

本論文獲得法務部司法官學院110年傑出碩博士犯罪防治研究論文獎

中央警察大學犯罪防治研究所

博士論文

指導教授：鄧煌發 博士

黃俊能 博士

臺北市住宅竊盜犯罪群聚及區位因素
之研究

**A Study of the Clustering and Ecological
Factors of Burglary in Taipei City**

研 究 生：廖劍峯

中華民國 109 年 6 月

謝誌

撰寫論文是印證所學的最後試煉，今日得以完成，首先承蒙指導教授鄧煌發與黃俊能老師提攜教導，在您的春風化雨下，資質駑鈍如我，終能發榮滋長，而弘毅寬厚之言傳身教，更是學生終生心靈寶藏，何其有幸成為兩位老師入門弟子，從今爾後必當謹記教誨、堅守初心。

六年多的學習，在鑠懿淵積的許春金老師、黃富源老師、張平吾老師、蔡田木老師、謝文彥老師、逐步引領下，終能升堂入室，一窺犯罪學領域奧妙殿堂，而論文口試時，您們不辭諸煩，斧正論文之繆誤，修正觀念之偏頗，使臻完備，感念於心。林增祥、陳玉書老師的扎實統計教導及訓練，猶如打通量化研究任督二脈，讓我望見未來曙光；郭佩茶老師、董正談老師，由於您們的啟迪，開啟我涉獵資訊領域的契機。這篇論文，俯拾皆是各位恩師們授業解惑苦心，師恩浩蕩，永銘於心。

「一朝同學一世友，一輩同窗三輩親」；研究期間，葉碧翠、簡鳳容、鄭勝天、張庭毓及邱嘉祥同學，與您們共同討論，相互砥礪，堅定大家一定要完成學業的信念，也恭喜碧翠畢業後擔任犯防所教師，持續貢獻母校，更特別感謝妳初稿發表時擔任與談人，提供修正建議，受益良多。

論文能夠如期完成，學長紀致光、方嘉鴻與蔡文瑜助教猶如及時雨，讓我能按部就班，循序漸進，減少摸索時間，避免誤入偏差，您們的經驗與熱情，讓我能依時程順利畢業，永難忘懷。

最後，感謝辦公室長官、同仁們您們對我的寬容與鼓勵，讓我能朝著目標逐步邁進；也感謝我親愛的家人，您們默默關懷、循循包容，從無怨悔，特別是妻子楊妙娜，在我困圍愁城時，妳給我最大的力量，讓我破除陰霾、重新出發，山高海深，無以回報！

「就讀警大是幸福，順利畢業是榮耀」，感謝上天圓滿安排，在人生旅程中能與您們相遇！

廖劍峯 謹誌

中華民國 109 年 6 月 于新店

摘要

隨著犯罪地理學與環境犯罪學的理论軌跡匯合，以地理資訊系統製作犯罪地圖日臻普及與成熟，但如何將時間因素同時納入分析卻是棘手難題。為此，本研究使用掃描統計技術並配合地理資訊系統的製圖功能，針對臺北市 2015 至 2017 年住宅竊盜犯罪的空間、時間及時空分布現象進行分析。研究結果顯示，在空間分析方面，無論是使用圓形或橢圓形視窗掃描，臺北市住宅竊盜犯罪群聚熱區以中山區、萬華區、大同區及士林區較為顯著，且橢圓視窗掃描精確度略優於圓形視窗；在時間群聚部分，時間序列掃描顯示，歲末年終與跨年假期為住宅竊盜高發生時段；而在時空掃描方面，本研究使用「回顧性」與「前瞻性」掃描進行分析，在「回顧性」掃描部分，犯罪群聚的空間分布仍集中在中山區、萬華區及士林區等，但增加群聚時間的訊息後能更清楚判別群聚之特性；「時間趨勢的空間變化」與「前瞻性」掃描則發現在住宅竊盜犯罪冷區中，南港、內湖等部分地區有異常群聚現象，應特別予以關注，以達到「防微杜漸」的效果。

在了解時空中犯罪的分布特性後，本研究主要依據社會解組與新機會理論並參考相關文獻，選取各類社經人文變項進行區位分析，以釐清犯罪群聚形成之因素。在分析方法上，首先運用群聚共變項分析，以確認所選取之變項與本研究之群聚具有相關；其次，採用資料探勘技術與羅吉斯迴歸，篩選出犯罪群聚區位重要變項共計 8 項，結果顯示臺北市住宅竊盜犯罪冷區之特徵為高所得及房價較高地區有較低住宅竊盜犯罪；空屋率與單獨住戶比例較高地區，因監控力降低，易導致住宅竊盜犯罪；而在抑制犯罪的監控作為方面，犯罪熱區中有較多的警力配置，而路燈與監視器對於防竊的效果在本研究中並不明顯。

綜整前述之研究成果，本研究建議應針對不同熱區特性研擬具特色警務規劃，運用前瞻性掃描統計建構住宅竊盜犯罪及時監測預警系統，並根據犯罪群聚最大概率與被害風險數據合理配置警政資源，以提升犯罪預防之成效。

關鍵字：掃描統計、環境犯罪學、住宅竊盜犯罪、資料探勘

Abstract

With the integration of the theory of criminal geography and environmental criminology, using geographic information systems to make crime maps has become increasingly popular and mature, but how to analyze time and space at the same time is a difficult problem. For this reason, this study used scan statistics and geographic information system to analyze the spatial, temporal and Spatiotemporal distribution of burglary in Taipei city from 2015 to 2017. The results show that, in terms of spatial analysis, whether it is using round or elliptical window scanning the hot spots of burglary are Zhongshan District, Wanhua District, Datong District and Shilin District, and the scanning accuracy of the elliptical window is slightly better than the circular window. In the time cluster, the time series scanning shows that the year-end and New Year holidays were hot time of burglary. And in terms of space-time scanning, this study was analyzed using both "retrospective" and "prospective" scans; In the "retrospective" scanning, the spatial distribution of burglary hotspots was still concentrated in the Zhongshan district, Wanhua District and Shilin District. However, the characteristics of clustering can be determined more clearly by increasing the information of time. "Spatial variation in the time trend" and "prospective" scanning found that in cold spots of burglary, some areas such as Nangang District, Neihu District have abnormal clustering phenomenon, which should be paid special attention to, so as to achieve the effect of "nip in the bud".

After understanding the distribution characteristics of crime in time and space, this study selects various socioeconomic variables for ecological analysis based on criminology theory and relevant literature, so as to clarify the formation factors of crime clustering. Firstly, cluster covariant analysis was used to confirm the correlation between the selected variables and the clustering in this study. Secondly, data mining and logistic regression were used to screen out a total of 8 important ecological variations of crime cluster. The results show that the characteristics of burglary coldspots in Taipei city are in high-income areas. The vacancy rate and the proportion of individual households are relatively high, resulting in reduced surveillance and easy to become burglary hotspots. There are more police forces in crime hotspots, while street lamps and monitors have no obvious effect on burglary prevention.

To sum up the above research results, this study suggests that special police planning should be developed for different crime hotspots, then prospective scan statistics should be used to construct a burglary timely monitoring and early warning

system, and based on the maximum probability of crime and victim risk data to allocate police resources to enhance the effectiveness of crime prevention.

Keywords : scan statistic, environmental criminology, residential burglary, data mining

目次

謝誌	i
摘要	iii
Abstract.....	v
第一章 緒論	1
第一節 研究背景.....	1
第二節 研究動機.....	5
第三節 研究目的.....	8
第四節 名詞釋義.....	9
第二章 文獻探討	13
第一節 犯罪活動的時空特性.....	13
第二節 犯罪學相關理論.....	21
第三節 犯罪學理論與區位分析.....	35
第四節 資料探勘與犯罪分析.....	44
第五節 實證研究探討.....	49
第六節 綜合評析.....	57
第三章 研究設計	65
第一節 研究流程與概念架構.....	65
第二節 研究範圍與資料蒐整.....	69
第三節 變項選取與分析策略.....	73
第四節 研究方法與工具.....	84
第五節 研究倫理.....	116
第四章 臺北市住宅竊盜時空分析	119
第一節 臺北市住宅竊盜犯罪資料探索分析.....	119
第二節 竊盜犯罪資料空間探索分析.....	138
第三節 空間掃描統計分析.....	146
第四節 時空掃描統計分析.....	173
第五章 臺北市住宅竊盜區位分析	195
第一節 區位分析資料前置處理.....	195
第二節 臺北市住宅竊盜區位變項分析.....	200
第三節 模型評比與變項篩選.....	221

第四節 羅吉斯迴歸統計分析.....	228
第五節 區位變項理論意涵與實務意義.....	237
第六章 研究結論與建議.....	247
第一節 研究結論.....	247
第二節 研究建議.....	258
第三節 研究限制.....	262
第四節 結語.....	263
參考文獻.....	265
中文部分.....	265
外文部分.....	272

表次

表 2-5-1	相關實證研究文獻整理摘要表.....	50
表 3-2-1	臺北市政府住宅竊盜犯罪資料示意表.....	71
表 3-3-1	犯罪區位理論自變項表.....	80
表 3-4-1	二元分類常用的指標名稱和計算方式.....	111
表 4-1-1	2015-2017 年臺北市住宅竊盜發生數	120
表 4-1-2	2015-2017 年臺北市住宅竊盜月發生數	121
表 4-1-3	2015-2017 年臺北市住宅竊盜星期別發生數	122
表 4-1-4	2015-2017 年臺北市住宅竊盜每日時段發生數	124
表 4-1-5	2015-2017 年臺北市住宅竊盜各行政區發生數	126
表 4-1-6	2015-2017 年臺北市各行政區住宅竊盜數常態分布檢定	128
表 4-1-7	2015-2017 年臺北市住宅竊盜數變異數同質性檢定	128
表 4-1-8	2015-2017 年臺北市住宅竊盜數量變異數分析	128
表 4-1-9	2015-2017 年臺北市各行政區住宅竊盜數多重比較	129
表 4-1-10	2015-2017 年臺北市住宅竊盜發生數、里數量分布統計	131
表 4-1-11	2015-2017 年臺北市住宅竊盜各行政區失竊率.....	132
表 4-1-12	2015-2017 年臺北市住宅竊盜各里失竊率占比	134
表 4-2-1	2015-2017 年臺北市住宅竊盜犯罪全域式群聚分析表	144
表 4-3-1	2015 年臺北市住宅竊盜圓形空間掃描摘要表.....	150
表 4-3-2	2015 年臺北市住宅竊盜圓形空間群聚區(里).....	150
表 4-3-3	2016 年臺北市住宅竊盜圓形空間掃描摘要表.....	153
表 4-3-4	2016 年臺北市住宅竊盜圓形空間群聚區(里).....	153
表 4-3-5	2017 年臺北市住宅竊盜圓形空間掃描摘要表.....	155
表 4-3-6	2017 年臺北市住宅竊盜圓形空間群聚區(里).....	155
表 4-3-7	2015-2017 年臺北市住宅竊盜圓形空間掃描摘要表	157
表 4-3-8	2015-2017 年臺北市住宅竊盜圓形空間群聚區(里).....	157
表 4-3-9	2015 年臺北市住宅竊盜橢圓形空間掃描摘要表.....	159
表 4-3-10	2015 年臺北市住宅竊盜橢圓形空間群聚區(里).....	160
表 4-3-11	2016 年臺北市住宅竊盜橢圓形空間掃描摘要表	162
表 4-3-12	2016 年臺北市住宅竊盜橢圓形空間群聚區(里).....	162
表 4-3-13	2017 年臺北市住宅竊盜橢圓形空間掃描摘要表.....	164

表 4-3-14	2017 年臺北市住宅竊盜橢圓形空間群聚區(里).....	164
表 4-3-15	2015-2017 年臺北市住宅竊盜橢圓形空間掃描摘要表	166
表 4-3-16	2015-2017 年臺北市住宅竊盜橢圓形空間群聚區(里).....	166
表 4-3-17	臺北市住宅竊盜熱點分類標準表.....	170
表 4-3-18	臺北市住宅竊盜熱點性質分類.....	171
表 4-4-1	臺北市住宅竊盜「年」掃描摘要表.....	174
表 4-4-2	各年度「月」掃描摘要表.....	174
表 4-4-3	2015-2017 年「月」掃描摘要表	175
表 4-4-4	各年度「日」掃描摘要表掃描摘要表.....	175
表 4-4-5	2015-2017 年「日」掃描摘要表	176
表 4-4-6	臺北市住宅竊盜季節性「月」掃描摘要表.....	177
表 4-4-7	臺北市住宅竊盜季節性「日」掃描摘要表.....	177
表 4-4-8	2015-2017 年臺北市住宅竊盜圓形時空掃描摘要表	181
表 4-4-9	2015-2017 年臺北市住宅竊盜時空群聚區(里).....	182
表 4-4-10	異常趨勢之群聚摘要表.....	185
表 4-4-11	異常趨勢之群聚區(里).....	185
表 4-4-12	2017 年 9 月 30 日預警發布結果.....	189
表 4-4-13	2017 年 9 月 30 日預警發布群聚區(里).....	189
表 4-4-14	2017 年 10 月 31 日預警發布結果.....	190
表 4-4-15	2017 年 10 月 31 日預警發布群聚區(里).....	190
表 4-4-16	2017 年 11 月 30 日預警發布結果.....	191
表 4-4-17	2017 年 11 月 30 日預警發布群聚區(里).....	191
表 4-4-18	2017 年「前瞻性」時空掃描結果.....	192
表 4-4-19	2017 年「前瞻性」時空掃描群聚區(里).....	192
表 4-4-20	2017 年「回顧性」時空掃描結果.....	193
表 4-4-21	2017 年「回顧性」時空掃描群聚區(里).....	194
表 5-1-1	2015-2017 年臺北市住宅竊盜冷熱群聚區摘要表	196
表 5-1-2	2015-2017 年臺北市住宅竊盜冷熱群聚區(里).....	196
表 5-1-3	臺北市住宅竊盜區位變項敘述統計表.....	198
表 5-1-4	臺北市住宅竊盜群聚共變項檢定.....	199
表 5-1-5	區位變項共線性檢定.....	200

表 5-2-1	臺北市住宅竊盜冷熱區決策樹分類規則表.....	202
表 5-2-2	ROC 操作特徵曲線評估表.....	204
表 5-2-3	調整後決策樹分類規則表.....	205
表 5-2-4	隨機森林相對重要變項排序表.....	210
表 5-2-5	隨機森林訓練樣本分類錯誤率評估表.....	212
表 5-2-6	調整後隨機森林訓練樣本分類錯誤率評估表.....	212
表 5-2-7	調整後隨機森林相對重要變項排序表.....	214
表 5-2-8	XGBOOST 相對重要變項排序表.....	219
表 5-2-9	XGBOOST 疊代次數與訓練錯誤率關係表.....	220
表 5-3-1	傳統決策樹模型錯誤矩陣次數表.....	222
表 5-3-2	傳統決策樹模型錯誤矩陣百分比表.....	222
表 5-3-3	條件推論決策樹模型錯誤矩陣百分比表.....	223
表 5-3-4	隨機森林模型錯誤矩陣百分比表.....	223
表 5-3-5	ADABOOST 模型錯誤矩陣百分比表.....	223
表 5-3-6	XGBOOST 模型錯誤矩陣百分比表.....	224
表 5-3-7	各模型 AUC 值比較.....	224
表 5-3-8	各類樣本集錯誤矩陣評估比較.....	225
表 5-3-9	各類樣本集 AUC 值評估比較.....	226
表 5-3-10	各模型相對重要變項篩選結果.....	227
表 5-3-11	變項篩選後各模型 AUC 值評估.....	228
表 5-4-1	COX & SNELL RSQUARE 與 NAGELKERKE R QUARE 適配度檢定.....	229
表 5-4-2	整體模型 HL 適配度檢定.....	230
表 5-4-3	整體模型變項統計檢定表.....	232
表 5-4-4	變項篩選統計檢定表.....	233
表 5-4-5	變項篩選前模型錯誤矩陣表.....	234
表 5-4-6	變項篩選後模型錯誤矩陣表.....	234
表 5-4-7	各類樣本集錯誤矩陣評估比較.....	235
表 5-4-8	各類樣本集錯誤矩陣評估比較.....	235
表 5-5-1	區位變項相關矩陣.....	239
表 6-1-1	住宅竊盜犯罪相關實證研究與本研究結果比較分析表.....	254

圖次

圖 1-1-1	臺灣地區 2014 年至 2019 年竊盜發生數.....	2
圖 1-1-2	臺灣地區 2019 年各類竊盜發生比例.....	3
圖 2-1-1	空間點型態分布.....	17
圖 2-2-1	犯罪基本三角.....	25
圖 2-2-2	新機會理論關係圖.....	28
圖 2-2-3	BURGESS 的都市同心圓圖.....	30
圖 2-2-4	破窗理論流程圖.....	32
圖 2-2-5	被害者的生活方式暴露模型.....	35
圖 2-3-1	資料探勘的步驟.....	48
圖 3-1-1	研究流程圖.....	66
圖 3-1-2	研究概念架構圖.....	68
圖 3-3-1	研究概念架構圖(含變項).....	82
圖 3-3-2	分析策略圖.....	83
圖 3-4-1	空間掃描視窗變化.....	88
圖 3-4-2	區域中心點的逐步掃描方式.....	88
圖 3-4-3	時間掃描方式.....	89
圖 3-4-4	時空掃描方式.....	91
圖 3-4-5	時空掃描示意圖.....	91
圖 3-4-6	決策樹節點圖.....	97
圖 3-4-7	條件推論決策樹圖.....	101
圖 3-4-8	隨機森林 BAGGING 原理圖.....	104
圖 3-4-9	BOOSTING 分類器模型權重波動圖.....	106
圖 3-4-10	推進法分類器組合圖.....	107
圖 3-4-11	靈敏性與特異性對應值.....	112
圖 3-4-12	ROC 曲線與 AUC 面積圖.....	113
圖 4-1-1	2015-2017 年臺北市住宅竊盜年分布圖.....	120
圖 4-1-2	2015-2017 年臺北市住宅竊盜月分布圖.....	121
圖 4-1-3	2015-2017 年臺北市住宅竊盜周分布圖.....	123
圖 4-1-4	2015-2017 年臺北市住宅竊盜時段分布圖.....	124
圖 4-1-5	2015-2017 年臺北市各行政區住宅竊盜數.....	127

圖 4-1-6	2015-2017 年臺北市各行政區住宅竊盜數	131
圖 4-1-7	2015-2017 年臺北市各行政區住宅竊盜失竊率	133
圖 4-1-8	2015-2017 年臺北市各行政區住宅竊盜失竊數主題圖	135
圖 4-1-9	2015-2017 年臺北市各行政區住宅竊盜失竊率主題圖	135
圖 4-1-10	2015-2017 年臺北市各里住宅竊盜失竊數主題圖	136
圖 4-1-11	2015-2017 年臺北市各里住宅竊盜失竊率主題圖	136
圖 4-2-1	2015-2017 年臺北市住宅竊盜點位圖	142
圖 4-2-2	2015-2017 年臺北市住宅竊盜熱區圖	143
圖 4-3-1	2015 年臺北市住宅竊空間群聚圖	151
圖 4-3-2	2015 年臺北市住宅竊空間群聚圖	152
圖 4-3-3	2016 年臺北市住宅竊空間群聚圖	154
圖 4-3-4	2017 年臺北市住宅竊空間群聚圖	156
圖 4-3-5	2015-2017 年臺北市住宅竊空間群聚圖	158
圖 4-3-6	2015 年臺北市住宅竊橢圓空間群聚圖	161
圖 4-3-7	2016 年臺北市住宅竊橢圓空間群聚圖	163
圖 4-3-8	2017 年臺北市住宅竊橢圓空間群聚圖	165
圖 4-3-9	2015-2017 年臺北市住宅竊橢圓空間群聚圖	167
圖 4-3-10	臺北市住宅竊圓形與橢圓比較圖	168
圖 4-4-1	2015-2017 年「月」掃描群聚月份圖	175
圖 4-4-2	2015-2017 年「日」掃描群聚月份圖	177
圖 4-4-3	2015-2017 年「日」掃描群聚月份圖	178
圖 4-4-4	2015-2017 年時空掃描群聚圖	182
圖 4-4-5	監測預警系統架構圖	188
圖 5-2-1	臺北市住宅竊盜冷熱區決策樹分類規則圖	201
圖 5-2-2	調整後決策樹分類規則圖	205
圖 5-2-3	條件推論決策樹分類規則圖	208
圖 5-2-4	隨機森林相對重要變項排序圖	211
圖 5-2-5	調整前隨機森林袋外樣本 ROC 圖	213
圖 5-2-6	調整後隨機森林袋外樣本 ROC 圖	213
圖 5-2-7	隨機森林建樹量與錯誤率分析圖	215
圖 5-2-8	ADABOOST 推進法疊代次數與錯誤率相關圖	216

圖 5-2-9	ADABOOST 推進法相對重要變項圖	217
圖 5-2-10	XGBOOST 相對重要變項圖	219
圖 5-2-11	XGBOOST 疊代次數與訓練錯誤率關係圖.....	220
圖 5-4-1	各模型 ROC 曲線圖	236
圖 5-5-1	群聚冷熱區之區位變項圖.....	246

第一章 緒論

第一節 研究背景

竊盜犯罪 (larceny) 屬財產犯罪之一種，一般認為其雖不若強盜與搶奪犯罪對於被害人造成立即身體危害，但事實上此類犯罪形成之財物損失與被害人之心理影響，卻可能十分重大與深遠，同時亦可能衍生強盜、強制性交，甚或殺人案件，故各界對於此類犯罪之研究與防制從未懈怠。

長久以來，竊盜犯罪率在世界各國始終高居不下，在我國亦占全般刑案之多數，但在 2015 年聯合國第一 3 屆預防犯罪刑事司法會議上 (Crime Prevention and Criminal Justice Convention) 發布會員國的犯罪率調查結果卻顯示 2003 年至 2013 年間全球各國暴力、財產、毒品等犯罪率整體均呈現下降趨勢，其中最為明顯的是汽機車竊盜罪，降幅達二分之一，住宅竊盜罪也下降四分之一(周愷嫻，2017)，與此同時我國竊盜犯罪率亦有同步減緩趨勢，根據許春金教授等學者統計，在 2004 至 2013 年間，臺灣地區汽車竊盜犯罪率降低 87%，機車竊盜降低 86%，一般竊盜 (含住宅竊盜) 降低 53%，比起世界各國下降趨勢更為明顯(許春金、陳玉書、蔡田木、洪千涵、白鎮福，2015)。2013 年後迄今，我國竊盜犯罪數量持續下降，在數量方面，2014 年發生件數為 76,330 件，2015 年為 66,255 件，2016 年為 57,606 件，2017 年為 52,025 件，2018 年 47,591 件，2019 年更持續下降至 42,272 件(如圖 1-1-1)，但在全般犯罪數量中，仍占極高比例，在 2019 年約為 15.75%，僅次於酒後駕車的 19.94% 與違反毒品危害防制條例的 17.53% (警政署，108 年警政統計年報)。

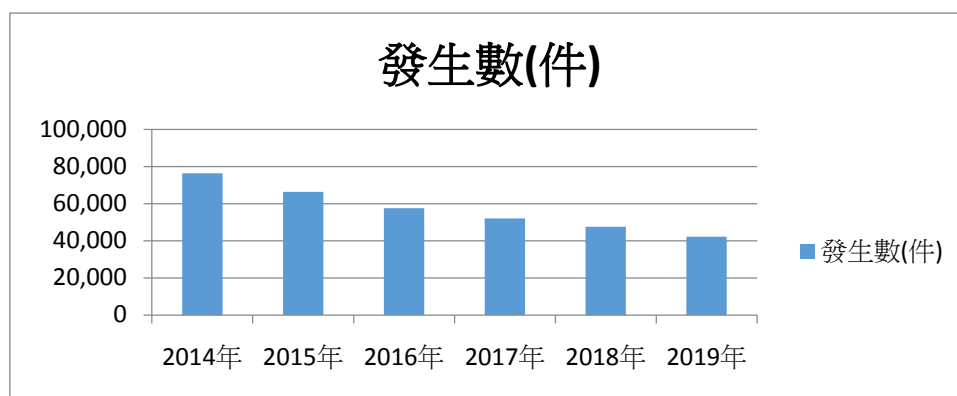


圖 1-1-1 臺灣地區 2014 年至 2019 年竊盜發生數

資料來源：警政統計查詢網

在竊盜種類方面，現階段警政署將竊盜類型區分為 4 類，分別為「重大竊盜」(serious larceny)、「普通竊盜」(general larceny)、「機車竊盜」(motorcycle theft)、「汽車竊盜」(motor vehicle theft)，相關定義如次(內政部警政署警政統計名詞定義, 2019)：

一、重大竊盜案件指具有下列情形之一者：

- (一)、失竊物總值 100 萬元以上竊案。
- (二)、竊盜槍械、軍火、爆裂物、或國防上、交通上、學術上之重要設施、器材。
- (三)、被竊人係具外交身分之外籍人員，或來訪之外籍貴賓。
- (四)、竊盜重要儀器、文件等影響國家與社會安全情節重大之竊案。

重大竊盜自 2016 年 3 月 24 日起，修正失竊物金額總價值，由 50 萬元提高至 100 萬元以上之竊案，並刪除竊盜保險箱、櫃內之財物總值 10 萬元以上竊案。

二、普通竊盜案件：指意圖為自己或第三人不法之所有，而竊取他人動產之犯罪行為。不含重大竊盜及車輛竊盜。

三、汽車竊盜案件：指意圖為自己或第三人不法之所有，而竊取他人之汽車者。
(不含機車竊盜)

