的迴歸模式，可得到一旅行時間預測值，將此預測的旅行時間輸出，並將此旅行時間存入資料庫中，提供給下一個時段的旅行時間預測模組使用。



## 4.4 預測精確度之衡量

本研究採用之預測精確度衡量準則，是以平均絕對誤差百分比(Mean Absolute Percentage Error, MAPE)為評估指標，來衡量預測值與實際值之差異，其計算公式如式 4-13 所示：

$$MAPE = \frac{1}{M} \sum_{k=1}^M \left| \frac{x(k) - x'(k)}{x(k)} \right| * 100\% \quad (4-13)$$

其中

$x(k)$ 為實際值， $x'(k)$ 為預測值，M：樣本數

MAPE 之值愈小，表示模式正確預測能力愈強，精確度愈高，根據 Lewis (1982) MAPE 值之大小，將模式預測能力分為四種等級，如表 4.3 所示，當 MAPE 小於 10% 時，表示模式之預測能力高度精確，MAPE 介於 10%~20%，預測能力為良好。

表 4.3 MAPE 預測能力之等級

MAPE 值	預測能力
< 10%	高度精確
10%~20%	良好
20%~50%	合理
> 50%	不正確

## 第五章 數值測試

根據前面三、四章所構建之 k-NN 與 TVC 高速公路旅行時間預測模式，本章將以國道一號高速公路作為測試對象，測試在實際高速公路上，兩種模式的預測績效，並對兩種模式的預測績效進行評估。

### 5.1 模式構建與測試

#### 5.1.1 測試路段

本章將以國道一號高速公路南下泰山收費站(35.3 km)到楊梅收費站(71.4 km)作為測試路段，此路段總長為 36.1 公里，泰山收費站設為起點位置，楊梅收費站為迄點，此路段上共有 15 個固定點，因此每五分鐘可收到一組 14 筆兩兩固定點之間的旅行時間資料，表 5.1 為各固定點編號名稱及位置。

表 5.1 各固定點編號名稱及位置

固定點編號	1	2	3	4	5
固定點名稱	1S32.7	1S37	1S39	1S42	1S47
固定點位置(公里)	32.764	37.76	39.256	42.75	47.507
固定點編號	6	7	8	9	10
固定點名稱	1S50	1S51	1S52	1S54	1S55
固定點位置(公里)	50.01	51.09	52.84	54.014	55.77
固定點編號	11	12	13	14	15
固定點名稱	1S57	1S60	1S63	1S65	1S68.1
固定點位置(公里)	57.13	60.873	63.26	65.598	68.05

#### 5.1.2 預測時段

本研究將利用 2008 年 1、2 的歷史資料庫來預測以下兩個時段的旅行時間，每五分鐘預測一次，以比較 k-NN 與 TVC 的預測結果。

(一) 2008 年 3 月份每周四早上 8:00~9:00。

(二) 2008 年 3 月份每周六早上 10:00~11:00。

### 5.1.3 蔊集資料

本研究的 k-NN 旅行時間預測模式，利用兩兩固定點之間的旅行時間資料作比對，上述提到泰山到楊梅收費站路段總共有 15 個固定點，因此每五分鐘可收到 1 組 14 筆同時段、兩兩固定點之間的旅行時間，由於要利用 1、2 月份的歷史資料來作預測，本研究假設所蒐集這兩個月份的資料時皆無缺漏及異常的情況發生，其旅行時間資料庫格式如表 5.2 所示，圖 5.1 為旅行時間資料庫實際圖。

表 5.2 旅行時間資料庫格式

欄位名稱	資料內容	資料格式
VD_UP	上游固定點	S:南下，N:北上
VD_Down	下游固定點	S:南下，N:北上
Travel_Time	偵測器間旅行時間	單位:秒
TT_DateTime	日期、時間	年/月/日 時:分:秒
Weekday	星期別	1:星期日，2:星期一 依此類推
Section_NO	路段編號	ex:29，編號第 29 路段
Direction	方向	S:南下，N:北上
VD_Up_AveSpeed	上游固定點的平均速度	hr/km
VD_Down_AveSpeed	下游固定點的平均速度	hr/km
Probe_AveSpeed	探針車平均速度	hr/km
Probe_Speed_1	大客車探針車平均速度	hr/km
Probe_Speed_2	小客車探針車平均速度	hr/km

	VD_Up	VD_Down	Travel_Time	TT_DateTime	Weekday	Section_NO	Direction	VD_Up_AveSpeed	VD_Down_Ave...	Probe_AveSpeed	Probe_Speed_1	Probe_Speed_2	▲
1547	1550	109.693939393...	2008/1/3 上午 07:05:00	5	29	5	89.666664	79	83.166664	96	66		
1528.8	1529	23.5047601379...	2008/1/3 上午 07:15:00	5	17	5	95.666664	92.666664	89.666664	88	93		
1529	1530	21.7636949827...	2008/1/3 上午 07:40:00	5	18	5	94.333336	0	92.199997	90	94		
1518.1	1518.7	45.6879581151...	2008/1/3 上午 08:15:00	5	3	5	44	51.5	0	0	0	0	
1550	1551	47.9423113704...	2008/1/3 上午 08:45:00	5	30	5	76	89.5	77	81	74		
1522	1524.9	150.582657142...	2008/1/3 上午 09:20:00	5	10	5	52.5	52.5	0	0	0	0	
1520	1520.6	30.0759493670...	2008/1/3 上午 09:45:00	5	7	5	79	79	0	0	0	0	
1519.9	1520	5.08799999999...	2008/1/10 上午 07:05:00	5	6	5	75	75	0	0	0	0	
1518.7	1519.4	38.0712328767...	2008/1/10 上午 07:30:00	5	4	5	73	73	0	0	0	0	
1537	1539	61.953897097554	2008/1/10 上午 08:00:00	5	26	5	85	73.333336	87.5	84	85		
1521	1522	59.7913043478...	2008/1/10 上午 08:35:00	5	9	5	86	75	0	0	0	0	
1524.9	1525	5.02641509433...	2008/1/10 上午 09:15:00	5	11	5	70	89	0	0	0	0	
1520	1520.6	27.9529411764...	2008/1/10 上午 09:40:00	5	7	5	85	85	0	0	0	0	
1554	1555	64.7741569363...	2008/1/17 上午 07:05:00	5	33	5	86.333336	83	98.714287	98	101		
1528.1	1528.8	30.7544041283...	2008/1/17 上午 07:40:00	5	16	5	0	97.666664	82.75	78	92		
1529	1530	23.1389810404...	2008/1/17 上午 08:20:00	5	18	5	91.333336	0	86.666664	85	90		
1525.4	1526	56.7107569721...	2008/1/17 上午 08:45:00	5	13	5	83.666664	83.666664	0	0	0	0	
1568.1	1574	362.729503151...	2008/1/17 上午 09:15:00	5	39	5	74.333336	0	82.428574	65	88		
1525	1525.4	17.3126654064...	2008/1/17 上午 09:40:00	5	12	5	92	84.333336	0	0	0	0	
1525.4	1526	56.2624505928...	2008/1/17 上午 09:40:00	5	13	5	84.333336	84.333336	0	0	0	0	
1526	1527	31.0766798418...	2008/1/17 上午 09:40:00	5	14	5	84.333336	84.333336	0	0	0	0	
1527	1528.1	32.8199749668...	2008/1/17 上午 09:40:00	5	15	5	0	0	72.333336	62	78		
1528.1	1528.8	30.0587279686...	2008/1/17 上午 09:40:00	5	16	5	0	100.66666	85	78	92		
1528.8	1529	23.0932098162...	2008/1/17 上午 09:40:00	5	17	5	100.66666	88.666664	92.333336	92	93		
1529	1530	22.1728260869...	2008/1/17 上午 09:40:00	5	18	5	88.666664	0	92	92	92		
1530	1530.1	5.07096774193...	2008/1/17 上午 09:40:00	5	19	5	0	0	93	93	93		
1530.1	1530.5	15.668085106383	2008/1/17 上午 09:40:00	5	20	5	0	86.666664	93.666664	93	94		
1530.5	1531	42.424312578941	2008/1/17 上午 09:40:00	5	21	5	86.666664	90	94.666664	94	96		
1531	1532	15.6604969106...	2008/1/17 上午 09:40:00	5	22	5	90	0	97.5	96	99		

圖 5.1 旅行時間資料庫

本研究利用 ETC 資料作為實際值，與預測結果作比對，ETC 旅行時間資料庫格式如表 5.3 所示：

表 5.3 ETC 旅行時間資料庫格式

欄位名稱	資料內容	資料格式
DetectionDatetime	日期、時間	年/月/日 時:分:秒
Start_TollNo	收費站起點代碼	14003: 泰山收費站 22002: 楊梅收費站
End_TollNo	收費站迄點代碼	同上
Dir	方向	S:南下，N:北上
Cartype	車種	31:小客車，41:大客車
Travel_Time	旅行時間	分鐘
Weekday	星期別	1:星期日，2:星期一 依此類推

	DetectionDatetime	Start_TollNo	End_TollNo	Dir	Cartype	Travel_Time	Weekday
	2008/1/1 上午 09:55:00	22002	14003	N	31	21.8212128	3
	2008/1/1 下午 07:55:00	22002	14003	N	31	26.1690483	3
	2008/1/2 上午 05:55:00	22002	14003	N	31	21.0166664	4
	2008/1/2 下午 03:55:00	22002	14003	N	31	24.785	4
	2008/1/3 上午 01:55:00	22002	14003	N	31	21.3666668	5
	2008/1/3 上午 11:55:00	22002	14003	N	31	22.8083344	5
	2008/1/3 下午 09:55:00	22002	14003	N	31	21.9125	5
	2008/1/4 上午 07:55:00	22002	14003	N	31	29.1333332	6
	2008/1/4 下午 05:55:00	22002	14003	N	31	57.8807678	6
	2008/1/5 上午 03:55:00	22002	14003	N	31	21.166666	7
	2008/1/5 下午 01:55:00	22002	14003	N	31	23.2016659	7
	2008/1/5 下午 11:55:00	22002	14003	N	31	0	7
	2008/1/6 上午 09:55:00	22002	14003	N	31	22.9888878	1
	2008/1/6 下午 07:55:00	22002	14003	N	31	39.2129631	1
	2008/1/7 上午 05:55:00	22002	14003	N	31	21.35	2
	2008/1/7 下午 03:55:00	22002	14003	N	31	24.4424248	2
	2008/1/8 上午 01:55:00	22002	14003	N	31	0	3
	2008/1/8 上午 11:55:00	22002	14003	N	31	22.55	3
	2008/1/8 下午 09:55:00	22002	14003	N	31	21.7606068	3
	2008/1/9 上午 07:55:00	22002	14003	N	31	30.6452389	4
	2008/1/9 下午 05:55:00	22002	14003	N	31	34.6666679	4
	2008/1/10 上午 03:55:00	22002	14003	N	31	26.1	5
	2008/1/10 下午 01:55:00	22002	14003	N	31	24.3904762	5
	2008/1/10 下午 11:55:00	22002	14003	N	31	21.6444435	5
	2008/1/11 上午 09:55:00	22002	14003	N	31	23.1846161	6
	2008/1/11 下午 07:55:00	22002	14003	N	31	35.33205	6

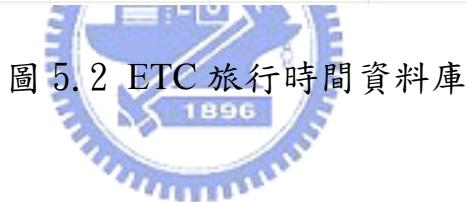


圖 5.2 ETC 旅行時間資料庫

## 5.2 建立 k-NN 預測模式

依照本研究的 k-NN 模式構建流程，建立完整預測模式。

### (1) k-NN 資料分群

將資料依照星期別作分群，因此每周的資料可分成七群，當收到即時資料，能立即判定屬於哪一分群，在分群的資料庫中比對資料。

### (2) 比對時間範圍

由於考量預測路段為泰山收費站到楊梅收費站，此路段總長為 36.1 公里，以時速 80 公里作計算，其旅行時間大約 30 分鐘，因此設置比對時間範圍為前後半小時歷史資料。

### (3) k 值設定

k 值設定方式為先設定各門檻值，以 50 為一間距單位，當距離量度小於此門檻值，則判定與即時資料相近，可進行選取。檢視各門檻值內各預測時段(每五分鐘為一預測時段)的 k 值及對應的誤差值，選取當時段誤差最小的 k 值解，以此方式設定 k 值。

### (4) 驗證預測結果

以泰山收費站到楊梅收費站的 ETC 旅行時間資料作為實際值，與預測結果作比對，以 MAPE 結果判定預測績效，檢視預測能力。

## 5.3 k-NN 預測結果

### (一) 2008 年 3 月份每周四早上 8:00~9:00。

在設定 k 值部分，門檻值最小設置為 100，因為門檻值小於 100 時會有某些時段選取不到歷史資料，門檻值最大為 1200，當門檻值為 1200 時，每個時段的歷史資料都已全部選取完畢，因此沒有必要再增加門檻值作選取，各門檻值對應 k 值表如表 5.4 所示，因礙於篇幅故僅列出部分時段。



表 5.4 門檻值對應 k 值(1)

時間	門檻值 100	門檻值 150	門檻值 200	…	門檻值 1200
2008/3/6 AM 08:00:00	87	108	108	…	117
2008/3/6 AM 08:05:00	82	107	107	…	117
2008/3/6 AM 08:10:00	80	106	106	…	117
2008/3/6 AM 08:15:00	76	105	105	…	117
2008/3/6 AM 08:20:00	72	104	104	…	117
2008/3/6 AM 08:25:00	69	104	104	…	117
2008/3/6 AM 08:30:00	64	101	104	…	117
2008/3/6 AM 08:35:00	61	97	104	…	117
2008/3/6 AM 08:40:00	60	93	104	…	117
2008/3/6 AM 08:45:00	59	89	104	…	117
2008/3/6 AM 08:50:00	60	85	104	…	117
2008/3/6 AM 08:55:00	72	81	103	…	117
2008/3/6 AM 09:00:00	51	79	103	…	117
2008/3/13 AM 08:00:00	60	106	108	…	117
2008/3/13 AM 08:05:00	59	106	107	…	117
2008/3/13 AM 08:10:00	57	106	106	…	117
2008/3/13 AM 08:15:00	57	105	105	…	117
2008/3/13 AM 08:20:00	57	104	104	…	117
2008/3/13 AM 08:25:00	61	104	104	…	117
2008/3/13 AM 08:30:00	60	104	104	…	117
2008/3/13 AM 08:35:00	59	104	104	…	117
2008/3/13 AM 08:40:00	57	104	104	…	117
2008/3/13 AM 08:45:00	45	104	104	…	117
2008/3/13 AM 08:50:00	16	76	104	…	117
2008/3/13 AM 08:55:00	14	89	104	…	117
2008/3/13 AM 09:00:00	3	37	89	…	117
2008/3/20 AM 08:00:00	105	108	108	…	117
2008/3/20 AM 08:05:00	105	107	1007	…	117
2008/3/20 AM 08:10:00	82	106	106	…	117
⋮	⋮	⋮	⋮	⋮	⋮
2008/3/28 AM 08:50:00	52	104	104	…	117
2008/3/28 AM 08:55:00	49	104	104	…	117
2008/3/28 AM 09:00:00	24	40	98	…	117

從表 5.5 得知各門檻值所對應的 k 值結果，接著必須要知道當各時段選取其 k 值結果時，所對應的絕對誤差值，如表 5.6 所示：

表 5.5 k 值對應絕對誤差百分比(1)

時間	門檻值 100 (%)	門檻值 150 (%)	門檻值 200 (%)	…	門檻值 1200 (%)
2008/3/6 AM 08:00:00	3.3122	0.0313	0.1373	…	2.9167
2008/3/6 AM 08:05:00	7.9029	4.3257	4.2248	…	6.1097
2008/3/6 AM 08:10:00	8.5474	4.7389	4.6387	…	5.5016
2008/3/6 AM 08:15:00	8.7632	4.2922	4.1920	…	4.5773
2008/3/6 AM 08:20:00	8.7054	3.0971	2.9956	…	2.4249
2008/3/6 AM 08:25:00	6.1503	0.2184	0.3225	…	1.1862
2008/3/6 AM 08:30:00	4.6972	2.4876	3.5486	…	4.3400
2008/3/6 AM 08:35:00	2.9441	4.6972	6.4024	…	7.4087
2008/3/6 AM 08:40:00	5.1685	1.9844	4.0514	…	5.2920
2008/3/6 AM 08:45:00	3.4620	3.5241	5.9078	…	7.4996
2008/3/6 AM 08:50:00	0.5965	6.4690	9.2052	…	11.3174
2008/3/6 AM 08:55:00	4.4415	6.8553	9.4976	…	11.9408
2008/3/6 AM 09:00:00	8.1666	6.3207	9.3469	…	12.1619
2008/3/13 AM 08:00:00	3.5689	0.2726	0.1329	…	2.9208
2008/3/13 AM 08:05:00	7.2782	3.7953	3.8373	…	1.7937
2008/3/13 AM 08:10:00	6.8343	3.1353	3.2438	…	2.3095
2008/3/13 AM 08:15:00	5.0664	2.4919	2.5992	…	2.1867
2008/3/13 AM 08:20:00	5.5232	3.5107	3.6191	…	4.2286
2008/3/13 AM 08:25:00	6.8881	4.5884	4.6971	…	5.5984
2008/3/13 AM 08:30:00	6.6542	5.2566	5.3647	…	6.1699
2008/3/13 AM 08:35:00	9.5166	8.6964	8.8072	…	9.8362
2008/3/13 AM 08:40:00	3.5855	3.2087	3.3134	…	4.5453
2008/3/13 AM 08:45:00	6.0823	3.3203	3.4249	…	4.9793
2008/3/13 AM 08:50:00	17.9408	11.6544	7.1529	…	9.2253
2008/3/13 AM 08:55:00	20.8247	11.8058	9.5287	…	11.9532
2008/3/13 AM 09:00:00	24.0018	17.0215	9.5356	…	9.3428
:	:	:	:	:	:
2008/3/28 AM 08:50:00	5.6681	5.7277	5.6326	…	3.8075
2008/3/28 AM 08:55:00	3.6933	3.8802	3.7839	…	1.6540
2008/3/28 AM 09:00:00	4.8455	3.1527	10.9651	…	9.6364

