第二章、文獻回顧

由於本研究為二期計畫中之第一期,目的是為了進行旅行時間模式資料的前處理與發展旅行時間預測模式,因此本節先對旅行時間模式相關文獻進行回顧，接著再對資料處理相關方法進行回顧。

收集用路人資訊及道路交通資料,是預測路段旅行時間時重要的一環,現階段交通參數資料的蒐集方式上,大致可分為偵測器(VD)、探針車(probe vehicle)以及自動辨識(AVI)等三種類型，其中 AVI 與 ETC 類似,都是利用比對車牌或是車上機編號來計算旅行時間。對於旅行時間之預測其概念大致上可分為兩種來做說明:第一種是利用模擬來分析駕駛人行為之假設性資料,利用資料進行旅行時間的推導;第二種是利用即時或事後所偵測到的交通參數資料,用不同的方法來進行資料分析與旅行時間推估預測,一般而言可以分成模擬資料推估、時間序列分析、車輛辨識方法、迴歸分析、KNN 法、類神經網路、模糊理論與考量路口延滯方法等,各種方式都有學者提出相關的研究。分述如下：

2.1 旅行時間預測模式

2.1.1 模擬(simulation)資料推估

Chang 等人(1994)建構一套巨觀粒子模擬(MPSM)系統，分別採用 MPSM,修正的 MPSM(M-MPSM),和微觀(micro)三種車流模擬模型，用於先進交通管理系統的應用，並以 MPSM 模擬高速公路上的車流，M-MPSM 模擬都市街道車流系統，而微觀模型則用來針對已產生壅塞的街道模擬,則可以模擬結果推導出包含旅行時間等各項車流系統資訊。Johnston 等人(1999)建構一套巨觀的車流模擬系統，以平行計算的方式來運作，並將明尼阿波理斯市(Minneapolis)的公路路網的車流資訊代入以驗證此模型的準確度與計算模擬速度。Kachani 等人(2001)利用巨觀模式將車流視為液體流動，使二階多項式旅行時間(PTT)和指數旅行時間(ETT)等模式來模擬出駕駛對上游的擁塞與否所產生的反應行為與路段密度或鄰近路段密度所造成的車流效應這兩種現象，藉以估計旅行時間，可增加等候線來將此類模型用於使用者動態均衡的問題上。交通部運研所(2004)，提出一套先進旅行者資訊系統(ATIS)，透過蒐集路況資料，建立「靜態」及「動態」資料庫的旅行時間推估模式,並開發資訊模擬顯示系統及路況回報資訊統,以不特定點對點之邏輯,建立一結合地理資訊系統的旅行時間資訊查詢網頁將預估之旅行資訊提供給用路者,使用路者有更多的交通資訊

以進行旅運決策，進而減少用路者的旅運成本。Kiesling 等人(2005)提出微觀的車流模擬系統平行時間(time-parallel)模擬的方式將各節點時間區隔成好幾個區段(interval)，此模擬方法有兩個優點， 其一為即使不存在適當的變數仍可以使用此模擬方法來模擬，其二為每節點可以單獨模擬出其時間間距，而不需要考慮其它節點。經過驗證此模型後得到，在低密度時使用有較佳的適用性。

2.1.2 時間序列分析

Hellinga 等人(1999)以等候理論為基礎，得到平均延滯的抽樣估計中，車輛到達時間的分配，發現抽樣偏差及調查比例會造成旅行時間計算結果的誤差。在假設路徑旅行時間為各路段之旅行時間和，旅行時間預測方法多採用各種時間序列方式或是類神經網路模式，資料來源多採用速度、流量、佔有率等交通參數推算旅行時間；而在擁擠車流的情況下可能會產生較差的預測結果