四、機車竊盜案件：指意圖為自己或第三人不法之所有，而竊取他人之機車者。

此 4 類案件中，以 2019 年統計為例，以普通竊盜 34,725 件，約占 82% 最高，

其次為機車竊盜 6,058 件，約占 14% 次之，汽車竊盜 1,456 件，約占 4% 第三，重大竊盜 33 件，則不足 0.01%（如圖 1-1-2）。

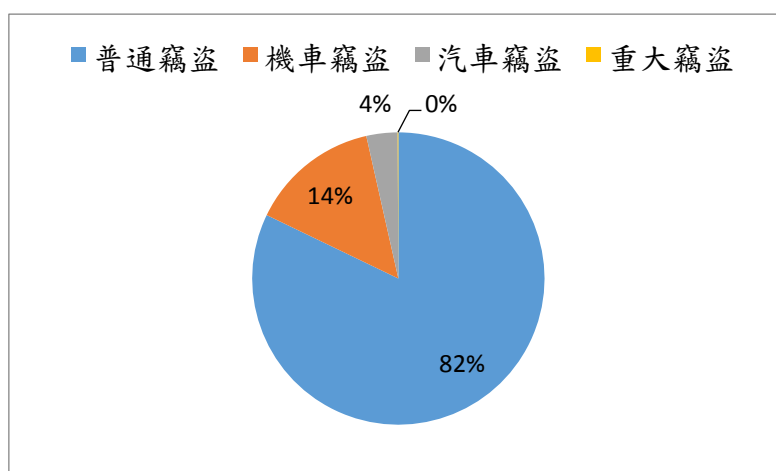


圖 1-1-2 臺灣地區 2019 年各類竊盜發生比例

資料來源：內政部警政署統計年報(2019 年)

根據由警政署刑事警察局出版的 2016 年中華民國刑案統計資料分析，其對於竊盜犯罪之分類主要分為 3 大類，分別為「一般竊盜」、「汽車竊盜」與「機車竊盜」，其中以「一般竊盜」為最多，而所謂「一般竊盜」係指汽、機車以外之竊盜，其財產損失之金額（價）額以未滿新臺 100 萬元為限(內政部警政署刑事警察局, 2016)，其與上述普通竊盜意義相當。而在各類型竊盜案件分類中，以「民生竊盜」(living larceny) 為最直接影響民眾的竊盜案件。所謂「民生竊盜」包括住宅竊盜、公用設施、車輛零件及車內物品、電機及農漁牧機具、農漁牧產品、電纜線 6 項¹；觀察 2007 年至 2019 年間發展趨勢主要有 2 大特點(內政部警政署警政統計通報，2019)：

- (一) 10 餘年來民生竊盜案件發生數呈現逐年下降趨勢，2007 年發生 35,416 件最多，2019 年 8 月止已降至 4,047 件，較上年同期續減 495 件(-10.84%)。
- (二) 依案類別觀察，發生數以「住宅竊盜」為最多，歷年占總數超過 5 成 4 以上，其次則為「車輛零件及車內物品竊盜」。

¹ 自 2006 年 3 月 15 日起實施「改善治安強化作為」專案，即擇定相關條件統計民生竊盜，另自 2009 年 1 月 10 月 1 日起擇取條件已修正，公用設施、車輛零件及車內物品、電機、農具等項與以前擇取項目不同。

綜上所知，可清楚瞭解竊盜犯罪雖在全球及我國皆有大幅下降趨勢，但在各類犯罪類型中仍占有相當高比例；此外，在民生竊盜犯罪中「住宅竊盜」最為嚴重，故多年來警政署除將民生竊盜一直列為重點工作外，並要求所屬持續執行「提升住宅竊盜偵防效能執行計畫」加強偵防住宅竊盜，同時落實住宅防竊諮詢工作，以提升民眾居家安全防範意識及防竊能力。經多年來之努力，住宅竊盜之防治已逐漸顯現成效，根據內政部警政署統計數據指出，2018 年我國住宅竊盜發生數持續呈逐年減少趨勢，為 3,682 件，與 2014 年相較減低 1,797 件，減幅達 32.1%，而破獲率則是逐年提升，由 2014 年的 84.14% 上升至 2018 年的 92.53%（如圖 1-1-3）。

總之，保護社會安全，防止一切危害，使民眾免於恐懼，不但是民眾所期盼，更是政府重要施政工作，而民眾對治安的感受與評價，主要取決於自身週遭生命財產權益是否被侵害而定，而減少住宅竊盜犯罪以落實居住安全更是重中之重，臺北市政府警局甚至成立「治安風水師專區」，提供市民住宅防竊安全檢測的服務，足見對於此類型犯罪的重視，故本研究即以臺北市住宅竊盜犯罪作為分析對象。

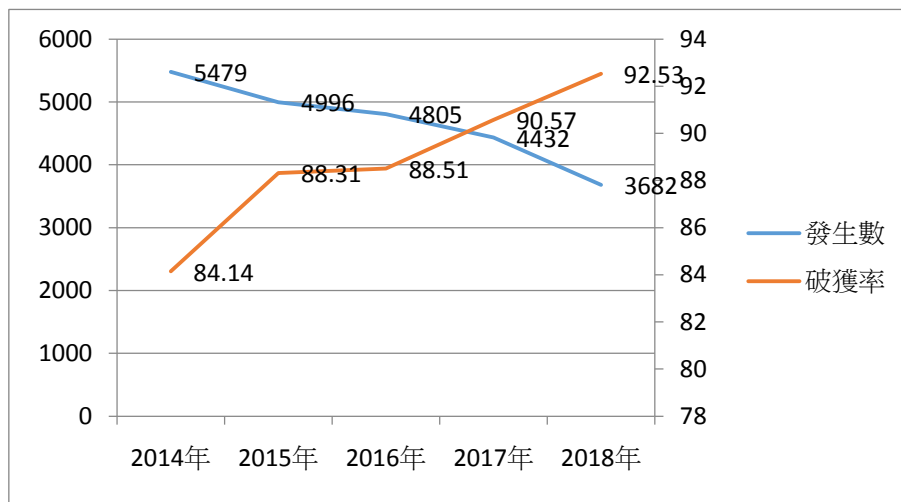


圖 1-1-3 臺灣地區 2014 年至 2019 年住宅竊盜發生數及破獲率

資料來源：警政統計通報(2019 年)

第二節 研究動機

一、地理資訊系統與資料探勘的整合

關於犯罪研究，自十八世紀以來歷經許多階段之演變。從古典犯罪學派伊始，歷經實證犯罪學派、新古典犯罪學派，各學派之爭議，莫不在於找尋犯罪預防之最佳方法；惟其關注焦點多在於刑事政策及刑事司法層面問題。迨 1980 年後，因歐美政治、經濟、社會結構產生巨大變化，刑事政策之典範遂由犯罪後刑事司法處理，轉移至注重各人（社區）犯罪事前預防系統之建置，加以統計技術日益發達，故犯罪學研究焦點由犯罪個人與環境交互影響之綜合因素，朝向「情境犯罪」研究，配合數據資料分析，成為犯罪學近年研究之趨勢。

近年，各國政府大量開放各類官方統計資料，加以電腦軟硬體日益發達，越來越多的犯罪研究傾向尋求與犯罪相關之環境硬體及社經因素，使得犯罪學分析更邁向重視數據之研究與科技之運用；在時代科技推動下，犯罪學分析技術不再囿於傳統報表解釋或統計分析，呈現出百花齊放的榮景，其中地理資訊系統與資料探勘技術的運用日益普及，除開拓各類犯罪議題的研究視野外，對於犯罪防治也帶來巨大貢獻。

首先，就地理資訊系統發展沿革來看，地理資訊系統(geographic information system, GIS)萌芽於 1960 年代，經美國政府大力推動，技術日趨成熟。上世紀末，美國政府在投入大量經費在政府部門犯罪製圖的發展；其中，成立於 1997 年的美國國家司法學院(The National Institute of Justice, NIJ)犯罪製圖研究中心 CMRC(Crime Mapping Research Center)在 2002 年改制為公共安全犯罪製圖與分析計劃(Mapping and Analysis for Public Safety Program, MAPS)，MAPS 對於美國公共治安及犯罪製圖影響深遠，除美國外亦影響其它國家與地區，MAPS 除定期舉行研討會外，亦研發許多與犯罪製圖相關之應用與軟體，協助各地警方勤務的派遣、打擊犯罪及減少犯罪發生(黃俊能，2008)

在我國方面，地理資訊系統運用於警政治安略晚於歐美國家，但自 2008 年起警政署及各縣市警察局陸續規劃建置「犯罪地圖資訊管理系統」，結合犯罪資料庫及地理資訊系統，以「強盜」、「搶奪」及「竊盜」3 項案類作分析比較，製作犯罪地圖，充分掌握地區刑案發生的空間分布型態，辨識治安熱區，及各轄區案件變化趨勢及區位特性，進而研析轄區治安狀況，藉以強化犯罪熱區加強勤務

部署及警力巡邏，有效提升犯罪防制成效(許華孚、吳吉裕，2015)。此外，我國也仿效外國做法，將地理資訊系統技術導入 e 化社區警政，結合電子地圖與警政機關資料庫，已經可以針對轄區內所發生的各種犯罪案件加以管理、統計、分析，同時民眾可以透過網路查詢，清楚掌握住宅周邊竊案發生分布，進而採取防盜措施，有效降低竊盜案件，為民眾居家安全提供更多保障。

值得注意的是，除了地理資訊系統的快速發展外，自上世紀末以來，隨著科技進步所帶動的相關資訊發展，各類結構化與非結構化資料正在迅速膨脹增加，「大數據」(big data) 概念孕育而生，相關分析技術的運用為各行各業與學術領域帶來無限機遇，但同時也形成諸多挑戰。

所謂的「大數據」或稱之為「巨量資料」，這個詞最早由 IBM 公司提出，2010 年才真正開始受到注目，並成為專業用語登載於《維基百科》，算是「大數據」的正式問世；2012 年，《紐約時報》(The New York Times) 的專欄文章” The Age of Big Data” 更是宣告「大數據時代」的來臨(Lohr, 2012)。一般來說，「大數據」涵蓋的範圍很廣，定義也各家歧異，目前多將其特性歸類為「3Vs」或「4Vs」，即：資料量(volume)、資料傳輸速度(velocity)、資料類型(variety)，以及後來提出的第四個 V—真實性(veracity)(鍾孝宇，2017)；目前，許多人所談論的「大數據」不僅只是資料，也泛指用來分析、處理巨量資料的新興科技技術，如：資料倉儲(data warehouse)架構的建置、雲端計算(cloud computing)、資料探勘(data mining) 等分析技術。

在用途方面，根據調查顯示，「大數據」目前最常使用在電信、金融、政務、交通和醫療 5 大領域(袁帥，2017)，但近年來亦被廣泛應用到安全執法的過程當中，在犯罪學等各領域已有顯著研究成果，如：預測犯罪熱點、視覺化組織犯罪人脈圖、辨識連續犯、犯罪特性分群、預測網路犯罪、改善鑑識成效、辦案筆錄文本分析，與使用社群網路或手機資料即時預測犯罪等；實際成效如：亞利桑那州的圖森(Tucson)警察局使用 IBM 公司所設計的 COPLIN 系統，分析自 1986 年到 1999 年約一百五十萬筆犯罪資料，有效防治犯罪(Nissan, 2012)；而洛杉磯警局與加州大學洛杉磯分校合作，採集分析 80 年來約一千三百萬起犯罪案件，用於進行犯罪行為的大型研究，透過大數據演算法預測，成功將相關區域的犯罪率降低 36%(張瑋倫，2016)。因此，可說只要有大量資料的地方，大數據相關分析技術即有運用餘地。

目前，許多人對於大數據與資料探勘並沒有太嚴格的區分，事實上兩者概念仍有若干差異。所謂大數據正如上所述，由於資料量規模巨大，致使無法透過人工，在合理時間內達到擷取、管理、處理、並整理成為人類所能解讀的資訊。在總資料量相同的情況下，與個別分析獨立的小型資料集(data set)相比，將各個小型資料集合併後進行分析可得出許多額外的資訊和資料關聯性，可用來察覺商業趨勢、避免疾病擴散、打擊犯罪或測定交通路況等；這樣的用途正是大型資料集成為風潮的原因。

而資料探勘則是從大數據資料庫發現知識(knowledge-discovery in databases, KDD)中的一個步驟及技術。資料探勘一般是指從大量的資料中自動搜尋隱藏於其中的有著特殊關聯性(association rule learning)的資訊的過程，透過統計、線上分析處理、情報檢索、機器學習、專家系統(依靠過去的經驗法則)和模式識別等諸多方法來實現上述目標。

由於地理資訊系統與資料探勘技術近年來均已被證明對於犯罪防治有明顯成效，故本研究之動機之一，就在於結合此 2 項技術之優點，希藉由資料探勘的技術，對於住宅竊盜犯罪形成之各種社經因素提出模型分類；換言之，除辨識犯罪熱區外，再加上對於犯罪發生因素的分析，對住宅竊盜犯罪之防治將有更多助益。

二、犯罪空間與時間群聚分析的整合

目前在地理資訊系統的運用中，其對於犯罪案件的分析普遍聚焦在犯罪的空間分布，分析事件是否呈現群聚及其發生的頻率、強度等關係(溫在弘、劉擇昌、林民浩，2010)，因而犯罪是否呈現群聚的「犯罪熱區」(hot spots)成為地理資訊系統分析相當重要一環，但卻常常忽略時間與空間的交互作用問題，即便若干研究將時間因素納入，也往往只是當成一個「變項」處理，雖進行統計檢定(如卡方適合度檢定)，但在分析層面上只能顯示犯罪的時間分布是否有顯著差異，而若干研究甚至將時間與空間資料分開處理，僅將犯罪時間分析作為研究的背景資料，造成此類分析通常僅能分階段展示不同時間的空間群聚現象，未能提供進一步的資訊，如案件是否有明顯集中於地區、某個時段，案件的空間分布與時間的關係等(Grubestic & Mack, 2008；廖興中、徐明莉，2017)，本質上仍屬於空間性質之分析，與本研究所稱之「時空分析」(spatio-temporal analysis)有所不同。故

目前越來越多的研究文獻以 Hägerstrand 於 1970 年提出的時間地理學(time geography)做為研究架構，企圖同時由時間與空間面向切入，發展出新式的時空分析方法來研究人類行為(Bhunja & Shit, 2019)，各類地理資訊系統軟體亦致力於時空模型開發，顯示對此問題之興趣與重視，也同時凸顯時空整合研究的困難；揆諸國內外運用地理資訊系統的犯罪學研究，「時空分析」仍屬少見，故如何將時間因素同時納入到空間分析中，並觀察空間在時間趨勢的變化，為本文另一研究動機。

第三節 研究目的

由前述得知，我國竊盜犯罪雖逐年下降，但住宅竊盜依然相對顯得嚴峻，近年來警政單位提出各項防竊肅竊政策與方案，對於防制住宅竊盜確獲致一定成效，但若能夠對於住宅竊盜的時空分布與特性有更多的瞭解，相信對於提升警政效能並減輕警力的負擔必能有所裨益。基此，本研究欲達成之研究目的如下：

一、運用掃描統計與 GIS 有效辨識犯罪時空熱區

犯罪分析之主要目的在於系統地研究犯罪型態及趨勢，以協助司法警政系統打擊犯罪並建立預防策略；然而，過去許多有關犯罪群聚的研究多偏向空間面向的研究，壓縮或忽略時間的重要性，導致許多犯罪資訊的流失。故本研究將以臺北市政府公開資料平臺 2015 年至 2017 年的住宅竊盜統計數據為依據，運用時空掃描統計(space-time scan statistics)方法進行時空群聚分析，並引用 GIS 製圖技術對臺北市住宅竊盜犯罪分布情形予以視覺化，目的在於協助警方迅速有效辨識犯罪時空熱區，俾利於將有限的警力資源配置在合適的地點與正確的時間以嚇阻犯罪，藉此提升執法與犯罪預防工作效能。

二、應用資料探勘技術探討犯罪熱區之區位因素

目前國內外整合地理資訊系統與資料探勘技術於社會科學領域面的研究雖日益增加，但在犯罪防治領域仍相對較為較少，對於造成犯罪熱區之區位因素分析多仍運用傳統統計方式，著重在因素間的因果關係，與資料探勘重視因素間的關聯性有所不同；因此，本研究將根據犯罪學理論與過去相關文獻指引，蒐集相關之各類人文社經變項，除採用統計分析方式外，另將應用資料探勘技術對臺北

市住宅竊盜進行區位分析，希能協助警方有效掌握高犯罪率地區的不同區位因素，進而能制定更有效的犯罪預防策略。

三、計算熱區風險值俾於警政資源與勤務規劃

不同犯罪型態皆有其不同之時空特性與發生原因。在住宅竊盜方面，根據過去的研究顯示，竊賊利用民眾上班、上學或外出後，乘隙侵入行竊的情形居多；而從季節性觀察，元旦、春節及元宵節期間連續長假期間，都會區民眾返鄉及出遊人潮湧入鄉間，皆會造成都會市區入宅竊盜案增加；因此，本研究另一目的乃針對住宅竊盜發生密集區域，除分析其犯罪時空與區位特性，研提著重犯罪預防的對策外，且本研究所採之研究方法，經由計算相對風險值，可進一步確認各區域住宅竊盜風險排序；換言之，根據對此風險值大小之評估，警方對於勤務與投入資源規劃的輕重緩急將更能有效掌握，及時回應民眾需求，希對於降低住宅竊盜的發生有實質裨益。

第四節 名詞釋義

一、住宅竊盜犯罪

根據我國《刑法》第三 20 條第一項規定「意圖為自己或第三人不法之所有，而竊取他人之動產者，為竊盜罪」；若由發生場所區分，竊盜案件依發生場所主要可區分為「住宅」、「市街商店」、「特定營業場所」、「交通場所」、「文教衛生機構」、「金融證券機構」、「郊區及其他場所」等 7 大類範疇。而所謂住宅竊盜，係指刑案紀錄表中竊盜案件發生場所為住宅，其範圍為普通住宅(1、2 層樓)、公寓(3 至 5 層建築物)、大廈(6 層以上建築物)、別墅(獨立式、含雙併)、透天厝、農家住宅、出租公寓(套房)、宿舍、空屋、集合住宅及其他住宅 11 項(行政院主計總處，2018)。

本研究所指之「住宅竊盜犯罪」乃依據上述我國法律定義與官方統計資料分類方式，係指發生於 11 項屬於「住宅」場所範疇內，其行為人意圖為自己或第三人不法之所有，而竊取他人之動產者之犯罪行為。

二、犯罪時空熱區

美國地理學家 Waldo R. Tobler 於 1970 年提出「地理學第一定律」(Tobler's

First Law 或 Tobler's First Law of Geography, TFL), 認為任何事物都相關, 相近的事物關聯更緊密², 換言之, 人類行為與所在地區有絕對的影響, 此種影響隨著距離的加大而遞減, 如果加進時間的維度, 此種影響有時間間距, 不是立即發生, 此即「空間擴散作用」(Miller, 2004)。而此距離產生影響之空間因素表現在「空間自相關」(spatial auto-correlation)與「空間異質」(spatial heterogeneity)兩個概念上, 空間自相關是認為「鄰近地區的影響」(neighborhood effect)的大小, 空間異質性指空間位置差異造成的觀察行為相異現象。

基於對「空間自相關」與「空間異質」的理解, 一般對於犯罪熱區的定義, 指的是犯罪地點呈現高度群聚趨勢, 表示在空間上有顯著地集中分布於某一處或是多處, 從而顯現與其他地區不同的異常現象; 由此可知, 熱區與群聚實為一體兩面, 犯罪熱區的形成為犯罪地點群聚的結果, 前者可視為一種現象, 後者著重在形成的過程, 故在本研究中熱區與群聚在描述時經常交互使用。

因此, 犯罪地點是否為熱區, 可藉由對於「空間自相關」的統計檢定, 評估鄰近地區屬性是否有顯著相似的特性, 判斷空間群聚之趨勢, 從而辨識具統計意義之犯罪空間熱區。值得注意的是, 上述所提到的「空間的擴散作用」已指出空間的變化實無法忽略時間的維度的影響, 故若再加上時間面向, 本研究指稱之「犯罪時空熱區」是指「犯罪地點與時間經統計檢定同時發生群聚現象」, 所使用的統計檢定方式為「時空掃描統計」。

三、時空掃描統計

自從時空群聚分析的概念被提出以來, 多年來各種方法不斷推出, 依據各類偵測群聚方法的差異, 現有的時空群聚分析方法大致可以分為 3 種類型: (一) 時空掃描統計; (二) 時空密度群聚(density-based clustering); (三) 時空混合距離(hybrid distance)(唐建波、鄧敏、劉啟亮, 2013)。由於本研究係使用時空掃描統計方法, 故其餘 2 種不擬多做介紹。

時空掃描統計旨在探測一定時空範圍內的群聚性與隨機分布模式, 經由統計檢定, 是否達到顯著水準, 以確定可能異常的群聚性時空事件。時空掃描最早之用途係由哈佛大學教授 Kulldorff 運用於探測傳染病的時空群聚模式; 該方法首先定義一系列以空間距離為半徑, 時間間隔為高的圓柱型時空掃描視窗; 然後, 針對每個掃描視窗, 依據人口數(母群體)與實際案例數計算理論案例數, 進行

² 原文為"Everything is related to everything else, but near things are more related to each other".

統計上之檢定（相關論述與統計原理請參閱第二章第一節及第三章第二節）。目前時空掃描已逐漸擴展到各領域使用，本研究主要以此方式用來偵測臺北市住宅竊盜之時空群聚現象。

四、資料探勘

資料探勘在本質上屬於跨學科的電腦科學分支，它的出現主要是因應資訊大量出現，形式不一致，難以統一處理與消化而出現的一種技術，故資料探勘與資料倉儲技術的發展與有著密切的關係；換言之，資料倉儲的發展是促進資料探勘越來越受重視的原因之一，過去資料倉儲系統可以很有效率實現資料的登錄、查詢、統計等功能，但無法發現數據中存在的關係和規則，故資料探勘就是為此運應而生。

資料探勘演算法經過多年的發展目前已經成為一種成熟、穩定，且易於理解和操作的技術，且不一定需依賴資料倉儲的資料來源；亦即，資料倉儲並不是資料探勘存在的先決條件，因為目前資料探勘可直接從各種資料來源中挖掘資訊。例如，在本研究中所使用的資料探勘技術主要以分類技術為主，資料來源為政府公開資料平臺的人文社經統計資料，但由於目前政府公開資料庫均為各機關將其可公開之統計資料建置於上，缺點為格式不一，無法具備如倉儲資料的相關功能，使用時仍必須仰賴人工整理，故本研究之資料探勘所指涉之範圍主要係資料探勘的演算法技術。

五、犯罪區位因素分析

區位學(ecology)為研究動植物間的競合與環境互動之關係的學門，美國芝加哥學派將區位學的方法使用在研究都市生態上，觀察都市居民與其生存環境的互動關係，建構出具有空間意識的人文區位學(human ecology)；其中，Park與Burgess於1921年以城市生態學(urban ecology)的視角，運用製圖方式，發現城市中的犯罪分布與環境有密切的相關；Shaw與McKay(1942)則承襲Burgess都市成長理論研究青少年犯罪問題，探討人文社經因素的區域性差異與犯罪率之關聯，因而產生「社會解組理論」，可視為犯罪區位學(criminological ecology)之代表。

將犯罪區位學與統計學等分析技術相結合即本研究所稱之「犯罪區位因素分析」(本研究簡稱為「犯罪區位分析」)，運用此分析方式，可得知某犯罪類型的

發生與犯罪區域之人文社經因素是否相關或相互影響，進而根據區位因素，擬定有效犯罪防治策略。本研究犯罪區位因素除依據社會解組理論外，另將以「新機會理論」做為主要理論，藉由兩項理論指引，選取區位變項，在社會解組理論方面包括社經地位，如收入、教育程度等，而新機會理論則從犯罪被害與阻絕犯罪機會等因素，如空屋率、獨居率、房地產價格及監視器、警力配置等進行分析。

第二章 文獻探討

第一節 犯罪活動的時空特性

任何一種犯罪皆為犯罪人在特定時空下所進行，且不同犯罪型態皆具有不同的時空特性，例如搶奪犯罪常發生在人潮與商業活動較多，多元土地使用地區，以及可及性較高的道路(張淑貞、李素馨，2012)，通常土地使用強度較高地方也意味著有較多受害目標物可供選擇和隱匿機會，便捷度高的街道則提供搶匪一個高可及性和逃逸機會；但土地使用強度高的地區和便捷度高的道路通常也表示有較高的自然監視性，而抑制犯罪發生，且有時擁塞的交通流量可能會影響逃逸(Bellamy, 1996)。此外，相關文獻亦指搶奪案件群聚在商業區、市場區、金融證券區及人口群聚複雜地區，發生搶奪案件時段以夜晚及下午最多(林進發，2005)。另有研究指出白天的暴力犯罪主要集中於商業購物區，夜晚的暴力主要集中在娛樂區，白天搶奪案件與金融密度有顯著正相關，此犯罪分布現象與店家營業時間及日常活動型態有關(湯巧君，2007)。

相對地，住宅竊盜與搶奪犯罪雖然均為財產犯罪之一種，因兩者在犯罪性質有所差異，故其具有之時空特性與分布便有所不同。首先，住宅的分布多非商業區與娛樂區，故在地理分布上可能就與搶奪犯罪不同；此外，搶奪犯罪為犯罪者與受害者直接接觸所造成，而住宅竊盜的犯罪者則採取隱匿的行動，故住宅竊盜犯罪較易發生之時段也將與搶奪犯罪形成差異，國內許多研究均指出，都會住宅竊盜發生，易發於民眾外出工作時段，此乃住宅缺乏監控力所致(陳惠貞，2008；彭福康，2015；劉擇昌，2011)。