已從表 5.5 得知各時段選取 k 值時所得到誤差結果，因此從各時段的誤差值當中，挑選最小的誤差結果，其所選取的 k 值，即設定為當時段的 k 值最佳解。各時段選取 k 值所得到預測結果如表 5.6 所示，最大一筆誤差在 3/28 早上 8:10，其誤差值約為 9 %，整體預測 MAPE 值約為 3.4 %。圖 5.3 為 k-NN 預測結果，圖 5.4 至圖 5.7 為預測 3 月份各禮拜四早上 8:00~9:00 預測結果。

表 5.6 k-NN 預測結果(1)

時間	ETC (分鐘)	k-NN (分鐘)	k 值	絕對誤差值 (%)
2008/3/6 AM 08:00:00	25.8389	24.9830	108	0.0313
2008/3/6 AM 08:05:00	27.1378	26.2916	109	3.1182
2008/3/6 AM 08:10:00	27.3242	26.3392	108	3.6048
2008/3/6 AM 08:15:00	27.3236	26.4531	107	3.1861
2008/3/6 AM 08:20:00	26.9896	26.4519	106	1.9924
2008/3/6 AM 08:25:00	26.3046	26.3621	104	0.2184
2008/3/6 AM 08:30:00	25.7704	25.6512	111	0.4625
2008/3/6 AM 08:35:00	25.2730	25.3653	111	0.3652
2008/3/6 AM 08:40:00	25.9573	25.9218	107	0.1368
2008/3/6 AM 08:45:00	25.5788	25.4115	57	0.6539
2008/3/6 AM 08:50:00	24.9000	25.0485	60	0.5965
2008/3/6 AM 08:55:00	24.9722	26.0814	72	4.4415
2008/3/6 AM 09:00:00	25.1767	26.7680	79	6.3207
2008/3/13 AM 08:00:00	25.8400	25.8744	108	0.1329
2008/3/13 AM 08:05:00	25.0308	25.9808	59	3.7953
2008/3/13 AM 08:10:00	25.2381	25.7194	108	1.9071
2008/3/13 AM 08:15:00	25.5150	25.8332	107	1.2472
2008/3/13 AM 08:20:00	25.2667	25.8320	106	2.2375
2008/3/13 AM 08:25:00	25.2056	25.8980	105	2.7473
2008/3/13 AM 08:30:00	25.3262	26.1935	105	3.4244
2008/3/13 AM 08:35:00	24.7144	26.6188	104	7.7053
2008/3/13 AM 08:40:00	26.1427	26.7366	104	2.2718
⋮	⋮	⋮	⋮	⋮
2008/3/28 AM 08:50:00	28.8152	28.0725	109	2.5800
2008/3/28 AM 08:55:00	28.4242	28.3685	104	0.2000
2008/3/28 AM 09:00:00	31.25	30.2648	111	3.1500
平均				3.39

### k-NN 預測(1)

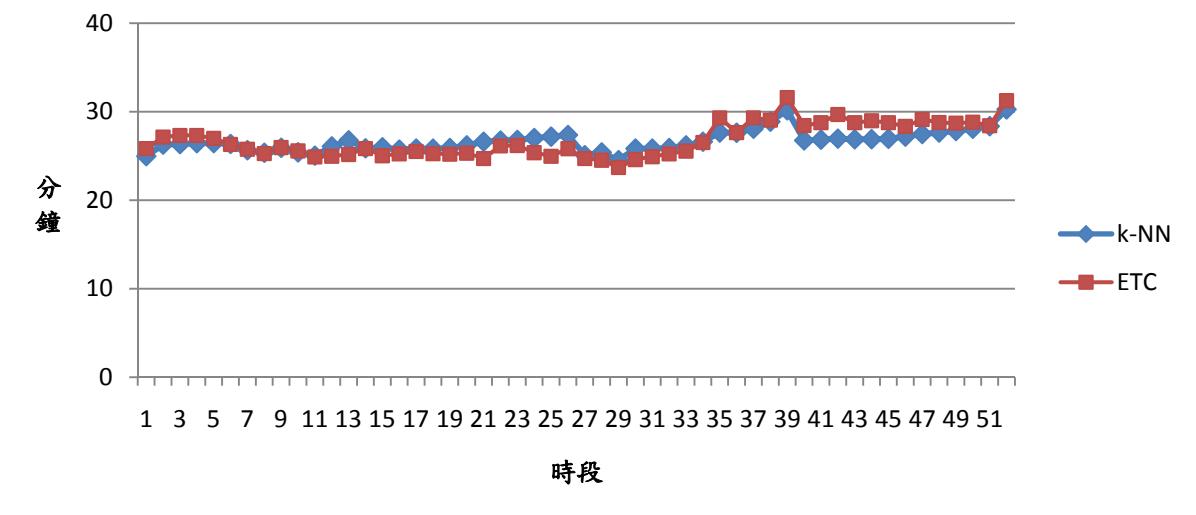


圖 5.3 k-NN 預測結果(1)



### 3/6 k-NN預測

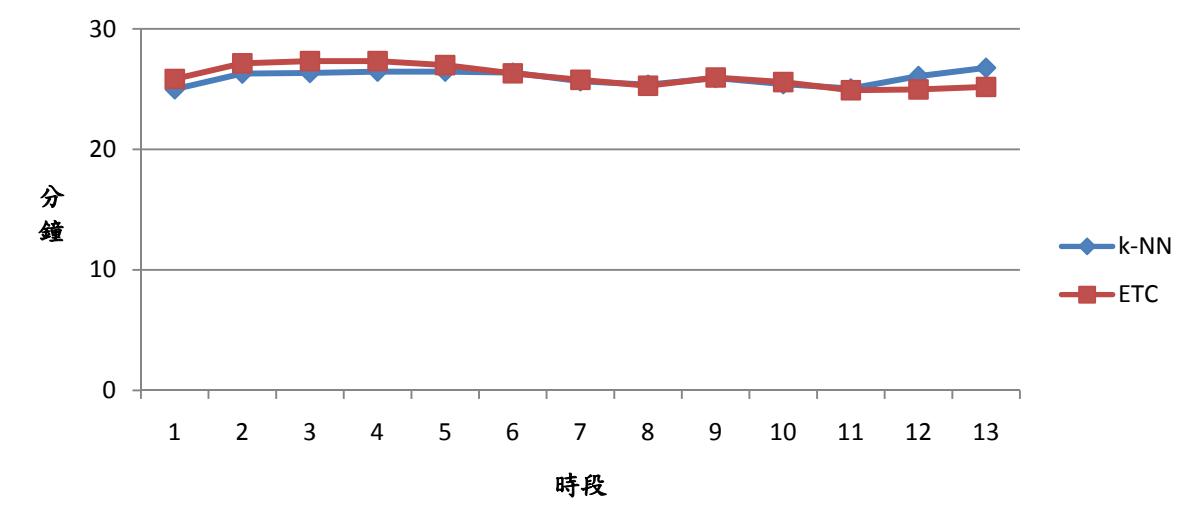


圖 5.4 3/6 k-NN 預測結果

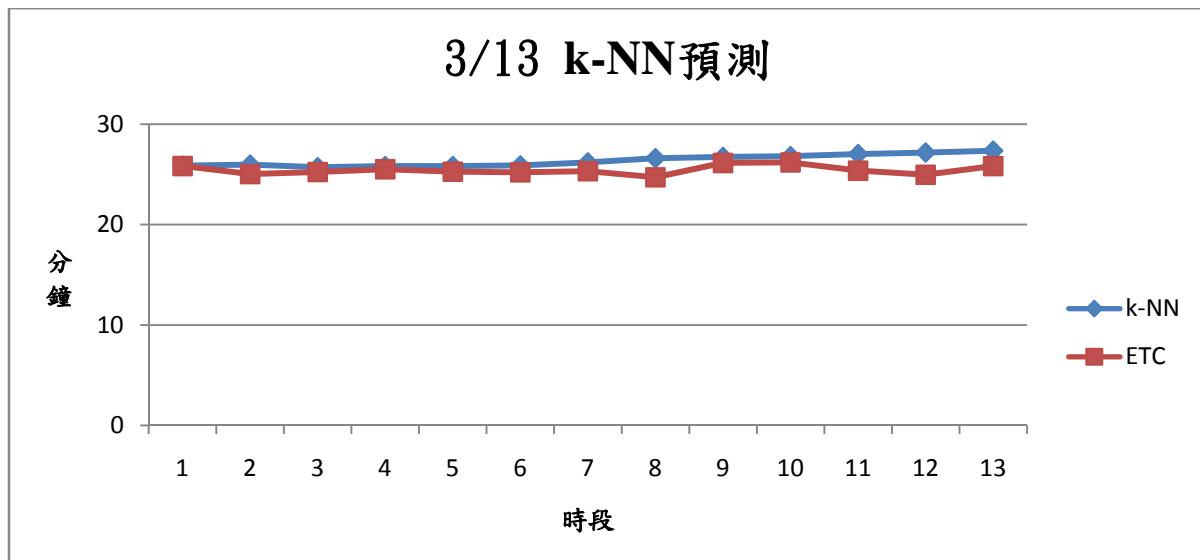


圖 5.5 3/13 k-NN 預測結果

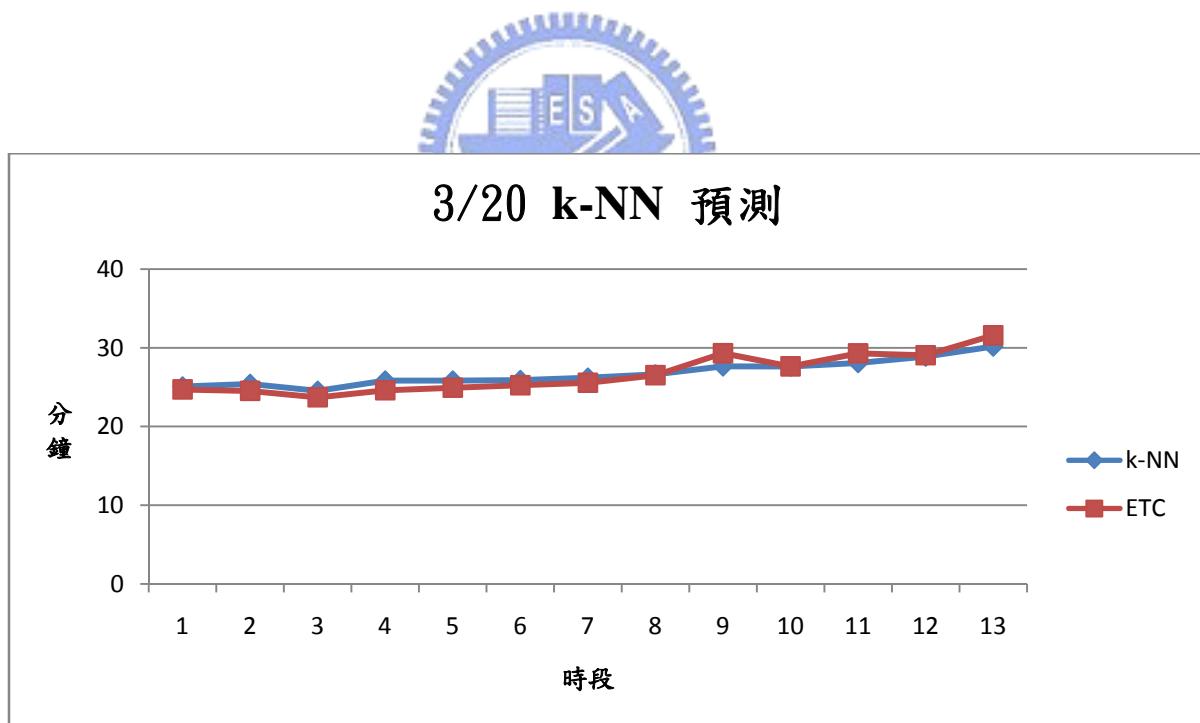


圖 5.6 3/20 k-NN 預測結果

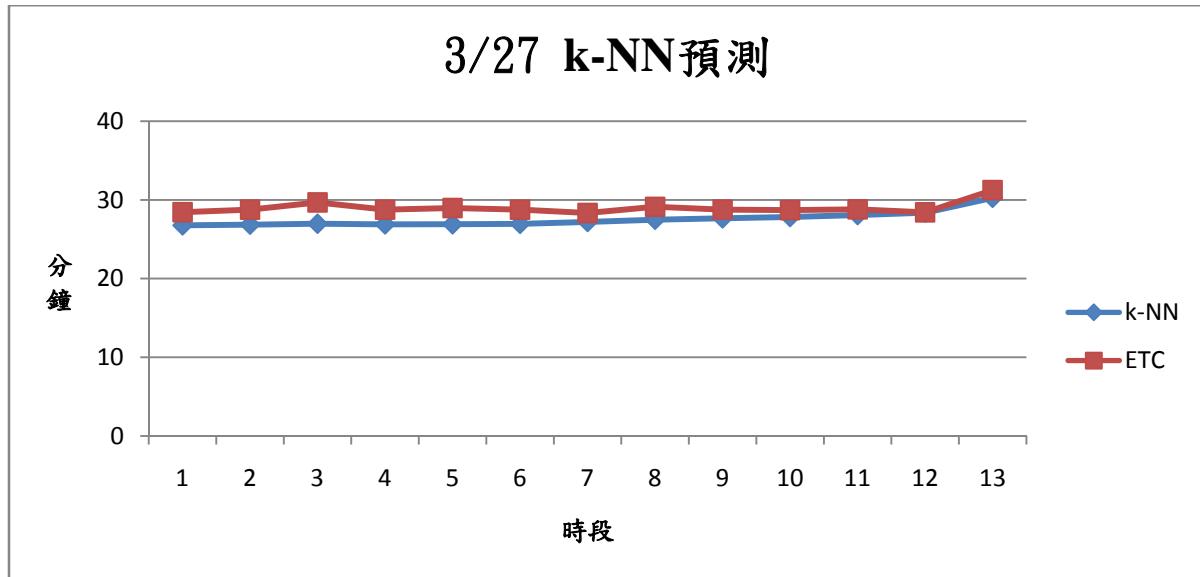


圖 5.7 3/28 k-NN 預測結果

(二) 2008 年 3 月份每周六早上 10:00~11:00。

門檻值最小設置為 600，最大為 1100，各門檻值對應 k 值表如表 5.7 所示，因礙於篇幅故也僅列出部分時段。

表 5.7 門檻值對應 k 值(2)

時間	門檻值 600	門檻值 650	門檻值 700	…	門檻值 1100
2008/3/1 AM 10:00:00	91	99	103	…	104
2008/3/1 AM 10:05:00	57	87	101	…	104
2008/3/1 AM 10:10:00	63	88	101	…	104
2008/3/1 AM 10:15:00	67	90	101	…	104
2008/3/1 AM 10:20:00	95	99	101	…	104
2008/3/1 AM 10:25:00	96	99	101	…	104
2008/3/1 AM 10:30:00	82	91	96	…	104
2008/3/1 AM 10:35:00	97	101	103	…	104
2008/3/1 AM 10:40:00	82	90	95	…	104
2008/3/1 AM 10:45:00	83	88	96	…	104
2008/3/1 AM 10:50:00	85	93	95	…	104
2008/3/1 AM 10:55:00	89	93	96	…	104
2008/3/1 AM 11:00:00	2	15	41	…	104

2008/3/8 AM 10:00:00	102	104	104	...	104
2008/3/8 AM 10:05:00	102	103	103	...	104
2008/3/8 AM 10:10:00	102	103	103	...	104
2008/3/8 AM 10:15:00	97	103	103	...	104
2008/3/8 AM 10:20:00	96	102	103	...	104
2008/3/8 AM 10:25:00	95	102	103	...	104
2008/3/8 AM 10:30:00	101	103	103	...	104
2008/3/8 AM 10:35:00	97	102	103	...	104
2008/3/8 AM 10:40:00	99	103	103	...	104
2008/3/8 AM 10:45:00	103	103	103	...	104
2008/3/8 AM 10:50:00	101	103	103	...	104
2008/3/8 AM 10:55:00	102	103	103	...	104
2008/3/8 AM 11:00:00	101	103	104	...	104
2008/3/15 AM 10:00:00	100	103	103	...	104
2008/3/15 AM 10:05:00	101	102	102	...	104
2008/3/15 AM 10:10:00	99	101	102	...	104
:	:	:	:	:	:
2008/3/29 AM 10:50:00	94	100	103	...	104
2008/3/29 AM 10:55:00	100	101	101	...	104
2008/3/29 AM 11:00:00	100	100	101	...	104

從表 5.7 得知各門檻值所對應的預測結果，如表 5.8 所示：

表 5.8 k 值對應絕對誤差百分比(2)

時間	門檻值 600 (%)	門檻值 650 (%)	門檻值 700 (%)	...	門檻值 1100 (%)
2008/3/1 AM 10:00:00	1.4604	0.1478	1.3451	...	1.8094
2008/3/1 AM 10:05:00	2.9486	1.3445	1.9995	...	2.1367
2008/3/1 AM 10:10:00	0.8989	0.6098	3.6083	...	3.8286
2008/3/1 AM 10:15:00	5.5029	2.6941	0.2220	...	0.0603
2008/3/1 AM 10:20:00	0.8369	0.0587	0.8387	...	2.0400
2008/3/1 AM 10:25:00	2.0892	3.1918	3.9893	...	5.4720
2008/3/1 AM 10:30:00	0.9109	3.3531	4.1012	...	7.4666
2008/3/1 AM 10:35:00	3.5564	4.9933	5.5639	...	5.9168
2008/3/1 AM 10:40:00	2.1606	4.7698	5.2557	...	8.4736