Yang et.al(2006)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。其方法是將蒐收所得資料視為時間序列，以 ARIMA 模式進行旅行時間預測。對明尼蘇達 194 號高速公路作實測，結果顯示此方法能夠有效預測短期內的旅行時間。Ichiro et.al(1998)時間的預測包括延滯時間，且預測方法可以定期由時間序列資料中獲得。因此欲建立一含有歷史交通流量與先前旅行時間序列資料庫，希望能在相近的資料中準確預測未來旅行時間值。系統所需的要素有 AVI、雙向通訊、紅外線車輛偵測器與上述資訊資料庫。

Nagaoka 等人(1999)提出了以下的方法求得旅行時間。根據相關係數及統計方法以及由偵測器所得到的平均速率及平均旅行時間，可以得到以下兩種方向對於旅行時間的預測。

1. 相關係數修正

   由平均速率估計隨時段改變而變化的速率函數，再由該速率函數預測速率即求出預測的旅行時間。

2. 統計方法修正

   由平均的旅行時間估計旅行時間的函數，再由該函數預測旅行時間。

2.1.3 車輛辨識方法

車輛辨識方法包括透過自動車輛辨識(automatic vehicle identification, AVI)、影像處理、利用裝有車上單元之探針車等方式來辨識車輛進而推估旅行時間(Chung et al., 2003; Bakata and Takeuchi, 2004; Yang, 2005)。Chen 等人(2001)指出利用探針車蒐集旅行時間資料可能會有較佳的效果，但因車輛派遣數量的限制可能導致其他問題發生。該研究以卡門濾波法為理論基礎，利用路徑或是路徑中各路段探針車旅行時間資料，預測下一期路徑旅行時間。測試結果發現在一般車流情況下，直接應用計算路徑旅行時間預測下一期路徑旅行時間較為精準；在車流狀況接近飽和或擁擠情形下，如果探針車的數量比例過低，預測效果會受影響。Yamane 等人(1999)利用汽車牌照辨識之 AVI 及 UVD 收集旅行時間所需資料，以提供用路人即時之旅行時間資訊。Bae 等人(1995)針對使用自動車輛定位(AVL)資料推估技術的公車建立一套模型來估計旅行時間，透過此套模型發展出公車在到達站點位置時間的預測模型。Sherali 等人(2006)利用線性重劃(reformulation-linearization)的方法建立一個線性的混合整數規劃(MIP)模型來最大化 AVI 使用者在高需求的道路上受到各種車流行為包括合併車流、道路阻塞等狀況時的旅行時間效益，用分枝界限法求解推估最後得到的旅行時間。

除了透過上述方法來辨識車輛外，亦可只透過迴圈偵測器或路側偵測器來進行車輛或車隊辨認進而推估旅行時間，Sun 等人(1999)透過比對車隊的波形(Waveform)來判斷同一個車隊經過上下游偵測器間的旅行時間，但此波形要透過另外一套獨立的速度偵測設備來標準化。Coifman(1999)透過比對車隊的長度來辨識車隊，但也需要另一套設備來測量速度。Karric Kwong 等人(2008)，則利用偵測器來辨識車輛以求得旅行時間分佈。

M.Ndoye 等人(2008)與 Oh 等人(2002)也利用迴圈偵測器來辨識車輛或車隊，進而推估其旅行時間。但在將速度標準化的過程中，其假設速度為一常數，因此只要車輛在兩偵測器間的速度有變化，亦即加速或減速，則此假設就不成立，且其只有展示離峰時段的旅行時間推估結果。

2.1.4 迴歸分析

迴歸分析是統計方法中廣為使用的分析方法，其運用涉足各個領域，在交通領域中，Kwon et.al(2000)用 I880 資料庫來建立旅行時間

預測模式，資料庫中包含單一線圈所偵測到的佔有率以及流量資訊，及探針車所得到的旅行時間。模式中將偵測到的流量、佔有率、及當時的星期別及日期都視為一個集合，並對探針車得到的實際旅行時間做線性迴歸，最後以 MSPE 來檢定預測績效。