由上可知，不同的犯罪型態有其相異的時空分布特性；相對地，不同時空環境下也將造成迥異的犯罪型態，時空與犯罪的各項因素相互交錯，更顯犯罪分析問題的複雜性。

一、犯罪熱區的意義與類型

針對犯罪時空特性之分析過程中，如何辨識「犯罪熱區」(crime hot spots)為首重議題。所謂犯罪熱區也有稱之為犯罪群聚(crime cluster)，通常指犯罪集中且受到關注之區域，其可為特殊之地址、街區亦或是連續犯罪之發生地區(劉擇昌，2011)。Lersch(2004)則將重複受害稱為「熱點」，並將熱點和鄰近重複組合

成「熱區」(Gilbertson, 2007)。換言之，即是犯罪活動的數量與平均值相比明顯偏多或受害風險與期望值相比明顯偏高的區域(Eck, Chainey, Cameron, Leitner, & Wilson, 2005)。如由時空角度來劃分，犯罪熱區不限於空間區域，更可分為時間熱區及時空熱區；而從時空相對關係觀察，熱區概可分為以下 3 種：

- (一) 高頻率熱區：是從犯罪活動的本身來分析，即犯罪活動頻率發生較高的區域或時段。
- (二) 高危險熱區：是從潛在受害物件的角度來分析，即潛在受害物件遭受侵害可能性高的區域或時段。
- (三) 異常熱區：是從犯罪活動時空變化的角度來分析，指犯罪活動的數量有異常的區域或時段。

舉例而言，若甲區與乙區為面積相若兩地，甲區發生住宅竊盜為 100 起，區內戶數為 5,000 戶，乙區發生住宅竊盜 50 起，區內戶數為 1,000 戶，則從犯罪活動的空間觀點而言，甲區為高頻率熱區；但從家戶受害風險的機率來看，則乙區顯為高危險熱區；而若以時間來分析， T_1 與 T_2 為兩個不同時段但同為 1 個月，若在 T_1 時間內發生 50 起，而 T_2 時間內發生 40 起，則單以月統計而言， T_1 可能是時間熱區，惟若 T_1 與上月相比數量並無太大差別，但 T_2 則是突然暴增，或許 T_2 更可能是時間異常熱區。若再以甲區、乙區及 T_1 、 T_2 時空交互分析，則熱區又將呈現不同面貌。總之，熱區是相對的，與研究的時間及空間尺度（範圍）相關。街道尺度的犯罪熱區在整個城市尺度上不一定是熱區；同樣，某月犯罪案件高發時段，在一年的尺度中，亦可能不是一個熱點(Eck et al., 2005)。此外，對於犯罪熱區的認識，更可進一步深入瞭解在熱區內何種犯罪多於其他類型犯罪（熱區類型），及在熱區內多次或者更容易受害的目標（熱區目標）。

此外，根據犯罪熱區產生的機制，可分成以下 3 種類型(Clarke & Eck, 2005)：

(一) 犯罪創造地 (crime generators)

某些地方容易吸引大量的人，原本這些人聚集原因與犯罪動機並無直接無關，但卻為罪犯提供大量的目標與機會，當時間和地點聚合時，將容易會導致犯罪或混亂。例如：購物區、交通樞紐、節日和體育賽事場地等；犯罪或混亂事件主要是由於大量的人群聚集和犯罪目標。

(二) 犯罪吸引地(crime attractors)

此地是提供罪犯者所熟知許多犯罪機會的地方，有犯罪動機的人被吸引至此。在短期內，罪犯者可能來自該地區以外，時日一久，在特定情況下，罪犯者可能會遷往這些地區，色情行業和販毒區就是明顯的例子；同樣地，若干娛樂場所或許因允許越軌活動而聞名。這些地方一開始可能只有當地人知道，但隨著名聲的傳播，越來越多的罪犯被吸引進來，從而增加犯罪和騷亂事件的數量。

(三) 犯罪促進地 (crime enablers)

當某些地方對行為幾乎沒有規範或規範沒有得到執行，就會發生促進犯罪發生的種情況。例如，停車場服務員的離開讓人們可以在停車場閒逛，這導致車輛盜竊的增加，這是因移除管理產生的犯罪。此外，原本的管理機制因隨著時間的推移而逐漸消失，從而導致問題的增長；例如，從小父母親對小孩呵護備至，處處關照，但隨著小孩日益成長，逐漸讓孩子們有越來越多的自由活動空間，小孩們受害者或成為罪犯者的風險就會增加。

整體而言，近年來隨著犯罪地圖研究方法的發展，熱點的分析可以是任何尺度，從點到面均可(Eck et al., 2005)，而對於從微觀尺度研究高犯罪集中的地點的重要性也有相當共識，但相對是否會因此忽略非熱點或無犯罪的空間特質；因此，Swartz (2000)建議在犯罪熱點架構中應加入「冷點」(cold spots)的分析，此對犯罪的空間分布的全面瞭解更有所幫助，本研究即是採取此觀點，在後續的犯罪區位分析中，即是探討造成犯罪群聚冷熱區的區位因素。

二、犯罪熱區分析方法

犯罪熱區時空分析內容主要集中在「熱區」、「熱時」以及「熱區時空互動」等方面，研究方法融合犯罪學、環境地理學、經濟學、地理資訊科學以及製圖學等多學科理論及相關技術，可改善單一時空面向與研究方法無法有效分析所有犯罪類型的侷限。目前識別熱區方法有許多，歸納起來可分為地圖製圖方法和資料(數據)分析方法兩大類，但鑑別熱區若僅透過犯罪製圖技術，純粹以視覺化表達犯罪資料之空間特徵雖簡潔易懂，然若未經相關統計檢定恐造成精確度不足，僅能是「犯罪概述(crime summary)」層次，未能達到找出「區域內之某個地方具有相對顯著的犯罪風險群聚趨勢」(溫在弘等，2010)；故以下概述空間熱區與時

空熱區在數據分析的研究方式。

(一) 空間熱區分析

刑案資料通常會記載犯罪地點、犯罪時間等，其中犯罪地點即為空間熱區分析最重要之依據。隨著 GIS 發展，目前已可透過位址定位(address matching)的方法，可將文字位元址轉換成 X-Y 地理座標(geographic coordinates)或因研究需求進一步再轉換為經緯度，便可以點資料(point data)形式呈現其空間分布。然而，在實際情況下，由於犯罪資料存在地址遺漏、地址錯誤等問題，將造成有些犯罪事件不能位址定位，從而影響犯罪分析的正確性，故 Ratcliffe 指出，為能使地圖正確反映犯罪分布的空間模式，犯罪案件的定位率(定位的案件數與案件總數比)至少要達到犯罪案件資料的 85%(Ratcliffe, 2004)，而此點資料若予以加總即可成為面資料(polygon data)。

空間熱區分析首要係瞭解其犯罪的空間分布、其案件是否呈現群集及其發生的頻率及強度等關係(溫在弘等，2010)，藉由點資料在地理上分布，可區分為分散(dispersed)、隨機(random)、群集(clustered)3 大類，如圖 2-1-1：

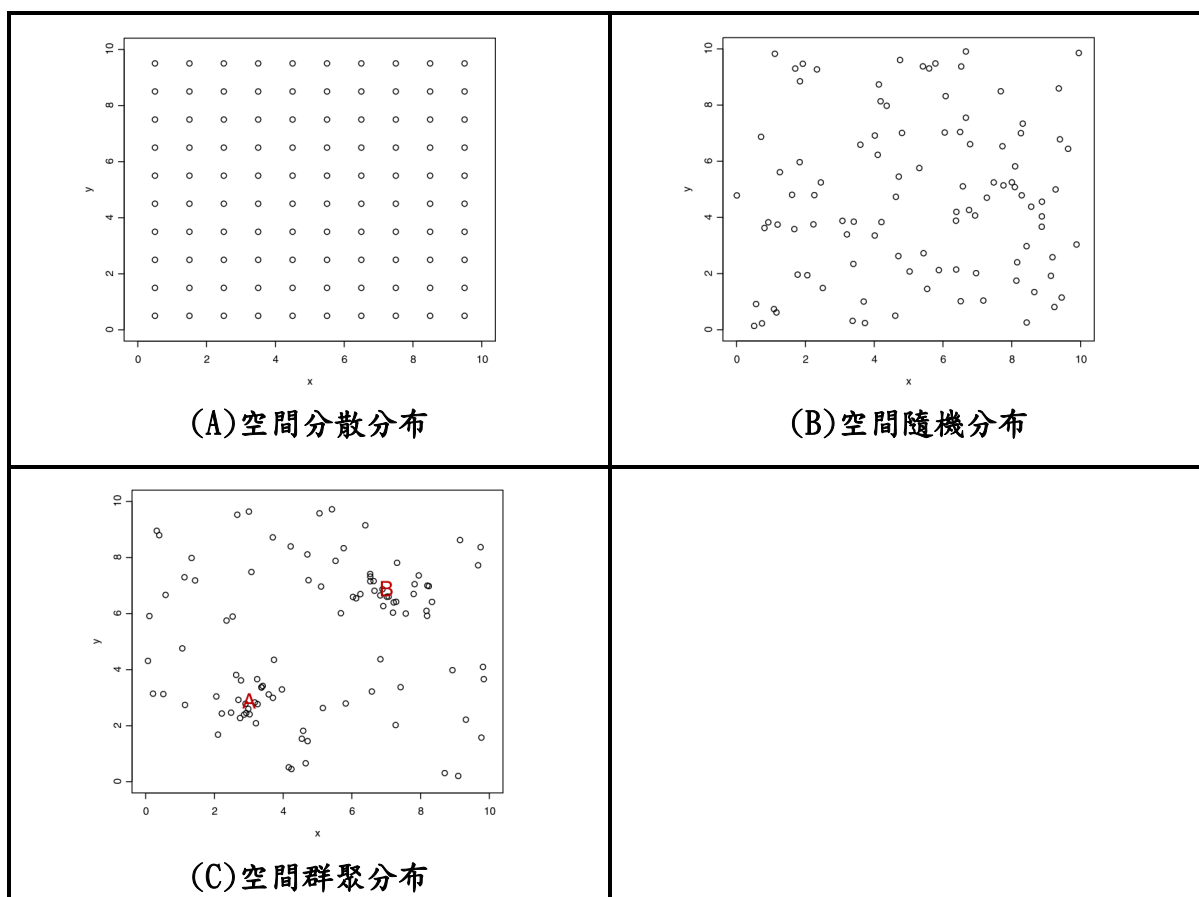


圖 2-1-1 空間點型態分布

資料來源：Toshiro Tango (2010)

從圖 2-1-1(A)可看出，點在平面上分布呈現規則排列，此即分散分布，又可稱為完全規則型態(completely regular pattern)(Toshiro Tango, 2010)，顧名思義，即是點資料位置具有規則可循；(B)是隨機分布，此分布常被用於群聚分布在統計檢定的虛無假設 H_0 ；(C)則為統計的對立假設 H_1 ，也就是判定事件是否為熱區的檢定目標，圖中 A 與 B 為可能群聚區。

而在的統計檢定方法分類上，Besag 等依據檢定目的不同，分成 2 類(J. Besag & J. Newell, 1991)：

- 1、一般檢定(general tests)：此種檢定方法目的在於瞭解「事件是否在研究區域內發生群聚」。
- 2、焦點檢定(focused tests)：此種檢定方法目的在於瞭解「某一特定的位置周圍是否有顯著較高的事件發生率」。

上述的「一般檢定」繼而被 Kulldorff (1998)分成 2 種型態：

- 1、全域式群聚檢定(global clustering tests)：此種檢定方法目的在於評估事件是否傾向於群聚，但無法知其確切群聚區域。
- 2、聚類偵測檢定(cluster detection tests)：適用於檢測局部群聚和評估其重要性，與較為所知的區域式群聚檢定(local clustering tests)相當。

從上可知，學者對於群聚之認知與分類非不相同，差異在於使用不同名詞詮釋各類型態之群聚，故群聚之類型主要可歸納為焦點檢定、全域式群聚檢定及區域式群聚檢定 3 種。

通常熱區在數據資料的分析方式可分為空間分布描述、空間群聚分析與空間自相關分析 3 大部分(溫在弘，2015)。其中空間分布描述的主要分析方法又可分為空間分布中心與離散趨勢分析 2 種，在中心趨勢方面，方法為平均中心點(mean center)、中位數中心點(median center)、中心地理物件(central feature)等；在離散趨勢方面，則有標準距離偏差(standard distance deviation)、標準偏差橢圓(standard deviational ellipse)等。而在空間群聚方面，主要方法有最近鄰階層群聚(nearest neighbor hierarchical clustering)、核密度推估(kernel density estimation)等。在空間自相關方面有全域分析的 Moran's I 及區域分析的 Anselin's LISA、Getis-Ord's G_i^* 等。

然而，在應用上述空間分析的方法時需注意資料型態的使用，否則將造成分析錯誤或 GIS 軟體無法執行。進而言之，犯罪資料型態通常呈現為「地理位置資料」(geographical location data)和「區域統計資料」(regional count data)2 種(Toshiro Tango, 2010)；所謂「地理位置資料」即前述之座標資料，為「點資料」(point data)，以點分布型態呈現，而「區域統計資料」即是將點資料依據行政區劃如鄉鎮、里等加總，屬於「面資料」(polygon data)，選擇之尺度依據研究目的而定。以上熱區分析方法中，面資料僅可使用在空間自相關分析，再以點方式呈現，並不適用於刑案個案為單位的點資料。

上述熱區分析方法基本上均可藉由 GIS 加以實現；具體來說，目前常被使用的 GIS 軟體，如 Geo Da、ArcGis、CrimeStat 及 Qgis 等均有內鍵統計分析功能，使用者將資料整理為軟體可讀取格式，軟體讀取後即可進行分析，分析結果除產生報表外，直接可加以視覺化，若熟悉軟體使用功能，對於犯罪製圖十分便利。

(二) 時空熱區分析

除上述較常被使用的空間分析方法外，近年來更多的研究注意到事件發生的時間面向問題(Townsley, 2008)，在許多統計學家的努力下，各種不同偵測時空群聚的方法相繼被提出，如「廣義加成模型」(generalized additive models, GAM)、「貝式時空方法」(bayesian spatio-temporal approach)等，其中「掃描統計」(scan statistics)的發展頗受矚目，並已普遍使用在流行病學、人口學、地理學等各領域(Nakaya & Yano, 2010)，近年也逐漸受到犯罪學研究重視並加以運用。

掃描統計推出之初，仍有許多不足之處，惟經過不斷研究，目前掃描統計除能偵測空間群聚外，也能將時間因素納入分析，且偵測熱區能力效果亦佳，普及率越來越高，故本文擬以此為主要時空熱區研究方法。

掃描統計在上世紀 60 年代由 Joseph Naus 在其所發表的"Clustering of Random Points in Two Dimensions"首先提倡(Naus, 1965)，當時主要目的係運用於公共衛生與流行病疫情偵測領域(Tonini, Tuia, & Ratle, 2009)，開啟偵測群聚方法論新紀元。Naus 利用 Barton 和 Mallows 的推論，找到事件均勻分布下，掃描統計量超過單個臨界值的確切概率，惟 Naus 的方法尚未成熟，其侷限為：1、該方法係以固定大小的矩形視窗進行掃描，但大多數情形是人們並無法事先獲知事件群聚的範圍，且不同事件群聚的形狀也不是固定；例如，傳染病可能是沿著河流呈現線狀群聚等。2、其概率模型是針對均勻卜瓦松分布(homogeneous poisson distribution)資料，不能用於其他類型的資料；然而，無論是疾病傳染強度或犯罪被害風險，經常是取決於處於危險中的人群密度，而呈現非均勻的卜瓦松分布(inhomogeneous poisson distribution)(Toshiro Tango, 2010)；此後，陸續有許多的掃描統計相繼推出。

首先，Openshaws 等提出的「地理分析機」(geographical analysis machine, GAM)，從統計方法上進行突破，其原理是在不同的尺度上檢測大量重疊的掃描範圍，並評估偶然發生的事件數量的統計機率(Openshaw, Charlton, Wymer, & Craft, 1987)，但此方法的缺點可能因為圓形視窗內人口數的不同及重覆檢定產生問題(蔡丞庭，2011)，Besag 和 Newell(1991)特別發表論文對此方法提出許多評論。隨後，在掃描視窗方面，Loader 提出利用大小變化的矩形視窗(Loader, 1991)；Alm 則使用圓形、橢圓形及三角形視窗(Alm, 1997, 1998)進一步改善先前缺點。

約在同時期，哈佛大學醫學教授 Kulldorff 推出連續掃描的圓形視窗方法，並開發統計軟體 SaTScan 後，受到廣大歡迎，惟當時缺點為對於非圓形與不規則的區域誤差較大(Kulldorff, 1997)，不過後來此缺點也經過修正，並加入橢圓形掃描視窗(Kulldorff, Huang, Pickle, & Duczmal, 2006)。而為改良 SaTScan 對於掃描區域的侷限，Tango 和 Takahashi(2005)提出「彈性空間掃描統計」(flexible spatial scan statistic)與 FLeXScan 軟體，專用於檢測不規則形狀，但由於計算量太大，故必須將掃描限制在若干最近鄰的群集(T.Tango & Takahashi, 2012)，由於 FLeXScan 只能對空間進行掃描，缺乏時空分析，普及程度不如 SaTScan 廣泛。

SaTScan 推出後，許多研究紛紛以其與其他空間統計方法進行比較，例如 Esra Ozdenerol 等人，利用 SaTScan 軟體中的空間掃描統計模型與廣義加成模型、「疾病映射模型」(disease mapping model)進行比較，檢測整個挪威死亡率高的空間群集，結果顯示 Kulldorff 的空間掃描統計較適用於低風險地區，因其不過度受到周圍地區的影響，雖然它的敏感度比廣義加成模型低一些，故當在高風險與正常風險地區，SaTscan 的空間掃描統計的相對風險會比廣義加成模型的相對風險值較低；反之，廣義加成模型若在周圍是無人煙且低死亡率的影響，則可能導致無法完全找到死亡率較高的群集(Aamodt, Samuelsen, & Skrondal, 2006)。

其次，在處理資料類型方面，Turnbull 和 Kulldorff 等人，為改良掃描統計只能運用在均勻卜瓦松分布(homogeneous poisson distribution)資料的缺點，陸續提出非均勻卜瓦松分布的處理方法及伯努力模型(bernoulli model)、多項式模型(multinomial model)等(Kulldorff, 1997; Turnbull, J. Iwano, S. Burnett, L. Howe, & C. Clark, 1990)不斷豐富掃描統計的內容。而最值得關注的是，時間與空間結合的時空群聚掃描統計，在 Kulldorff 的研究團隊持續研發後，終於獲得關鍵性突破；爾後，又陸續推出「回顧性時空掃描」(retrospective time-space scan)與「前瞻性時空掃描」(prospective time-space scan)等。由於 SaTscan 在時空分析上的良好探測性，美國國家司法研究所(national institute of justice)也對此掃描統計方法與軟體加以推薦(Jefferis, 1998)。整體而言，掃描統計具有以下優點(Kulldorff, 2001)：

- 1、可對非均勻的人口密度和不同數量的共變項進行調整。
- 2、不須指定偵測範圍的大小或位置，即可逐步搜索以偵測群聚位置，此方法避免人為主觀設定與預選偏差(preselection bias)問題。
- 3、基於對數概似率的檢驗統計量考慮到多重檢驗，並顯示拒絕虛無假設的

P 值。

4、如果拒案例為隨機分布的絕虛無假設，即可根據輸入資料的地區行政編碼，明確辨識群聚的位置。

相較於犯罪熱區分析，犯罪時空熱區研究可以更加瞭解不同犯罪行為的時空分布特性，以協助司法系統更有效率投入警政資源與人力，提升打擊和預防犯罪效能。透過高頻率熱區分析，有利於司法人員精確投放和部署警力；而高危險熱區往往是和一定的空間環境有所關聯，可進一步考察犯罪發生和空間環境的關係，如犯罪與土地使用、社會經濟、人口以及其他空間環境特徵的關係以鑑別誘發犯罪的環境因素，從而改變環境不利因素以預防和阻止犯罪；對異常熱區的分析與監測則有助於對偶突發或短期特殊犯罪的爆發進行預測和預警，以防患於未然。

第二節 犯罪學相關理論

一、環境犯罪學

傳統犯罪學領域主要在探討生物、社會等各種因素如何塑造個人人格特質，從時間觀點觀察人如何發展成罪犯，換言之，犯罪者是研究的主體；在此思維下，犯罪防治必須由源頭著手，改善潛在犯罪群體的社會不利處境，降低犯罪者的犯罪動機。然而，在上世紀 60 至 70 年代，當美國社會歷經劇烈的城市化擴張，城市衰敗、社會解組和失序引發大量的犯罪發生後，學界開始湧現以改善城市規劃和空間環境設計的方式以應對犯罪的思潮。在此思潮下 Newman(1970)提出「防衛空間」(defensible space)、Jeffery(1971)的經由環境設計預防犯罪 (crime prevention through environmental design, CPTED)、Wilson 與 Kelling(1982)的破窗理論(broken windows theory)、Clarke 和 Cornish(1985)的情境預防理論(situational crime prevention)相繼提出，構成環境犯罪學的主要經典理論。

自環境犯罪學興起後，採取不同的視角，探討為何某些人在某些特殊時間和空間犯了某種特殊類型之犯罪；亦即，環境犯罪學並不去解釋犯罪行為發生的根本原因，而是著重在探討犯罪要素與周遭地理時空的互動關係，企圖藉由改善環境因素達到預防犯罪的目的(Brantingham & Brantingham, 1991)。而 Wortley 與 Ann Mazerolle(2009)則更進一步對環境犯罪學提出 3 大論述：

- (一)環境的性質對於犯罪行為有直接重大影響。環境視角所依賴之原則，即所有的行為都源於一個人與環境的互動，環境不僅是犯罪行為的背景，而且在犯罪的發生和發展過程中扮演基礎性的作用。因此，犯罪事件不僅源於犯罪的個人，它們同樣是由犯罪現場的犯罪因素造成。環境犯罪學解釋環境如何影響行為以及為什麼有些環境會造成犯罪。
- (二)犯罪在時空上的分布並非隨機的。由於犯罪行為依賴於情境因素，犯罪模式是根據犯罪發生環境的位置而定，犯罪將集中於犯罪機會和其他有利於犯罪活動的環境；犯罪率因地區、地方而異，並可能在不同時間達到高峰，犯罪分析的目的是識別和描述這些犯罪模式。
- (三)瞭解犯罪環境的作用和犯罪模式是調查、控制和預防犯罪的有力武器，這方面的知識使司法人員和研究者能夠將資源集中在特定地點的特定犯罪問題上；改變目標環境的犯罪因素可以減少該地區的犯罪發生率。環境犯罪學與犯罪分析相結合，為解決犯罪問題提供切實可行的方法。

事實上，在 Wortley 等人闡述環境犯罪學的核心概念前，犯罪與環境之間的關係更早在 1831 年就引起 Guerry 等法國學者的關注並經由其研究顯示，財產犯罪多發生於高所得地區，而暴力犯罪則易發生於低所得地區，此不只指出環境因素在犯罪的重要性，亦隱含環境與犯罪機會之間的連帶關係(Shane D. Johnson et al., 2007)。其後，20 世紀中葉前，美國芝加哥大學的社會學者 Shaw 與 McKay 等人首次將犯罪現象列為研究的主體，針對芝加哥城市犯罪區位特徵進行研究，深入分析環境特徵與犯罪的關係(Shaw & McKay, 1942)，並聚焦於空間特性對犯罪影響之犯罪理論，其對「少年犯罪區研究」等論點之「社會解組理論」(social disorganization theory)，係犯罪生態(criminal ecology)領域中首先聚焦於空間特性與過程對犯罪影響之犯罪理論，相關研究也對於犯罪製圖與空間分析之發展產生巨大影響。

在犯罪區位研究的影響下，陸續有許多研究試圖從城市、社區的角度，透過分析區域內的統計調查資料以分析犯罪與環境特徵之間的關係，因而可稱為巨觀(macro)環境犯罪學研究。近年來，由於 GIS 的發展，犯罪學者對於犯罪與空間環境的關係多從微觀(micro)角度出發展開研究，如針對住宅、街道、特殊營業場所等地點進行分析(Shane D. Johnson, Bowers, Birks, & Pease, 2009)，為城市和建築的規劃設計提供建議，這些研究使得環境犯罪學理論從巨觀向微觀逐漸發展，

並從環境、城市規劃等角度研究犯罪發生的原因。

目前，環境犯罪學領域提出多項犯罪學理論已被廣泛運用，從社會解組理論到日常活動理論(routine activity theory)、理性選擇理論(rational choice theory)，再到破窗理論等，各犯罪學理論分別從不同的角度對犯罪與時空環境之間的相互作用關係進行闡述，不僅解釋犯罪發生的機制，也提出「犯罪機會」應該也是犯罪發生的一項「根本原因」(root cause)(許春金，2009)，此對於警力部署、犯罪原因分析等具有重要的參考價值。

二、新機會理論

1998年，Felson和Clarke發表「新機會理論」(new opportunity theory)，此乃為區別由Cloward與Ohlin於1960年所提出之「機會理論」(opportunity theory)。「新機會理論」主要由3個理論所構成；包括：「日常活動理論」、「犯罪型態理論」(crime pattern theory)及「理性選擇理論」，這些理論均隱含「機會」之概念，亦即以機會的變化來解釋犯罪型態及數量的變化；而所謂的「機會」包括合適標的物(suitable targets)、方法或工具(means)、情境(situations)等「機會」的變化(許春金，2009)，而上述3個理論均可說明相關變化情形。簡述如下：

(一) 日常活動理論

「日常活動理論」起始於Cohen和Felson(1979)針對1947至1974年間美國犯罪趨勢的研究所發表的”Social Change and Crime Rate Trends: A Routine Activity Approach”研究論文。該研究以「人文區位學」(human ecological theory)作為理論建構基礎，並沒有強調罪犯的特點，而是將焦點放在掠奪性犯罪行為(predatory criminal acts)並由巨觀(macro level)與微觀(micro level)兩層面研究犯罪現象。