2008/3/1 AM 10:45:00	2.7371	3.9676	5.5980	...	8.2275
2008/3/1 AM 10:50:00	3.7809	4.8422	5.7386	...	8.6763
2008/3/1 AM 10:55:00	1.2276	2.4551	3.6280	...	5.9343
2008/3/1 AM 11:00:00	22.3189	22.3275	9.9374	...	2.1385
2008/3/8 AM 10:00:00	0.3467	0.0582	0.0582	...	0.0582
2008/3/8 AM 10:05:00	1.1726	1.2106	1.2106	...	0.4232
2008/3/8 AM 10:10:00	9.9921	10.0377	10.0377	...	9.6480
2008/3/8 AM 10:15:00	2.6440	2.0773	2.0773	...	1.6801
2008/3/8 AM 10:20:00	4.3731	3.4780	3.5611	...	3.1947
2008/3/8 AM 10:25:00	4.1653	2.5161	2.5896	...	2.2382
2008/3/8 AM 10:30:00	5.3962	5.2370	5.2370	...	4.9087
2008/3/8 AM 10:35:00	5.2985	4.5809	4.6698	...	4.3511
2008/3/8 AM 10:40:00	4.2803	4.2613	4.2613	...	3.9512
2008/3/8 AM 10:45:00	2.4614	2.4614	2.4614	...	2.1531
2008/3/8 AM 10:50:00	3.0776	3.1941	3.1941	...	2.8927
2008/3/8 AM 10:55:00	2.8235	2.8451	2.8451	...	2.5505
2008/3/8 AM 11:00:00	1.8946	1.6909	1.3991	...	1.3991
2008/3/15 AM 10:00:00	4.1252	3.9953	3.9953	...	4.4718
2008/3/15 AM 10:05:00	3.3526	3.3168	3.3168	...	4.5994
2008/3/15 AM 10:10:00	2.8956	3.2064	3.6867	...	4.5713
:	:	:	:	:	:
2008/3/29 AM 10:50:00	2.3652	4.3598	5.1761	...	5.5036
2008/3/29 AM 10:55:00	0.8883	0.9284	0.9284	...	1.7319
2008/3/29 AM 11:00:00	0.3324	0.3324	0.5826	...	1.3751

從表 5.8 得知各時段選取 k 值時所得到誤差結果，從各時段的誤差值當中，挑選最小的誤差結果，各時段選取 k 值所得到預測結果如表 5.9 所示，最大一筆誤差在 3/1 上午 11:00，其誤差值約為 11%，整體預測 MAPE 值約為 3.7 %。圖 5.8 為 k-NN 預測結果，圖 5.9 至圖 5.13 為預測 3 月份各禮拜六上午 10:00~11:00 k-NN 預測結果。

表 5.9 k-NN 預測結果(2)

時間	ETC (分鐘)	k-NN (分鐘)	k 值	絕對誤差值 (%)
2008/3/1 AM 10:00:00	43.5396	44.1252	103	1.3451
2008/3/1 AM 10:05:00	46.3500	45.4232	103	1.9995
2008/3/1 AM 10:10:00	48.2021	46.4628	103	3.6083
2008/3/1 AM 10:15:00	47.3067	47.4117	103	0.2220
2008/3/1 AM 10:20:00	47.2083	47.6042	102	0.8387
2008/3/1 AM 10:25:00	46.4322	48.2845	102	3.9893
2008/3/1 AM 10:30:00	46.0625	47.9516	98	4.1012
2008/3/1 AM 10:35:00	47.1667	49.7910	103	5.5639
2008/3/1 AM 10:40:00	46.4208	48.8606	97	5.2557
2008/3/1 AM 10:45:00	46.8111	49.4316	98	5.5980
2008/3/1 AM 10:50:00	46.7864	49.4712	98	5.7386
2008/3/1 AM 10:55:00	48.3056	50.0581	100	3.6280
2008/3/1 AM 11:00:00	50.3500	55.3535	65	9.9374
2008/3/8 AM 10:00:00	44.3532	44.3274	104	0.0582
2008/3/8 AM 10:05:00	45.5524	45.0009	104	1.2106
2008/3/8 AM 10:10:00	51.3067	46.1566	104	10.0377
2008/3/8 AM 10:15:00	48.0860	47.0871	104	2.0773
2008/3/8 AM 10:20:00	49.7611	47.9891	103	3.5611
2008/3/8 AM 10:25:00	50.0942	48.7969	103	2.5896
2008/3/8 AM 10:30:00	52.0571	49.3309	103	5.2370
2008/3/8 AM 10:35:00	52.2300	49.7910	103	4.6698
2008/3/8 AM 10:40:00	52.4258	50.1917	104	4.2613
2008/3/8 AM 10:45:00	51.7773	50.5028	104	2.4614
2008/3/8 AM 10:50:00	52.3603	50.6879	104	3.1941
2008/3/8 AM 10:55:00	52.5115	51.0175	104	2.8451
2008/3/8 AM 11:00:00	52.1565	51.4268	104	1.3991
2008/3/15 AM 10:00:00	42.4300	44.1252	104	3.9953
2008/3/15 AM 10:05:00	43.3651	44.8034	103	3.3168
2008/3/15 AM 10:10:00	44.3302	45.9645	103	1.3451
:	:	:	:	:
2008/3/29 AM 10:50:00	48.1933	50.6879	103	5.1761
2008/3/29 AM 10:55:00	50.3010	50.7680	102	0.9284
2008/3/29 AM 11:00:00	50.7292	51.0247	102	0.5826
平均				3.6817

## k-NN 預測(2)

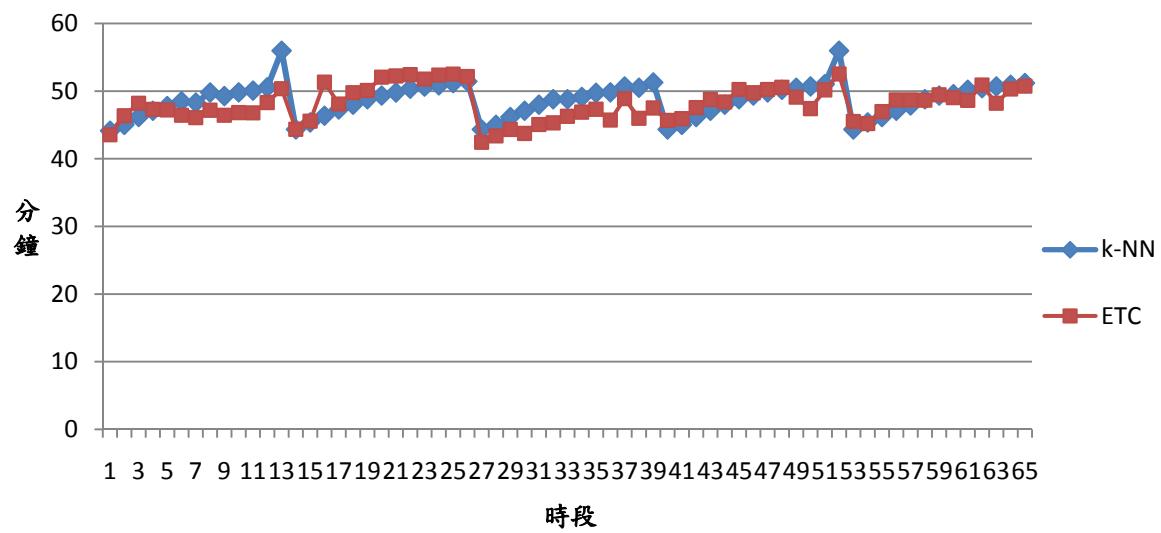


圖 5.8 k-NN 預測結果(2)



## 3/1 k-NN預測

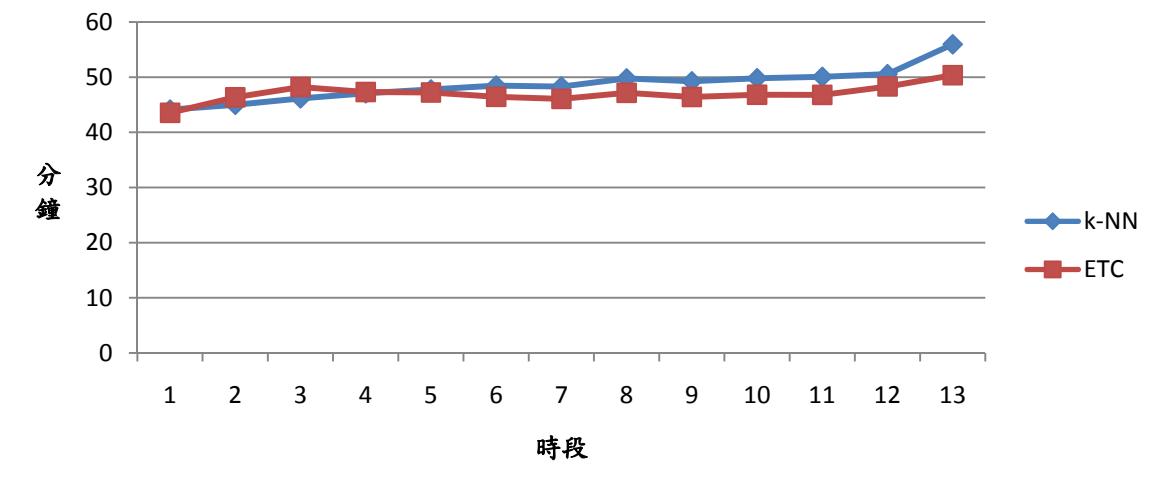


圖 5.9 3/1 k-NN 預測結果

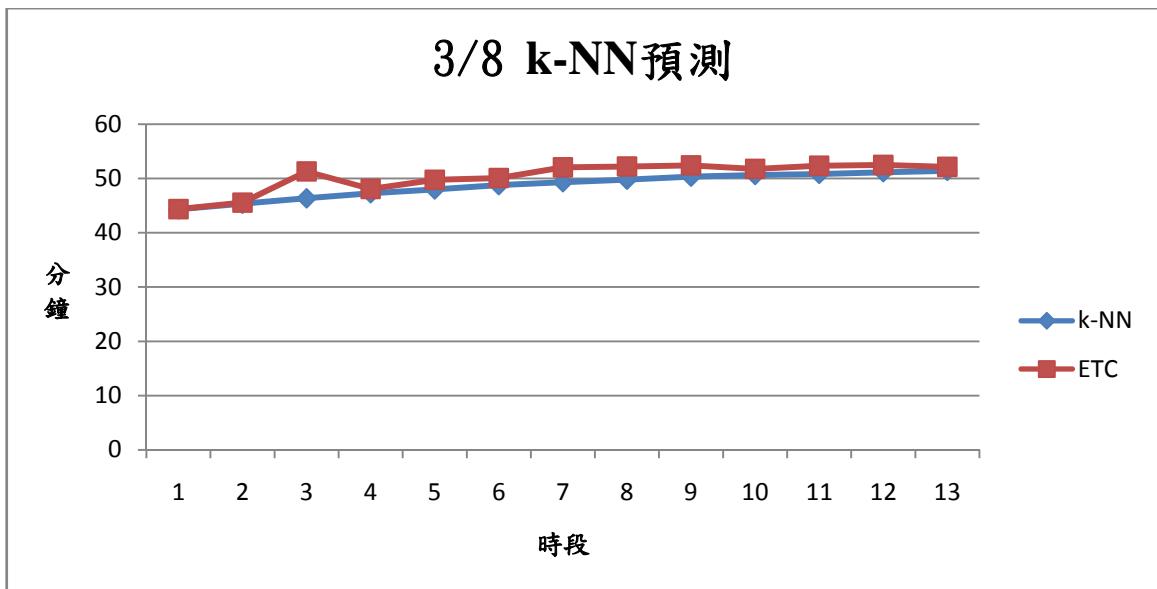


圖 5.10 3/8 k-NN 預測結果

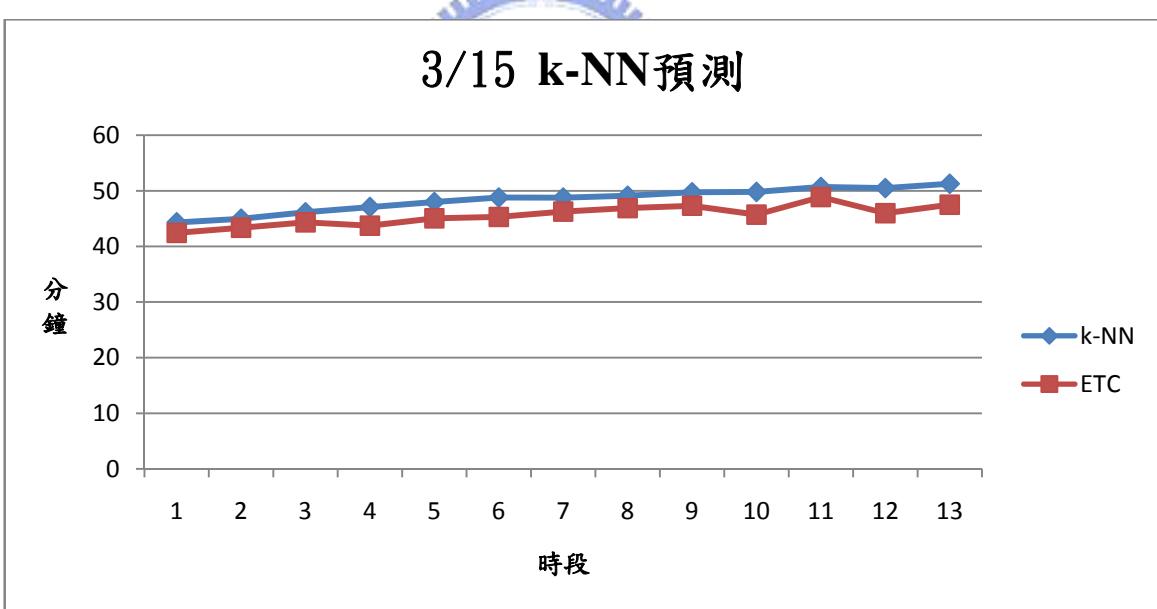


圖 5.11 3/15 k-NN 預測結果

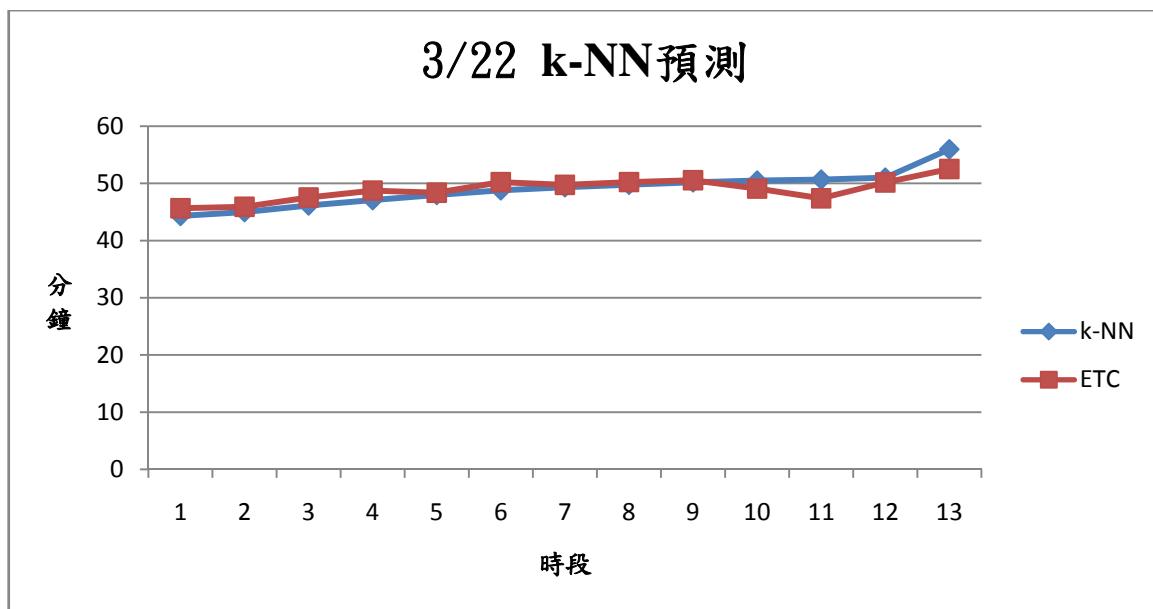


圖 5.12 3/22 k-NN 預測結果

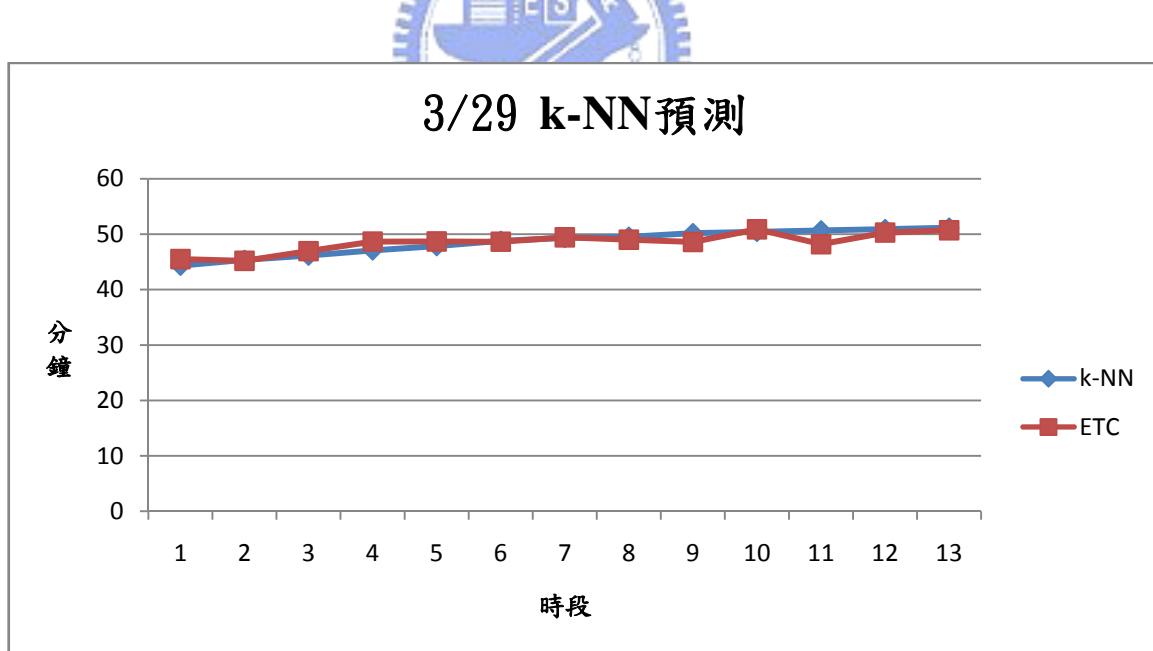


圖 5.13 3/29 k-NN 預測結果

## 5.4 建立 TVC 預測模式

一般在迴歸研究中，迴歸式所選用之解釋變數必須是有意義的，而且變數之間必須互相獨立，與應變數間的相關程度應該要高。而迴歸式的各項參數項需通過 F 檢定及 t 檢定，才能確保所建立之模式是有意義的，而且較估出來解數變數之參數值不能為 0，最後再由判定係數  $R^2$  來判斷所建立之迴歸試中對應變數解釋能力的高低與否，經過以上個步驟，才能建立一個完整的迴歸式，因此，本研究在線性迴歸的檢定上，使用 SPSS 統計軟體，進行各種模式的檢定，其作法將說明如下。