Rice et.al(2004)使用線性迴歸、主成分(principal component)及鄰域(nearest neighborhood)三種方法，應用在 PeMS 計劃中預測旅行時間。其中迴歸方式是先以上下游透過單一線圈所得到的流量與佔有率的資料藉由 Jia et al.(2001) 文獻找出的 g-factor 求出速度後，推估目前的旅行時間，再代入歷史資料庫所求得的線性迴歸方程式中預測某一時間差(time lag)的旅行時間，此迴歸式以時間別及星期別做為 pattern 的區分，最後預測結果以 root MSE 來檢定預測績效。

You 等人(2000)以無母數迴歸(non-parametric regression)統計方法作為核心的演算工具，並結合圖形資訊系統(GIS)發展出另外一套混合(hybrid)車輛旅行時間預估模型。

Sen 等人(1997)引用美國芝加哥的 ADVANCE 計劃所收集到車輛於路段旅行時間資料，討論車輛偵測器不完整資訊推估旅行時間之研究中，各迴歸式的參數是否具有顯著性，並估算各參數。該研究假設依照在某路段上有無佈設偵測器的兩種情況，採用不同方式來進行旅行時間預估。在都市幹道的網路上，受到號誌系統控制與環境因素之影響下，路段旅行時間之不確定性較高速公路系統來的複雜。

支持向量機迴歸(support vector regression, SVR)是由 Vapnik's (1995, 1997, 2003)提出，並應用於時間序列的預測上。Wu(2004)嘗試將其運用在旅行時間預測上，用中研院的 ITWS 計劃在中山高速公路部份路段上以單一迴圈偵測到的每三分鐘更新一次的速度資訊，使用支持向量機迴歸(support vector regression)法、當前旅行時間預測法(current travel time prediction method)、歷史平均旅行時間預測法(historical mean prediction method)來預測旅行時間，以 RME 和 RMSE 來評估績效，認為 SVR 有較佳的績效。以上的方法主要是根據偵測器所收集的旅行時間樣本資料做分析，由於偵測器每三分鐘更新一次，所以造成資料有中斷或是錯誤的現象發生，因此將有效的資料收集後，再根據不同的估計方法預測旅行時間的長度。

2.1.5 KNN 法

k 最近鄰法(k-Nearest Neighbor, KNN)，是一種利用歷史資料的特性推估未來資料的方法。最初是由 Benedetti(1977)、Stone(1977)及 Tukey(1977)這些學者提出了近鄰法(nearest neighbor)的概念。這些學

者利用一元位置估計(univariate location estimators)，由平均數(mean)和中位數(median)建立了無母數迴歸式的模型，進而引申出最近鄰法的概念。

Altman(1992)將以上的研究進行整理，將一元位置估計引申至多元位置估計(multivariate location estimators)並且提出了 KNN 法，將目前的輸入資料與歷史資料相比對，找出最接近的 k 筆歷史資料，將這 k 筆資料的輸出值進行平均或加權平均後，當作目前輸入資料的輸出預測值。Smith、Demetsky(1997)對 KNN 法進行績效評估，分析比較以下四種交通流量的預測方法：歷史平均法，時間序列法、類神經網路法與 KNN 法。根據歷史的流量資料來預測未來的流量，結果發現當歷史資料量大時，KNN 法所預測的誤差結果比其他三種方法所得到的結果誤差來得小，因此可以針對流量進行預測。

Clark(2003)除了流量外，試著利用 KNN 法試著針對其他可收集到的交通資訊進行預測，如速度和佔有率。接著進行一一分析和交叉分析，結果發現當同時利用流量、佔有率和速率這三項變數進行分析時，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。Rice、Zwet(2004)則是利用兩種特徵來計算 KNN 之距離，一種為偵測器所測得的速度值，另一種為旅行時間，而特徵向量除了包含當下時間 t 的資料外，還加入前幾個時間(t-1、t-2、…、t-w)的資料，即是利用一個時間窗的資料來計算 KNN 之距離，最後再取最接近的 k 個資料來預測旅行時間。Robinson、Polak(2005)提出了四點建立 KNN 模型時應該要注意的條件，包括：特徵向量該包含哪些屬性、建立適當的距離數值(distance metric)、決定每次取最近鄰的數量(k)、利用加權法減少 KNN 模式的誤差。