在巨觀層面，日常活動理論探討社會變遷對於犯罪機會的影響。根據Cohen等的研究指出，二次世界大戰後，美國民眾的日常生活型態遽變，家庭成員暴露在家以外的時間增加，導致社會環境中的犯罪機會大增，從而使得整體犯罪率上升。Cohen研究結論認為，遠離家庭和家庭的活動將增加犯罪的機會，從而產生較高的犯罪率，而主要的因素係因勞動參與率(labor force participation)和單身成年家庭(single-adult households)等變項產生變化所導致(E. Cohen & Felson,

1979)。

在微觀層面，Cohen 的「日常生活理論」則關注在例行生活中，犯罪機會的產生原因。對此，Cohen 等提出誘發犯罪機會的 3 個要素，建立犯罪基本三角概念(basic crime triangle)(如圖 2-2-1)：

- 1、有動機的嫌疑犯(likely offender)：即為具有犯罪意圖的潛在犯罪者，是一切犯罪的動力源頭，為必備要素，沒有此項要素，犯罪就不會發生，故日常活動理論的重點在強調其他 2 項要素。
- 2、適合的目標(suitable target)：在日常生活理論中，受害目標可能是人(person)，也可能是物品(object)；當下的時空位置使其受侵害的可能性加大。Felson 指出，影響標的物風險的要素有 4 項(V.I.V.A)：
 - (1)價值(value)：V 代表目標的價值。
 - (2)慣性(inertia)：I 代表目標的可移動性。
 - (3)可見性(visibility)：V 代表目標的可見度。
 - (4)可接近性(access)：A 代表目標的可接近性。
- 3、有能力監控者(capable guardian)：在「日常生活理論」中，監控者不僅限於員警或是監視器等制式編制，而是泛指身處案發現場或是鄰近地區，且能發揮影響力抑止犯罪發生的任何人或物(M. Felson, Clarke, Great Britain. Home Office. Research, Policing, & Unit, 1998)。

「日常生活理論」主張，當這 3 項要素在時間與空間中交集時，犯罪才會發生，缺一不可。

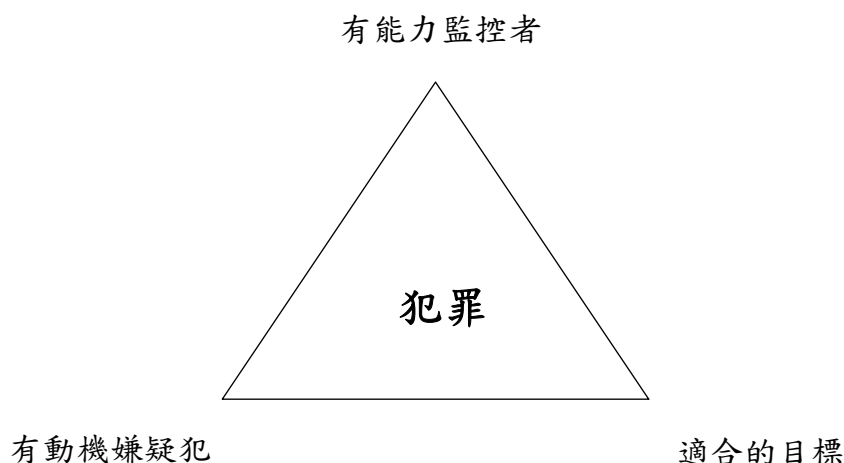


圖 2-2-1 犯罪基本三角

資料來源：M. Felson et al. (1998)

(二) 犯罪型態理論

犯罪型態理論亦稱為犯罪搜尋理論(crime search theory)，結合理性選擇理論、日常活動理論和環境犯罪學來解釋人(people)和物(thing)如何在某社區中的時、空移動而發生犯罪。此理論認為每個犯罪事件都是法律、犯罪動機和目標特徵在時空特定點的環境背景的偶然交集產物，犯罪事件中的每個要素都有一定的歷史軌跡，這些軌跡由過去的經驗和未來的意圖、日常活動和生活節奏(rhythm)以及環境的制約所形成。此理論有 3 個核心觀念：節點(node)、路徑(path)與邊緣(edge)(許春金，2010)。路徑為個人日常活動路線，節點指人們日常活動的起點或終點，例如家庭、學校、工作場所等長時間停留的地方，而邊緣係指人們工作、居住、休閒等活動之邊界地帶，此一地帶因人們彼此較不熟悉，因而較易產生犯罪。

整體而言，此理論係透過現場、情況、活動空間、觸發事件和潛在動機相關的過程，探索犯罪發生和犯罪行為模式。首先，以個人作為分析的最基本單位，人透過決策過程決定日常活動，當犯罪誘因出現時，人即可能轉變成潛在的犯罪者，企圖進行犯罪行為；其次，人與人之間的互動會形成網絡(network)，網絡內的成員會互相影響，包含犯罪行為；在物理空間的範疇內，犯罪型態理論將個人生活中的活動範圍定義為活動空間(activity space)；另一方面，將移動的路線區分為節點與路徑兩類屬性，節點為路徑間交接的場所，當網絡中的犯罪者與受害者，彼此間的活動空間交疊時，犯罪機會就會大量產生，且往往集中於活動節點，

造成附近群聚現象。

值得注意的是，在犯罪型態理論中，「地點」的觀念是不可或缺的元素，亦即犯罪者必定是在某一個地點實施犯罪行為，同時其特徵更會影響犯罪發生的可能性。根據日常活動理論所強調的地點特徵，包含監控者、管理者的出現及效能，以及有能力守衛者的出現。犯罪型態理論，藉由針對某地點如何被潛在犯罪者所發現或注意的解釋，而將具有合適標的物的地點與該地點的背景環境相連結。

犯罪型態理論和日常活動理論在許多方面是彼此相互支持的。但是，這兩個理論對於某些發生在特定地點的犯罪，似乎有不同的解釋。例如，就某些具高犯罪率的地點而言，犯罪型態理論，會把解釋重點放在犯罪者如何發現及接近該地點；而日常活動理論，則把解釋焦點集中在標的物的型態或行為、以及防衛者不在現場的原因上；換言之，對犯罪型態理論而言，地點的問題，主要是其座落的位置以及和環境之間的關係，而日常活動理論指出地點的問題，主要是某些特定人的出現及消失。因此，可以看出兩種理論在不同的背景及情境下，均有其解釋力(孟維德，2001)。

(三) 理性選擇理論

長久以來，理性選擇理論被廣泛應用於研究政治、經濟學及其他社會學科學上，以建立解釋人類行為的模型。此理論觀點可以追溯到「理性人」的經濟模型和貝加利亞(Cesare Beccaria)和邊沁(Jeremy Bentham)的古典學派。1968年，美國經濟學者貝克(Gary Becker)出版一篇重要論文：「犯罪與懲罰」(Crime and Punishment: An Economic Approach)，他認為有用的犯罪理論「只是延伸經濟學對於選擇之分析」，犯罪決定機制與購買物品或從事某項活動的決定機制相當，這種過程稱之為「期望利益模式」(expected utility model)，主張人即使在不確定情況與無法掌握完全資訊情況下，也期待做出對自己最有利之決定；此論點可說是為理性選擇犯罪理論立下重要根基(許春金，2010)。1980年代，犯罪學者 Clarke and Cornish 引入理性選擇的方法在犯罪學領域以解釋犯罪行為，除學界內激起廣泛的迴響外，也激起「嚇阻理論」(deterrence theory)與「情境犯罪預防」(situational crime prevention)的發展。

理性選擇理論採用決策方法來解釋犯罪它是一種自願的、功利的行為理論，其中犯罪和犯罪行為被視為選擇的結果；相對地，對行動方案所涉及的努力、報

酬和代價的合理考慮又影響到是否犯罪選擇。簡單地說，犯法行為乃是犯罪者在考量個人因素（如金錢、報復、享樂等）及情境因素（如目標物受保護程度、員警執法效能等）所決定的行為，即理性犯罪者會評估犯罪風險高低、處罰的嚴厲性、犯罪利益等，若犯罪利益大於風險，便會從事犯罪，反之則會放棄。故 Clarke and Cornish 針對前述，提出理性選擇 3 個概念：1、罪犯是理性的，他們做出的選擇和決定有利於自己；2、犯罪重點明確；3、犯罪與犯罪性是有區別的，亦即「理性選擇」與個人在整個生命歷程中，對於犯罪風險認知，也會影響其選擇(R. V. Clarke & Cornish, 1985)。

在犯罪學理論中理性選擇理論經常被用來檢測竊盜犯罪，而結論也大致上均支持竊盜犯罪行為是經由理性選擇的過程，包括對目標物是否容易得手、犯罪利益是否可觀、被捕風險高低、地理環境合適性以及考量個人的偷竊技術能力等，而常業竊盜犯比偶發隨機的竊盜犯更具有理性選擇的傾向，主要原因在於常業竊盜犯以此犯罪做為收入來源並維持生計，故必須加強技術熟練以取得最大的利益避免被捕的風險與延長「職業生涯」。因此，理性選擇理論在竊盜犯罪研究中，可視為此領域的共通理論。

綜合前述，日常活動理論強調大社會變化而改變犯罪機會；犯罪型態理論強調社區的差異和變化而影響犯罪機會；理性選擇理論指出個人所處場域犯罪機會不同，個人決定是否犯罪亦有所差異(許春金，2010)，3 理論均聚焦於「機會」，所以稱之為「新機會理論」，Felson 等人更進一步將新機會理論根據研究尺度大小分類—日常生活理論屬整體社會層級；其次，犯罪型態理論研究地區尺度，最後，理性選擇理論分析個人層次。其關係圖如下：

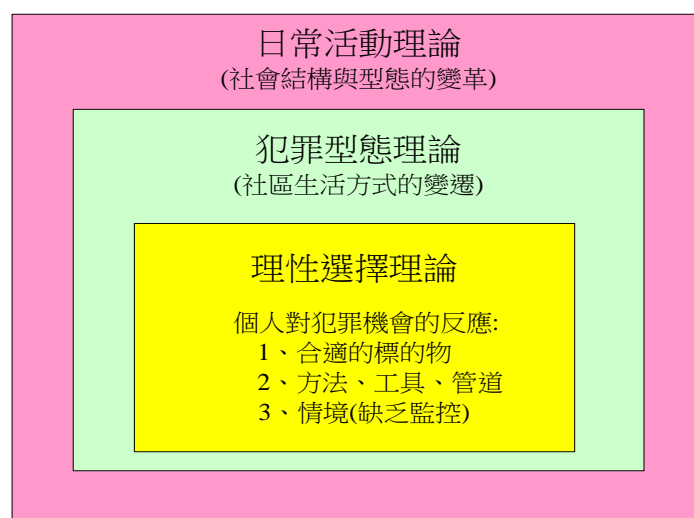


圖 2-2-2 新機會理論關係圖

資料來源：許春金(2010)。

三、社會解組與集體效能理論

社會解組的基礎概念承接自涂爾幹的思想，涂爾幹認為快速的社會變遷造成傳統秩序脫序，而都市化造成社區控制力崩潰，犯罪滋長(周愨嫻、曹立群, 2007)。社會解組理論主張用組織程度衡量社區結構，將社區分類為社會組織社區與社會解組社區兩者：前者具有凝聚力(solidarity)、整合(cohesion)、團結(integration)等特性能凝聚社區意識，抵抗犯罪入侵；後者缺乏上述特性，社會控制失調導致解組現象(Kubrin, 2009)。

以環境犯罪學觀點而言，社會解組理論係以地方為基礎(a place-based)的犯罪學理論。Park 與 Burgess 於 1921 年以城市生態學(urban ecology)的視角，運用製圖方式，發現城市中的犯罪分布與環境有密切的相關(如圖 2-2-3)。研究指出城市就像是一個自然系統(natural system)，每一個區域都會成長，從中心開始朝向下一個區域移動，同心圓的最中心為 Zone I，為商業中心；Zone II 稱為過渡區域，一般為城市中最老舊的地區，為貧窮民眾聚集區；Zone III 為工人的住宅區，居住此區域通常在 Zone II 工作；Zone IV 為住宅區，為較昂貴的公寓；Zone V 為通勤區域，整體而言，居住在 Zone IV 與 Zone V 地區的都是經濟狀況較為好的居民(蔡德輝、楊士隆, 2012)。根據此都市成長模型，這 5 個區域會如同自然環境中的野生植物的擴張過程一般，當內城區域治安開始惡化時，有經濟能力的居民向城外移動尋求更優質的居住品質，城內遺留社會弱勢族群，無力改變環境生

活品質，最終導致社會解組現象(Burgess et al., 1967)。

基於 Burgess 之基礎，Shaw 與 McKay(1942)對芝加哥地區進行偏差行為研究，得出偏差行為受到經濟水準之影響，從而提出 3 項解釋因素：第一，地區經濟較差，會有較高的犯罪率；第二，較高犯罪率地區，會有更多弱勢居民集中情形，為的是領取政府補助；第三，犯罪率高的地區，會有較多的移民，造成更大的人口異質性。Shaw 等人同時指出，由於環境特徵的影響，即便居民自由流動，犯罪率始終是穩定的。

社會解組理論的出現對於環境犯罪學為重要里程碑，但在學界眾多的檢視下，也暴露其理論之不足。Kornhauser 在其發表的” Social Sources of Delinquency: An Appraisal of Analytic Models”(Kornhauser, 1984)指出，社會解組理論應該要包含 2 個不同的機制以解釋地區犯罪，此 2 機制分別是文化偏差(culture deviance)，以及社會解組(social disorganization)。文化偏差模式認為，造成偏差與犯罪行為的主因是社區內的次團體文化與主流社會規範的差異所造成；而社會解組模式則是認為，社區內結構的衰頹，阻礙社區居民實現共同目標的能力，因而產生犯罪行為。

除 Kornhauser 外，也有其他的學者針對社會解組理論提出實證研究上的質疑；Bursik 與 Webb(1982)針對 Shaw 與 McKay 所提出的有關犯罪率穩定問題產生質疑，故同樣使用 Shaw 與 McKay 的原始資料重新檢驗在 1940 年 1970 年間偏差行為的分布模式是否會類似，結果發現芝加哥地區鄰里環境結構的穩定性僅限於 1930 年 1940 年間，在經歷二次世界大戰以及黑人人口的移入後，人口生態產生結構性變化；因此 1950 年 1960 年間犯罪率不再呈現穩定狀態，不同地區的偏差行為改變甚鉅(朱群芳，2015)。

另 Sampson 等人則是從相反面向對社會解組之意涵進行討論。Sampson 與 Groves(1989)針對英國 238 個社區，進行社會解組與社會控制力量對於暴力犯罪及財產性犯罪實證研究，其研究假設係基於社會解組理論出發，認為某地區如果社經地位較低、人口異質性較高，及人口移動率較高，將造成這個地區居民人際關係不穩定，連帶影響參與地方組織與事務意願低落，進而阻礙居民共同維護社區秩序的社會控制力量，也就會造成犯罪與偏差行為。研究結果發現社區中人際網絡與居民參與地方組織呈現正相關，使得暴力犯罪與財產犯罪降低，證明社區

的社會連結網絡對於影響社區犯罪率之影響力。然而，Sampson 的研究結論隨著都市化的發展，鄰里關係疏離，也逐漸受到挑戰，於是 Sampson 便將其焦點轉移至集體效能概念(Sampson, Raudenbush, & Earls, 1997)。

所謂集體效能指的是社區奠基於社區居民的凝聚力與互相信任，以及樂於參與公共事務，而產生自身維持公共秩序的能力。如果社區有較多的資源、同質性較高，人口移動率也較低，則社區便較容易產生高的集體效能，居民會較願意參與公共事務、當社區中有問題發生時會更願意去共同面對並加以解決。1979 年 Sampson 對芝加哥地區的 343 個鄰里進行實證研究，並將集體效能分成非正式社會控制(informal social control)，及社會凝聚與信任(social cohesion and trust)兩部分進行討論，研究結果發現緊密的親屬連結關係、組織參與等對於殺人罪等並無顯著影響，但居民社會凝聚與信任越高，非正式社會控制越強，則可降低暴力犯罪的發生(Sampson et al., 1997)。因此，一般認為，強調社會凝聚及相互信任的集體效能概念比起傳統社會注重的人際關係聯繫的社會連結模式，或許更能符合現代都市的生活型態。

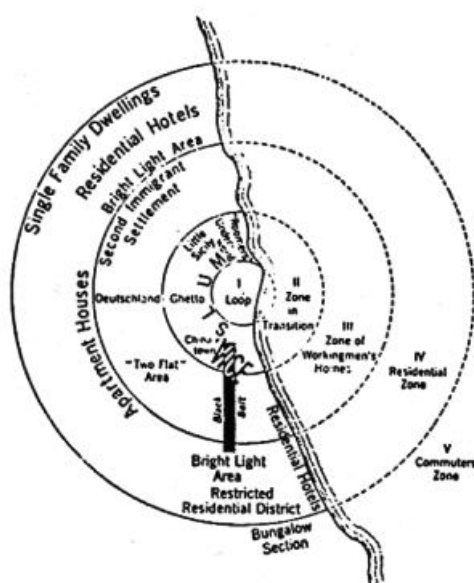


圖 2-2-3 Burgess 的都市同心圓圖

資料來源：Burgess et al. (1967)

四、破窗理論

1969 年，史丹佛大學(stanford university)的社會心理學家 Philip Zimbardo 進

行一項實驗，實驗內容係將 2 輛沒有車牌並打開引擎蓋的車棄置在紐約布朗克斯 (Bronx) 的貧民區與加州一處名為帕洛阿爾托的中產階級區 (Palo Alto)，此目的主要提供車輛被破壞的「機會」。實驗結果為停在紐約的車就在 10 分鐘後，汽車開始被拆除零件，3 天之後車就被破壞殆盡，但停在加州的車 1 周後卻完好無缺；其後，Zimbardo 對加州的車輛窗戶加以破壞，結果與紐約情形相同，原本完好車輛也發生同樣遭受破壞的狀況 (Wilson & Kelling, 1982)。

1982 年，Wilson 和 Kelling 以 Zimbardo 實驗為基礎，在《大西洋月刊》(the Atlantic Monthly) 上發表一篇頗有影響力的文章—“Broken Windows: The Police and Neighbourhood Safety”，通稱「破窗理論」(broken windows theory)。Wilson 和 Kelling 指出，城市居民在日常生活中主要的恐懼來源，非僅是犯罪行為，更多來自周遭環境的失序行為騷擾，如街頭的行乞、酗酒、吸毒、當眾喧鬧及製造髒亂者等行為 (Wilson & Kelling, 1982)。

在文章中，Wilson 和 Kelling 援引 1970 年代中期，美國紐澤西州宣布一項「安全與清潔社區計畫」(safe and clean neighborhoods program)，指派步行警員巡邏維持社區秩序，以改善 28 個城市的社區生活品質為例，說明這項施政雖然對於降低重大犯罪成效有限，但事實證明那些經常步行巡邏的社區的居民似乎比其他地區的居民更有安全感，他們傾向於相信犯罪已經減少，而且採取更少的措施來保護自己不受犯罪的侵害(例如，鎖著門待在家裡)；此外，在徒步巡邏區域的居民對警員的看法更為正向，與被派往巡邏車巡邏的警官相比，步巡警員的士氣更高，工作滿意度更好，對鄰居的態度也更友好 (Wilson & Kelling, 1982)。此項政策凸顯巡邏員警的功能主要在維護與社區居民所形成的秩序共識，並約束脫序者必須遵守的規則；換言之，這些社區共識所形成的公眾秩序維護或許與社區民眾恐懼的暴力性犯罪無關，亦可能不符合嚴謹法律規定與程序，但不可輕忽都市中社區居民對員警維護公眾秩序的期待以及在警方採取相關措施後得到的安全感受。

此外，Wilson 和 Kelling 同時認為在 Zimbardo 實驗中 Palo Alto 那臺車被打破窗戶卻未修補，代表「無人在意」(no one cares) 之情形，再加以破壞也不會受到譴責；所以，車輛被破壞核心問題是沒有去修補的行動。因此，Wilson 等認為，「無人照管」(untended) 的行為也會導致社區鄰里控制的崩潰，一個非正式控制減弱的鄰里，代表居民在控管上是無能力，即所謂集體效能下降，此將導致犯罪

開始入侵，而輕微的脫序行為將導致更多嚴重的犯罪行為，因此破窗理論強調「防微杜漸」，主張著力打擊輕微罪將有助減少更嚴重犯罪，故應該以「犯罪零容忍」的態度面對失序行為。Taylor 和 Harrel 將社區失序導致犯罪率上升之過程歸納為如下幾點(圖 2-2-4)：

- (一) 由於對失序行為的冷漠，部分社區鄰里民眾開始遷出。
- (二) 因故無法遷出之民眾則獨善其身，對周遭事物漠不關心。
- (三) 社區集體效能下降，治安情形惡化。
- (四) 形成破窗效應，失序行為導致犯罪孳生。

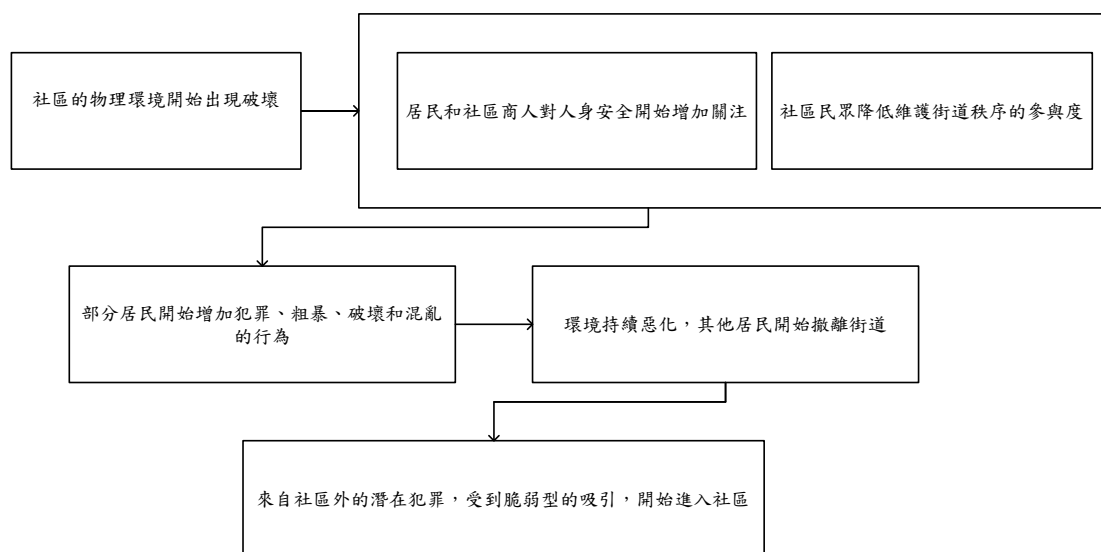


圖 2-2-4 破窗理論流程圖

資料來源：Taylor and Harrell(1996)

受到破窗理論的影響，20 多年後，美國人口最多的 3 個城市—紐約、芝加哥與洛杉磯，都至少採納了 Wilson 和 Kelling 理論的某些方面，主要是透過更積極地執行輕罪法以防止犯罪的發生與蔓延(Harcourt & Ludwig, 2006)但此舉卻也引起許多質疑聲浪，最主要在於當都市或社區產生失序並導致犯罪開始蔓延時，人們卻經常將遊民、乞丐及低收入者等視為可能禍首，進而導致具有歧視性的政策出現，反造成人權的戕害，同時也反映治理階級在道德、秩序的單一向度，對相同族群、相近階級或類似興趣的同好，固然容易形成共同的價值觀、凝聚共識，但相對地也容易傾向於產生社會排除的作用，甚至透過治理行為與懲罰機制，轉化成對弱勢社群、少數民族與外來者的壓制與迫害。但無論如何，破窗理論的提

出凸顯維護社區秩序的集體意識對於犯罪預防的重要性，此與社會解組與集體效能之論點有異曲同工之妙，對於社會治理與警政措施的規畫仍具重要參考價值。

五、受害者學與生活暴露理論

直至今日為止，犯罪學研究仍多將焦點集中在犯罪加害者方面之研究，故可稱之為「加害者學」(offenderology)；然而，從上述的新機會理論中明白指出，犯罪的發生除加害者外，必有受害者，否則犯罪無從發生。傳統的犯罪學者試圖探究犯罪原因與加害者動機，並從社會文化及政經因素探索犯罪行為，而受害者相對地在犯罪過程中亦擔負某種程度之責任，扮演從屬角色；所以，兩者在犯罪行為中是相對的一體兩面(dyad) (張平吾、黃富源，2012)。

在 1970 年代，受害者學在犯罪學中成為特殊領域；然而，在受害者學發展的歷程中，過去數十年有卻許多爭論，包括有學者認為加害者本身即是受害者，因其生命歷程中可能遭受到貧窮、暴力等惡劣條件，導致其由受害者轉變為加害者，甚至有學者認為受害者學缺乏一套理論體系與概念，只是一種想法或概念，或強調有重複或多次受害經歷的個人的分布和特徵，頂多只能稱之為受害者類型學(typology)不能視之為學術領域之科學學科(Meier & Miethe, 1993)；對於各項爭論與質疑，學者 Viano(1992)等人認為，受害者學的作用在於研究受人忽視的受害者問題，並非自詡為新興學科以證明新領域的誕生，主要目的在於提出超越現有學科的科際整合研究與解決策略(張平吾、黃富源，2012)，故對於受害者學之認識，宜從務實角度加以理解，更能體會其在犯罪學之貢獻。