### (1) TVC 資料分群

將資料依照星期別作分群，再將每周資料區分成尖、離峰交通趨勢，當收到即時資料，能立即判定屬於哪一分群，由於假日的交通狀況與平日狀況不同，因此將獨立出來考慮。

### (2) 資料整理

將蒐集到的資料依照 4.7 所介紹資料整理方式進行整理。

- (一) 首先要預測 2008 年三月份各禮拜四早上 8:00~9:00，因為該時段屬於尖峰時段，因此必須整理資料庫中 1、2 月份的禮拜四早上 7:00~9:00 尖峰時段的歷史資料，以此作迴歸模式。
- (二) 預測 2008 年三月份各禮拜六早上 10:00~11:00，將整理資料庫中 1、2 月份 9:30~11:30 的歷史資料，以此作迴歸模式。

### (3) 驗證預測結果

以泰山收費站到楊梅收費站的 ETC 旅行時間資料作為實際值，與預測結果作比對，以 MAPE 結果判定預測績效，檢視預測能力。

## 5.4.1 測試資料整理

模式構建前，將蒐集到的資料按照 4.7 所介紹資料整理方式進行整理，以泰山收費站到楊梅收費站間的 ETC 旅行時間資料作為迴歸模式的依變數 Y，兩兩固定點之間的旅行時間當作自變數 X，因此自變數有 X1、X2、...、

X14，將扣除掉事故資料部分進行迴歸分析，各上、下游固定點所對應的變數名稱如表 5.10 所示。

表 5.10 上、下游固定點變數名稱對照

變數名稱	X1	X2	X3	X4	X5	X6	X7
上游固定點	1S32.7	1S37	1S39	1S42	1S47	1S50	1S51
下游固定點	1S37	1S39	1S42	1S47	1S50	1S51	1S52
變數名稱	X8	X9	X10	X11	X12	X13	X14
上游固定點	1S52	1S54	1S55	1S57	1S60	1S63	1S65
下游固定點	1S54	1S55	1S57	1S60	1S63	1S65	1S68.1

(一) 預測 2008 年三月份各禮拜四早上 8:00~9:00。

迴歸 標準化殘差 的常態 P-P 圖

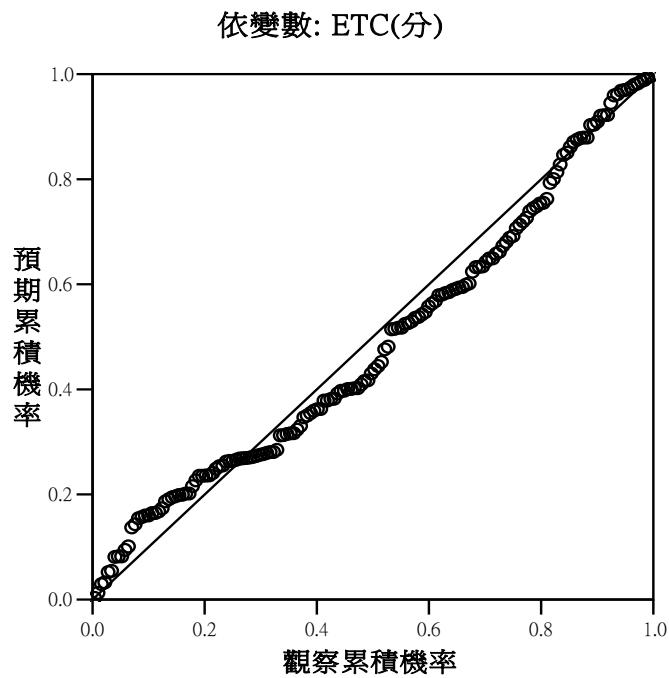


圖 5.14 迴歸標準化殘差 P-P 圖(1)

## Step 1. 強迫進入法進行迴歸模式之建立

► 報表分析結果如下：

表5.11 模式摘要(1)

模式	R	R 平方	調過後的 R 平方	估計的標準誤	R 平方改變量	變更統計量			
						F 改變	分子自由度	分母自由度	顯著性 F 改變
1	.739(a)	.545	.503	1.1275872019	.545	12.941	14	151	.000

a 預測變數：(常數), X1, X2, X3, X4, X5, X6, X7, X8, X9, X10, X11, X12, X13, X14  
 b 依變數：ETC(分)

模式摘要表，有負相關係數R=0.739，R平方=0.545，調整過後的R平方=0.503，估計的標準誤1.128，R平方改變量=0.545，F值的改變=12.941，分子自由度=14，分母自由度=151，F值改變的顯著性=0.000。

表5.12 變異數分析(1)

模式		平方和	自由度	平均平方和	F 檢定	顯著性
1	迴歸	230.360	14	16.454	12.941	.000(a)
	殘差	191.989	151	1.271		
	總和	422.349	165			

a 預測變數：(常數), X1, X2, X3, X4, X5, X6, X7, X8, X9, X10, X11, X12, X13, X14  
 b 依變數：ETC(分)

變異數分析摘要表，有SSR=230.36，SSE=191.989，SST=422.349，SSR+ SSE = SST，F值等於12.941(F=MSR/MSE=16.454/1.271=12.941)，p=0.000<0.05，達顯著水準。

表5.13 系數(1)

模式	未標準化係數		標準化係數 Beta	t	顯著性	迴歸係數 B 的 95% 信賴區間		相關			共線性統計量	
	B 之估計值	標準誤				下限	上限	零階	偏	部分	允差	VIF
1 (常數)	20.277	1.667		12.162	.000	16.983	23.572					
X1	0	.007	.000	.001	.999	-.014	.014	.422	.000	.000	.183	
X2	-.053	.018	-.345	-2.912	.004	-.089	-.017	.047	-.231	-.160	.215	
X3	.018	.011	.163	1.626	.106	-.004	.040	.337	.131	.089	.301	
X4	.012	.011	.181	1.151	.251	-.009	.034	.000	.093	.063	.122	
											8.177	

X5	-.030	.028	-.195	-1.048	.296	-.086	.026	-.009	-.085	-.057	.087	11.448
X6	-.090	.031	-.354	-2.884	.005	-.152	-.028	-.038	-.228	-.158	.199	5.013
X7	.076	.019	.534	4.090	.000	.040	.113	-.097	.316	.224	.177	5.652
X8	-.031	.020	-.166	-1.559	.121	-.070	.008	-.082	-.126	-.086	.267	3.746
X9	-.080	.031	-.340	-2.593	.010	-.142	-.019	-.009	-.206	-.142	.175	5.711
X10	.128	.025	.759	5.162	.000	.079	.177	.146	.387	.283	.139	7.189
X11	-.019	.007	-.290	-2.711	.007	-.033	-.005	.022	-.215	-.149	.263	3.797
X12	.038	.017	.256	2.250	.026	.005	.070	.479	.180	.123	.232	4.309
X13	.033	.008	.556	4.171	.000	.017	.048	.589	.321	.229	.169	5.903
X14	-.008	.016	-.068	-.488	.626	-.038	.023	.411	-.040	-.027	.157	6.362

a 依變數：ETC(分)

迴歸分析的各係數值，常數項等於20.277，未標準化的迴歸係數X1=0，標準化的迴歸係數X1 Beta值=0.000，t值=0.001，p=0.999，未達到顯著水準。

---

## Step 2. 刪除最不顯著的變數 X1，再作一次迴歸分析。

► 報表分析結果如下：

表5.14 模式摘要(2)

模式	R	R 平方	調過後的 R 平方	估計的標準誤	變更統計量				
					R 平方改變量	F 改變	分子自由度	分母自由度	顯著性 F 改變
1	.739(a)	.545	.507	1.1238719143	.545	14.029	13	152	.000

a 預測變數：(常數), X2, X3, X4, X5, X6, X7, X8, X9, X10, X11, X12, X13, X14

b 依變數：ETC(分)

模式摘要表，有負相關係數R=0.739，R平方=0.545，調整過後的R平方=0.507，估計的標準誤1.124，R平方改變量=0.545，F值的改變=14.029，分子自由度=13，分母自由度=152，F值改變的顯著性=0.000。

表5.15 變異數分析(2)

模式		平方和	自由度	平均平方和	F 檢定	顯著性
1	迴歸	230.360	13	17.720	14.029	.000(a)
	殘差	191.989	152	1.263		
	總和	422.349	165			

a 預測變數：(常數), X2, X3, X4, X5, X6, X7, X8, X9, X10, X11, X12, X13, X14

b 依變數：ETC(分)

變異數分析摘要表，有SSR=230.36，SSE=191.989，SST=422.349，SSR+SSE=SST，F值等於14.029(F=MSR/MSE=17.720/1.263=14.029)，p=0.000<0.05，達顯著水準。

表5.16 係數(2)

模式	未標準化係數		標準化係數	t	顯著性	迴歸係數 B 的 95% 信賴區間		相關			共線性統計量	
	B 之估計值	標準誤	Beta 分配			下限	上限	零階	偏	部分	允差	VIF
1 (常數)	20.278	1.625		12.482	.000	17.068	23.487					
X2	-.053	.016	-.345	-3.272	.001	-.085	-.021	.047	-.257	-.179	.269	
X3	.018	.010	.163	1.884	.061	-.001	.037	.337	.151	.103	.402	
X4	.012	.011	.181	1.164	.246	-.009	.034	.000	.094	.064	.124	
X5	-.030	.027	-.194	-1.117	.266	-.082	.023	-.009	-.090	-.061	.099	
X6	-.090	.030	-.354	-2.988	.003	-.149	-.030	-.038	-.236	-.163	.213	
X7	.076	.019	.534	4.112	.000	.040	.113	-.097	.316	.225	.178	
X8	-.031	.019	-.166	-1.633	.105	-.068	.006	-.082	-.131	-.089	.291	
X9	-.080	.031	-.340	-2.620	.010	-.141	-.020	-.009	-.208	-.143	.178	
X10	.128	.024	.759	5.243	.000	.080	.176	.146	.391	.287	.143	
X11	-.019	.007	-.290	-2.722	.007	-.033	-.005	.022	-.216	-.149	.264	
X12	.038	.017	.256	2.260	.025	.005	.070	.479	.180	.124	.232	
X13	.033	.008	.556	4.229	.000	.017	.048	.589	.324	.231	.173	
X14	-.008	.015	-.067	-.524	.601	-.036	.021	.411	-.042	-.029	.180	

a 依變數：ETC(分)

迴歸分析的各係數值，常數項等於20.278，未標準化的迴歸係數X14=-0.008，標準化的迴歸係數X14 Beta值=-0.067，t值=-0.524，p=0.601，未達到顯著水準。

Step 3. 刪除最不顯著的變數 X14，再作一次迴歸分析。

► 報表分析結果如下：

表5.17 模式摘要(3)

模式	R	R 平方	調過後的 R 平方	估計的標準誤	變更統計量				
					R 平方改變量	F 改變	分子自由度	分母自由度	顯著性 F 改變
1	.738(a)	.545	.509	1.1212049239	.545	15.248	12	153	.000

a 預測變數：(常數), X2, X3, X4, X5, X6, X7, X8, X9, X10, X11, X12, X13

b 依變數：ETC(分)

模式摘要表，有負相關係數R=0.738，R平方=0.545，調整過後的R平方=0.509，估計的標準誤1.121，R平方改變量=0.545，F值的改變=15.248，分子自由度=12，分母自由度=153，F值改變的顯著性=0.000。

表5.18 變異數分析(3)

模式		平方和	自由度	平均平方和	F 檢定	顯著性
1	迴歸	230.013	12	19.168	15.248	.000(a)
	殘差	192.336	153	1.257		
	總和	422.349	165			

a 預測變數：(常數), X2, X3, X4, X5, X6, X7, X8, X9, X10, X11, X12, X13

b 依變數：ETC(分)

變異數分析摘要表，有SSR=230.013, SSE=192.336, SST=422.349, SSR + SSE = SST, F值等於15.248(F=MSR/MSE=19.168/1.257=15.248), p=0.000<0.05, 達顯著水準。

表5.19 係數(3)

模式	未標準化係數			標準化係數 Beta 分配	t	顯著性	迴歸係數 B 的 95% 信賴 區間		相關			共線性統計 量	
	B 之 估計 值	標準 誤	標準 誤				下限	上限	零階	偏	部分	允差	VIF
1 (常數)	20.028	1.549		12.927	.000		16.967	23.089					
X2	-.052	.016	-.337	-3.238	.001		-.083	-.020	.047	-.253	-.177	.275	3.633
X3	.018	.009	.167	1.953	.053		.000	.037	.337	.156	.107	.406	2.464
X4	.013	.011	.185	1.199	.233		-.008	.034	.000	.096	.065	.125	8.023
X5	-.030	.026	-.199	-1.148	.253		-.083	.022	-.009	-.092	-.063	.099	10.112
X6	-.089	.030	-.352	-2.976	.003		-.149	-.030	-.038	-.234	-.162	.213	4.694
X7	.075	.018	.522	4.091	.000		.039	.111	-.097	.314	.223	.183	5.478
X8	-.036	.016	-.195	-2.330	.021		-.067	-.006	-.082	-.185	-.127	.423	2.362
X9	-.077	.030	-.327	-2.574	.011		-.137	-.018	-.009	-.204	-.140	.184	5.434
X10	.127	.024	.753	5.230	.000		.079	.175	.146	.389	.285	.144	6.961
X11	-.020	.007	-.298	-2.838	.005		-.033	-.006	.022	-.224	-.155	.270	3.709
X12	.037	.016	.250	2.223	.028		.004	.069	.479	.177	.121	.235	4.257
X13	.030	.005	.505	5.681	.000		.019	.040	.589	.417	.310	.376	2.659

a 依變數：ETC(分)

迴歸分析的各係數值，常數項等於20.028，未標準化的迴歸係數X5=-0.03，標準化的迴歸係數X5 Beta值=-0.199, t值=-1.148, p=0.253，未達到顯著水準。

#### Step 4. 刪除最不顯著的變數 X5，再作一次迴歸分析。

► 報表分析結果如下：

表5.20 模式摘要(4)

模式	R	R 平方	調過後的 R 平方	估計的標準誤	變更統計量				
					R 平方改變量	F 改變	分子自由度	分母自由度	顯著性 F 改變
1	.735(a)	.541	.508	1.1223652428	.541	16.480	11	154	.000

a 預測變數：(常數) X2, X3, X4, X6, X7, X8, X9, X10, X11, X12, X13

b 依變數：ETC(分)

模式摘要表，有負相關係數R=0.735，R平方=0.541，調整過後的R平方=0.508，估計的標準誤1.122，R平方改變量=0.541，F值的改變=16.480，分子自由度=11，分母自由度=154，F值改變的顯著性=0.000

表5.21 變異數分析(4)

模式		平方和	自由度	平均平方和	F 檢定	顯著性
1	迴歸	228.355	11	20.760	16.480	.000(a)
	殘差	193.994	154	1.260		
	總和	422.349	165			

a 預測變數：(常數) X2, X3, X4, X6, X7, X8, X9, X10, X11, X12, X13

b 依變數：ETC(分)

變異數分析摘要表，有SSR=228.355，SSE=193.994，SST=422.349，SSR+ SSE = SST，F值等於16.480(F=MSR/MSE=20.760/1.26=16.480)，p=0.000< 0.05，達顯著水準。

表5.22 系數(4)

模式		未標準化係數		標準化係數 Beta 分配	t	顯著性	迴歸係數 B 的 95% 信賴區間		相關			共線性統計量	
		B 之估計值	標準誤				下限	上限	零階	偏	部分	允差	VIF
1	(常數)	20.112	1.549		12.982	.000	17.051	23.172					
	X2	-.053	.016	-.347	-3.349	.001	-.085	-.022	.047	-.261	-.183	.277	3.604
	X3	.021	.009	.195	2.363	.019	.004	.039	.337	.187	.129	.440	2.273
	X4	.004	.007	.051	.506	.613	-.010	.017	.000	.041	.028	.289	3.465
	X6	-.103	.027	-.408	-3.777	.000	-.158	-.049	-.038	-.291	-.206	.256	3.903
	X7	.072	.018	.506	3.982	.000	.037	.108	-.097	.306	.217	.185	5.408

X8	-.038	.016	-.204	-2.440	.016	-.069	-.007	-.082	-.193	-.133	.427	2.343
X9	-.092	.027	-.390	-3.395	.001	-.146	-.039	-.009	-.264	-.185	.226	4.427
X10	.140	.021	.832	6.587	.000	.098	.183	.146	.469	.360	.187	5.352
X11	-.020	.007	-.298	-2.830	.005	-.033	-.006	.022	-.222	-.155	.270	3.709
X12	.030	.016	.208	1.952	.053	.000	.061	.479	.155	.107	.263	3.796
X13	.029	.005	.494	5.579	.000	.019	.039	.589	.410	.305	.381	2.623

a 依變數：ETC(分)

迴歸分析的各係數值，常數項等於20.112，未標準化的迴歸係數X4=0.04，標準化的迴歸係數X4 Beta值=0.051，t值=0.506，p=0.613，未達到顯著水準。

---

Step 5. 刪除最不顯著的變數 X4，再作一次迴歸分析。

► 報表分析結果如下：

表5.23 模式摘要(5)

模式	R	R 平方	調過後的 R 平方	估計的標準誤	變更統計量				
					R 平方改變量	F 改變	分子自由度	分母自由度	顯著性 F 改變
1	.735(a)	.540	.510	1.1196691124	.540	18.189	10	155	.000

a 預測變數：(常數) X2, X3, X6, X7, X8, X9, X10, X11, X12, X13

b 依變數：ETC(分)