Chang(2006)將旅行時間預測分為兩個階段，旅行時間推估與旅行時間預測。在旅行時間推估階段中，其模式採用線性迴歸與以速度軌跡為基礎的混和模式，藉此將偵測器所偵測到的即時交通資訊轉換為路段的旅行時間。有了該時間點下各路段的旅行時間後，接著再採用 K 最近鄰點法(K-Nearest Neighbor model, KNN)來預測路徑的旅行時間。

2.1.6 類神經網路

一般而言，在做靜態的旅行時間預測，大多以統計的方法，如無母數迴歸或是時間序列分析等來做預測。但於動態預測方面，這些統計方法於捕捉動態號誌控制系統下的車流資訊顯的相當不足。Palachara 等人(1999)提出了以模糊系統及類神經網路的方法來進行

這類型的研究。Yoshikazu 等人(1998) 利用架設在路段的 AVI 系統蒐集車輛資料，應用混合式類神經網絡方法解釋每個路段的旅行時間與整個路徑的旅行時間之間的關係。Fu 等人(1999)以人工類神經網路(ANN)的方式來模擬路網中的車輛旅行時間，並運用於車輛定線派遣問題。

國內相關文獻有李季森等人(2001)探討國內高速公路駕駛人變換車道行為與變換車道時間，進而研究於不同預測時間、流量、探針車混合比例與區段長度等相關參數之實驗組合，並透過類神經網路進行旅行時間之預測。張修榕等人(2001)透過類神經網路模式來進行雙階段高速公路旅行時間之預測。針對感應線圈偵測器可蒐集車流速度及流量的特性，利用模擬的方式產生所需之交通資料並作驗證；接著是預測部分，採用倒傳遞(feed-forward back propagation)類神經網路模式來建立不同車流型態下之旅行時間預測模式。黃裕文等人(2003)以微觀的角度探討國內高速公路施工路段的車流變化，同樣以上述的方法建立旅行時間預測模式。溫志元等人(2002)係針對高速公路進口匝道匯流路段之變換車道行為與加速車道變換車道匯入主線行為動機與條件進行界定，透過類神經網路進行旅行時間預測。此外，路段線形(Road Profile)亦可能造成旅行時間之推算誤差，林士傑等人(2001)以中華顧問工程司交通千里眼(e-traffic)所提供之即時交通播報資訊，再加上高速公路幾何、交通量調查與客運車輛 GPS 等資料，運用類神經網路準確預測高速公路旅行時間，來供用路人參考以降低不確定性。鑒於國內偵測器普遍設置不足，吳佳峰等人(2001)希望透過 GPS 車輛歷史旅行資料預估車輛旅行時間，為了能夠正確預估車輛旅行時間，該研究設定了車輛運行路線分段以及車輛歷史旅行資料劃分時段之準則。近年來有許多研究利用多項偵測單元進行資料融合，藉以提升旅行時間推算之準確率，如李穎等人(2002)融合國道客運班車 GPS 資料、車輛偵測器資料、事件資料等真實資料，以類神經網路法尋找各項資料來源其參數與旅行時間之關係。張慶麟等人(2002)以 AVI 方式針對高速公路平常日之車流情形，先行應用車流模擬方式考量不同資料輸出時距、偵測器佈設間距及 AVI 辨識率等產生相關資料，配合簡單指數平滑法、Holt's 指數平滑法、ARIMA 模式及倒傳遞神經網路構建四種旅行時間預測模式，分別進行預測績效分析。黃文鑑等人(2007)則是利用車輛偵測器與探針車類神經於中正機場至台北之間有探針車輛的路段使用類神經進行旅行時間預測。