過去的受害者研究顯示，一旦被害，再一次被害的風險就會增加(S. Lloyd, Farrell, & Pease, 1994)，對於再次被害，一般稱為「重複被害」(repeat victimization)，意指某種特殊類型的標的物在特定時間內反覆被害的現象，此標的物可以是人，亦可為地方，而有關「地方」之重複被害則稱之為「犯罪熱區(點)」，在住宅竊盜中經常出現此情形(許春金，2010)，且許多研究也均已證實此一論點(Hoppe & Gerell, 2019; Townsley, Homel, & Chaseling, 2000；張平吾、蔡德輝，2005)；而在住宅竊盜重複被害原因方面，被害人本身不注意、進出管道容易與否、環境因素及被害人的生活型態等均是造成犯罪被害的原因(張平吾、蔡德輝，2005)。

因此，從受害者相關的研究中可得知，至少兩個理論與受害者學密切相關，其一，為「日常活動理論」，另一理論則為「受害者生活暴露理論」

(lifestyle-exposure theories of victimization)(以下簡稱「生活方式暴露理論」)。從「日常活動理論」方面來看，主要是以相對的視角檢視被害人如何成為犯罪標的；而「生活方式暴露理論」則是觀察個體本身是否具有某些特性，導致被害危險性增加，乃至成為被害人。由於「日常活動理論」前已論及，不再重述，以下概述「生活方式暴露理論」在被害人學之觀點。

最早以「生活暴露理論」來解釋犯罪被害人現象的學者為 Hindelang, Gottfredson, and Garofalo(1978)，這一理論最初是為瞭解釋在社會中暴力被害人風險的差異而提出，但目前已被擴展到包括財產犯罪等不同類型犯罪的運用，並為進一步對犯罪目標選擇過程之相關理論奠定基礎(Meier & Miethe, 1993)。「生活方式暴露理論」的基本前提是，被害可能性(機率)在統計學差異主要來自於被害人個人生活方式的差異，而生活方式的動態變化是重要指標，因為被害人在「不恰當的時間」接觸到危險地點，亦即在時間與空間的交集下，特定的時空往往具有較高的被害風險；換言之，個人的生活方式是決定犯罪被害風險的關鍵因素，而所謂的生活方式被則被定義為「日常活動」，包括職業活動(工作、學校等)和休閒活動等，人們的日常活動自然會使他們接觸到犯罪，或者是增加被害的風險，若進一步推論，在家中通常會降低被害人的風險，而處於公共場所則會升高風險，且隨著時間的增加，風險機率越高。據此，Hindelang 等人提出 8 項假設，說明生活方式與暴露被害間之關聯：(一) 被害之機率與暴露在公共場所時間長短成正相關；(二) 特定生活方式使個人置身公共場所可能性有所差異；(三) 生活方式相近者，彼此接觸互動機會較多；(四) 個人與犯罪者之人口特性越相似，成為被害人機率越高；(五) 個人與家庭外成員互動時間長短，隨其生活方式差異而有所不同；(六) 個人成為被害人的機會，和其與家庭外成員互動情形相關；(七) 生活方式不同與個人阻絕具犯罪特性之人接觸的能力有關聯；(八) 生活方式不同與個人成為被害的方便性、誘發性及易被侵害性有關聯(蔡德輝、楊士隆，2019)。

另 Meier 等認為，個體在生活方式與日常活動的差異是由於個人對在社會中所扮演角色的期望和結構約束的集體反應或適應所決定，從而被賦予實現的身分特徵(例如，年齡、性別、種族、收入、婚姻狀況、教育、職業)都與成為犯罪目標有重要關聯(如圖 2-2-5)，因為身分屬性使個體行為選擇受到限制，導致生活方式與日常活動模式的固化，並進一步增加接觸危險或易受傷害的機會。

整體而言，「日常活動理論」與「生活方式暴露理論」都強調日常活動或生活方式的模式如何為犯罪提供機會，強調犯罪機會的可獲得性是如何決定的，此兩種理論之間的根本區別除在用語外，生活方式理論主要從概率的角度考慮風險（例如，某些行為提高個體成為被害者的可能性），而日常活動理論則描述被害事件本身構成要件（例如，如果3個關鍵因素通時出現被害就會發生，但如果其中一個因素缺失，則犯罪難以成立）；兩理論對於犯罪發生的關注焦點略有不同。

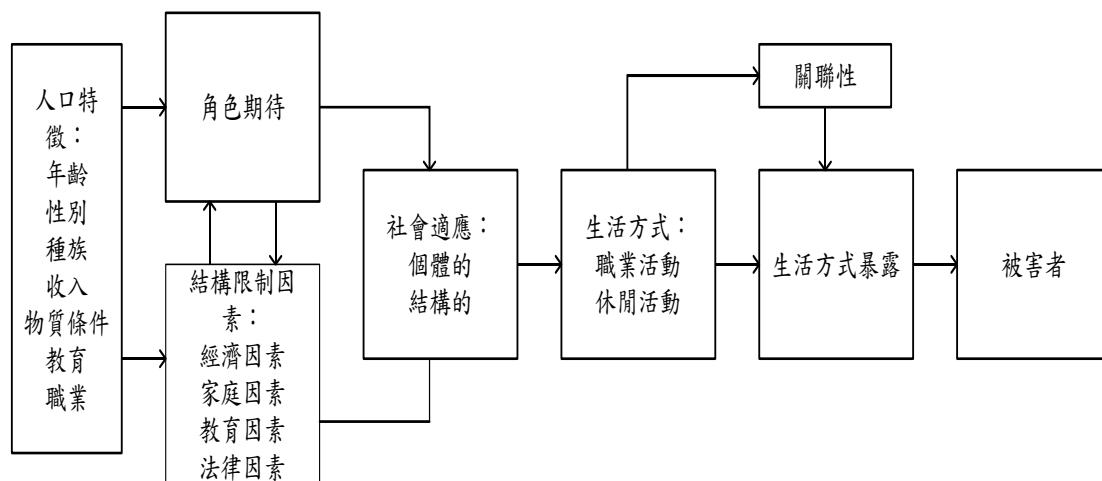


圖 2-2-5 被害者的生活方式暴露模型

資料來源：Meier and Miethe (1993)

第三節 犯罪學理論與區位分析

一、犯罪學理論與區位分析的關聯性

環境犯罪學已明確指出地理環境是影響犯罪的重要因素；換言之，犯罪空間分布是特定地理環境內多種因素誘發犯罪行為的綜合結果。綜觀各犯罪學理論可以看出，社會解組理論注重失控的時空環境對於區域範圍內犯罪形成的影響，其分析尺度為城市、地區。日常活動理論、生活型態理論、理性選擇理論等注重分析人的行為與周圍空間環境的關聯，尤其在於犯罪機會，故稱為「新機會理論」，其分析尺度是以人的生活範圍大小進行分析。然而；犯罪分析的目的，不只是在於發現問題，而是如何提出犯罪防治策略方為首要，故在理論的指引下，如何找出影響犯罪的因素（變項），進而針對犯罪危險因素研擬抑制或消除對策亦甚為重要；因此，運用「人文區位學」探討區域特徵的研究方法，結合相關犯罪學理

論，將有助於篩選出影響犯罪空間型態與其區位的重要因素。

人文區位學是研究人類如何與環境相互作用的學科，其起源乃基於「生物生態學」(biological ecology)之概念，而應用在犯罪學領域中，前述 Quetelet & Guerry、Park & Burgess 及 Shaw & McKay 等人都是犯罪地理與區位學的先驅者；而隨著犯罪資料及其他相關人文社經資料相對過去容易取得，以及電腦技術等條件更加成熟的支援下，當前的犯罪地理區位研究尺度逐漸由巨觀尺度，如國家、區域、城市等向微觀尺度，如街道、公共設施及商店等單位擴展。近年來的相關研究顯示，微觀尺度空間犯罪數量的變化可能引起整體都市犯罪型態的改變(Weisburd, Bushway, Lum, & Yang, 2004)；因此，從微觀尺度研究犯罪變化趨勢與發生原因，在犯罪預防的實務運作上可能更貼近於民眾切身生活。

犯罪區位學演變至今，鄰里、社區的人口特性、經濟概況、土地利用等社經人文因素是歷時不衰的重要研究變項，而物質空間環境如何為犯罪主體提供犯罪機會也成為近年來微觀犯罪研究的重要焦點。相關研究顯示交通線路(Liggett, Loukaitou-Sideris, & Iseki, 2003)、街道類型等均對犯罪主體的決策有顯著影響(Shane D. Johnson & Bowers, 2010)，此呼應理性選擇理論之觀點；而公共與私人設施，如公園、娛樂場所、交通運輸節點等也被做為研究都市犯罪熱點的重要因素，運用與犯罪相關的變項，並結合犯罪群聚地圖，進而犯罪預測，此方法更有助於警政單位有效地識別潛在犯罪高發區並投入警務資源(Kennedy, Caplan, & Piza, 2011)。

當前關於犯罪區位分析涵蓋巨觀與微觀面向，兩者對於犯罪發生的區位因素都具有部分解釋力，相關實證研究也相當豐富，但目前似乎尚無任何單一面向的理論能做出完整解釋。故多年來，犯罪區位分析與犯罪學理論發展踏著同樣步伐，逐漸由「理論建構」走向「理論整合」，呈現各理論間除相互競爭外，也存有許多的整合情況，其目的無非是找出最佳能夠解釋更多犯罪現象理論，以對於犯罪現象產生的原因做出更好解釋，並對其未來發展做出正確預測。因此，立基於犯罪學理論的犯罪區位分析，也期望能藉由理論之整合，以便能從巨觀到微觀層面，由不同的視角與面向探討犯罪發生的區位因素。

在上一節中本研究臚列出經常被用於討論住宅竊盜的相關犯罪學理論，就單一理論而言，均能對住宅竊盜發生原因做出部分合理的解釋，而若能將這些理論加以整合並萃取重要變項，相信更能有助於後續實證分析的發現。惟值得一提是，

理論之整合並非恣意而行、隨興所至，必須有其整合條件與適當方法；例如，理性選擇理論係基於古典犯罪學派與經濟觀點之理性選擇，假設人是具有自由意志 (free will)，以經濟學理論的預期效果原則 (expected utility principle) 為根基，所謂的預期效果原則即是人們根據其預測最大利益與最小成本或損失的機率之內容來做理性的決定 (江振亨，2002)。基本上，理性選擇理論對於人是否犯罪之推論是基於對於人性之「假設」，但此種對人是理性的假設也受到許多實證派之批評，甚至也有學者認為人在犯罪時是非理性、認知扭曲，並從理性選擇理論的基本假設、認知能力提出反駁；此外，也有許多理論探討犯罪發生時，並未將人是否有理性選擇列入考量之列，如社會解組理論；另從理論層次的視角來看，巨觀理論與微觀的理論是否具有整合條件？如何整合？也是學界長期以來討論的課題。故以下首先概述理論整合的意涵與方式；其次，探究哪些理論適於整合及其條件；最後，本研究希能藉由理論的整合之引導，研提適切的研究架構，俾利於住宅竊盜犯罪之區位分析。

二、犯罪學理論整合的意涵與方式

(一) 犯罪學理論整合的意涵

所謂「理論整合」，依據美國學者 Farnworth 的見解，係指以理論間共通相似之部分為基礎，組合兩種或兩種以上先前存在之理論；另外，Thornberry 認為整合理論系指將兩組或兩組以上具邏輯相關之命題予以組合，形成較大型之相關命題，便以對獨特現象提供較為周延之解釋 (許福生，2007)；Pearson 和 Weiner 認為，理論整合就是要尋找一個能放之四海而皆準的共通理論 (Akers & Sellers, 2013)。Steven F. Messner, Marvin D. Krohn 和 Allen E. Liska (1989) 認為，理論整合就是將兩個或更多現已存在、具有共同點的理論組合成單一且重新構造的理論模式；曹立群等認為，整合就是最大限度地解釋依變項，找到最能解釋依變項的幾個關鍵變項的組合 (曹立群、周素嫻，2007)。上述觀點的共同點是：第一，理論整合目的是形成解釋力更強大的理論；第二，被整合的部分具有共同相似性，或者在邏輯上相關；第三，整合形成的理論具有完整性，在內部結構上無衝突。

事實上，在犯罪學理論中，許多新理論的提出，或多或少都吸取過去理論的精華部分而加以修正或賦予新的詮釋；例如，美國社會學家 Agnew 修正 Merton 之古典「緊張理論」 (strain theory) 又稱「文化無規範理論」 (anomie theory) 而提出

「一般化緊張理論」(general strain theory)。兩者之差異主要在於探討犯罪角度之不同，Merton 是採取巨觀的層面探討犯罪問題，不強調個別差異性，而 Agnew 的緊張理論強調微觀的層面，指出個別的差異性，如男性犯罪率為何高於女生，少年犯罪為何多於成年。此外，兩者對於「緊張」的定義與來源也不同，Merton 的緊張理論，主張文化目標與追求文化目標手段間的落差，導致低階層青少年產生「身分挫折」的問題，而以犯罪手段作為達到文化目標的替代手段；Agnew 的一般緊張理論，除將「緊張理論」的犯罪手段吸收整合外，並另外提出 3 種會促使一般人產生負面影響狀態的壓力來源。

基於以上可知，沒有一種理論是全然可取或不可取，每一種理論都或多或少有其合理的成分，故犯罪學理論整合，就是在保持理論科學性的基礎上，尋找一種能夠解釋盡可能多的犯罪現象的理論；其不必包含被整合的理論所有的概念和變項，而是有選擇地選取適合自身研究可運用的因素，其結果宜應包含其他理論的重要成分，又能再補充其不足或賦予其符合時代背景之新意義。

然而，學界對犯罪學理論整合的必要性在認知上仍不一致。一種觀點認為，理論的繁多源於犯罪原因的多樣化，是犯罪學發展的自然產物；與之相對的觀點則認為，理論的紛繁複雜反映犯罪學尚非成熟的科學，故有些學者基於 Popper (1959) 的立論，主張理論只能以證偽(falsifiability)³加以檢驗，對理論整合持直接反對或強烈質疑的論調；他們認為，不同理論背後的哲學基礎、發展邏輯、概念變項等方面都或多或少有差異，強加整合容易造成邏輯上的錯誤或概念上的模糊，最終消弱理論對犯罪現象的解釋力(李波, 2011)。相較於理論整合與證偽的主張，部分學者則採取折衷立場，認為證偽法並未過時，整合則需進一步探索；犯罪學中紛繁複雜層出不窮的理論學說既需要證偽也需要整合，既對整合不須排斥但也不能整合過度。整合過度，可能會使理論的論述更加模糊不清；但是如果完全忽視整合，則許多看似矛盾的理論中的共同點就可能被忽略掉，產生見樹不見林的結果，導致減少理論對犯罪現象的解釋力(曹立群、周素嫻, 2007)。

雖然證偽法反對理論整合，但證偽通常是用以檢驗理論的整體層面，它只能驗證理論，對於理論中有用的變項卻無法將其提取出來。因此，即使反對整合的聲浪如此之大，理論整合還是發展出諸多成果，例如澳洲犯罪學家 Braithwaite

³ 證偽性(falsifiability)，又稱可反證性、可否證性，為 Karl Raimund Popper 所提出，在科學和科學哲學中用來表示由經驗得來的表述所具有的一種屬性，並使用嚴格否證法來判別一個理論是否科學，即這些結論必須容許邏輯上的反例的存在。

的「明恥整合理論」(theory of reintegrative shaming)即整合標籤理論 (labeling theory)、犯罪副文化理論(subcultural theory)、控制理論(control theory)、機會理論(opportunity theory)及學習理論(learning theory)等理論中互補與共通部分而形成的整合理論(Braithwaite, 1989)，進而對修復式司法 (restorative justice) 制度的建構重大影響。

(二) 犯罪學理論的整合方式與類型

1、Hirschi 之觀點

Hirschi(1979)指出理論整合的類型具有「上下整合」(up-and-down)、「並排整合」(side-by-side)與「首尾相連整合」(end-to-end)3種。簡述如下：

(1)「上下整合」型

此方式係先尋找出具有類推特性(generality)之主要概念層級，再吸收其他理論之概念，而加以融合，此種整合方式經常可以產生更多的解釋，因為不同理論的解釋方式，整合後可能發現對於某些結果有共同影響，此種方法有如統計上的「多元迴歸」(multiple regression)，藉由增加更多變項(許多變項原本看似無關)，以增加預測能力；但這種方法也有其缺陷，採用此方法之整合論者易傾向於毫無疑問地接受理論整合後每一部分真實性，且因核心概念是從別的理论鑑借而來，所以在發展新理論後不免受到前理論的制約。

(2)「並排整合」型

將理論放在一起，找出理論共通(重疊)部分(如解釋目的相通)，適當加以分類，並予以抽離，形成目的一致之整合，可避免理論不相容性的問題，並回答相對重要性的問題，此方法可避免理論整合後過大的問題，但在解釋犯罪行為的意義上有很大的不同。Hirschi 認為這是一種很適當的理論整合方法，這使得每種理論都可以「用自己的術語」(in its own terms)來定義犯罪行為。例如 Bernard 和 Snipes(1996)以 A 理論解釋暴力犯罪和 B 理論解釋貧困人口犯罪為例，綜合理論解釋貧困人口暴力犯罪(Benefiel, 2014)。

(3)「首尾相連整合」型

是指將不同理論的部分頭尾相接，使理論成為序列性發展，前面理論的依變項成為後續理論的自變項。通常情況下，首尾相連整合型理論所解釋的內容要少於由其個別組成理論所解釋的內容的總和，亦即吾人無法期待理論整合後解釋力必然會提高，其原因在於前面的理論的解釋力可能被後面理論所吸收。

2、Liska 與我國學者之觀點

美國犯罪學者 Liska 指出，理論整合可用概念整合(conceptual intergration)或命題整合(propositional intergration)方式進行(Messner, Krohn, & Liska, 1989)。概念整合指各犯罪理論間具有共通意涵、概念之關鍵變項，將類似概念變項予以吸收，如「社會學習理論」對犯罪有利或不利的定義，吸收「控制理論」的信仰變項即為例證(許福生，2007)；命題整合係將不同犯罪理論間，根據其目的加以整合，而不論其主張為何(蔡德輝，楊士隆，2019)，如「無規範理論」與「衝突理論」皆可解釋中下階層的高犯罪而加以整合；又如從「社會解組理論」出發，結合「控制理論」的家庭依附變項為另一例證。Liska 等人進一步認為，理論的整合形式雖有不同，均可運用於微觀、巨觀或是跨越兩者間之層級，而在理論整合過程之要務在於精心設計以求理論間邏輯一致性，故拋開理論間之藩籬，實為理論整合成功之關鍵。

(三) 犯罪區位分析主要理論與整合

1、犯罪區位分析主要理論

犯罪學理論的發展過程，就像任何社會科學一樣，在於試圖回答一個問題假設的形成。對於犯罪學家來說，問題可能是為什麼有些人從事犯罪行為而另一些人不從事呢？為什麼不是每個人都參與犯罪？什麼環境因素導致或加劇犯罪發生的可能性？就犯罪區位分析而言，關注焦點在第三個問題—「什麼環境因素導致或加劇犯罪發生的可能性」，主要目的除發現犯罪地理分布的特性問題外，並尋求犯罪學理論對於相關發生因素的解釋。

在探討犯罪地理區位的領域中，社會解組理論與日常活動理論為主要 2 項重要理論，Miethe 和 Meier 認為社會解組理論和日常活動理論兩者是相輔相成(Miethe & Meier, 1994)，Bursik 和 Webb 則進一步主張兩種理論可相互整合(Bursik & Webb, 1982)。而 Rice 等學者將兩種理論加以整合後，並以汽車竊盜犯罪為例，

指出理論整合後具有更多的解釋力(Rice & Smith, 2002)，而國內許有犯罪研究也同樣採行此兩種理論對各類型犯罪進行分析，尤其是對住宅竊盜犯罪的研究幾乎無法忽略此兩大理論的重要性。

就理論整合的方式而言，微觀層面的日常活動理論⁴與巨觀層面的社會解組理論的整合，為典型的「並排整合」型，後者正好為前者提供一個宏觀的理論背景。該整合模式簡便易行，能夠彌補單一理論的不足，增強理論的解釋力，且每個理論的獨特性不會受到影響(李波，2011)，在 Rice 的研究也曾證實此項說法。

值得一提的是，國內外許多研究在引用此兩項理論時，常未對此兩理論如何整合的機制有所論述，研究過程僅是將涉及兩理論的相關變項納入，再藉由統計技術之運算軟體執行變數檢定或迴歸等相關性或因果關係的分析（屬區位分析），研究結果或能顯示多項人文環境因數與犯罪率表現具有顯著關係，並依據結果提出建議，但較缺乏對理論如何連結與整合機制的探討，可能因而忽略重要的概念與變項；例如：社會解組理論在 Kornhauser(1978)的著作發表之前，犯罪次文化(subculture of delinquency)一直被社會解組理論家認為是犯罪和犯罪動機起源的關鍵，因此早期的社會解組理論並沒有在犯罪的起源和動機多有著墨(Matsueda, 2017)，但 Kornhauser 的著作發表後，對於該理論的解釋中「犯罪動機」的概念開始被「假定」，其提出的「社區控制模式」(community control model)認為社會解組的發生是因為缺乏社區控制使得犯罪動機可以在行為中表現出來。而後經由 Sampson 等人從相反面向對社會解組之意涵進行討論後，以解釋社區結構以及社會控制對犯罪率的影響，並進而發展出社會凝聚及相互信任的集體效能概念；但許多引用社會解組理論進行測量的研究，經常忽略集體效能對於犯罪動機與機會抑制的作用。

而當社會解組理論藉由「社區控制模式」衍伸出「犯罪動機」此概念與假設後，其與日常活動理論至少在兩個方面有部分重疊，即它們對社區社會控制的處理以及對犯罪動機的假設，此提供兩理論異中求同的整合契機，但由於雙方對於犯罪控制與動機問題的觀點相當不同，在整合前必須對此加以討論，以釐清雙方概念差異，進而尋求其整合方式。

⁴ 日常活動理論 1979 年首先提出時主要在探討美國「直接接觸掠奪性犯罪」及「住宅被竊」犯罪率的增加。認為人們生活方式的改變、犯罪標的物的變化及監控型態的變化正反應了「日常活動」的變化，因此乃以「日常活動理論」稱之，屬巨觀的犯罪（被害）理論；目前許多研究也將其應用至微觀犯罪事件的分析。

2、社會解組理論與日常活動理論的概念差異

日常活動理論強烈認為動機是存在且是此理論構成的重要因素之一，但動機在社會解組理論中可能被假設為社區特點的產物（貧困、種族異質性、人口流動）。換言之，在研究實際操作中，日常活動理論通常假設有足夠數量的具有犯罪動機者，故操弄的關鍵變項是機會和監控；而社會解組理論的社會結構因素產生了有利於犯罪動機行為出現的環境，即地區特性（區位特性）造成犯罪動機，從某種程度而言，社會解組理論更強調犯罪動機的起源。

在控制概念方面，日常活動理論中的有能力監控概念和社會解組理論中的初級（家庭）、次級（鄰居）和3級（警員）控制概念不同處在於日常活動理論從微觀層面（定義為直接可見周圍環境潛在的犯罪事件）觀察當下監控情況，而社會解組理論則從更宏觀取向，關注「鄰里嵌套」(nested neighborhoods)的概念，亦即「鄰里」的小單位的社會控制機制被嵌入到大單位的社會控制機制中，這使得社會解組理論關注在低度社區控制（如異質性）的原因，而日常活動理論則不重視環境異質等特性等問題(Rice & Smith, 2002)。

日常活動理論與社會解組理論除在控制與動機有所不同外，在犯罪「機會」方面，兩項理論也有明顯的歧異。社會解組理論隱含地做出假設認為：對於那些有動機的人來說，犯罪的機會是恆定，並且（或者）這種機會是犯罪決定的附帶條件，社會解組理論事實上忽略刺激個人成功將動機轉化為行動的機會。而日常活動理論則認為，大部分犯罪是由潛在罪犯遇到的情況機會所引發，犯罪者在特定情境下，判斷何為合適的機會，並將動機轉化為實際行動，有動機的犯罪者故意尋找目標，更多的機會也將為他們提供有利於犯罪活動的環境(Cohen, Felson, & Land, 1980)，犯罪機會可說是相當充足(Mayhew, 1990)，故當代 Clarke「情境犯罪預防」(situational crime prevention)乃藉由消除犯罪機會來控制犯罪。

綜合兩理論間差異可知，巨觀的社會解組理論強調社會脈絡情境與地區特性，貧窮、異質性等造成集體效能低落，導致犯罪的發生，但其卻不關心個別的潛在犯罪人犯罪之機會與受害者風險；而微觀角度的日常活動理論，雖關注於個人犯罪的當下原因，與被害可能風險，卻忽略社會結構與區位因素對於犯罪發生的整體環境，故兩者皆有理論均有欠缺周延之處，但也凸顯相互整合以形成互補的價

值性。

3、具體整合方式

過去許多學者曾對社會解組與日常活動理論的整合提出許多建言，並進一步以實證研究加以驗證，已奠定相當基礎。若干學者建議，在對巨觀的鄰里分析中，可將較常用且具代表性的社會解組變項，採用日常活動機會觀點的思維，探討結構對機會的影響與對受害者之影響(Rice & Csmith, 2002; Smith, Frazee, & Davison, 2006; Vesey & Messner, 1999)，具體來說，就是運用兩理論具有共通概念的變項做為界接的橋樑。例如，Miethe 和 Meier(1994)將社會解組理論和日常活動理論結合為整合模型，並以其探討汽車竊盜犯罪的區位因素；此模型假設這兩項理論都提出它們與犯罪事件之間關係的脈絡背景，Miethe 等人認為動機（或至少是有利於動機表現的條件）和機會可能對犯罪地點有直接影響，但它們對解釋犯罪率差異（或犯罪熱區）的大部原因是由環境特性所造成。而在整合的方式上，Miethe 等人則先以概念做為平臺，再以變項加以連結；例如，在「社區控制」方面，認為無論是社會解組理論或日常活動理論，「住宅自有率」都是一項重要的指標，屋主對自己的房產及其周邊地區擁有既得利益，因此與非住宅擁有者相比，自有住宅之屋主更可能對違法者採取舉報或潛在違法者採取行動(Rountree, Land, & Miethe, 1994)，而社區鄰里間自有住宅率越高，共同利益越多，更易產生集體效能。