模式摘要表，有負相關係數R=0.735，R平方=0.540，調整過後的R平方=0.510，估計的標準誤1.120，R平方改變量=0.540，F值的改變=18.189，分子自由度=10，分母自由度=155，F值改變的顯著性=0.000

表5.24 變異數分析(5)

模式		平方和	自由度	平均平方和	F 檢定	顯著性
1	迴歸	228.032	10	22.803	18.189	.000(a)
	殘差	194.317	155	1.254		
	總和	422.349	165			

a 預測變數：(常數) X2, X3, X6, X7, X8, X9, X10, X11, X12, X13

b 依變數：ETC(分)

變異數分析摘要表，有SSR=228.032，SSE=194.317，SST=422.349， $SSR + SSE = SST$ ，F值等於18.189( $F=MSR/MSE=22.803/1.254=18.189$ )， $p=0.000 < 0.05$ ，達顯著水準。

表5.25 係數(5)

模式	未標準化係數		標準化係數 Beta 分配	t	顯著性	迴歸係數 B 的 95% 信賴 區間		相關			共線性統計 量	
	B 之 估計 值	標準 誤				下限	上限	零階	偏	部分	允差	VIF
1 (常數)	20.188	1.538		13.125	.000	17.150	23.226					
X2	-.051	.015	-.333	-3.349	.001	-.081	-.021	.047	-.260	-.182	.301	
X3	.022	.009	.200	2.454	.015	.004	.040	.337	.193	.134	.447	
X6	-.104	.027	-.410	-3.815	.000	-.158	-.050	-.038	-.293	-.208	.257	
X7	.072	.018	.505	3.984	.000	.036	.108	-.097	.305	.217	.185	
X8	-.039	.015	-.209	-2.517	.013	-.069	-.008	-.082	-.198	-.137	.432	
X9	-.085	.023	-.360	-3.667	.000	-.131	-.039	-.009	-.283	-.200	.308	
X10	.138	.021	.817	6.680	.000	.097	.179	.146	.473	.364	.199	
X11	-.018	.006	-.274	-2.911	.004	-.030	-.006	.022	-.228	-.159	.335	
X12	.029	.015	.196	1.891	.060	-.001	.059	.479	.150	.103	.277	
X13	.030	.005	.500	5.728	.000	.019	.040	.589	.418	.312	.390	

a 依變數：ETC(分)

迴歸分析的各係數值，常數項等於20.188，未標準化的迴歸係數X12=0.029，標準化的迴歸係數X12 Beta值=0.196，t值=1.891，p=0.060，未達到顯著水準。



Step 6. 刪除最不顯著的變數 X12，再作一次迴歸分析。

► 報表分析結果如下：

表5.26 模式摘要(6)

模式	R	R 平方	調過後的 R 平方	估計的標準誤	變更統計量				
					R 平方改變量	F 改變	分子自由度	分母自由度	顯著性 F 改變
1	.728(a)	.529	.502	1.1288817418	.529	19.491	9	156	.000

a 預測變數：(常數) X2, X3, X6, X7, X8, X9, X10, X11, X13

b 依變數：ETC(分)

模式摘要表，有負相關係數R=0.728，R平方=0.529，調整過後的R平方=0.502，估計的標準誤1.129，R平方改變量=0.529，F值的改變=19.491，分子自由度=9，分母自由度=156，F值改變的顯著性=0.000

表5.27 變異數分析(6)

模式		平方和	自由度	平均平方和	F 檢定	顯著性
1	迴歸	223.547	9	24.839	19.491	.000(a)
	殘差	198.802	156	1.274		
	總和	422.349	165			

a 預測變數：(常數) X2, X3, X6, X7, X8, X9, X10, X11, X13

b 依變數：ETC(分)

變異數分析摘要表，有SSR=223.547, SSE=198.802, SST=422.349, SSR + SSE = SST，F值等於19.491(F=MSR/MSE=24.839/1.274=19.491)，p=0.000<0.05，達顯著水準。

表5.28 係數(6)

模式	未標準化係數		標準化係數 Beta 分配	t	顯著性	迴歸係數 B 的 95% 信賴 區間		相關			共線性統計 量	
	B 之 估計 值	標準 誤				下限	上限	零階	偏	部分	允差	VIF
1 (常數)	20.772	1.519		13.672	.000	17.771	23.773					
X2	-.053	.015	-.343	-3.433	.001	-.083	-.022	.047	-.265	-.189	.302	
X3	.027	.009	.247	3.149	.002	.010	.044	.337	.244	.173	.492	
X6	-.095	.027	-.373	-3.501	.001	-.148	-.041	-.038	-.270	-.192	.266	
X7	.065	.018	.455	3.643	.000	.030	.101	-.097	.280	.200	.193	
X8	-.032	.015	-.172	-2.113	.036	-.062	-.002	-.082	-.167	-.116	.458	
X9	-.089	.023	-.374	-3.794	.000	-.135	-.042	-.009	-.291	-.208	.310	
X10	.138	.021	.820	6.657	.000	.097	.179	.146	.470	.366	.199	
X11	-.013	.006	-.202	-2.328	.021	-.025	-.002	.022	-.183	-.128	.400	
X13	.036	.004	.611	9.375	.000	.028	.044	.589	.600	.515	.710	
											1.408	

a 依變數：ETC(分)

迴歸分析的各係數值，常數項等於20.772，未標準化的迴歸係數與標準化的迴歸係數皆已達到顯著水準。

(二) 預測2008年三月份各禮拜六早上10:00~11:00。

迴歸標準化殘差的常態 P-P 圖

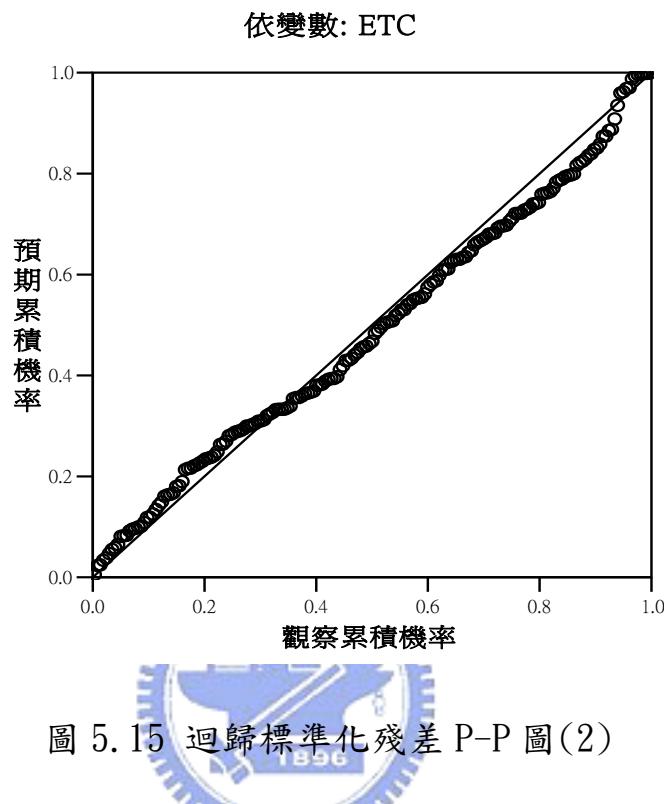


圖 5.15 迴歸標準化殘差 P-P 圖(2)

### Step 1. 強迫進入法進行迴歸模式之建立

► 報表分析結果如下：

表5.29 模式摘要(7)

模式	R	R 平方	調過後的 R 平方	估計的標準誤	變更統計量				
					R 平方改變量	F 改變	分子自由度	分母自由度	顯著性 F 改變
1	.839(a)	.704	.681	4.4743361588	.704	29.965	14	176	.000

- a 預測變數：(常數), X1, X2, X3, X4, X5, X6, X7, X8, X9, X10, X11, X12, X13, X14  
b 依變數：ETC

模式摘要表，有負相關係數R=0.839，R平方=0.704，調整過後的R平方=0.681，估計的標準誤4.474，R平方改變量=0.704，F值的改變=29.965，分子自由度=14，分母自由度=176，F值改變的顯著性=0.000

表5.30 變異數分析(7)

模式		平方和	自由度	平均平方和	F 檢定	顯著性
1	迴歸	8398.547	14	599.896	29.965	.000(a)
	殘差	3523.464	176	20.020		
	總和	11922.011	190			

a 預測變數：(常數), X1, X2, X3, X4, X5, X6, X7, X8, X9, X10, X11, X12, X13, X14

b 依變數：ETC

變異數分析摘要表，有SSR=8398.547, SSE=3523.464, SST=194.317, SSR+SSE=SST，F值等於29.965(F=MSR/MSE=599.896/20.02=29.965)，p=0.000<0.05，達顯著水準。

表5.31 緣數(7)

模式	未標準化係數		標準化係數 Beta	t	顯著性	迴歸係數 B 的 95% 信賴區間		相關			共線性統計量	
	B 之估計值	標準誤				下限	上限	零階	偏	部分	允差	VIF
1 (常數)	20.218	12.902		1.567	.119	-5.243	45.680					
X1	.006	.007	.063	.793	.429	-.009	.020	.313	.060	.032	.263	
X2	.016	.036	.029	.456	.649	-.055	.088	.213	.034	.019	.406	
X3	-.004	.012	-.020	-.353	.724	-.027	.019	-.030	-.027	-.014	.532	
X4	.011	.007	.135	1.649	.101	-.002	.024	.288	.123	.068	.252	
X5	-.013	.009	-.196	-1.450	.149	-.032	.005	.454	-.109	-.059	.092	
X6	.067	.027	.363	2.474	.014	.014	.121	.425	.183	.101	.078	
X7	-.055	.014	-.340	-3.961	.000	-.082	-.027	.143	-.286	-.162	.228	
X8	.029	.013	.182	2.238	.026	.003	.054	.452	.166	.092	.254	
X9	.021	.008	.291	2.834	.005	.006	.036	.681	.209	.116	.160	
X10	-.009	.013	-.072	-.730	.467	-.034	.016	.577	-.055	-.030	.172	
X11	.019	.007	.375	2.503	.013	.004	.033	.707	.185	.103	.075	
X12	-.038	.014	-.445	-2.768	.006	-.064	-.011	.633	-.204	-.113	.065	
X13	.087	.019	.581	4.517	.000	.049	.125	.686	.322	.185	.101	
X14	.004	.087	.003	.041	.967	-.168	.175	.504	.003	.002	.397	

a 依變數：ETC

迴歸分析的各係數值，常數項等於20.218，未標準化的迴歸係數X14=0.004，標準化的迴歸係數X14 Beta值=0.003，t值=0.041，p=0.967，未達到顯著水準。

Step 2. 刪除最不顯著的變數 X14，再作一次迴歸分析。

► 報表分析結果如下：

表5.32 模式摘要(8)

模式	R	R 平方	調過後的 R 平方	估計的標準誤	變更統計量				
					R 平方改變量	F 改變	分子自由度	分母自由度	
1	.839(a)	.704	.683	4.4617006497	.704	32.453	13	177	.000

a 預測變數：(常數), X1, X2, X3, X4, X5, X6, X7, X8, X9, X10, X11, X12, X13

b 依變數：ETC

模式摘要表，有負相關係數R=0.839，R平方=0.704，調整過後的R平方=0.683，估計的標準誤4.462，R平方改變量=0.704，F值的改變=32.453，分子自由度=13，分母自由度=177，F值改變的顯著性=0.000。

表5.33 變異數分析(8)

模式		平方和	自由度	平均平方和	F 檢定	顯著性
1	迴歸	8398.513	13	646.039	32.453	.000(a)
	殘差	3523.499	177	19.907		
	總和	11922.011	190			

a 預測變數：(常數), X1, X2, X3, X4, X5, X6, X7, X8, X9, X10, X11, X12, X13

b 依變數：ETC

變異數分析摘要表，有SSR=8398.513, SSE=3523.499, SST=11922.001, SSR + SSE = SST, F值等於32.453(F=MSR/MSE=646.039/19.907=32.453)，p=0.000<0.05，達顯著水準。

表5.34 紣數(8)

模式		未標準化係數		標準化係數	t	顯著性	迴歸係數 B 的 95% 信賴區間		相關			共線性統計量	
		B 之估計值	標準誤				下限	上限	零階	偏	部分	允差	VIF
1	(常數)	20.735	3.259	Beta 分配	6.363	.000	14.304	27.167					
	X1	.006	.007				.804	.423	-.008	.020	.313	.060	.033
													3.761

X2	.016	.036	.029	.456	.649	-.055	.088	.213	.034	.019	.406	2.462
X3	-.004	.011	-.020	-.361	.719	-.027	.018	-.030	-.027	-.015	.538	1.860
X4	.011	.006	.135	1.722	.087	-.002	.024	.288	.128	.070	.270	3.705
X5	-.013	.009	-.196	-1.454	.148	-.032	.005	.454	-.109	-.059	.092	10.834
X6	.067	.027	.363	2.500	.013	.014	.120	.425	.185	.102	.079	12.592
X7	-.055	.014	-.340	-3.972	.000	-.082	-.028	.143	-.286	-.162	.228	4.393
X8	.029	.013	.182	2.245	.026	.003	.054	.452	.166	.092	.254	3.932
X9	.021	.007	.292	3.017	.003	.007	.035	.681	.221	.123	.178	5.606
X10	-.009	.012	-.073	-.789	.431	-.033	.014	.577	-.059	-.032	.193	5.174
X11	.019	.007	.378	2.771	.006	.005	.032	.707	.204	.113	.090	11.113
X12	-.038	.013	-.447	-2.891	.004	-.063	-.012	.633	-.212	-.118	.070	14.300
X13	.087	.019	.582	4.589	.000	.050	.124	.686	.326	.188	.104	9.638

a 依變數：ETC

迴歸分析的各係數值，常數項等於20.735，未標準化的迴歸係數X3=-0.004，標準化的迴歸係數X3 Beta值=-0.020，t值=-0.361，p=0.719，未達到顯著水準。

### Step 3. 刪除最不顯著的變數 X3，再作一次迴歸分析。

► 報表分析結果如下：

表5.35 模式摘要(9)

模式	R	R 平方	調過後的 R 平方	估計的標準誤	變更統計量				
					R 平方改變量	F 改變	分子自由度	分母自由度	顯著性 F 改變
1	.839(a)	.704	.684	4.4507850877	.704	35.319	12	178	.000

a 預測變數：(常數), X1, X2, X4, X5, X6, X7, X8, X9, X10, X11, X12, X13

b 依變數：ETC

模式摘要表，有負相關係數R=0.839，R平方=0.704，調整過後的R平方=0.684，估計的標準誤4.451，R平方改變量=0.704，F值的改變=35.319，分子自由度=12，分母自由度=178，F值改變的顯著性=.000。

表5.36 變異數分析(9)

模式		平方和	自由度	平均平方和	F 檢定	顯著性
1	迴歸	8395.922	12	699.660	35.319	.000(a)
	殘差	3526.089	178	19.809		
	總和	11922.011	190			

a 預測變數：(常數), X1, X2, X4, X5, X6, X7, X8, X9, X10, X11, X12, X13

b 依變數：ETC

變異數分析摘要表，有SSR=8395.922,SSE=3526.089,SST=11922.001,SSR+SSE=SST,F值等於35.319(F=MSR/MSE=699.66/19.809=35.319)，p=0.000<0.05，達顯著水準。

表5.37 紳數(9)

模式		未標準化係數		標準化係數 Beta	t	顯著性	迴歸係數 B 的 95% 信賴區間		相關			共線性統計量	
		B 之估計值	標準誤				下限	上限	零階	偏	部分	允差	VIF
1	(常數)	20.362	3.083		6.605	.000	14.279	26.446					
	X1	.006	.007	.067	.847	.398	-.008	.020	.313	.063	.035	.269	3.721
	X2	.014	.035	.025	.397	.692	-.056	.084	.213	.030	.016	.421	2.376
	X4	.010	.005	.119	1.875	.062	-.001	.020	.288	.139	.076	.415	2.411
	X5	-.013	.009	-.182	-1.413	.159	-.030	.005	.454	-.105	-.058	.100	9.963
	X6	.067	.027	.361	2.499	.013	.014	.119	.425	.184	.102	.079	12.586
	X7	-.056	.014	-.345	-4.096	.000	-.082	-.029	.143	-.293	-.167	.234	4.275
	X8	.029	.013	.183	2.269	.024	.004	.054	.452	.168	.092	.255	3.924
	X9	.021	.007	.292	3.029	.003	.007	.035	.681	.221	.123	.178	5.605
	X10	-.009	.012	-.072	-.773	.441	-.033	.014	.577	-.058	-.032	.194	5.160
	X11	.019	.007	.375	2.765	.006	.005	.032	.707	.203	.113	.090	11.093
	X12	-.038	.013	-.449	-2.916	.004	-.064	-.012	.633	-.213	-.119	.070	14.276
	X13	.087	.019	.584	4.618	.000	.050	.124	.686	.327	.188	.104	9.623

a 依變數：ETC

迴歸分析的各係數值，常數項等於20.362，未標準化的迴歸係數X2=0.014，標準化的迴歸係數X2 Beta值=0.025，t值=0.397，p=0.692，未達到顯著水準。

Step 4. 刪除最不顯著的變數 X2，再作一次迴歸分析。

► 報表分析結果如下：

表5.38 模式摘要(10)

模式	R	R 平方	調過後的 R 平方	估計的標準誤	變更統計量				
					R 平方改變量	F 改變	分子自由度	分母自由度	顯著性 F 改變
1	.839(a)	.704	.686	4.4402991073	.704	38.698	11	179	.000

a 預測變數：(常數), X1, X4, X5, X6, X7, X8, X9, X10, X11, X12, X13

b 依變數：ETC

模式摘要表，有負相關係數R=0.839，R平方=0.704，調整過後的R平方=0.686，估計的標準誤4.44，R平方改變量=0.704，F值的改變=38.698，分子自由度=11，分母自由度=179，F值改變的顯著性=0.000。

表5.39 變異數分析(10)

模式		平方和	自由度	平均平方和	F 檢定	顯著性
1	迴歸	8392.801	11	762.982	38.698	.000(a)
	殘差	3529.210	179	19.716		
	總和	11922.011	190			

a 預測變數：(常數), X1, X4, X5, X6, X7, X8, X9, X10, X11, X12, X13

b 依變數：ETC

變異數分析摘要表，有SSR=8392.801, SSE=3529.210, SST=11922.001, SSR+SSE=SST, F值等於38.698(F=MSR/MSE=762.982/19.716=38.698)，p=0.000<0.05，達顯著水準。