2.1.7 模糊理論

Li 等人(2002) 發展一套利用單一探針車作旅行時間估計。首次引入了駕駛行為變數，分別為快速、中等以及慢速；駕駛行為的分類是透過探測車的測量值與平均旅行時間，經模糊理論比較所得。當探針車的測量值經比較後被歸類後，若被判定為中等速度的車輛，其測量值就被視為該路段實際旅行時間，否則實際的旅行時間需要由探針車測量值乘以系數作調整後求得。經實際驗證，在非擁擠的情況下得出足夠準確的結果，而在擁擠情況下表現則未如理想，仍有待改善。

2.1.8 其它推估方法

國外期刊研究有 Choi 等人(1998)利用衛星定位系統(GPS)及電子地圖來計算及蒐集市區路段之動態旅行時間；另外，該文獻提到為最常用的方法是以浮動車輛法(floating car method) 來得到路段旅行時間，缺點在於需要蒐集較多的交通資料，例如每個車道、方向、時段的各種不同的交通參數資料等。Yoshikazu 等人(1998)研究高速公路線上型態(on-line)之旅行時間預測模式，提出需要必要的條件因素才能預測高速公路旅行車輛偵測器不完整資訊推估旅行時間，其預測之困難在於需考慮交通車流之動態變化。

國內有李俊賢等人(1996)研究隨機性動態旅行時間，以 Fu 等人(1999)所提出之動態隨機最短路徑問題(DSSPP)為基礎，建立隨機性動態旅行時間(SDTT)模式。卓訓榮等人(2003)以參考擬最鄰近法（pseudo-nearest-neighbor）的概念，以最鄰近參考數列之對應數值進行不完整資料之差補，並應用灰關聯度函數作為衡量兩不完整數列間之鄰近程度以彌補交通資料集合必須符合 Gaussian 隨機分配之限制，再透過模糊類神經網路之倒傳遞網路學習機制推估旅行時間。王晉元等人(2005) 則利用靜態路段流量守恆之觀點，在偵測器佈設不足之前題下，推論資料不完整路段之流量可能範圍，若假設已知路口的轉向比、路段容量、偵測器的佈設位置，則可以縮小路段流量不確定的範圍。

預測方法	理論內容	優點	缺點
模擬	將真實世界狀況模式化，輸入各種情境，以模擬真實情況。	1. 可模擬多種交通情境 2. 有套裝軟體可使用，如：FRESIM、PRAMICS	需要有較多的資料
時間序列	利用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。
車輛辨識	利用同一車輛行經路徑，所計算出的路徑旅行時間。	在一般車流情況下，計算路徑旅行時間預測下一期路徑旅行時間較為準確。	車流狀況接近飽和或擁擠情形下，如果探針車的數量比例過低，會降低預測結果。
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，解釋變數與應變數之關係，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	1. 係數固定，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。

		程式。	
KNN	利用歷史資料的特性推估未來資料的方法。將目前的輸入資料與歷史資料相比對，找出最接近的 k 筆歷史資料，將這 k 筆資料的輸出值進行平均或加權平均後，當作目前輸入資料的輸出預測值。	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。
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。 3. 最佳隱藏層數目及神經元數目決定無規則可循。

表 2.8-1 旅行時間推估方法整裡

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，KNN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KNN模式的預測準確度高，因此本研究選擇使用KNN作為另一預測模式，並當KNN方法比對不到合適資訊時則透過流量、速度及旅行時間等迴歸參數來計算旅行時間以補足KNN的缺點，並將兩種方

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。

2.2 資料插補模式

資料遺漏的定義可以解釋為，在現實生活中可能確實存在這筆資料，可能在蒐集或傳輸的過程中不見了，以致資料有遺漏現象，我們稱此現象為遺漏資料值(Missing Value) (D. Plye, 1999)。