另 Triplett 等人從社會解組理論出發，探討「社會控制」與「集體效能」對犯罪率影響。在其研究中將社會解組常用的傳統變項列入第一層次測量其與犯罪率關係，如：住宅自有率、女性人口比例、年齡等做為測量變項，另將酒吧、賣酒商店列入第二層次測量(Triplett, Sun, & Gainey, 2005)，而酒吧等變項則是經常被做為新機會理論之變項。

根據上述理論整合方式之討論可知，透過變項的整合為廣泛並可行應用方式，由於日常活動理論包含於「新機會理論」中，且日常活動理論與生活型態理論，甚至生活暴露理論在許多方面互為相通且變項亦可交互使用，故本研究即以強調機會的「新機會理論」與「社會解組理論」做為犯罪區位分析的基礎研究概念架構。

第四節 資料探勘與犯罪分析

一、犯罪資料的探勘技術

在犯罪防治研究領域，除對犯罪資料進行分析外，如何運用分析結果進而對犯罪預測並加以預防為最重要課題；然而，犯罪問題近來深受網路科技發展的影響，如行動裝置的普及化、社群媒體平臺的發達等，使得犯罪活動的觸角延伸至世界各地，更趨向國際化、科技化快速發展，造成現今犯罪環境日趨多元複雜，引發層出不窮的新興犯罪問題，使得犯罪分析、預防與偵查對策的挑戰日趨嚴峻。因此，面對犯罪問題的發展趨勢，現今執法人員必須具備更前瞻的思維、運用最新科技與嶄新研究方法來管理與應用資訊科技，包括資訊的取得、分析與應用，讓犯罪研究與實務工作能在先進科技的支援下，掌握時代脈動、與時俱進，而資料探勘技術對於當前犯罪之多元複雜特性，堪稱一大利器。

資料探勘技術用於發掘大量資料中「未預期的」資料模式及「隱藏的」規則，其使用多種先進的演算法(algorithm)，以充分探索並且找出一個或多個數據源的數據集特徵，以及找出資料的重要可辨識的模型、趨勢、與關連性。資料探勘源自於統計原理，但和一般用統計方法來進行資料分析有很大的不同，資料探勘分析前不需要對資料提出假設，統計分析則藉由對假設進行檢定等方式，希望能發現單一規則或單一模式來顯示資料的性質，但現實上大多數的資料是由多條規則或多個模型混合產生，故面對資料的複雜性，傳統的統計分析技術不易發掘隱藏的規則；此外，資料探勘除運用統計的概念來判讀資料的性質外，更結合現代電腦的高運算效能，來歸納規則或模型，如此可使資料的詮釋更具彈性，也較能發掘出真正隱藏在資料背後的資訊。因此，簡單來說，資料探勘就是以統計學和電腦科學為基礎，所發展出來能快速分析資料的方法(翁慈宗，2009)。

結合上述，犯罪資料的資料探勘，即是運用各種演算法針對現有之刑案紀錄、法院判決書等，在加上與犯罪資料看似無關聯的各類結構化與非結構化資料⁵，如：日常消費資料、社會經濟統計資料、氣象資料、時間資料等，將資料歸一化後⁶，執行演算分析並建立模型，進而加以預測，協助執法機關在有限資源的前

⁵ 簡單地說，能夠用數據或統一的結構加以表示，稱之為結構化數據，如數字、符號；反之，非結構化數據指的是如文本、圖像、聲音、媒體等資訊。

⁶ 資料歸一化（與標準化概念相似）問題是資料探勘中特徵向量(features vector)表達時的重要問題，當不同的特徵成列在一起的時候，由於特徵本身表達方式的原因而導致在絕對數值上的小資料被大資料掩蓋的情況，這個時候我們需要做的就是對抽取出來的特徵向量進行歸一化處理，以

提下，主動進行犯罪行為的掌握與預防，藉以發展有效的策略提升執勤效率。然而，犯罪分析與預測僅是犯罪預防的一個過程，資料探勘的結果的詮釋，仍要仰賴專業工作人員具備相關背景知識來予以解讀結果；換言之，資料探勘運用在實務工作上，絕非少數人所能完成，其乃集合相關領域之專業人士以達眾志成城的團體作戰。

二、資料探勘主要技術

資料探勘有兩個主要的目標：分析與預測。前者是指描述在這複雜的資料中，到底發生了什麼事，後者指的是根據資料屬性之過去觀察值來推估該屬性未來之值；因此，目標不同須採不同的資料探勘技術。然而，資料探勘的技術相當繁多，至今各種演算法仍不斷推陳出新，例如使用 R 語言進行程式撰寫，進入微軟公司的軟體原始碼代管服務平臺「GitHub」(<https://github.com/>)，對類似問題的不同解決方法與相關套件源源不絕，但這多如牛毛的技術如何劃分類型，目前還沒有一致的見解，惟群集分析(clustering analysis)、分群(類)分析(classification analysis)及關聯規則探勘(association rule mining)3種類型的資料探勘方法，應是一般學習資料探勘必備之基礎知識，坊間相關教科書中也均會對此3部分進行討論，且3種方法對於犯罪資料處理亦相當適用；故以下簡介此3種資料探勘技術概念。

(一) 群集分析

群集分析的分群方式是將資料按照某類標準歸為幾個類別，其中特徵較相近者聚為同類，不相同者另歸為一類，其原理是將組間差異最大化與組內差異最小化。目前有關分群分析的演算法相當多，一般來說，可分成4大類：分割式、階層式、密度基礎和格子基礎的分群方法(陳榮昌，林育臣，2003)，較具代表性的有 K-平均值分群(k-means)、k-中心點分群(k-medoids)、密度分群(density-based spatial clustering of applications with noise, DBSCAN)、階層分群(hierarchical clustering, HC)、期望值最大化分群(expectation-maximization, EM)等(黃文、王正林，2016)，其中 k-means 演算的邏輯簡單易懂、可以接受的時間複雜度的特性，應用最為廣泛，在 GIS 群集分析也經常被使用。

保證每個特徵被分類器平等對待。

k-means 演算法是屬於分割式分群方法，它以重心點或是中心點為基礎的方式，將資料群體進行分群，因為各群體的代表點不一定是群集中的一點，所以可以在多數的情況下找到最佳群集，它是一個非常普遍且廣泛被使用的資料分群方法，將所有樣本點分割為 K 個原始群集，此 K 個群集重心稱為種子點(seed point)。重複針對每一樣本，計算比較該樣本在該群中之離差平方和(sum of squares of deviations)以及失去該樣本後該群之離差平方和。若失去該樣本後可使離差平方和降低，便將該樣本改為指定至其他群集中，重複上述之計算，目的是要使各群中之離差平方和最小化，直到各樣本不需重新指派到其他群集中，便可完成分群。此方法缺點為容易受到離群值(outliers)所影響，優點在對於小型的資料集合有著不錯的處理能力，但是處理大型資料庫就宜另採以取樣為基礎的 CLARA(clustering algorithm based on randomized search)(陳榮昌、林育臣, 2003)。

(二) 分群分析

在資料探勘領域中，分類分析被視為一種監督式學習(supervised learning)的知識發掘⁷，這類型的探勘方式是具有目標導向性的，在執行分類前有特定的變數是希望被推測的，且希望將某組特定的分類套用至原始資料，並找出特定的關係。根據 Han 等學者的定義，分群是透過分類演算法建立分類模型或分類器(classifier)，以預測研究者對於觀察目標所關心的類別標籤(categorical labels)，例如：在住宅竊盜犯罪資料中，用此分析可區辨造成地區性竊盜風險「高」或「低」的各相關規則；這些類別可以用離散的數值來表示（如「高」以 1 表示，「低」以 0 表示），其數值順序與大小並無比較上的意義(Jiawei Han, Kamber, & Pei, 2011)。目前在資料探勘演算法中可以做分類的方法有很多，例如：貝式分類法(bayes classifier)、分類與迴歸樹(classification and regression tree, CART)簡稱決策樹(decision tree)、支援向量機(support vector machine)、類神經網路(neural network)

⁷ 監督式的統計學習，是在已知的資料輸入項目後，能透過模型與對應關係的建構，得到可以預期或是有預測能力的特定資料輸出，這樣的學習過程被稱之為監督式的統計學習，因為我們能夠透過模型與變數間相對關係的界定去解變化的情況；簡單的說，就是透過這樣的資訊解析過程，瞭解其中資訊的變化，並進行相關資訊的萃取與解讀。

非監督式(unsupervised)的統計學習，則是在給定相關資料輸入後，透過適當的資料處理與聚合，讓資料替自己說話，呈現出資料之間彼此相關的程度，但是對於資料輸出的情況是無法預測，也不可能在前因為對過往模型的熟悉而有相關的預測能力者，資料最後輸出的型態完全取決於目前手邊資料自身的特性與型態，而我們只是透過特定的邏輯與方法，來呈現資料自身原來的樣貌，這樣的學習過程被稱之為非監督式的統計學習。

等。

以上方法中，決策樹是一個視覺化清晰，容易一目了然的分類規則的方法。決策樹是一個類似流程圖的樹狀結構，包含根節點(root node)、子結點(child node)，與葉節點(leaf node)，節點與節點之間以分支連結。決策樹的建構與分類過程相當簡單快速，產生的規則也容易被研究者所解讀吸收，分類圖形容易判讀，且大致擁有不錯的準確度，故獲得廣泛的使用，常被當作與其他分類方法的評比優劣之參考對象。但是決策樹的建構需要品質良好的資料，否則會因為雜訊的影響產生許多小分支，必須經過修剪(tree pruning)才能增進準確度；由於決策樹易於解讀的特性，衍生出更多基於此種分類方式的改良演算法，如條件推論決策樹等，本研究在資料探勘的分析中，主要便是以決策樹系列（族）的方法進行。

（三）關聯規則探勘

關聯規則探勘又稱購物籃分析(market basket analysis)，是資料探勘的核心技術之一，其連結規則模型及演算法為 IBM 公司所提出，其中最經典的 apriori 演算法在關聯規則分析領域具有巨大影響力，其特性為簡單、容易瞭解及實作。此項技術已頻繁使用在各領域，最常見到就是瀏覽購物網站時，出現「購買此商品的顧客也同時購買」等提示語。簡單地說，關聯分析係探討大量資料中變數間相互連結程度；此外，關聯規則探勘類似統計的相關性(correlation)，但兩者計算方式不同，不可混為一談。

形成關聯的 3 個主要概念為支持度(support)、信心可靠度(confidence)、與增益度(lift)3 項指標。支持度是建立關聯性原則的第一指標，透過設定最小門檻值，篩選出頻繁組合；信心可靠度是當某先決條件(X)發生下，對應結果(Y)發生的機率；提升度為信心可靠度的互補指標，評估 XY 同時出現與 Y 出現的機率，避免高估其關聯性。支持度、信心可靠度及增益度 3 項指標分別代表關聯準則的顯著程度(Significance)，正確程度及價值。其機率表示如下：

- 1、支持度：衡量前提項目 X 和結果項目 Y 一起出現的機率； $\text{Prob}(X \cap Y)$ 。
- 2、信賴度：測量前提項目 X 發生時，結果項目 Y 也出現的條件機率； $\text{Prob}(X|Y)$ 。
- 3、增益度：測量比較信賴度與結果項目單獨發生時的大小； $\text{Prob}(Y|X)/\text{Prob}(Y)$ 。

在資料輸入方面，關聯規則探勘必須輸入離散的屬性值，如果該屬性為連續性的數值，必須將其離散化方可進行分析。

三、資料探勘的步驟

為了能從資料中挖掘出有用且隱藏的資訊，資料探勘的技術除上述探勘方法外，相關前置步驟亦不可輕忽，否則將嚴重影響資料品質。

首先，設定明確主題與目標，清楚研究目的，因為在目的不明的狀況下，無從蒐集哪些資料來做分析，或該使用哪一類型的分析方式來進行探勘，否則分析結果並無任何意義。而且在資料量大的情況下，資料探勘所耗費時間或財務成本比使用統計方法或資料檢索的成本高；換言之，資料探勘的技術並非唯一分析資料的方法，研究者應就成本較低的分析方法先行確認可否替代資料探勘，例如：在確認統計方法和資料檢索等都無法產生所需要的資訊後，再進行資料探勘。

其次，在目標明確之後，接著便是蒐集及整理出適合資料探勘的資料格式。由於現今資料的來源有很多種方式，如商業資料、政府公開資料、甚至是語音及影像資料等，這些儲存方式均有所不同，例如：電腦中一筆筆循序擺放的檔案、資料庫中彼此有關聯性的檔案，或是放在資料倉儲中的大量歷史資料，更有許多是來自於網頁資料，這些資料可能是集中存放於雲端資料庫或是散布在多部電腦主機中；資料探勘的第二個步驟，就是把這些不同來源和格式的資料蒐集起來，並用適當的格式來存放。須注意是，資料的蒐集必須考慮倫理與法律問題，尤其是利用「網站擷取」(web scraping)技術從網路自動蒐集資料，是否絕對合法，為蒐集資料時必須注意的問題。

資料蒐整好後，並非立即可使用，必須詳細檢視是否有異常值或遺漏值等；此外，還必須視使用資料探勘的方法，決定資料是否必須進行轉換及標準化；另有些欄位是否必須加以分割或合併也須詳加考慮，如此才能得出較佳的資料探勘品質，進而提高對研究的詮釋程度。

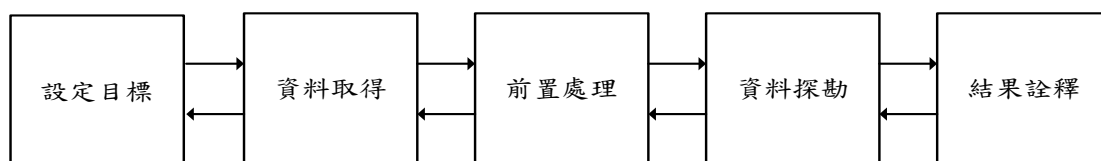


圖 2-3-1 資料探勘的步驟

第五節 實證研究探討

一、住宅竊盜相關研究綜整

國內外有關住宅竊盜犯罪的相關研究相當豐富，研究方法與分析方式也相當多元，成果豐碩。檢視以統計為主要之住宅竊盜犯罪研究取向大多聚焦於犯罪現狀分析、危險因素研究、犯罪成因探討、犯罪者特質分析等，著重於「犯罪前因論」與「犯罪者」之研究(Snook, Dhimi, & Kavanagh, 2011; Stibbs, 1994; Townsley, Birks, Ruitter, Bernasco, & White, 2016; Triplett et al., 2005; 王翔正, 2014; 余任晴, 2019; 邱靖方, 2008; 邱豐光, 2008; 洪百亮, 2009; 陳瑞基, 2010; 楊武德, 2006)，探討導致人們成為犯罪者的心理和社會驅力，希望能找到改變這些因素的方法，並發展出諸多可以解釋住宅竊盜犯罪或偏差行為原因之論點(劉擇昌, 2011)，亦有研究並不使用 GIS，僅使用統計方式對可能造成住宅竊盜的環境變項進行分析(Xu, Fu, Kennedy, Jiang, & Owusu-Agyemang, 2018)；近年來，由於空間資料分析在 GIS 中的應用越來越受到重視，故有關住宅竊盜的空間分析研究也不斷增加；但相對地，以掃描統計結合 GIS 或(並)運用資料探勘技術探討住宅竊盜的文獻相形少見，故以下蒐整國內外以 GIS、掃描統計及資料探勘技術為主要研究方法之住宅竊盜研究進行綜合討論。

表 2-5-1 相關實證研究文獻整理摘要表

作者 (年分)	研究主題	研究資料	研究方法
李珀宗(2005)	社區犯罪基圖在員警機關防制住宅竊盜犯罪之應用—以臺北市松山區為例	2003 年臺北市松山區住宅竊盜資料	GIS 製圖：疊圖分析；實地觀察法
賴致瑜(2006)	臺北市住宅竊盜犯罪地圖製作與犯罪區位分析	2001 年至 2004 年臺北市住宅竊盜統計資料；2000 年戶口及住宅普查資料	GIS 製圖：犯罪區位分析；核密度推估法
陳陸毅(2008)	臺中市住宅竊盜空間分布之研究	2001 年及 2006 年臺中市警察局提供臺中市住宅竊盜案件資料	GIS 製圖：環域分析；網格分析；犯罪時間分析
黃映翎(2008)	臺北市士林區指標型犯罪之空間分析	1997 年至 2006 年臺北市士林區住宅竊盜、機車竊盜等	GIS 製圖：全域及局部群聚檢定；核密度推估法
陳惠貞(2009)	高雄市住宅竊盜犯罪地圖繪製與時空分析	1999 年至 2008 年高雄市住竊盜犯罪發生數、2006 年高雄市住宅竊盜犯罪案件發生時間與地點資料	GIS 製圖：全域及局部群聚檢定；住宅竊盜犯罪的發生時間分析
黃瀾嬋(2010)	社區犯罪基圖對於員警派出所預防住宅竊盜犯罪之研究-以臺北市中正第一分局為例	2008 年至 2009 年臺北市政府警察局中正第一分局轄區住宅竊盜資料	GIS 製圖：治安斑點圖；犯罪基圖
劉擇昌(2011)	住宅竊盜犯罪熱區空間分析與環境特性之研究—以臺北市大安區為	1999 年至 2009 年臺北市大安區住宅竊盜資料	GIS 製圖：探索性空間分析；全域及局部群

	例		聚檢定；犯罪熱區與冷區分析；田野實地調查
邱奕堯(2013)	臺北市犯罪現象之時空分析	2010年臺北市住宅竊盜及強奪犯罪資料	運用掃描統計進行時空分析，並以GIS視覺化呈現
王占宏(2013)	基於掃描統計方法的上海犯罪時空熱點分析	2006年至2010年上海搶奪與住宅犯罪資料	運用掃描統計進行時空分析，並以GIS視覺化呈現
彭福康(2015)	監視影系統與竊盜犯罪率之時空分析	2008年至2014年臺北市竊盜犯罪案件(普通及住宅竊盜、機車竊盜、汽車竊盜)資料	GIS製圖：平均最近鄰、核密度分析、地理權重分析；住宅竊盜犯罪的發生時間分析
周德芳(2016)	基隆市住宅竊盜犯罪空間與區位分析	2008至2012年基隆市住宅竊盜案件	GIS製圖：犯罪熱區與冷區分析
林維真(2017)	以時間與空間觀點探討住宅竊盜的發生	2014至2016年臺灣北部特定區域之住宅竊盜犯罪資料	GIS製圖：環域分析、疊圖分析
林宛宣(2019)	應用資料探勘於臺北市住宅竊盜環境特性之關聯研究	2015年至2017年臺北市政府開放平臺住宅竊盜案件資料	Google地圖與地標地圖；資料探勘之群集與關聯規則分析

林建隆(2010)	刑案隱性鏈結關聯模式之研究-以臺北市搶奪與住宅竊盜案為例	2004年至2008年臺北市臺北市搶奪與住宅竊盜資料	資料探勘之鏈結分析
徐晨剛(2017)	應用資料探勘技術建構竊盜犯罪預測模型	2012至2014年臺北市竊盜犯罪案件(住宅竊盜、機車竊盜、汽車竊盜)	GIS犯罪熱點製圖；資料探勘之分類樹與隨機森林分析
Townsley et al. (2000)	住宅竊盜重複受害：空間與時間模式	澳洲昆士蘭 Beenleigh 警察局所提供 1995 年 6 月至 1996 年 12 月計 18 個月之竊盜報案紀錄	GIS 犯罪熱區製圖；存活分析；時空移動平均法
Ceccato, Haining, and Signoretta (2002)	運用空間分析工具於斯德哥爾摩之犯罪	瑞典斯德哥爾摩 1998 年住宅竊盜犯罪資料	GIS 製圖：Getis-Ord 法、資料探勘：群集分析
Martin (2002)	住宅竊盜的空間模式-評估鄰里社會資本的影響	1995 年至 1997 年底特律地區住宅竊盜資料	GIS 製圖：空間落遲迴歸分析、統計分析
Poulsen and Kennedy (2004)	比較 Dasymetric 和 Choropleth 繪圖差異，並分析住宅竊盜之犯罪熱點	2001 年美國麻薩諸塞州警察局提供之住宅竊盜資料	GIS 製圖：Dasymetric 製圖法、Choropleth 繪圖法
Malczewski and Poetz (2005)	安大略省倫敦區的住宅盜竊的社會經濟因素分析	1998 年至 2001 年安大略省倫敦區 8,534 筆住宅竊盜資料	GIS 製圖：以地理加權回歸 (GWR) 分析
Ratcliffe (2005)	區域內犯罪模式時空變化趨勢之偵測	2001 年澳洲雪梨 Canberra 區進行打擊住宅竊盜專案，實施前後資料	GIS 製圖：最近臨分析，蒙地卡羅檢定

P. Chen, Yuan, and Li (2013)	北京住宅竊盜的時空分析	2007 年北京朝陽區住宅竊盜	GIS 製圖：以「鄰近重複被害模式」進行分析
Mahfoud, Bhulai, and Mei (2017)	住宅竊盜的時空模型	2008 年至 2014 年土耳其警方提供之住宅竊盜資料	運用「廣義加成模型」(GAM)分析，再以 GIS 視覺化呈現
Awal, Rabbi, and Rana (2017)	運用資料探勘技術對孟加拉犯罪分析	以孟加拉警方網站公開資料(住宅竊盜及強盜等)預測 2018 犯罪趨勢	運用線性回歸及關聯規則技術進行分析
Hoppe and Gerell (2019)	馬爾默重複竊盜的模式：時間的穩定與變化	2009 年至 2014 年間瑞典第三大城市馬爾默 (Malmö) 住宅竊盜案件為分析資料	GIS 製圖：以「鄰近重複被害模式」進行分析

以上文獻基本上可分成 3 部分探討；首先，GIS 主要焦點在於找尋住宅竊盜的犯罪熱區與群聚，並經由地點分析結合相關理論，發現影響犯罪之相關因素，若干研究也會將時間因素在研究中呈現；而時空掃描統計除了偵測犯罪地點群聚外，更進一步將時間因素納入分析，凸顯犯罪空間在時間趨勢的變化；資料探勘技術目的則在發掘住宅竊盜的相關成因及區位因素。

二、GIS 於住宅竊盜之研究

GIS 在犯罪分析有多種方式，但對於運用 GIS 進行住宅竊盜熱區的研究而言，最基本也是最重要的第一步驟即是要先探測出群聚所在，否則後續研究將難以進行，而在發現犯罪群聚後，可再根據該群聚區的人文及環境等區位特性進一步分析造成犯罪群聚成因。

然而，目前對於住宅竊盜資料之取得，無論是由警方提供，或是來自於政府公開資料平臺，有關住宅竊盜犯罪者的個資與竊盜手法，較少列入資料內，故多數的研究主要以區域、都市、城鄉或鄰里為研究單位，屬於巨觀分析層次，而後

續的資料蒐集與分析也是依據選取的地理尺度，如：若以里為分析尺度，則研究者將依據社會解組理論、日常活動理論等選取衡量指標，蒐集各里的人口遷移率、離婚率、獨居與否及各類社經資料等，進一步加以分析。

在上述的文獻中，由於研究地區與時間不同，研究結果有時出現不一致現象。例如：有研究指出，住宅竊盜易發生於經濟水準與居住環境較佳地區，此現象與日常活動理論強調有較多的標的物易引發犯罪觀點符合(賴致瑜，2006)；然而，亦有研究指出，在較為貧窮區域中，潛在犯罪者較多，且住宅防衛功能低，反而較易產生住宅竊盜，社會解組理論似可對此提出說明(Malczewski & Poetz, 2005; Martin, 2002；黃瀾嬪，2010)。其餘在各項影響住宅竊盜因素方面，許多研究均指出人口密度與住宅竊盜有密切關係，人口密度高的住商混合區住宅竊盜發生率較高(李珀宗，2005；陳陸毅，2008；彭福康，2015)；獨居與否、單親家庭、失業率、經濟收入、離婚率、低教育人口比率、租屋率、男女性別比、老人人口率、監視器密度及人口流動性等因素對於住宅竊盜都具有若干之影響(Martin, 2002；周德芳，2016；林宛宣，2019；陳惠貞，2009；彭福康，2015；賴致瑜，2006)。整體來說，不同研究因所蒐集資料之差異，所得出的研究結果不盡相同，但均可以犯罪學的理論予以詮釋，其中社會解組理論與日常活動理論為經常被引用之理論，原因在於日常活動理論所提出「犯罪三角」易於研究設計、驗證與解釋；而社會解組理論從巨觀層面解釋犯罪發生，涵蓋範圍更大，且目前的政府公開資料庫中的人文社經資料，多與此理論有直接關聯，故在犯罪區位分析上，社會解組理論為不可或缺的基礎理論。

值得注意的是，上述研究中，若干研究也同時對住宅竊盜犯罪時間進行分析，目前較常見之做法係將每筆住宅竊盜發生之時間加以累計，得出次數最多時段，相關結論如：高雄市住宅竊盜高發時段約在 12 時至 17 時與 17 時至 19 時兩個時段(陳惠貞，2008)；臺北市住宅竊盜自上午 5 時至 6 時逐漸往上攀升，直到 18 時至 19 時達高峰後逐步下降，高發生率之時段為 8 時至 23 時，而非深夜 0 時至 8 時(彭福康，2015)，此外，以臺北市大安區 1999 年至 2008 年為例，亦得出類似結論(劉擇昌，2011)。然而，住宅竊盜與搶奪竊盜等不同，警方提供之報案時間紀錄經常是發現被竊後之時間，而非真正竊盜時間；因此，以小時進行分析，準確率仍待商榷，惟以上研究推論多指向都會住宅竊盜發生，高發於民眾外出工作時段，致住宅缺乏監控力所致，與日常活動理論觀點仍為符合。