表5.40 緣數(10)

模式		未標準化係數		標準化係數 Beta	t	顯著性	迴歸係數 B 的 95% 信賴區間		相關			共線性統計量	
		B 之估計值	標準誤				下限	上限	零階	偏	部分	允差	VIF
1	(常數)	20.978	2.658		7.891	.000	15.732	26.223					
	X1	.008	.006	.085	1.364	.174	-.003	.019	.313	.101	.055	.422	2.372
	X4	.009	.005	.116	1.845	.067	-.001	.020	.288	.137	.075	.421	2.376
	X5	-.013	.009	-.182	-1.414	.159	-.030	.005	.454	-.105	-.058	.100	9.963
	X6	.068	.026	.368	2.568	.011	.016	.120	.425	.189	.104	.081	12.418
	X7	-.056	.014	-.347	-4.127	.000	-.083	-.029	.143	-.295	-.168	.234	4.267
	X8	.029	.013	.180	2.248	.026	.003	.054	.452	.166	.091	.257	3.893
	X9	.022	.007	.299	3.144	.002	.008	.036	.681	.229	.128	.183	5.455
	X10	-.010	.012	-.075	-.819	.414	-.033	.014	.577	-.061	-.033	.196	5.107
	X11	.019	.007	.375	2.766	.006	.005	.032	.707	.202	.112	.090	11.090
	X12	-.038	.013	-.452	-2.948	.004	-.064	-.013	.633	-.215	-.120	.070	14.236
	X13	.087	.019	.583	4.620	.000	.050	.124	.686	.326	.188	.104	9.617

a 依變數：ETC

迴歸分析的各係數值，常數項等於20.978，未標準化的迴歸係數X10=-0.010，標準化的迴歸係數X10 Beta值=-0.075，t值=-0.819，p=0.414，未達到顯著水準。

Step 5. 刪除最不顯著的變數 X10，再作一次迴歸分析。

► 報表分析結果如下：

表5.41 模式摘要(11)

模式	R	R 平方	調過後的 R 平方	估計的標準誤	變更統計量				
					R 平方改變量	F 改變	分子自由度	分母自由度	
1	.838(a)	.703	.686	4.4362369765	.703	42.579	10	180	.000

a 預測變數：(常數), X1, X4, X5, X6, X7, X8, X9, X11, X12, X13

b 依變數：ETC

模式摘要表，有負相關係數R=0.838，R平方=0.703，調整過後的R平方=0.686，估計的標準誤4.436，R平方改變量=0.704，F值的改變=42.579，分子自由度=10，分母自由度=180，F值改變的顯著性=0.000。

表5.42 變異數分析(11)

模式		平方和	自由度	平均平方和	F 檢定	顯著性
1	迴歸	8379.576	10	837.958	42.579	.000(a)
	殘差	3542.436	180	19.680		
	總和	11922.011	190			

a 預測變數：(常數), X1, X4, X5, X6, X7, X8, X9, X11, X12, X13

b 依變數：ETC

變異數分析摘要表，有SSR=8379.576, SSE=3542.436, SST=11922.001, SSR+ SSE = SST，F值等於42.579(F=MSR/MSE=837.958/19.680=42.579)，p=0.000<0.05，達顯著水準。

表5.43 紳數(11)

模式		未標準化紳數		標準化紳數 Beta	t	顯著性	迴歸紳數 B 的 95% 信賴區間		相關			共線性統計量	
		B 之估計值	標準誤				下限	上限	零階	偏	部分	允差	VIF
1	(常數)	20.310	2.528		8.035	.000	15.322	25.297					
	X1	.009	.005	.101	1.702	.090	-.001	.020	.313	.126	.069	.466	2.145
	X4	.010	.005	.120	1.933	.055	.000	.020	.288	.143	.079	.425	2.354
	X5	-.013	.009	-.192	-1.501	.135	-.031	.004	.454	-.111	-.061	.101	9.871
	X6	.070	.026	.380	2.668	.008	.018	.122	.425	.195	.108	.081	12.290
	X7	-.058	.013	-.361	-4.409	.000	-.084	-.032	.143	-.312	-.179	.246	4.071

X8	.032	.012	.200	2.610	.010	.008	.056	.452	.191	.106	.282	3.551
X9	.018	.005	.241	3.748	.000	.008	.027	.681	.269	.152	.398	2.514
X11	.017	.006	.333	2.654	.009	.004	.029	.707	.194	.108	.105	9.556
X12	-.037	.013	-.438	-2.875	.005	-.062	-.012	.633	-.210	-.117	.071	14.046
X13	.087	.019	.584	4.633	.000	.050	.124	.686	.326	.188	.104	9.616

a 依變數：ETC

迴歸分析的各係數值，常數項等於20.310，未標準化的迴歸係數X5=-0.013，標準化的迴歸係數X5 Beta值=-0.192，t值=-1.501，p=0.135，未達到顯著水準。

### Step 6. 刪除最不顯著的變數 X5，再作一次迴歸分析。

► 報表分析結果如下：

表5.44 模式摘要(12)

模式	R	R 平方	調過後的 R 平方	估計的標準誤	R 平方改變量	變更統計量			
						F 改變	分子自由度	分母自由度	顯著性 F 改變
1	.836(a)	.699	.684	4.4515595848	.699	46.736	9	181	.000

a 預測變數：(常數), X1, X4, X6, X7, X8, X9, X11, X12, X13

b 依變數：ETC

模式摘要表，有負相關係數R=0.836，R平方=0.699，調整過後的R平方=0.684，估計的標準誤4.452，R平方改變量=0.699，F值的改變=46.736，分子自由度=9，分母自由度=181，F值改變的顯著性=0.000。

表5.45 變異數分析(12)

模式		平方和	自由度	平均平方和	F 檢定	顯著性
1	迴歸	8335.246	9	926.138	46.736	.000(a)
	殘差	3586.765	181	19.816		
	總和	11922.011	190			

a 預測變數：(常數), X1, X4, X6, X7, X8, X9, X11, X12, X13

b 依變數：ETC

變異數分析摘要表，有SSR=8335.246，SSE=3586.765，SST=11922.001，SSR+SSE=SST，F值等於46.736(F=MSR/MSE=926.138/19.816=46.736)，p=0.000<0.05，達顯著水準。

表5.46 係數(12)

模式		未標準化係數		標準化係數 Beta	t	顯著性	迴歸係數 B 的 95% 信賴區間		相關			共線性統計量	
		B 之估計值	標準誤				下限	上限	零階	偏	部分	允差	VIF
1	(常數)	20.962	2.499		8.389	.000	16.031	25.892					
	X1	.009	.005	.094	1.572	.118	-.002	.019	.313	.116	.064	.470	2.129
	X4	.006	.004	.075	1.368	.173	-.003	.015	.288	.101	.056	.560	1.786
	X6	.037	.014	.198	2.637	.009	.009	.064	.425	.192	.107	.294	3.397
	X7	-.045	.010	-.282	-4.499	.000	-.065	-.025	.143	-.317	-.183	.424	2.357
	X8	.021	.010	.134	2.128	.035	.002	.041	.452	.156	.087	.418	2.394
	X9	.020	.005	.269	4.356	.000	.011	.029	.681	.308	.178	.434	2.303
	X11	.016	.006	.321	2.552	.012	.004	.028	.707	.186	.104	.105	9.514
	X12	-.034	.013	-.401	-2.660	.009	-.059	-.009	.633	-.194	-.108	.073	13.686
	X13	.084	.019	.565	4.490	.000	.047	.121	.686	.317	.183	.105	9.521

a 依變數：ETC

迴歸分析的各係數值，常數項等於20.962，未標準化的迴歸係數X4=-0.006，標準化的迴歸係數X4 Beta值=0.075，t值=1.368，p=0.173，未達到顯著水準。



Step 7. 刪除最不顯著的變數 X4，再作一次迴歸分析。

► 報表分析結果如下：

表5.47 模式摘要(13)

模式	R	R 平方	調過後的 R 平方	估計的標準誤	R 平方 改變量	F 改變	變更統計量			
							分子 自由 度	分母 自由 度	顯著性 F 改變	
1	.834(a)	.696	.683	4.4621894278	.696	52.095	8	182	.000	

a 預測變數：(常數), X1, X6, X7, X8, X9, X11, X12, X13

b 依變數：ETC

模式摘要表，有負相關係數R=0.834，R平方=0.696，調整過後的R平方=0.683，估計的標準誤4.462，R平方改變量=0.696，F值的改變=52.095，分子自由度=8，分母自由度=182，F值改變的顯著性=0.000。

表5.48 變異數分析(13)

模式		平方和	自由度	平均平方和	F 檢定	顯著性
1	迴歸	8298.185	8	1037.273	52.095	.000(a)
	殘差	3623.826	182	19.911		
	總和	11922.011	190			

a 預測變數：(常數), X1, X6, X7, X8, X9, X11, X12, X13

b 依變數：ETC

變異數分析摘要表，有SSR=8298.185, SSE=3623.826, SST=11922.001, SSR+SSE=SST, F值等於52.095(F=MSR/MSE=1037.273/19.911=52.095)，p=0.000<0.05，達顯著水準。

表5.49 係數(13)

模式	未標準化係數		標準化係數 Beta	t	顯著性	迴歸係數 B 的 95% 信賴區間		相關			共線性統計量	
	B 之估計值	標準誤				下限	上限	零階	偏	部分	允差	VIF
1 (常數)	21.736	2.440		8.910	.000	16.923	26.550					
X1	.009	.005	.097	1.629	.105	-.002	.020	.313	.120	.067	.471	
X6	.046	.012	.248	3.754	.000	.022	.070	.425	.268	.153	.384	
X7	-.046	.010	-.285	-4.545	.000	-.066	-.026	.143	-.319	-.186	.425	
X8	.023	.010	.147	2.350	.020	.004	.043	.452	.172	.096	.427	
X9	.018	.004	.248	4.136	.000	.010	.027	.681	.293	.169	.463	
X11	.016	.006	.323	2.561	.011	.004	.028	.707	.187	.105	.105	
X12	-.034	.013	-.408	-2.698	.008	-.060	-.009	.633	-.196	-.110	.073	
X13	.086	.019	.577	4.583	.000	.049	.123	.686	.322	.187	.106	

a 依變數：ETC

迴歸分析的各係數值，常數項等於21.736，未標準化的迴歸係數X1=0.009，標準化的迴歸係數X1 Beta值=0.097，t值=1.629，p=0.105，未達到顯著水準。

Step 8. 刪除最不顯著的變數 X1，再作一次迴歸分析。

► 報表分析結果如下：

表5.50 模式摘要(14)

模式	R	R 平方	調過後的 R 平方	估計的標準誤	變更統計量				
					R 平方改變量	F 改變	分子自由度	分母自由度	顯著性 F 改變
1	.832(a)	.692	.680	4.4822977165	.692	58.629	7	183	.000

a 預測變數：(常數), X6, X7, X8, X9, X11, X12, X13

b 依變數：ETC

模式摘要表，有負相關係數R=0.832，R平方=0.692，調整過後的R平方=0.680，估計的標準誤4.482，R平方改變量=0.692，F值的改變=58.629，分子自由度=7，分母自由度=183，F值改變的顯著性=.000。

表5.51 變異數分析(14)

模式		平方和	自由度	平均平方和	F 檢定	顯著性
1	迴歸	8245.360	7	1177.909	58.629	.000(a)
	殘差	3676.652	183	20.091		
	總和	11922.011	190			

a 預測變數：(常數), X6, X7, X8, X9, X11, X12, X13

b 依變數：ETC

變異數分析摘要表，有SSR=8245.360，SSE=3676.652，SST=11922.001， $SSR + SSE = SST$ ，F值等於58.629( $F = MSR/MSE = 1177.909/20.091 = 58.629$ )， $p=0.000 < 0.05$ ，達顯著水準。

表5.52 係數(14)

模式	未標準化係數 B 之估計值	標準誤 標準誤	標準化係數 Beta	t	顯著性	迴歸係數 B 的 95% 信賴區間		相關			共線性統計量	
						下限	上限	零階	偏	部分	允差	VIF
1 (常數)	24.079	1.979		12.165	.000	20.174	27.984					
X6	.052	.012	.283	4.510	.000	.029	.075	.425	.316	.185	.429	2.331
X7	-.048	.010	-.301	-4.832	.000	-.068	-.029	.143	-.336	-.198	.435	2.298

X8	.027	.010	.171	2.808	.006	.008	.046	.452	.203	.115	.453	2.208
X9	.017	.004	.227	3.853	.000	.008	.025	.681	.274	.158	.487	2.052
X11	.021	.005	.430	3.980	.000	.011	.032	.707	.282	.163	.144	6.924
X12	-.039	.013	-.462	-3.118	.002	-.064	-.014	.633	-.225	-.128	.077	13.013
X13	.081	.019	.541	4.347	.000	.044	.117	.686	.306	.178	.109	9.193

a 依變數：ETC

迴歸分析的各係數值，常數項等於24.079，未標準化的迴歸係數與標準化的迴歸係數皆已達到顯著水準。

## 5.4.2 迴歸模式彙整

以下將迴歸檢定重要參考數值會整如表 5.5，根據上表可發現，當係數皆達到顯著水準，代表各係數能有效解釋 Y 的變動。

表 5.53 迴歸檢定彙整表

模式別	調整後 R <sup>2</sup>	預測變數	F 值	顯著性
模式 1	0.502	X2, X3, X6, X7, X8, X9, X10, X11, X13	19.491	0.000
模式 2	0.680	X6, X7, X8, X9, X11, X12, X13	58.629	0.000

模式 1 迴歸方程式為 5-1 所示：

$$Y=20.772 - 0.053 X_2 + 0.027 X_3 - 0.094 X_6 + 0.065 X_7 - 0.032 X_8 \quad (5-1) \\ + 0.017 X_9 + 0.138 X_{10} - 0.013 X_{11} + 0.036 X_{13}$$

模式 2 迴歸方程式為 5-1 所示：

$$Y=24.079 + 0.052 X_6 - 0.048 X_7 + 0.065 X_7 + 0.027 X_8 + 0.017 X_9 \quad (5-2) \\ + 0.021 X_{11} - 0.039 X_{12} + 0.081 X_{13}$$

### (一) 共線性問題

當迴歸模式中預測變數之間有太高的相關時，就會產生共線性的問題，造成迴歸係數與相關係數正負符號不同的問題發生，本研究暫且先忽略共線性問題的存在，主要檢測迴歸模式預測的能力是否準確。

## (二) 調整後 $R^2$ 方面

判斷係數  $R^2$  只在判斷依變數與自變數關係是否密切，即此迴歸式解釋能力之強弱。故可以得知在模式 1 之解釋能力達 0.502，模式 2 之解釋能力達 0.680。

## (三) F 檢定

F 檢定是以整體為主，表示此迴歸式中整體係數不全為零，換言之，在 F 檢定有顯著之前提下，再作個別係數的 t 檢定才有意義，意旨在於各個自變數與依變數是否有顯著關係，再經各個模式的 F 值整體考驗結果，各種模式的 F 值及顯著性，在迴歸效果中達到顯著水準，表示這模式皆具有統計上的意義，模式 1 的 F 值為 19.491,  $F > F_{0.05}(9,156) = 1.94$ ，表示檢定迴歸模式 A 的整體係數不全為零，模式 2 的 F 值為 58.629,  $F > F_{0.05}(7,183) = 2.06$ 。

## (四) t 檢定

模式一各項變數的 t 值若大於  $t_{0.05}(156) = 1.655$  或小於  $-t_{0.05}(156) = -1.655$ ，皆顯示達到顯著水準，因此從表 5.5 係數(6)中可得知所有變數及常數項皆達到顯著水準，表示各自變數  $X_i$  與應變數 Y 有顯著關係。

模式二各項變數的 t 值若大於  $t_{0.05}(183) = 1.653$  或小於  $-t_{0.05}(183) = -1.653$ ，皆顯示達到顯著水準，從表 5.5 係數(6)中可得知所有變數及常數項皆達到顯著水準，表示各自變數  $X_i$  與應變數 Y 有顯著關係。

## 5.5 TVC 預測結果

### (一) 2008 年三月份各禮拜四 上午 8:00~9:00 之 TVC 預測結果。

表 5.5 為 TVC 預測結果，圖 5.8 為 TVC 預測結果，其整體預測 MAPE 值約為 6.6%，最大一筆誤差在 3/28 早上 9:00 約為 16%。

表 5.54 TVC 預測結果(1)

時間	ETC (分鐘)	TVC (分鐘)	絕對誤差 (%)
2008/3/6 AM 08:00:00	25.8389	25.7209	0.4566
2008/3/6 AM 08:05:00	27.1378	26.1743	3.5504
2008/3/6 AM 08:10:00	27.3242	26.4343	3.2568
2008/3/6 AM 08:15:00	27.3236	26.3037	3.7325
2008/3/6 AM 08:20:00	26.9896	26.7571	0.8613
2008/3/6 AM 08:25:00	26.3046	27.0070	2.6703
2008/3/6 AM 08:30:00	25.7704	27.0070	4.7988
2008/3/6 AM 08:35:00	25.2730	27.0070	6.8611
2008/3/6 AM 08:40:00	25.9573	27.0070	4.0441
2008/3/6 AM 08:45:00	25.5788	27.0070	5.5837
2008/3/6 AM 08:50:00	24.9000	26.6039	6.8430
2008/3/6 AM 08:55:00	24.9722	27.2050	8.9411
2008/3/6 AM 09:00:00	25.1767	26.9261	6.9485
2008/3/13 AM 08:00:00	25.8400	26.7385	3.4773
2008/3/13 AM 08:05:00	25.0308	26.7385	6.8226
2008/3/13 AM 08:10:00	25.2381	26.7385	5.9451
2008/3/13 AM 08:15:00	25.5150	26.7385	4.7953
2008/3/13 AM 08:20:00	25.2667	26.7385	5.8253
2008/3/13 AM 08:25:00	25.2056	26.7385	6.0819
2008/3/13 AM 08:30:00	25.3262	26.7385	5.5766
2008/3/13 AM 08:35:00	24.7145	26.7385	8.1899
2008/3/13 AM 08:40:00	26.1427	26.7385	2.2791
2008/3/13 AM 08:45:00	26.1928	26.7385	2.0833
2008/3/13 AM 08:50:00	25.3769	29.0426	14.4452
2008/3/13 AM 08:55:00	24.9694	29.0426	16.3128
2008/3/13 AM 09:00:00	25.8257	28.5143	10.4105
2008/3/20 AM 08:00:00	24.7151	25.8899	4.7534
2008/3/20 AM 08:05:00	24.5273	26.2725	7.1152
2008/3/20 AM 08:10:00	23.7100	26.2790	10.8351
:	:	:	:
2008/3/28 AM 08:50:00	28.8152	26.7385	7.2067
2008/3/28 AM 08:55:00	28.4242	26.7385	5.9306
2008/3/28 AM 09:00:00	31.25	26.1076	16.4558
平均			6.5972