Little and Rubin（1987）將遺漏資料值分成三種：

1. 完全隨機遺漏（Missing Completely at Random, MCAR）

完全隨機遺漏是指資料的遺漏是完全隨機的，與其他變數無關。此種類型的遺漏值無法由資料中評估遺漏值的插補模式的好壞，但相對的各種演算法，皆可用來插補此類的資料。

2. 隨機遺漏（Missing at Random, MAR）

隨機遺漏是指資料是否遺漏與資料中其餘被觀察到的變數相關，而與遺漏部份的變數則無關。

3. 不可忽略的缺失（Nonignorable Nonresponse）

不可忽略的缺失是指當是否遺漏與資料中所有變數均相關時，則視此時遺漏值的發生為不可忽略的。我們可以藉助模型的幫忙，但資料的遺漏與遺漏的變數有關，因此此時所建立的模型並無法完全地呈現遺漏值的情況。

一般處理遺漏值的方式有很多種，但插補法則較多統計學者使用，插補法可分為兩類：

1. 單一插補法

單一插補法包括包括平均數插補與迴歸插補。當給定觀察資料，插補值是固定的，Rubin(1987)提到單一插補法(single imputation)主要有兩個優點：(1)插補資料後便有一組完整的資料來分析，(2)插補的方法可以結合資料收集者的訊息。相對的，單一插補法也有缺失：無法反應抽樣的變異，也就是會增加參數估計量過多的變異，造成信賴區間過於寬闊。且也忽略了完整資料與遺漏資料的多對一關係。

2. 多重插補法

Rubin於1978提出了多重插補(multiple imputation)的概念，主張應用各種插補方法和估計的數值，應該不限於一組。反之研究者對於某一特定變項之遺漏值的處置，可以插補(或估計)一系列的數值。由於每一個遺漏值皆有相對應的許多插補值或估計值，因此研究者可以比

較不同處置方法的差異，甚至估計插補的誤差，然後進一步模擬估計值的分佈。可是在實用的角度來看，由於多重插補必須產生許多的插補值，然後重複模型分析，自然也就會增加資料處理與分析的複雜性和成本。

Rubin提到多重插補法較有理論基礎且改善了單一插補法的缺點和承襲了單一插補法的優點。另外也增加了一些優點：

(1) 增加估計量的有效性(efficiency)

(2) 多重插補代表重複地隨機從模型中抽取出來，結合完整資料的推論會較有效

(3) 多重插補允許從多個模型中重複地隨機抽取插補值，並且允許對不同模式敏感性(sensibility)的推論。

趙民德與謝邦昌(1999)指出多重插補的好處是：

(1) 多重插補由抽樣單位統一執行

(2) 原先的公式(指的是估計參數的型態)不需改變

(3) 可以獲得較正確的統計推論

但是多重插補法也不是沒有缺點，相對於單一插補法，多重插補法存在一些問題：

(1) 增加計算上的困難

(2) 增加空間來儲存資料

(3) 分析上比單一插補法費力。

但是，以現今電腦的運算能力來看，相信這些都不是太大的問題。

一般來說常用的插補法有Random Hot-Deck 插補法、平均值插補法（Mean Imputation）、迴歸插補法（Regression Imputation）以及比例插補法（Ratio Imputation）、組內插補法(interpolation imputation)、組外插補法(extrapolation imputation)等、灰色插補（Grey Imputation）。以下分別說明之：

1. 比例插補法

比例插補法係利用兩個變數間所存在的比例關係進行插補，其中一個變數為具有缺失值的變數，也就是待差補的變數，其值為y_i，它在某些調查單位上是缺失的，缺失的部份記為y_i^*；另一變數為輔助變數，其值為z_i，它的資料在樣本中是完整而無缺失

的，如下表所示。

i	1	2	3	4
y_i	3	y_2^*	9	6
z_i	15	20	45	30

表 2.9- 1 比例插補法資料表

比例插補法係利用 y_i 與 z_i 的比例進行插補，所以 $y_2^* = 3/15 * 20 = 4$。

2. 平均數插補法

平均數插補法就是利用同一屬性(*X*)中的平均值計算出來取代其該屬性原有的遺漏值。這有明顯的缺點是它無法用在非數值的資料且可能扭曲X變項在樣本中的分配，因為所有出現遺漏值的觀察体，其*X*屬性的數值只有一個，就是平均數，這種事實，進一步也會降低、減少*X*屬性的變異量，造成變異量低估等問題。