從日分布來看，劉擇昌(2011)研究指出，臺北市大安區星期一至星期五非假日住宅竊盜犯罪發生率較高，其中又以星期一、星期五最高；另有研究以全臺北市進行分析，得到完全相同的結論(邱奕堯，2013)。但在月分布方面，則結論不一，例如：臺中市 2006 年的住宅竊盜犯罪，在春季(1 月至 3 月)發生率高於其他 3 季(陳陸毅，2008；陳惠貞，2008)，但其他研究則顯示 9 月、10 月為最高峰，2 月、8 月之發生率反而較低(劉擇昌，2011)，或 1、2、4、5 月為發生率較高月份(邱奕堯，2013)；究其原因，除地區、年份因素不同外，若能採用時間群聚檢定之分析，或可得到更準確之結果。

三、時空群聚分析於住宅竊盜之研究

過去，有關犯罪群聚研究多著重在空間或時間的單一維度(one-dimension)分析，較少關注什麼因素導致犯罪的時空群聚，及犯罪如何在這時空兩個維度同時發生變化(Shane D. Johnson et al., 2007)，例如：若僅用時間序列分析(time series analysis)，可能因忽略資料的空間架構，無法即時提出異常警訊，而若以純空間統計分析(pure spatial analysis)檢測資料，則可能隱藏事件間的時空關聯性，導致訊號誤判。因此，時空群聚分析(spatio-temporal clustering analysis)對聚焦於犯罪預防的決策與措施具有相當幫助，可有效降低警政資源的浪費。目前，時空群聚分析最常使用於流行病學領域，運用於犯罪學研究相對缺乏，主要原因在於兩者的特性差異，例如：疾病爆發通常具有特定潛伏期，但犯罪事件並無此特性，且疾病的傳染(contagion)與犯罪的擴散(diffusion)亦難完全比擬；因此，運用時空群聚分析於犯罪學研究，視犯罪型態特性，有時可能須進行資料調整，以提高檢定力。

國內以掃描統計對犯罪進行時空群聚分析研究至今仍偏少量，初步檢索 2 篇均為運用 Kulldorff 的時空分析方法之文獻；其一為對臺灣地區貪腐現象進行時空分析之研究，研究結論得出臺灣貪瀆起訴比率較高的風險區域並沒有長時間的固定性，而居民教育程度與當地經濟狀況可能是與貪瀆起訴比率相關的重要因素(廖興中、徐明莉，2017)。另一論文為針對 2010 年臺北市傷害罪與住宅竊盜罪的研究，研究發現住宅竊盜罪的時空群聚有由南向北逐漸轉移的現象，群聚持續的時間長度由長漸短，上半年度的群聚時間長達半年，但下半年群聚僅持續 2 個月(邱奕堯，2013)。

而國外對於住宅竊盜時空群聚研究方面相對較多，方法也較多元，其中 Johnson 等人以住宅竊盜重複受害者受害時間與竊盜熱點之分析成為後續相關研究重要參考；Johnson 等人以「重複被害」(repeat victimization, RV)及「臨近距離重複被害」(near repeat, NR)為核心概念出發，研究發現某地區住宅竊盜案件一旦發生，鄰近風險便開始升高，且容易重複被竊，顯示重複被害區其環境相對脆弱，易形成住宅竊盜的熱區(S. D. Johnson, Bowers, & Hirschfield, 1997)。Hino 和 Amemiya(2019)承繼上述研究方式，調查日本福岡市 2005 年至 2014 公共住宅戶風險，發現公共住宅的風險程度約高於全體住宅竊盜犯罪率的 2 倍。Hoppe 和 Gerell(2019)同樣採用「鄰近重複被害模式」(Near-repeat Victimization Patterns)以 2009 年至 2014 年間瑞典第三大城市馬爾默(Malmö)住宅竊盜案件為調查樣本，發現竊盜案件發生後的第一個星期，在發生點的周圍 100 公尺範圍內，為高風險地區與時段，故建議住宅竊盜發生後的兩周內，民眾與警方均應加強防備。

對岸中國大陸以時空群聚方法對住宅竊盜進行數量亦非常稀少，可能與資料取得困難有關⁸。在近期的研究方面，P. Chen 等人(2013)運用「鄰近重複被害模式」對 2007 年北京朝陽區住宅竊盜進行分析，結果顯示，首次發生住宅竊盜後，在 3 周內距離 200 公尺範圍為重複發生的高風險區。另一研究則運用 Kulldorff 的時空掃描統計，以上海市 2006 至 2010 年為研究區域，結果發現住宅竊盜聚集月分多在 11 月、12 月和 1 月，異常的持續時間較短，一般在 1 到 3 天之間，在熱點分析方面顯示熱點各處分散，而若改變掃描尺度，則全市幾被熱區所覆蓋(王占宏，2013)。

四、資料探勘於住宅竊盜之研究

資料探勘技術選擇預測變項時，適用自變項多、變數間相互作用與迴歸模式為未知之情況，特別適合建立巨觀犯罪預測模型。目前，國內以資料探勘進行住宅竊盜研究數量雖少，但研究方式與成果頗具多樣性與參考價值。

林建隆(2010)利用資料探勘的鏈結分析(link analysis)針對住宅竊盜的犯罪原因、犯罪習癖、準備措施、犯罪方法及犯罪工具 5 種變項進行分析，發現犯罪方法與犯罪工具為影響住宅竊盜最重要變項，而犯罪原因則相對較不重要；此外，

⁸ 相較於民主國家，中國大陸基於社會控制理由，學術研究要取得官方資料相當不易，尤其是社會科學等人文社經資料；通常研究須為政府批准或支持之項目，方能取得相關資訊，但資料正確性仍受質疑。

嫌犯的犯罪地點通常與其住居有「距離遞減」效應，亦即越遠離犯罪發生地，犯案次數越少，故住宅竊盜之發生地點越近時，可推估為同一犯罪者所犯案件之機率較高，而 36.14% 的案件為間隔在 1 公里以內。

另有研究以資料探勘之決策樹、隨機森林、支持向量機器技術進行分析，結果得出單獨生活戶越多、遷入人口數越多，監視器密度越高，共同生活戶越少，住宅竊盜越多之結論(徐晨剛，2017)；而國外結合資料探勘與熱點分析的研究顯示，隨著人口增加，包含住宅竊盜的各類型犯罪都會隨之增多(Awal et al., 2017)，此與前述以 GIS 得出人口密度高有較高住宅竊盜率的結論若合符節。

此外，亦有研究以「環境設計預防犯罪」(crime prevention through environmental design, CPTED)的觀點進行案件發生點位與都市環境間的資料探勘分析，運用群集與關聯規則技術，發現住宅竊盜犯罪多發生在面臨道路寬度為 10 公尺以上連通道路、住商混合使用型態地區、巷道乾淨未出現臨時停車與植栽過高等地區(林宛宣，2019)。

第六節 綜合評析

近年來國內關於住宅竊盜犯罪相關議題的研究，除對犯罪資料進行統計分析外，基於環境犯罪學概念，結合犯罪地點環境及警察機關執法策略的討論也日益增多。環境犯罪學的觀點認為，犯罪發生的原因並非全然取決於個體的動機，更重要是取決於可觀察的環境機會因素，故犯罪並非隨機，而是某一區域的環境結構在特定時間與犯罪者互動的產物。故在上述各節首先由辨識犯罪熱區出發，闡明犯罪熱區不僅囿於空間範疇，必須同時考量時間的重要性，進而提出分析時空熱區的方法，以補足過去實證研究較少觸及的問題；其次，藉由犯罪學理論之探討及其與區位分析的關係，進而探討理論與概念測量整合的可能性；最後，本研究列舉若干與住宅竊盜犯罪環境及區位分析相關研究，說明當前學界研究方式與成果。針對上述，以下擬由「理論與實證研究整合分析」及「住宅竊盜犯罪區位因素分析」兩方面加以評述，以做為本研究後續研究設計及實施方向之指引。

一、理論與實證研究整合分析

綜合上述相關之犯罪學論點及所列舉的國內外相關研究文獻，以下概略區分為巨觀與微觀分析層次，以進一步對住宅竊盜犯罪之發生原因歸納分析。

(一) 巨觀層次分析

巨觀層次的犯罪學理論與社會學理論有很深淵源，故常被稱為「犯罪社會學」，其主要特徵多以社會文化、結構與犯罪之關聯進行探討，如：「犯罪副文化理論」的解釋重心在於低產階級的犯罪行為，強調貧民區的居民之所以違反法律，是因為他們信奉於一套存在於低階層區域的特殊而獨立的副文化體系，而此文化體系卻使他們與中產階級的規範相衝突；另「緊張理論」認為，犯罪和偏差行為是人們在無法達到社會所界定的成功目標之時，對所產生的挫折或壓力的一種反應結果；兩項理論均為解釋低產階級者之犯罪(許春金，2010)；本研究選擇的社會解組理論，基本也是對低產階級犯罪因素進行解釋，但與前兩項理論最大差異在於此理論更聚焦於分析區域的生活條件、環境等如何影響犯罪的發生，與環境犯罪學直接契合。

以社會解組理論做為本研究巨觀理論，主要目的在於探討構成犯罪之群聚與區位因素，實證研究通常以 Shaw 和 McKay 提出的低收入戶多、族群差異大及人口流動率高是造成社會解組的三大結構原因做為構面，具體驗證方式包括：以人口流動率、居民遷移率、種族異質性、經濟收入、教育程度及家庭整合等社會結構因素加以衡量、解釋或預測區域間住宅竊盜犯罪之變異，進而了解這些因素如何引發住宅竊盜犯罪。如前述研究提到較為貧窮區域中，潛在犯罪者較多(Malczewski & Poetz, 2005; Martin, 2002; 黃瀾嬋，2010)；單親家庭、失業率、經濟收入、離婚率、低教育人口比率及人口流動性等因素對於住宅竊盜都具有若干之影響(Martin, 2002; 周德芳，2016; 林宛宣，2019; 陳惠貞，2009; 彭福康，2015; 賴致瑜，2006)。

在上述的文獻中，由於研究地區與時間不同，研究結果有時出現不一致現象。例如：部分研究指出，住宅竊盜易發生於經濟水準與居住環境較佳地區(賴致瑜，2006)；然而，亦有研究指出，在較為貧窮區域中，反而較易產生住宅竊盜(Malczewski & Poetz, 2005; Martin, 2002; 黃瀾嬋，2010)。其餘在各項影響住宅竊盜因素方面，許多研究均指出單親家庭、失業率、經濟收入、離婚率、低教育人口比率等因素對於住宅竊盜都具有若干之影響(Martin, 2002; 周德芳，2016; 林宛宣，2019; 陳惠貞，2009; 彭福康，2015; 賴致瑜，2006)。整體來說，社會解組理論從巨觀層面解釋犯罪發生，涵蓋範圍更大，且目前的政府公開資料庫

中的人文社經資料，多與此理論有直接關聯，故在犯罪區位分析上，社會解組理論為不可或缺的基礎理論。

（二）微觀層次分析

微觀之住宅竊盜犯罪地點分析通常以街道、案發地點及小尺度地區為主要研究單位，除探討研究區域內的人文因素，更重視實質物理環境特性對於犯罪的影響。在本研究與此相關之理論為新機會理論、破窗理論、生活暴露理論等；這些理論主要強調當某地點呈現如住宅防禦性低、區域物理環境監控力弱等有利於犯罪發生之情境或機會時，則會對於潛在犯罪者產生正面之心理增強效果進而促使其選擇偷竊行為；換言之，機會因素對於住宅竊盜犯罪之發生上從竊賊之動機、決意犯罪、目標出現、情境因素、侵入竊盜、停止或繼續犯罪等階段均扮演重要角色。具體而言，竊賊通常並不願涉險到不熟悉之地區尋找犯罪標的物，故與其居住、工作、休閒有關之生活型態則成為犯罪機會之開端；當犯罪者於生活型態範圍內發現低防衛性之區域或住宅，或其中具有符合高價值、易於移動攜帶、高度暴露、可接觸等具備吸引力之財物與機會時則更提高其犯罪傾向；反之，當竊賊之日常活動區域具有防衛空間意象，或住宅本身結構設計經過強化之設計時，則將因缺乏機會而放棄犯罪動機或轉移至他處犯罪。

相關研究指出，除人口密度、土地利用、租屋率、男女性別比、老人人口率等影響住宅竊盜犯罪外(李珀宗，2005；陳陸毅，2008；彭福康，2015)，額外的強固措施如，保全員、管理員、巡邏警察、監視系統路燈照明等監控機制均為有效降低住宅竊盜犯罪之重要因素(余任晴，2019；楊武德，2006)，凡此皆驗證新機會理論等對於犯罪發生的解釋。

比較巨觀與微觀層次對於住宅竊盜犯罪的分析除構成之因素不同外，對於時間因素，兩者也有相當的差異。巨觀層次的社會解組理論並沒有特別探討時間對與犯罪發生的重要性，而認為犯罪問題是長時間所形成的結構問題，無須特別予以處理。但在微觀的新機會理論等，除犯罪地點外，時間也是相當重要的因素，以日常生活理論來看，有動機犯罪者出現與監控的不在場，表示兩種情況出現的時間一致，再加上正好有合適之標的物，犯罪才會發生，故本研究認為「犯罪基本三角」宜在三角的核心中將時間納入。在上述的實證研究中，許多研究特別針對住宅竊盜犯罪發生的時間進行分析，歸納出住宅竊盜易於家中缺乏監控時間發

生，如每周上班日及白天工作時段等，也說明對於犯罪發生之分析，不可忽略時間之重要性。

二、住宅竊盜犯罪區位因素分析

回顧上述理論、觀點與相關實證研究可知，易發生住宅竊盜犯罪之地點與社會結構因素、周遭環境特徵、受監控與防衛程度，甚至時間等因素息息相關。以下擬針對上述巨觀與微觀分析及相關實證研究發現，抽出重要概念，以做為本研究區位分析變項選取之指引。

(一) 都市化因子—人口與經濟因素

世界各地不論貧富，人口由鄉村向都市移動的現象普遍存在，目前世界各國對於衡量都市規模通常以人口多寡為重要指標之一，而都市人口比率與密度則可做為檢視都市人口集中程度的指標。根據社會解組理論之主張，都市化導致人口激增、商業發達及社會組織的複雜性促成整個社會的變遷、解組，從而增加許多犯罪誘因，歸結而言，人口數量與密度的增加，使得犯罪問題愈趨嚴重。新機會理論則認為人口密度高造成受害者與犯罪者的空間距離愈近，犯罪可能性愈高(Rice, 2003)。在前述實證研究中，Awal(2017)等人認為隨著人口增加，包含住宅竊盜的各類型犯罪都會隨之增多；而陳陸毅(2008)針對臺中市進行住宅竊盜空間分析發現，市中心地區之住宅竊盜密度值高於其他地區；邱靖方(2008)運用犯罪被害經驗調查研究，發現人口密集、都市化程度高地區，住宅竊盜被害情形較嚴重。前述研究均指出人口密集處之住宅被竊風險較高，然亦有相關研究有不同之研究發現。陳惠貞(2009)針對高雄市住宅竊盜犯罪地圖繪製與時空分析之研究中，得出人口密度與住宅竊盜犯罪率呈現顯著負相關之結果；邱靖方(2009)則指出，各研究產生不同結果係因研究之地理尺度不同所致，以全國性或巨觀層次之調查研究多顯示人口密度高之區域與住宅竊盜事件發生呈現正相關，而在同一城市或區域中，位於人潮稀少之偏僻角落則相對成為被害高風險區域。

除人口密度外，人口流動也都市化的一項特徵，造成原因包括鄉村人口向都市流動，或都市中心因治安惡化導致人口外流。社會解組理論認為，高人口流動率使居民無法凝聚社區意識，從而缺乏集體效能與非正式監控，導致犯罪之發生；另從新機會理論觀之，高人口流動率使得潛在犯罪者具有高匿名性，致使犯罪者

可輕易進入或隱匿於該區域實施犯罪。楊武德(2006)進行住宅竊盜被害家戶訪談發現，社區或住宅位處出入複雜流動率高之處易遭竊；李珀宗(2005)針對臺北市松山區住宅竊盜進行研究後發現，異動頻繁的住宅區域，住戶品質參差不齊且互動不多易於遭竊；邱靖方(2008)研究發現，人口流動性與人口異質性是解釋區域住宅竊盜被害率的重要因素；同樣地，國外研究亦顯示人口流動率為解釋住宅竊盜之重要變項(Haining & Signoretta, 2002)。

衡量都市化程度除人口因素外，都市化程度與經濟發展也具有密切關聯，原因在於都市就業機會較多，致經濟因素一向是鄉村人口移向都市的主因，但也造成更多貧富不均之問題，呈現都市內不同地區的異質性；故社會解組理論認為，貧窮是造成犯罪的重要原因，而機會理論則認為，富裕地區家庭的高收入和高房價，代表家戶較有能力買得起高價值的產品，也較容易成為犯罪者下手的標的。關於地區的富裕程度和房價，與住宅竊盜被害的關聯，許多研究呈現結果不一，原因可能是缺乏納入家戶特性與防衛能力等中介影響；換言之，低收入者由於無法擁有較完善的防盜裝置，且多居住於犯罪率高的地方，造成貧窮家戶住宅竊盜被害風險較高(Kuo, 2015)。

(二) 社會及物理防衛能力

犯罪機會對潛在犯罪者而言，係指特定時間及地點，犯罪標的物的可見性和可接近性，強調標的物本身的屬性和特質，就住宅竊盜犯罪被害機會而言，主要關注在以家戶單位為基礎的社會及物理防衛能力。在社會防衛能力方面，無人在家狀態之住宅則為竊賊首選之標的；如：楊武德(2005)研究發現，被害家戶之居住人數偏少（戶量）、無人居住之空屋容易被侵入行竊，而無人在家時間長短也與住宅被竊機會大小有關；此外，獨居戶比率高、家戶人口密度低導致住宅監控力低亦為影響住宅竊盜發生之主因(陳惠貞，2009)；另 Martin(2002)研究指出，住戶年齡因素（老化指數）防衛力較低，為住宅竊盜主要解釋因素；歸納來說，住戶社會防衛力降低，不論是以家戶或是地區為測量單位，皆有較高的住宅竊盜被害。

物理防衛能力是指以防盜設備預防犯罪的效力，通常指鐵窗、監視系統、路燈照明、警衛及巡邏箱設置等。有關物理防衛能力與住宅竊盜被害的關係，並無一致性的研究結果；從地區層級來看，有研究顯示中低經濟等級或相對貧窮之住

宅區，由於防衛較弱且潛在犯罪者多居住於內而容易遭竊，而高級住宅多僱用警衛或裝設防竊措施反而較不容易遭竊(Malczewski & Poetz, 2005)；但也有研究得出相反結論，認為高級住宅區反易遭竊，但此項結論所涉及的已是家戶特性之「標的吸引力」，主要在於標的對於潛在犯罪者的吸引力愈高，影響被害風險的效果愈強；由此可知，地區經濟富裕或貧窮似並非影響住宅竊盜的決定因素，住宅防衛能力或為阻卻竊盜犯罪發生更重要之關鍵因子，特別是在建築物及周遭環境的設計上，對於竊盜犯罪的預防可能是更重要的關鍵因素，這也是情境犯罪預防特別強調的重點。

而在防竊設施的配置方面，若干研究指出，無管理員設置、缺乏防竊設備或安全設備若未妥善維護使用之住宅容易遭竊（李珀宗，2005；楊武德，2005；邱靖方，2008），故加裝監視器與增加警力配置等有助於預防竊盜犯罪發生(彭福康，2015)；但也有研究呈現相反的結果，物理防衛力的增加與住宅竊盜呈正相關(Kuo, 2015)，對於物理防衛力愈高，反而住宅竊盜風險愈高的結果，原因可能是先遭竊才加裝防盜設備，防盜設備數量的增加，是為了回應先前遭受被害的方式，通常稱此為「雙向因果」現象，但無論是事前或事後才增加物理防竊設施，此表明增加物理防衛能力確是預防竊盜犯罪的普遍共識。

(三) 居民集體(社區)意識

Sampson 等人的研究指出，居民社會凝聚與信任越高，非正式社會控制越強，則可降低暴力犯罪的發生；另環境設計預防犯罪與破窗理論觀點亦說明社區意識薄弱之地區，容易降低該區域之自然監控並形成缺乏管理與雜亂之環境意象，進而引發犯罪。相關研究指出，在社區組織上少有組成社區管理委員會與社區巡守隊之區域容易遭竊(楊武德，2005)，換言之，社區組織對於住宅竊盜防治具有正面的效果；但國內近期亦有研究發現，原本低收入及人口流動較高導致住宅竊盜犯罪率較高地區，並沒有因加入集體效能變項之而降低犯罪率，此研究結果與 Sampson 等人所提出的集體效能理論及前述研究並不一致(余任晴，2019)；而 Capowich(2003)針對芝加哥住宅竊盜劃分不同區域之研究中也發現，社區意識高之富裕區域可有效降低住宅竊盜率，然而對於貧窮地區之住宅竊盜率卻無顯著影響，是否意味集體效能在貧富不同地區產生增強或抵消作用，有待更進一步分析。而在集體效能的測量方面，通常以非正式社會控制及社會凝聚力與信任作為建構

指標以問卷進行測量外，在利用政府資料庫方面，則有以經常投票之選舉人比例、社區發展協會數量等測量集體效能，代表該區域居民參與公共事務的意願程度，凡此皆值得參考，並加以驗證(余任晴，2019；游士嫻，2010；Weisburd, Groff, & Yang, 2013)。

根據以上歸納，本研究在第三章中擬以巨觀的社會解組理論與微觀的新機會理論為基礎，將兩項理論加以連結，並參考實證研究的變項選取，建構研究概念架構，做為後續研究實施之指引。

第三章 研究設計

研究設計係針對研究問題或議題獲致答案的藍圖、結構與策略，為研究中最關鍵的部分，可說是研究的「戰略部署與規劃」，故須經反覆琢磨修正，以凸顯研究重要性與問題焦點，確保在研究實施過程分析方法之可行性，俾於提出創新成果與建議。基本上，廣義研究設計的內容包含研究流程與研究概念架構兩大部分；研究流程具體來說就是研究實施步驟，包含研究概念架構的建立；在本研究中將研究設計聚焦於研究概念架構上，主要考量在於研究流程通常較具有規範性，如質化研究方法論基礎多採歸納法(induction)，量化研究則多是演繹法(deduction)，研究流程的邏輯進行相對固定。而研究設計既然重點在於「設計」，即強調研究的變化與獨特性，此部分即是研究概念架構的功能所在；具體來說，研究概念架構乃是經由文獻回顧後，對於所欲探討問題進行理論選擇與建構分析過程，其中包含對變項選取及運用何種研究方法測量變項，以達成研究目的，同時此架構也應具備對研究發現的詮釋功能。由於研究者選擇之理論、變項及測量方式等難完全等同，形成研究概念架構風貌各異，也造就了研究的不斷推陳出新。以下就本研究之流程與架構分別說明。

第一節 研究流程與概念架構

一、研究流程

量化研究較強調「理論先於研究」(theory before research)，在研究程序上，多先根據理論邏輯思維而來，理論提供研究者命題與假設的陳述關係，研究者再從關係中，發展出對變項的界定與操作性定義，以進行分析與驗證(王麗雲，2005)。

本研究流程基本遵循以上研究步驟，可概分為5階段：首先，在研究緣起階段，係根據研究主題與背景，說明研究動機，並研擬所欲達到之研究目的。其次，在文獻回顧階段則選定「環境犯罪學」做為研究途徑(research approach)，同時檢視與研究主題相關之犯罪學理論及研究方法，另參考相關實證研究等文獻，做為下階段研究設計基礎。第三階段主要工作在於研究概念架構的成形並確認變項，同時進行相關資料蒐集、整理及提出研究方法與使用工具。第四階段為研究分析，本階段主要將所蒐整之資料透過研究工具以量化方式呈現研究成果，以檢視分析

結果是否達到預期目的，惟在進入下階段之前，有必要針對研究流程的每一步驟再行檢視，以確認研究正確性。最後階段則是針對研究分析結果進行討論，並提出建議，而對於研究未能完善之部分，應說明研究之限制，俾利於後續研究參考改進。綜上，本研究流程如圖(3-1-1)。