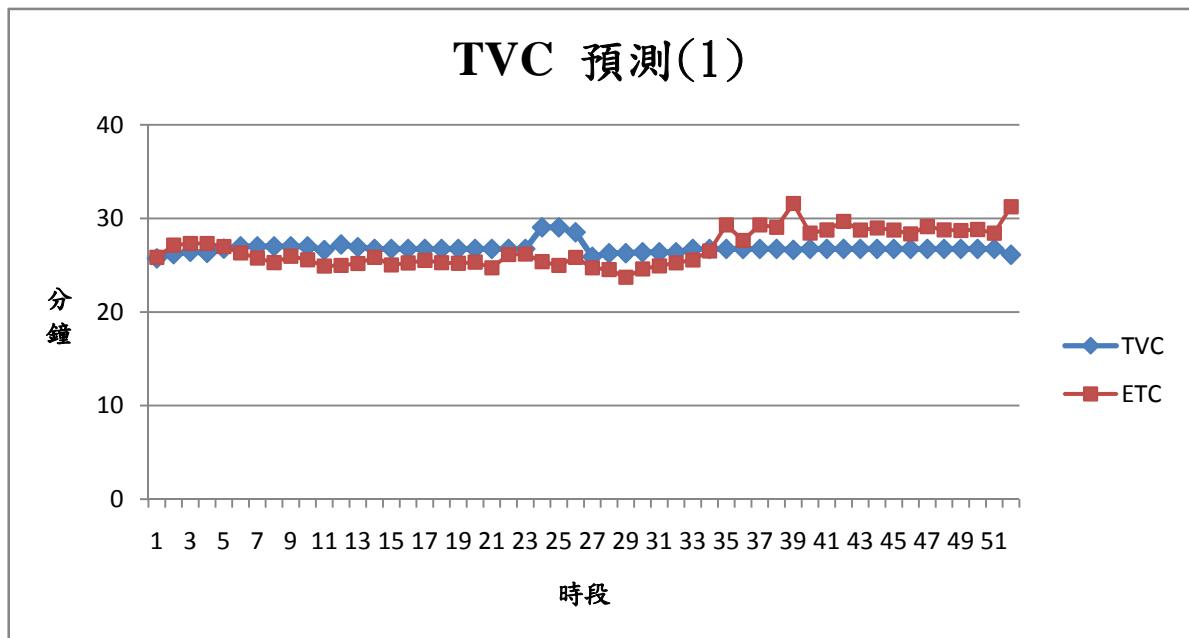


圖 5.16 TVC 預測結果(1)

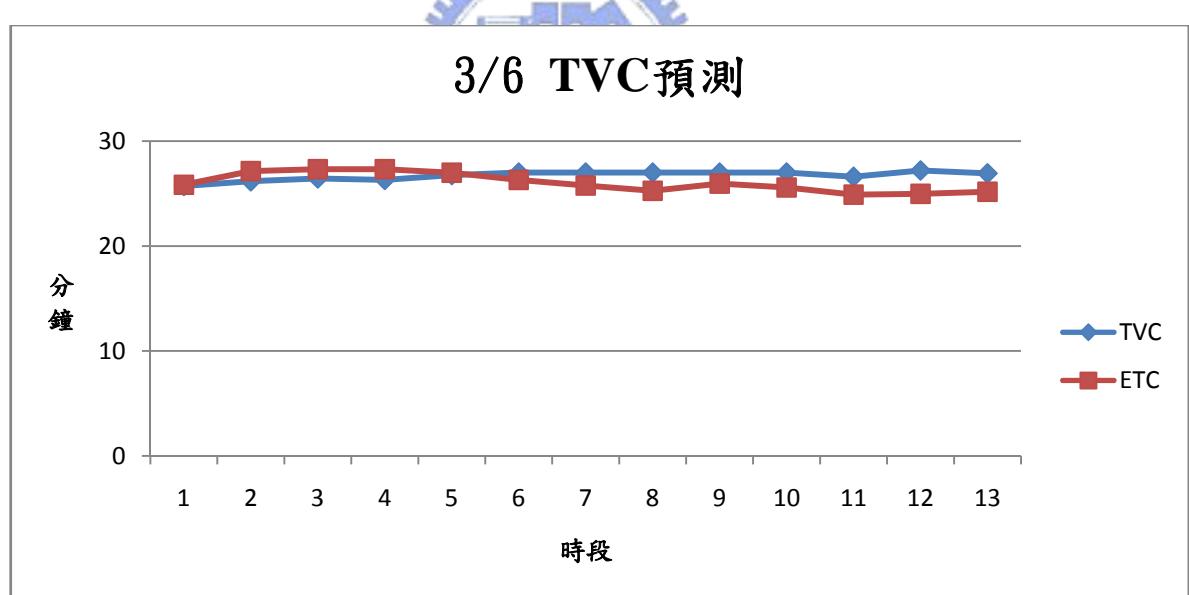


圖 5.17 3/6 TVC 預測結果

### 3/13 TVC測試

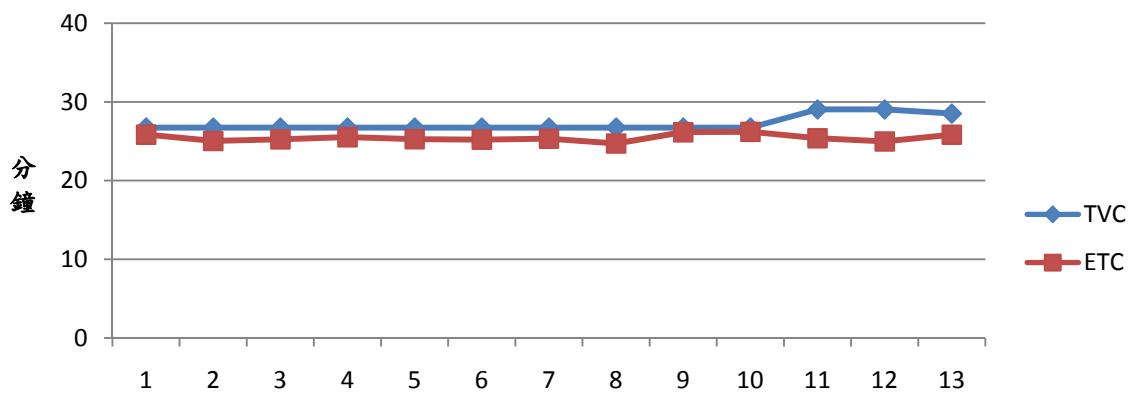


圖 5.18 3/13 TVC 預測結果



### 3/20 TVC預測

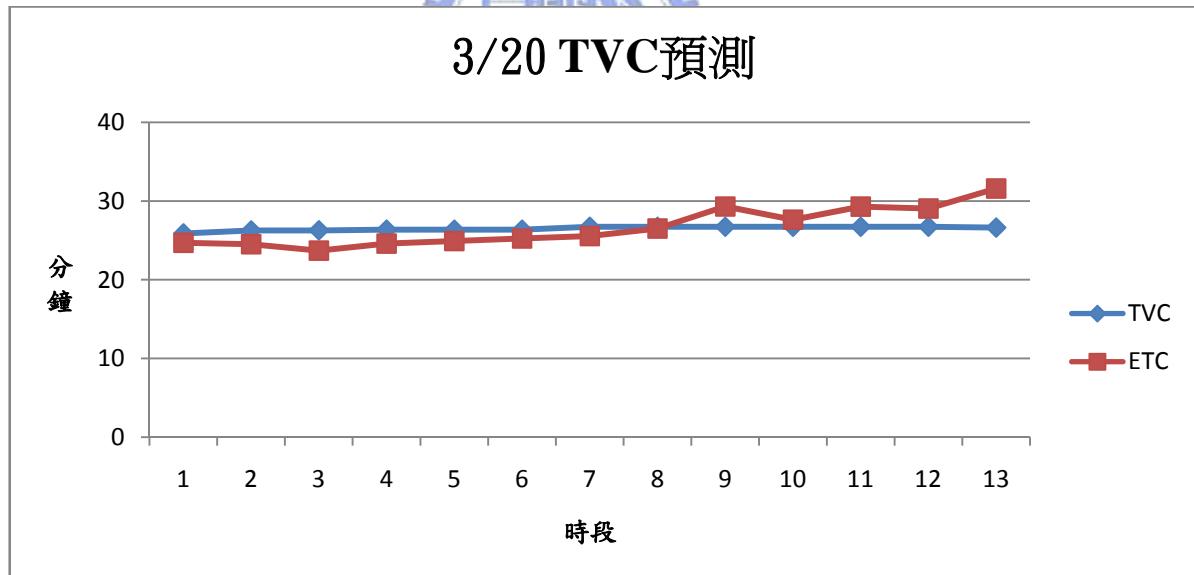


圖 5.19 3/20 TVC 預測結果

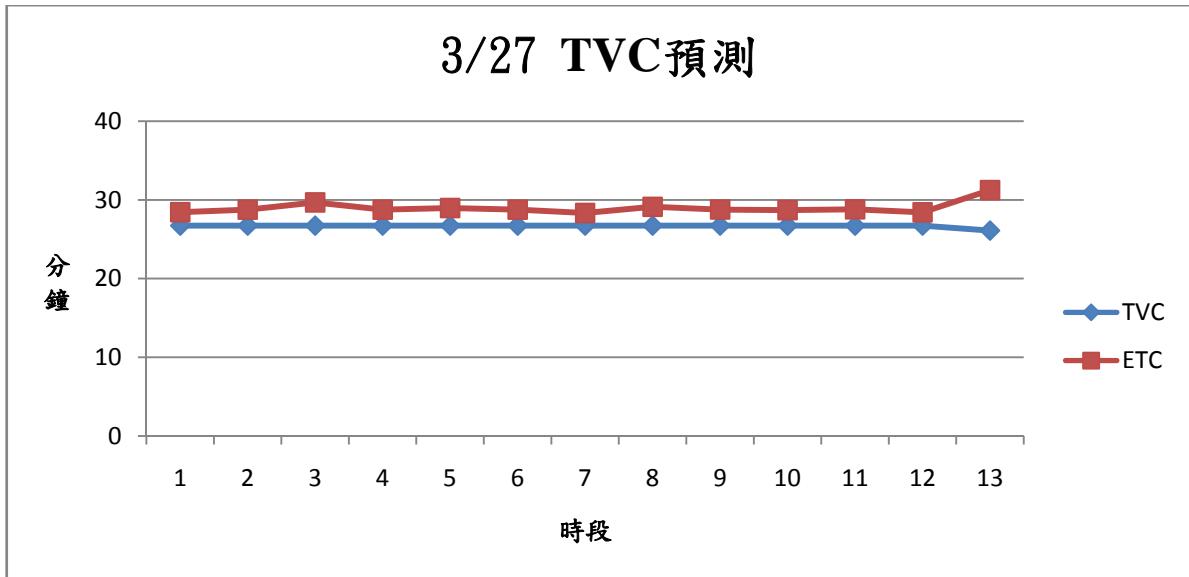


圖 5.20 3/27 TVC 預測結果

(二) 2008 年三月份各禮拜六 上午 10:00~11:00 之 TVC 預測結果。

表 5.5 為 TVC 預測結果，圖 5.8 為 TVC 預測結果，其整體預測 MAPE 值約為 5.96%，最大一筆誤差在 3/8 早上 10:00 約為 25%，圖 5.21 為 TVC 預測的結果，圖 5.22 至圖 5.27 為預測 3 月份各禮拜六的結果。

表 5.55 TVC 預測結果(2)

時間	ETC (分鐘)	TVC (分鐘)	絕對誤差值 (%)
2008/3/1 AM 10:00:00	43.5396	44.7620	2.8077
2008/3/1 AM 10:05:00	46.3500	49.7828	7.4063
2008/3/1 AM 10:10:00	48.2021	49.7828	3.2794
2008/3/1 AM 10:15:00	47.3067	49.7828	5.2342
2008/3/1 AM 10:20:00	47.2083	47.4540	0.5204
2008/3/1 AM 10:25:00	46.4322	49.1078	5.7622
2008/3/1 AM 10:30:00	46.0625	45.0812	2.1303
2008/3/1 AM 10:35:00	47.1667	53.3984	13.2122
2008/3/1 AM 10:40:00	46.4208	45.4943	1.9960
2008/3/1 AM 10:45:00	46.8111	46.1557	1.4000
2008/3/1 AM 10:50:00	46.7864	47.4540	1.4270
2008/3/1 AM 10:55:00	48.3056	49.1078	1.6607

2008/3/1 AM 11:00:00	50.3500	47.1771	6.3016
2008/3/8 AM 10:00:00	44.3532	55.5487	25.2418
2008/3/8 AM 10:05:00	45.5524	56.3681	23.7434
2008/3/8 AM 10:10:00	51.3067	55.6726	8.5095
2008/3/8 AM 10:15:00	48.0860	56.4530	17.4001
2008/3/8 AM 10:20:00	49.7611	55.3878	11.3075
2008/3/8 AM 10:25:00	50.0942	54.9435	9.6803
2008/3/8 AM 10:30:00	52.0571	54.9348	5.5278
2008/3/8 AM 10:35:00	52.2300	54.9768	5.2590
2008/3/8 AM 10:40:00	52.4258	55.7470	6.3352
2008/3/8 AM 10:45:00	51.7773	55.3536	6.9072
2008/3/8 AM 10:50:00	52.3603	54.8857	4.8232
2008/3/8 AM 10:55:00	52.5115	54.8857	4.5215
2008/3/8 AM 11:00:00	52.1565	56.3667	8.0722
2008/3/15 AM 10:00:00	42.4300	46.9181	10.5776
2008/3/15 AM 10:05:00	43.3651	47.5726	9.7027
2008/3/15 AM 10:10:00	44.3302	47.0651	6.1696
:	:	:	:
2008/3/29 AM 10:50:00	48.1933	53.8200	11.6751
2008/3/29 AM 10:55:00	50.3010	54.7305	8.8060
2008/3/29 AM 11:00:00	50.7292	54.7305	7.8876
平均			3.6817

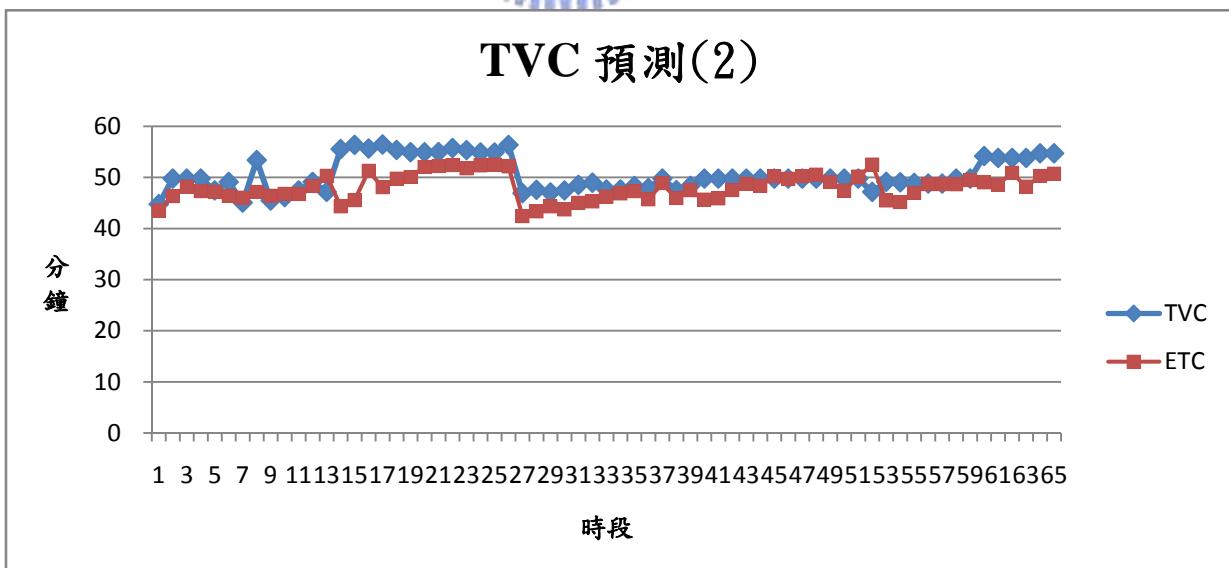


圖 5.21 TVC 預測結果(2)