3. 組內插補法(interpolation imputation)

與平均數插補不同在於其作法為將各遺漏值以其前後各幾個未遺漏的資料求其平均值替代之，為平均數插補法的變形。這種方法很難去決定要取幾個來平均呢？且有遺漏資料的這筆不見得與其前後資料有相關，一旦沒有任何關係的話，可能造成更大的偏差。

4. 組外插補法(extrapolation imputation)

利用遺漏值之前幾個或後幾個未遺漏資料的平均值替代之，為平均數插補法的變形，其可能產生的問題與組內插補法相同。

5. 熱卡插補法(hot-deck imputation)

熱卡插補法的基本精神，就是按照輔助變項的不同條件，將未出現遺漏值的觀察体分類成為若干的「插補空格」(imputation cell)，最後，每一個出現遺漏值的觀察体，依據其輔助變項的條件，從相對應的「插補空格」中找尋一個觀察体，以其觀測所得的變項數值代替其遺漏值。

Fellegi及Holt(1976)認為使用熱卡插補後的資料比平均數插補後的資料更能代表原始的母体分配行為；也就是說這些插補過後的值更具有代表性。主要是因為平均插補法在同一屬性中加入了相當大量的平均值，使得樣本分配的形狀受到不小的影響。

6. 冷卡插補法(Cold-deck imputation)

相對於熱卡插補法，若是區分或比較的對象是以前舊有的資料時，則稱為冷卡插補法。

7. 替代插補法(substitution imputation)

觀察變數間的相關性，對於高度相關的變數，其間之部分變數若存在著不完整資料，則可以利用相關變數(完整資料)之值推導出遺漏值的部分。

8. 迴歸插補法(Regression imputation)

迴歸插補法是一般插補法中較常用的方法，其迴歸模式通常表示如下：

$$\begin{bmatrix} Y \\ Z \end{bmatrix} = \begin{bmatrix} W \\ X \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$$

其中，Y 為已知值，Z 為遺漏值，X,W 分別為 Y,Z 的解釋變數，β 為估計的參數，ε_1 與 ε_2 分別為誤差項。利用已知的數據建立模型：$y = x\beta + \varepsilon_1$，利用最小平方法(LSE)估計出 $\hat{\beta}$，再把代入遺漏值模式：$z = w\hat{\beta} + \hat{\varepsilon}_2$ 則可得到一組插補值。這裡有關 $\hat{\varepsilon}_2$ 方面，若 $\hat{\varepsilon}_2 = 0$，則稱為確定迴歸插補法，否則稱隨機的迴歸插補法。迴歸插補法是用來預測某一屬性的遺漏值,其作法為藉由與其他屬性間建立關係。這是在統計上常用的方法，一般來說，預測遺漏值的效果還不錯，不過如果資料分佈太過分散，則其建立的關係可能就較差，而預測出的遺漏值也較較不準確，另外還有一點值得注意的就是迴歸法在處理大量資料時的執行效率。

9. 灰色插補（Grey Imputation）

有別於上述插補法不需符合統計分配即可插補，利用灰色理論中灰預測少量數據的優點（只要4個數據即可預測），常用於樣本數不足的情況下，利用剩餘完整資料推導出遺失值的部份，其優點為降低歷史資料量需求，所蒐集之資料也不需符合統計分配機率。

2.3資料過濾模式

利用高速公路上的交通資訊蒐集設備包括探針車(Probe Vehicle)、電子收費系統（ETC），並傳回該時點的速率與位置，因蒐集對象為國道客運本身為載客任務的關係，會有異於正常車流之行為產生，主要原因有下交流道載客後再上交流道、中途停靠休息站等因素，若是直接將所有傳回來的速度資料輸入旅行時間推估模式後，可能會使相

關模式結果產生不必要的誤差，以導致後續的旅行時間預測模式出現錯誤。一般常用方法以卡門濾波器法（Kalman Filter method）、平均數平滑法(Average Smooth method)以及自訂規則法等三種，以下分別說明之：