3/1 TVC預測

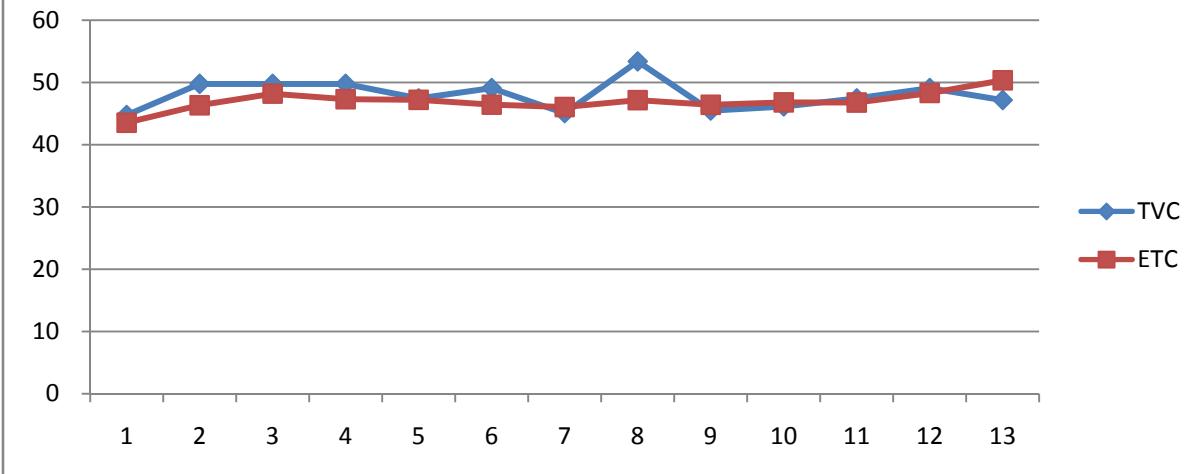


圖 5.22 3/1 TVC 預測結果

3/8 TVC預測

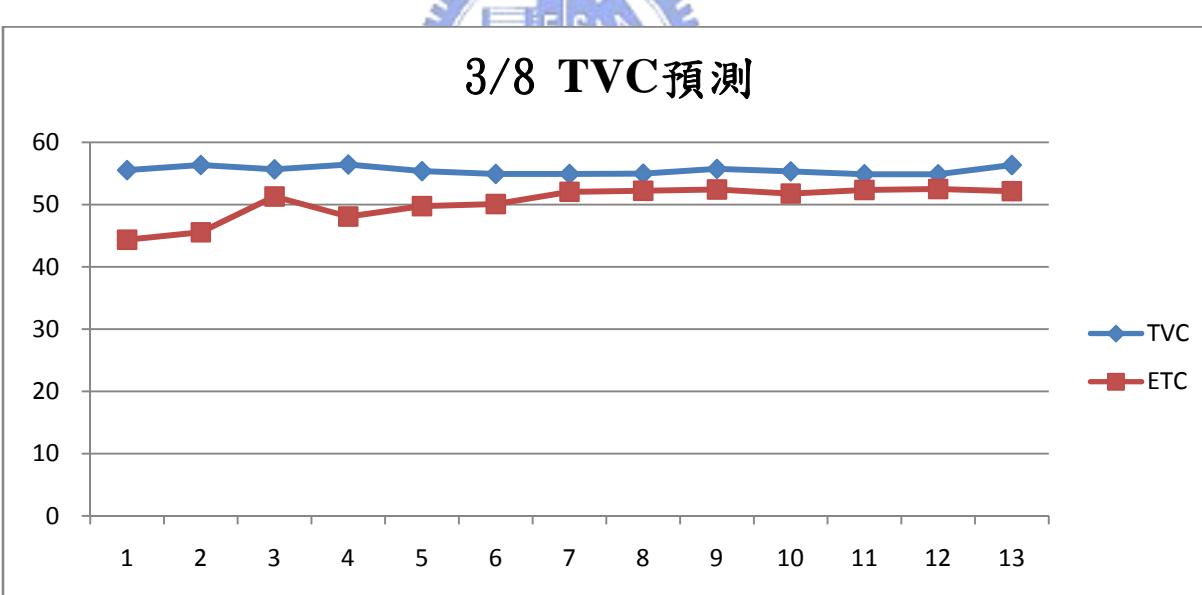


圖 5.23 3/8 TVC 預測結果

3/15 TVC預測

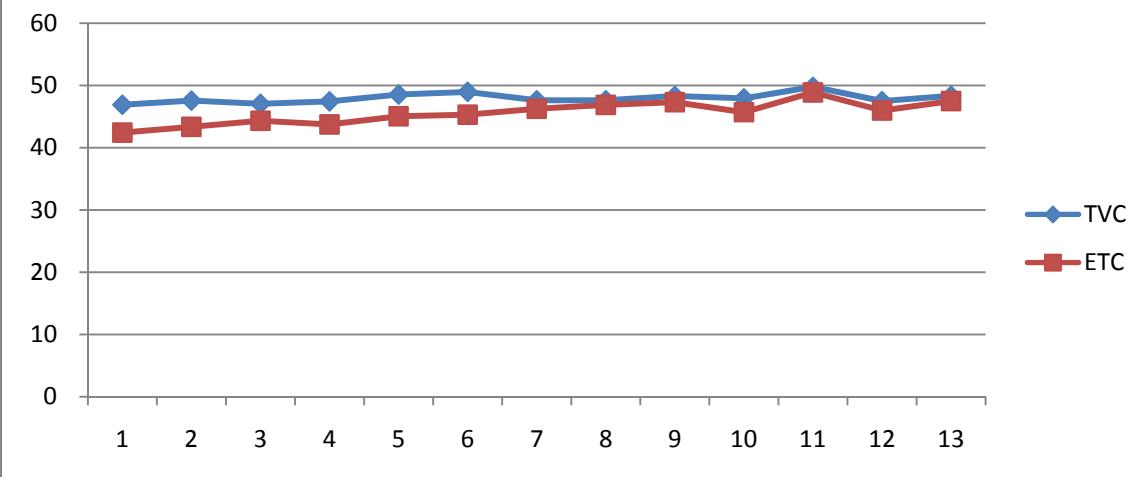


圖 5.24 3/15 TVC 預測結果



3/22 TVC 預測

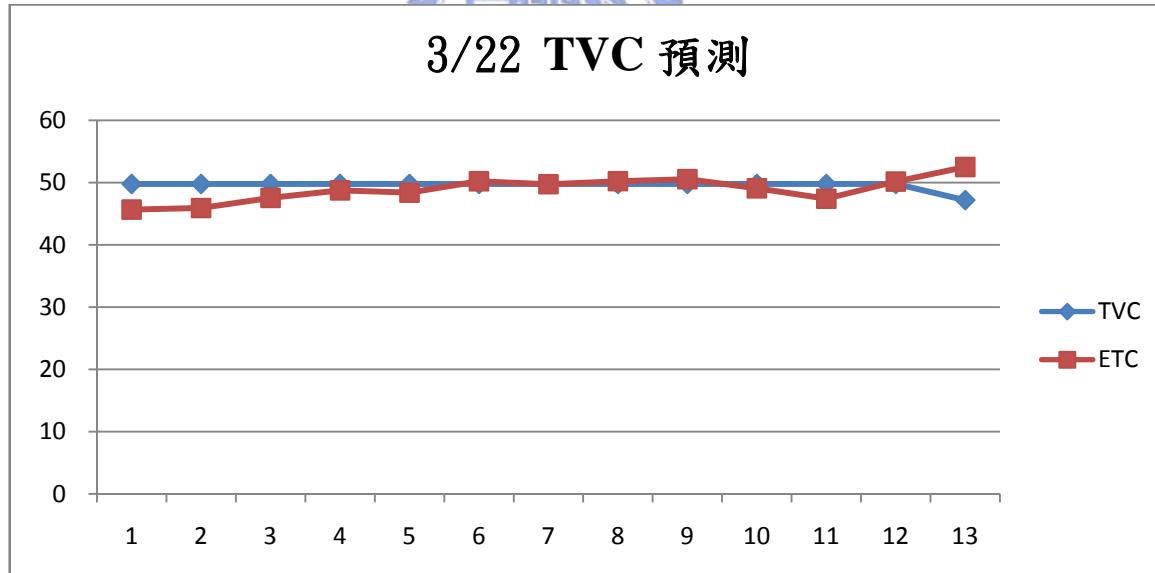


圖 3/22 TVC 預測結果

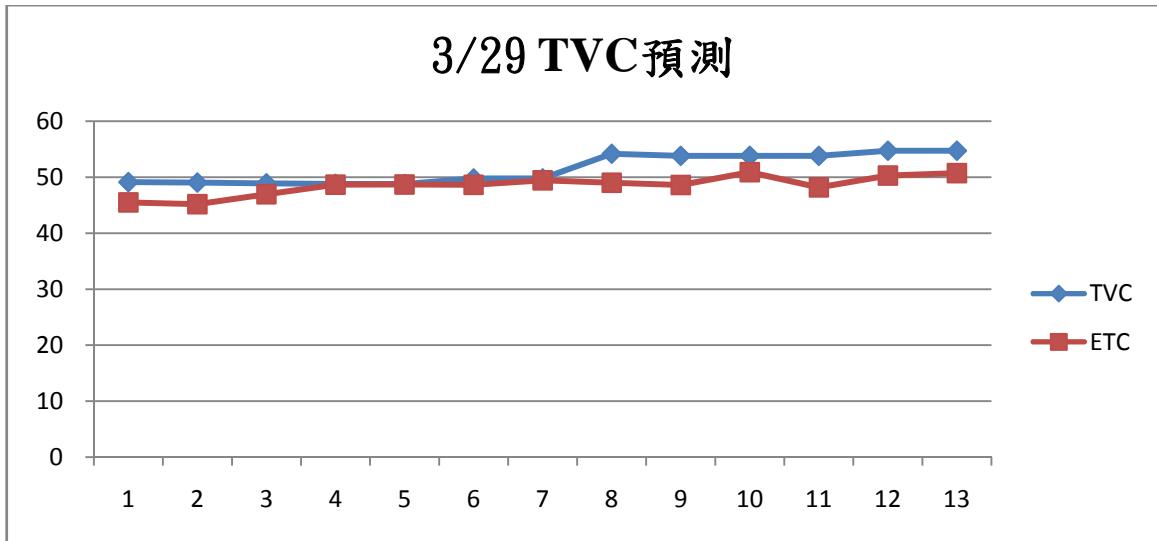


圖 3/29 TVC 預測結果

## 5.6 績效評估

### (一) k-NN 與 TVC 模式預測結果比較(1)



由以上的預測結果可得知，k-NN 的預測 MAPE 值為 3.4%，最大一筆誤差值為 9%，TVC 的預測 MAPE 值為 6.6%，最大一筆誤差值為 16%，利用統計檢定來比較模式預測誤差是否具有顯著差異。

$$u_1=3.4\% \text{ (k-NN MAPE)}; \quad u_2=6.6\% \text{ (TVC MAPE)}$$

$$n_1=52, \quad n_2=52, \quad \bar{X}=0.033, \quad \bar{Y}=0.066, \quad \sigma_1^2 = 0.00589, \quad \sigma_2^2=0.0012$$

$$(1) H_0: u_1 \geq u_2$$

$$H_1: u_1 < u_2$$

$$(2) \alpha=0.05$$

$$(3) C = \{Z \mid Z < -z_\alpha = -1.645\}$$

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{0.033 - 0.066}{\sqrt{\frac{0.00589}{52} + \frac{0.00124}{52}}} = -2.82$$

(4) 因  $Z=-2.82 \in C$ ，故知拒絕  $H_0$ ，亦及  $u_1$  顯著小於  $u_2$ 。

## (二) k-NN 與 TVC 模式預測結果比較(2)

$u_1=3.68\%$  (k-NN MAPE);  $u_2=5.96\%$  (TVC MAPE)

$n_1=65$ ,  $n_2=65$ ,  $\bar{X}=0.0368$ ,  $\bar{Y}=0.0596$ ,  $\sigma_1^2=0.0007$ ,  $\sigma_2^2=0.0025$

(1)  $H_0: u_1 \geq u_2$

$H_1: u_1 < u_2$

(2)  $\alpha=0.05$

(3)  $C = \{Z \mid Z < -z_\alpha = -1.645\}$

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{0.0368 - 0.0596}{\sqrt{\frac{0.0007}{65} + \frac{0.0025}{65}}} = -3.257$$

(4) 因  $Z=-3.257 \in C$ , 故知拒絕  $H_0$ , 亦及  $u_1$  顯著小於  $u_2$ 。



## 第六章 結論與建議

本研究主要利用 k-NN 模式與 TVC 模式預測高速公路旅行時間，並比較兩模式的預測績效，經過實例測試後，證實兩模式皆能夠準確作預測，且能實際運用在高速公路上，而比較兩模式的預測結果，發現 k-NN 模式的預測效果又比 TVC 模式較佳，可提供用路人更精確的預測結果。

### 6.1 結論

本研究之結論如下：

1. 由於本研究只利用探針車旅行時間資料作預測，即能得到精準的預測結果。因此在高速公路路網中，只要能完整蒐集探針車的旅行時間資料，即可利用本研究所構建的模式預測旅行時間。
2. 本研究利用 k-NN 與 TVC 兩種模式，分別實際預測在高速公路上尖峰時段的旅行時間，此時段有別於一般離峰時段有較穩定的旅行時間，是屬於較難預測的部份，而預測結果在 Lewis 對 MAPE 的分類上，k-NN 的 MAPE 為 3.4%、TVC 為 6.6%，兩者都屬高精準預測，顯示兩模式對旅行時間預測皆有不錯的效果，經由統計檢定兩模式預測誤差具有顯著差異，顯示利用 k-NN 模式作旅行時間預測較為準確。
3. 本研究 k 值的設定為每五分變化一 k 值解，有別於過去研究對 k 值設定方式大多採固定 k 值解，容易在某些預測時段資料選取太多或太少，產生預測誤差較大的問題，因此本研究採取變化 k 值設定的方式，能取得該時段較適當的歷史資料量。
4. 過去研究在 k 值取得方式，大多選擇測試選取各種 k 值，以得到 k 值最佳解，若遇到歷史資料量大時，此種作法就需花費較多時間去尋找 k 值，而本研究 k 值設定方式，為設置各種門檻值，作篩選歷史資料的條件，以此找出各時段的 k 值最佳解，此種作法能更方便取得 k 值，在預測準確度上也能達到一定的效果。

## 6.2 建議

本研究之建議如下：

1. 目前僅以探針車旅行時間來做資料比對，未來如能在比對時加入一些偵測器資料如：流量、速度、占有率作比對，應能在選取路段旅行時間資料時更嚴謹，同時增加預測之準確性。
2. 本研究對象為無事故發生之旅行時間預測，當事件或事故發生時，往往是旅行者最需要旅行時間預測的時機，卻也是旅行時間預測最困難的部份，未來如能建立高速公路交通事故資料庫，可以再針對事故發生時旅行時間變化作進一步研究，如遇到交通事故發生時，也能夠準確的預測旅行時間。



## 參考文獻

- [1]. Altman, N. S., "An Introduction to Kernel and Nearest Neighbor Nonparametric Regression," *The American Statistician*, Vol. 46(3), pp.175-185, 1992.
- [2]. Chang, G.L., Zou, N., and Wang, J., "Development and Field Evaluation of a Real-Time Travel Time Prediction System," Maryland Department of Transportation State Highway Administration, 2006.
- [3]. Choi, K., "An Algorithm for Calculating Dynamic Link Travel Times Using GPS and A Digital Road Map," 5<sup>th</sup> ITS World Congress, 1998.
- [4]. Clark, S., "Traffic Prediction Using Multivariate Nonparametric Regression," *Journal of Transportation Engineering*, Vol.129, No.2, pp.161–168, 2003.
- [5]. Kwon, J., Coifman, B., and Bickel, P., "Day-to-Day Travel Time Trends and Travel Time Prediction from Loop Detector Data," *Transportation Research Record* no.1717, Transportation Research Board pp. 120-129, 2000.
- [6]. Lam, W.H.K., Tang, Y.F., Chan, K.S., and Tam, M.L., "Short-term hourly traffic forecasts using Hong Kong annual traffic census," *Transp.*, Vol. 33, No. 3, pp. 291-310, 2006.
- [7]. Lewis, C. D. *Industrial and Business Forecasting Method*, 華泰書局, pp.38-41, 1982.
- [8]. Robinson, S., and Polak, J., "Modeling Urban Link Travel Time with Inductive Loop Detector Data by Using the k-NN Method," *Journal of the Transportation Research Board*, Vol. 1935, pp.47-56, 2005.
- [9]. Smith, B.L., and Demetsky, M.J., "Traffic Flow Forecasting: Comparison of Modeling Approaches," *Journal Transportation Engineering*, Vol.123, pp.261-266, 1997.
- [10]. Smith, B.L., Oswald, R.K., "Meeting Real-Time Traffic Flow Forecasting Requirements with Imprecise Computations," *Computer-Aided Civil and*

- Infrastructure Engineering, pp.201-213, 2003.
- [11]. Smith, B. L., Williams, B. M. and Oswald, R. K., "Comparison of Parametric and Nonparametric Models for Traffic Flow Forecasting," Transportation Research Part C, Vol. 10, No. 4, pp. 303-321, 2002.
- [12]. Weng, J.C., Hu, Z.W., Yu, Q., and Ren, F.T., "Floating Car Data Based Nonparametric Regression Model for Short-Term Travel Speed Prediction," Journal of Southwest Jiaotong University, Vol.15, No.3, Jul. 2007.
- [13]. Yang, D.B., "Application of The ARIMA Models to Urban Roadway Travel Time Prediction," Systems, Man and Cybernetics, SMC IEEE International Conference, pp.2529-2534, 2006.
- [14]. Zhang, X. and Rice, J. A., "Short-Term Travel Time Prediction," Transportation Research, Part C, Vol.11, pp.187-210, 2003.
- [15]. 吳佳峰，「有GPS資訊提供下之車輛旅行時間預估模式之研究」，國立交通大學運輸工程與管理系碩士論文，民國九十年。
- [16]. 李季森，「應用探測車法預測高速公路旅行時間」，國立中央大學土木工程研究所碩士論文，民國九十年。
- [17]. 林士傑，「高速公路旅行時間預測模式之研究-類神經網路之應用」，國立成功大學交通管理科學研究所碩士論文，民國九十年。
- [18]. 林傑斌、林川雄、劉明德、飛捷工作室，「SPSS 12 統計建模與應用實務」，博碩文化，民國九十三年。
- [19]. 陳建名，「有GPS資訊提供下之公車旅行時間之研究」，國立交通大學運輸科技與管理學系碩士論文，民國九十二年。
- [20]. 陳順宇，「多變量分析」，華泰文化，民國九十四年。
- [21]. 黃俊英，「多變量分析」，華泰文化，民國九十年。
- [22]. 楊奕農，「時間序列分析經濟與財務上之應用」，雙葉書廊，民國九十七年。
- [23]. 蔡百里，「資料融合技術應用於旅行時間推估之研究」，淡江大學運

輸管理學系運輸科學研究所碩士論文，民國九十五年。

- [24]. 張保隆，「管理數學」，華泰文化，民國九十四年。
- [25]. 蕭文龍，「多變量分析最佳入門實用書: SPSS+LISREL」，碁峰資訊，民國九十八年。