1. 卡門濾波器

最初由R.E.Kalman（1960）所提出，是一種藉由過去資訊不斷更新的遞迴(recursive)演算法，具有將資料平滑化的特性。它是利用間接衡量的狀態變數值及與觀察變數兩者的共變異訊息來遞迴更新系統狀態先前的估計，並對系統狀態作逐期的修正，其模式如下：

$$\zeta_{t+1} = F_{t+1}\zeta_t + \nu_{t+1}\text{（狀態模式）}$$

$$y_t = H'\xi_t + \omega_t\text{（觀察模式）}$$

模式中y_t是t時點回傳速率資料，ζ_t為每t時刻所要估計的平均速率，H'為係數。因此得到y_t之後，可利用狀態模式可推估下一時間點ζ_{t+1}，即可對ζ_{t+1}進行更新，以此在遞迴下去。Nanthawichit(2003)利用巨觀車流理論模式，將探偵針車所收集到的資料經由卡門濾波器以去掉極端值。此方法的車流、速率及密度預估的準確度較高，進一步可推估旅行時間等延伸資訊。Cathey and Dailey(2003)用AVL(Automatic Vehicle Location)自動車輛定位系統所收集到的資料，經由卡門濾波器去段極端值後，在將這些資料對應GIS(Geographical Information System)地理資訊系統上的各路段，加總各路段的「距離/速度」即可推估旅行時間。

2. 平術數平滑法

平均數平滑法之資料過濾模式擬定，假設取得之n筆速率資料為S_1~S_n，以S_1~S_n為輸入資料，則A_1~A_n為輸出之平滑結果，其模式如下：

$$A_j = \frac{S_1 + \cdots + S_j}{j}, j = 1\sim n$$

由模式可看出平滑的方法為以前j筆速率資料的平均數代表第j筆速率資料的平均值，而平均數為代表集中趨勢的統計量，也就是以前j筆速率資料的集中趨勢值來代表第j 筆的資料。

3. 自訂規則法

利用資料分析依照不同特性與範圍所自訂的過濾方法，如國道客運速率為0的資料，其原因可能為發生事故或是路況擁擠所影響。張惠汶(2002)以路口與站牌位置前後一範圍作為停等區，在根據公路客運GPS定位的位置資料、速率型態判別是否濾除該筆資料。在

停等區設定方面，主要收集實際公路客運GPS資料，求算公車自原行駛速率減速至完全停住所行駛之長度，及自停等加速至正常行駛速率之行駛長度，所求算出的長度設定為停等區長度，其設定方式會隨著地點之不同而重新修正。何佳儒(2010)以公路客運GPS所回傳速度資料，依照自訂事件型態，如:站牌上下車乘客、紅綠燈停等、急踩煞車之情形，所產生速度為零的資料過濾。因此本研究將所蒐集到的交通資料，依照資料本身特性發展自我過濾方式。

2.4 小結

根據旅行時間推估模式主要以KNN和多元線性迴歸模式作為本研究旅行時間預測模式之建構基礎。從上述文獻回顧中可發現線性迴歸方法是一種概念較簡單的方式，但又不失預測準確度的方法，可以運用在預測旅行時間週期性的變化上。在以往的簡單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，因此本研究採用的多元線性迴歸模式，能夠改善這項缺點。

此外本研究針對高速公路進行資料過濾，因此會先分析高速公路交通資訊的特性，接著針對不同交通資料來源的特性建立不同資料過濾模式;當過濾完異常資料後，異常資料即被刪除而造成資料遺漏。或者是偵測器設備、通訊問題而發生資料遺漏之情形，因此需要資料插模式插補遺漏之資料。依據上述文獻回顧結果，可發現多數方法皆需要完整的資料來源，但是實務上無法取得完整資料，因此資料插補法於旅行時間預測模式的實務應用上是不可或缺的

其作法為先將歷史資料做分群（星期別、各時點），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，因此本研究選擇線性迴歸方法模式作為旅行時間預測模式之一。