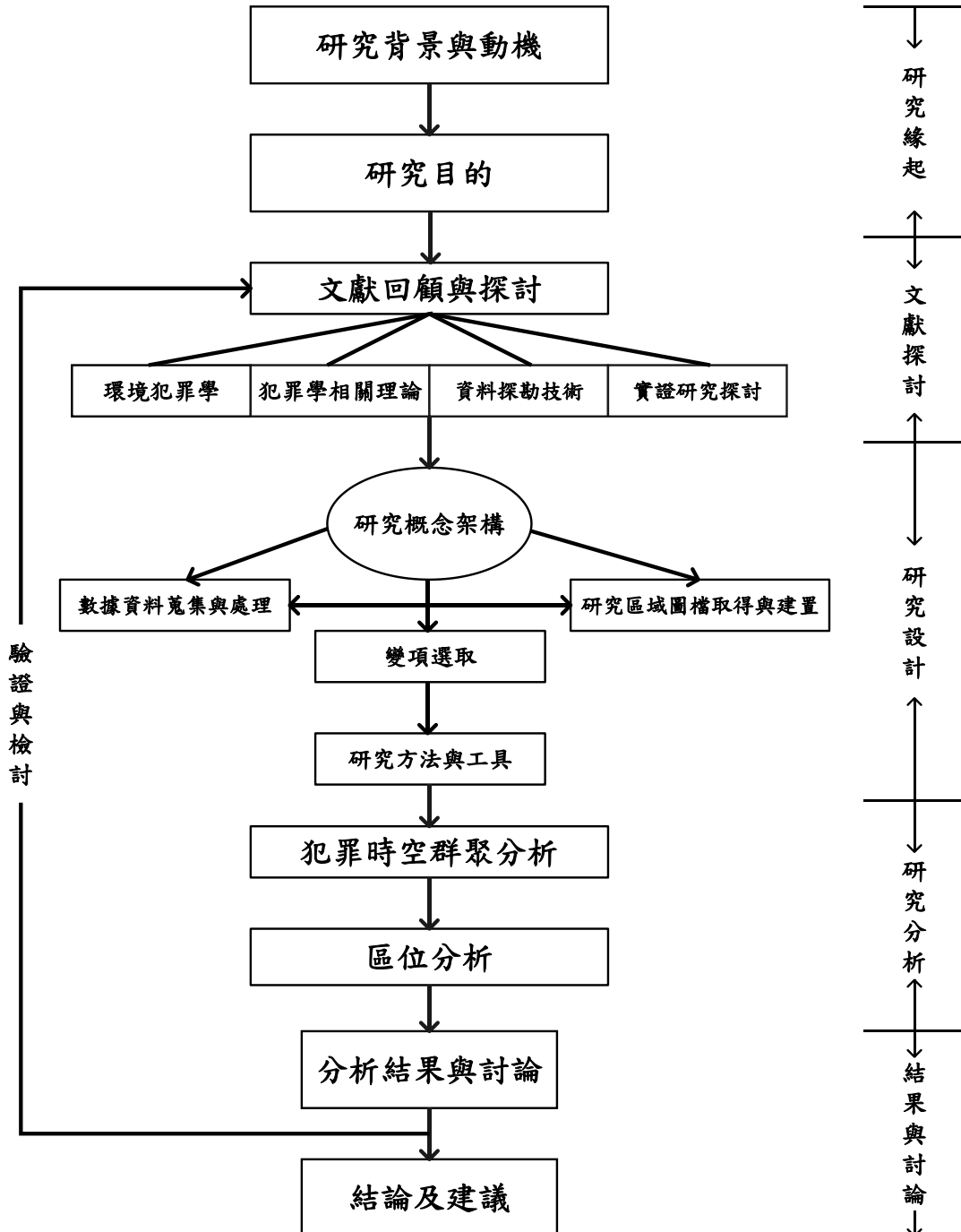


圖 3-1-1 研究流程圖

二、研究概念架構與假設

在本研究流程中，研究概念架構具有承上啟下的作用，發展自研究議題的知識（文獻）回顧，功能在提供研究的理論基礎與重要變項的界定，對研究資料蒐集方向等具有指導的作用，後續的研究方法亦是依據研究目的與資料特性所擬定。具體而言，本研究以社會解組理論與新機會理論做為主要理論基礎，主要在於此兩項理論除針對犯罪發生原因提供解釋外，對於犯罪發生地點也具有重要影響；而根據前述理論整合之探討，在研究實際操作中，新機會理論中的日常活動理論通常假設有足夠數量的具有犯罪動機者，故關鍵變項是機會和監控；而社會解組理論認為地區特性（區位特性）自然造成犯罪動機；據此，本研究將犯罪動機加以控制，探究重點在社會解組與犯罪發生（被害）機會相互關係，且透過集體效能與物理防衛能力（或監控）的中介機制後，對於住宅竊盜犯罪風險之影響。本研究概念架構(圖 3-1-2)與概念間關係如下。

(一) 研究概念架構圖

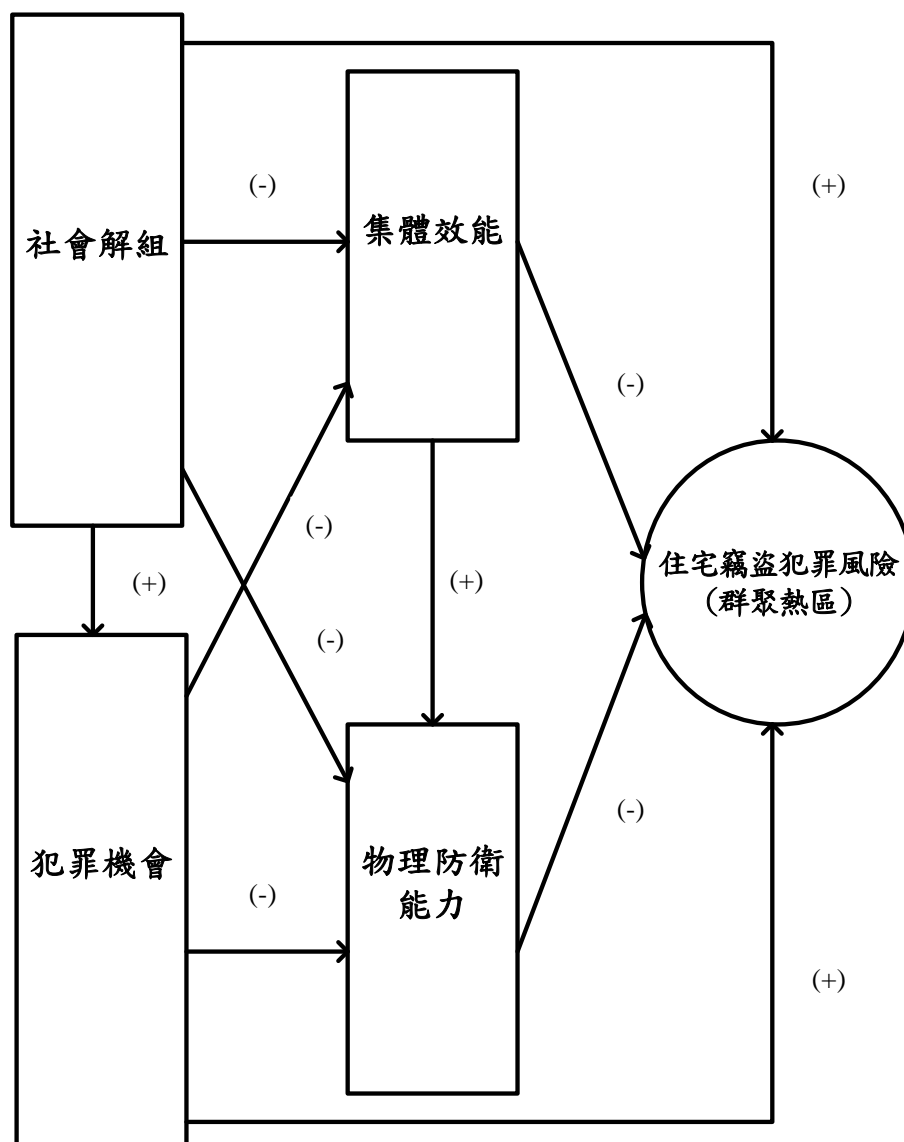


圖 3-1-2 研究概念架構圖

(二) 概念假設

本研究採取資料探勘屬於「探索導向」研究，與運用統計模型進行驗證的「假設導向」研究有所不同，故所列之假設主要在發現理論概念之關聯性，並擬以決策樹的條件式規則結果與統計相關分析進行判讀，而非統計的「路徑分析」(path analysis)驗證因果關係。根據以上研究概念架構圖，各概念間關係如下：

- 1、社會解組與住宅竊盜犯罪風險為正向關係，易形成住宅竊盜熱區。
- 2、犯罪機會與住宅竊盜犯罪風險為正向關係，易形成住宅竊盜熱區。

- 3、社會解組與犯罪機會為正向關係。
- 4、社會解組與集體效能為負向關係。
- 5、社會解組與物理防衛能力為負向關係。
- 6、犯罪機會與物理防衛能力為負向關係。
- 7、犯罪機會與集體效能為負向關係。
- 8、集體效能與住宅竊盜犯罪風險為負向關係，較不易形成住宅竊盜犯罪熱區。
- 9、物理防衛能力與住宅竊盜犯罪風險為負向關係，較不易形成住宅竊盜犯罪熱區。
- 10、集體效能與物理防衛能力為正向關係。
- 11、社會解組透過集體效能後降低住宅竊盜犯罪風險，較不易形成熱區。
- 12、犯罪機會透過物理防衛能力後降低住宅竊盜犯罪風險，較不易形成熱區。
- 13、社會解組透過物理防衛能力後降低住宅竊盜風險，較不易形成熱區。
- 14、犯罪機會透過集體效能後降低住宅竊盜犯罪風險，較不易形成熱區。

第二節 研究範圍與資料蒐整

一、研究範圍

根據研究目的，本研究主要在於分析臺北市住宅竊盜空間、時間及時空的分佈與區位特性，故在地理空間方面之範圍為臺北市，其中包含 12 行政區與 456 個里；由於本研究採用之掃描統計方法須以行政區之中心點做為掃描基礎點，且須有行政區代碼，故較為精確的「最小統計區」或「鄰」之劃分，並不適用於本研究，又若以臺北市 12 行政區為單位進行掃描，因範圍過大，無法有效捕捉各行政區社會經濟結構之確實差異，難以凸顯精確之犯罪群聚位置，故在研究尺度選擇以里做為研究單位，此亦為目前我國行政區域之空間劃分中最小的管轄單位，許多人文社經資料亦有以里為統計單位（如：人口特性統計、警力資源……等）；基此，本研究犯罪空間分析之尺度選擇以里為單位頗為適宜，同時里的資料特性亦與日常生活型態概念較為貼近，更能描繪住宅竊盜對於民眾之影響。

在時間方面，因臺北市政府資料公開平臺有關住宅竊盜地理資料近年才公布，公布時一度曾引起民眾疑慮，認為會對房價造成影響，故臺灣地區大部分縣市均

尚未公布此類型資料；而臺北市政府公布之住宅竊盜資料，首批資料時間為 2015 年 1 月至 2017 年 12 月，本研究即以此批資料為研究對象與時間範圍。

二、資料蒐集與處理

本研究在資料蒐集方面可分成 2 大部分。一為研究區域圖檔取得與建置；二是數據資料蒐集與處理；分述如下：

(一) 研究區域圖檔取得與基本資料建置

近年來政府為推動 GIS 研究，陸續開放各類地理圖資提供免費下載，本研究使用之「臺北市里界圖」取自臺北市政府所建置的「臺北市資料大平臺」資料庫，資料格式為向量圖形 shapefile 格式，此格式在 GIS 領域被廣泛使用多年，適用於大多數 GIS 軟體，通常以 X 與 Y 的方式來處理地理坐標，一般 X 對應經度，Y 對應緯度，但仍需進行座標轉換，目前在我國臺灣地區使用之座標為「TWD97 二度分帶坐標」(taiwan datum 1997)。shapefile 包含許多附屬檔案，均有其特殊功能，其中主要有 3 個必須的檔案，方可完整讀取圖資：

- 1、shp：圖形格式，用於儲存地圖元素的幾何資料。
- 2、shx：圖形索引格式，即幾何資料索引，記錄每一個幾何資料 shp 檔案之中的位置，能夠加快向前或向後搜尋幾何資料的效率。
- 3、dbf：屬性資料格式，以 dBase 的資料表格式儲存每個幾何形狀的屬性資料。臺北市的 shapefile 檔案中原始屬性資料欄位記載地圖圖徵的資訊，包括各里面積、市代碼、區代碼、里代碼、市名、行政區、里名、邊界座標。屬性資料對於若干強調視覺化的主題圖用途不大，但對於本研究而言，區里代碼、邊界座標，為進行掃描統計時計算里中心點座標與確認群聚區域時必須之欄位，而面積則是用於計算人口、監視器、路燈等密度時必備的分母數據，而進行區位分析時，則可將蒐整之資料合併至屬性表格中運用。

除 shapefile 圖檔外，住宅竊盜分布型態為本研究最基礎之資料，由臺北市政府所提供之住宅竊盜點位資料為「逗號分隔值」(comma-separated values, CSV) 格式的文字檔，此資料因審酌不違反《個人資料保護法》、《政府資訊公開法》等

相關規定，故刪去足以辨識民眾個資，僅提供案類(type)、發生(現)日期(date)、發生時段(time)、發生地點(location)資料欄位，位址也以區段標示方式進行模糊處理（表 3-3-1）。

表 3-2-1 臺北市政府住宅竊盜犯罪資料示意表

編號	案類	發生日期	發生時段	發生地點
1	住宅竊盜	1040101	07~09(時)	臺北市中山區新生北路一段 91~120 號

為能順利以研究軟體讀取資料，本研究對於住宅竊盜資料之處理方式如下：

- 1、有關犯罪發生時段的分析，對於犯罪偵防具有相當意義，Felson 和 Poulsen(2003)曾以統計學上的「4 分位法」，將每日犯罪發生的時間共 1,440 分鐘依發生先後排序，來分析犯罪時間的趨勢與變化，但 Felson 和 Poulsen 強調適合以 4 分位法分析之犯罪類型應符合報案時間與實際發生時間相近者；然而，實際上住宅竊盜犯罪發現與發生時段經常有所落差，並不適合採用此方法分析，以避免誤差過多產生偏誤之推論，故本研究以此欄位進行分析之結果恐較不精確，但仍具部分參考價值。
- 2、對於住宅竊盜發生地點模糊處理部分，本研究將區段門牌號碼前後範圍相加後取其中間值，以表 3-1-1 為例，將此門號定為 105 號，後續將各年度住宅竊盜犯罪地點資料予以地理編碼，使文字位元地址轉變為空間上具有座標的點位資料，而對於無法定位之地址（如中間值之門號恰巧位於十字路口等），則先予以刪除。
- 3、由於時空掃描統計在進行時間與時空分析時，必須同時輸入時間欄位，因此對案件發生日期，也須整理為合於規範之格式才能進行分析，如 1040101 須改為 2015/01/01。

以上臺北市住宅竊盜犯罪資料 CSV 檔案原有 1,832 筆，透過內政部地理資訊圖資雲服務平臺（簡稱 TGOS 平臺）進行批次門牌比對進行後為 1,759 筆，訂位率為 96%，符合 Ratcliffe 指出犯罪案件的定位率 85% 標準，故本研究空間地圖建置應能正確反映實際犯罪分布的空間模式，同時檢查時間欄位，也確認並無遺

漏。

(二) 數據資料蒐集與處理

以上研究區域圖資與基本資料建置主要用途是對臺北市住宅竊盜進行時空掃描的群聚分析的基礎工作，在確認群聚區域後，本研究即可根據群聚型態(冷、熱區)不同，進行區位分析。本研究在區位分析的數據資料蒐整主要方式為：第一，使用公開資料平臺的統計數據，直接以數學運算得出數值；例如，有關臺北市各里人口密度即是將各里人口數除以該里面積等；第二，使用 GIS 地理運算功能，例如：將臺北市路燈、監視器等點資料與臺北市里界圖套疊，計算出各里路燈數量等，再將其與面積相除，即可得出相關變項之密度；第三，若干變項資料，如社區發展協會名冊，並非數值及點位資料，所以仍必須先行轉換為地理座標後，再使用前述方式轉換為數值資料。除以上 3 種方式外，其餘如不動產相關資料等，則必須以逐筆查詢方式進行。

然而，無論是進行傳統的統計分析或資料探勘時經常會面對資料前置處理問題，在實務是指對於資料缺失(missing data)、雜訊(noise)、離群值(outliers)如何進行處理。在資料探勘領域中，資料前置處理泛指的是在分析演算法之前，先對資料進行處理，讓資料在格式上比較標準一致，降低數值差異過大的問題，為的是讓演算法不會因為資料產生的瑕疵而誤判，但不同的演算法對於資料前置處理的要求不一，例如支持向量機(SVM)對於資料的雜訊相當敏感，如進行標準化程序後，通常對於模型收斂(精確與穩定)具有幫助⁹，但大多數的資料探勘及機器學習並不特別強調資料標準化。

除資料標準化問題外，在資料蒐集過程中，資料可能會因某些原因造成遺失，此稱為遺漏值(missing data)。面對遺漏值問題，可以選擇捨棄此筆資料，抑或採用眾數、中位數或平均數來取代；在本研究中，由於全臺北市僅有 456 個里樣本，進行群聚分析後樣本必然更加減少，因此，對於缺失資料基本上均採取插補方式，若是連續變數以平均值替代，離散變數則用眾數或中位數替代；而前述的住宅竊盜點位資料方面，由於非屬於統計數值，無法進行插補，故採取捨棄方式。

⁹ 模擬中的不收斂性問題，是指即使執行程式也無法計算或者中途發生停止運算等問題，導致即使微小的條件變更也能引發結果很大的變化，出現不應該出現的結果等。簡而言之，就是分析錯誤和結果不穩定。

對於數據離群值處理方面，部分研究主張將離群值直接刪除，本研究認為離群值如非錯誤值，即是真實情況的表現，之所以在分析資料中呈現異常，可能係因數據量不足或特殊物理環境所造成，故對於資料出現異常，必須加以重視和分析，例如某一地點（同一座標）住宅竊盜突然在相同時段內升高，可能是由於大樓內遭到連續偷竊所致；然而，在使用資料探勘進行區位分析時，則有時必須考慮離群值對於模型敏感性的影響，若屬於對離群值較敏感的機器學習模型，在離群值的處理上就必須審慎考慮是否刪除。

第三節 變項選取與分析策略

區位分析旨在探討環境因素與犯罪分布型態之關聯性，實際操作上係以探討環境中的人文社經及物理變項等如何導致犯罪群聚的產生。本研究參考相關文獻並根據研究概念架構分別以社會解組理論與新機會理論選取變項，隨後進行區位分析。惟檢視過去若干使用 GIS 與統計方法結合之研究，在探討犯罪地理群聚情況後，於區位分析時並未考慮犯罪空間分布的特性，直接以研究區域的整體犯罪率做為依變項，將人文社經變項等做為自變項以進行統計分析，並未考量群聚內外之空間異質性，基本上這僅是將不同的研究合併，尤其是區位分析部分，與未使用地理資訊系統的研究並無不同，至多是在最後詮釋研究結果時，將空間分析做為輔助說明。相對地，部分研究則注意到空間特性問題，故從犯罪群聚內外或將群聚分為冷熱區加以比較兩者區位因素差異(邱奕堯, 2013; 張淑貞, 2008)；本研究在分析策略上採後者方式，將群聚分為冷熱區以進行區位分析。

一、變項選取

(一) 社會解組理論變項

社會解組理論強調犯罪在區域間的差異，並解釋各區域間犯罪率差異的原因，最早提出社會解組的 Shaw and Mackay(1942)指出，社經地位、人口流動率與族群異質性等區位特性與犯罪有強烈相關性；因此，此 3 項指標一般被視為測量社會解組的傳統指標，幾乎所有相關研究莫不以此為測量出發點，而 Sampson 則從集體效能概念(Sampson et al., 1997)出發，提出降低犯罪的要素，進一步對社會解組理論加以補充。基於以上，本研究考量我國社會組成特性與資料蒐集可能性，選取社會解組變項如下：

1、社經地位變項

評估社經地位的變項，常以家庭收入及教育水準等為代表，但如何以具體指標衡量，則有許多不同方式。在家庭收入方面，個人及家庭的所得包括工作收入、利息、股利、投資分紅、政府的社福津貼、私人的現金饋贈，以及買賣股票、基金、不動產的差價收入等都可做為測量經濟收入的變項，但這些變項過於瑣碎，且許多資料難以取得，故國際間多以家庭可以自由運用或支配的「可支配所得」(dsposable income)作為分析比較之基礎，其計算方式包括受僱人員報酬、產業主所得、財產所得、自用住宅設算租金收入、移轉收入等收入合計，再扣除利息支出及經常移轉支出，為家庭可以自由支配於消費或儲蓄之金額，但此項資料由於統計範圍的限制，不一定適用於各類研究；例如，臺北市的統計資料主要以行政區與家戶平均為單位，並無村里層級的統計數據，因此權變方式多以財政部財政資訊中心公布的「年度綜合所得稅申報初步核定統計專冊」的綜合所得項目做為衡量指標，亦不失為客觀的數據；但值得注意的是，有關綜合所得部分有許多的統計方法，其中包括所得平均數、中位數及變異數等，若干研究使用此資料時，選用平均數做為變項(邱奕堯，2013；劉擇昌，2013)，但以此變項是某真的能衡量地區所得狀況，實有值得商榷地方，主要在於收入的平均數極容易受到極端值影響，亦即地區中若有貧富差距極大情形下，平均數並無法反映多數人的所得狀況，所以目前許多研究遂改以所得中位數做為地區經濟狀況衡量標準。而判斷貧富差距的大小，可利用所得的變異係數為代表，亦即某鄉鎮市區的家庭所得分布很分散，變異係數的數值較高，可顯示貧窮差距較大情形。本研究採取家戶綜合所得中位數與變異數為衡量經濟狀況指標。

教育程度不僅是衡量個人社會地位的重要指標，也與經濟所得有密切關聯，依據我國定義，高等教育程度為大專以上之教育(內政部統計處，2018)，許多研究顯示，教育程度與收入有正相關(黃映翎，2008)；然而，正因為此特性，教育程度同時可做為社會解組與日常活動理論變項，原因在於高學歷者所得收入通常較高，住宅有價值之標的物也相對較多(周德芳，2016)；相反地，學歷較低者不但收入較低，許多研究也指出可能有更高的失業及犯罪率，故在本研究中將大專以下教育程度(簡稱「低教育程度率」)列為社會解組變項。

2、人口流動與族群異質性變項

人口流動反映社區人口結構穩定性，常以租屋人口數量、居民遷徙率等為指標，故又稱為人口遷徙。根據我國內政部定義，國內人口遷徙(internal migration)係指國人在本國鄉鎮市區戶籍管轄區域間之居住地變更或由某地區遷入另一地區，統計項目涉及遷出人口數(率)、遷入人口數(率)及住址變更之計算，而人口總移動率(total movement rate)，則是指1年內人口總移動人次對年中人口數的比率，亦即遷入人口數加上遷出人口數，再加上鄉鎮市區內住址變更人口數之兩倍的人口總數對年中人口總數之比率(內政部統計處，2017)。本研究根據 Shaw 與 McKay(1942)對人口流動性的測量及參考國內外研究以人口的社會增加及減少計算人口流動性，以人口「社會增加率」也稱「淨遷移率」，做為測量人口流動變項，主要計算人口遷入率與遷出率的差異，人口流動性愈高，當地居民不易形成凝聚力，使得社會控制力減弱，而易產生較高犯罪率。

國外研究對於族群異質性問題主要為種族所造成的地區差異，如黑人與白人是否因社經地位差異，導致犯罪率不同，若干研究也有改以移民人數為測量標準(Smith et al., 2006)；我國近年來因外籍配偶逐漸增多，臺灣新住民成為新興議題，但考量新住民人數在我國仍屬相對少數、臺北市亦無針對小範圍各里新住民數據統計，且顧及可能涉及族群偏見的研究倫理情況下，不列入對新住民變項的測量。此外，衡量社會解組程度常用之變項尚有離婚率及失業率，其中有關臺北市各里失業率在獲取上有實際困難，故僅將離婚率此一變項列入。

3、集體效能變項

前述文獻回顧中提到，1979年 Sampson 等人對芝加哥地區的343個鄰里進行實證研究中，以居民對於孩童偏差行為及鄰近消防局經費若被刪減的介入意願，測量，另對居民間互助及信任等的認同程度，測量社會凝聚力和信任概念，發現居民社會凝聚與信任越高，非正式社會控制越強，暴力犯罪的發生越低，由於非正式社會控制及社會凝聚力等兩種概念具有高度相關性，表示兩者在測量概念上相近，故將其稱為「集體效能」；Sampson 等人認為，因為集體效能的中介效果，居民可以透過增強集體效能，降低鄰里的犯罪問題，對於難以改變的社會解組狀態具有調節作用(Sampson et al., 1997)。隨後，有許多研究即針對集體效能此概念進行驗證，即便在社會結構迥異的國家，社會解組理論亦有相當程度解釋力；例

如，Zhang 等人(2007)從大陸天津市抽取 50 個「居民委員會」(neighborhood committee)共 2,474 名居民，以家庭變項(household variables)、鄰里結構因素(neighbourhood structural factors)和鄰里社會控制過程(neighbourhood social control processes)對住宅竊盜影響進行調查，研究發現集體效能中的社會凝聚力和信任概念對降低住宅竊盜有顯著關係，但「居民委員會」的半正式社會控制(semi-public control)¹⁰，並未對降低犯罪有顯著影響，Zhang 等人認為「居民委員會」在大陸社會中與西方的非正式社會控制性質上仍有所不同，故對於西方犯罪學理論的驗證，必須考慮社會制度與結構不同所造成的差異(Zhang, Messner, & Liu, 2007)。

臺灣與大陸地區雖實施不同政治制度，但在政府組織的架構及功能有許多類似之處，本研究以「里」為分析尺度，相當於大陸 Zhang 等人的「居民委員會」，具有半正式社會控制性質，但在服務功能與居民參與里活動時的自願性可能高於大陸，而游士嫻(2010)以臺灣地區各縣市每萬人社區發展協會數量，測量集體效能，數值愈高，表示該地居民「參與公共事務的意願」高，故本研究採為變項之一。此外，臺北市各里除設立社區發展協會外，尚有為民眾辦理活動所設置的里(區)民活動中心，此類地點設置即具有社區中心的機能，具有多功能用途，平時可做為里長的辦公處或社區居民聯絡感情的地方，對於強化民眾集體效能可能具有正向影響，亦將其列為集體效能變項。

以參與公共事務的意願高低衡量集體效能，在國外研究尚有以經常投票之選舉人(active voter)比例，做為集體效能之測量，並假設比例愈高，表示參與公共事務的意願越高(Weisburd, Groff, & Yang, 2013)。而現行臺灣地區地方公職人員選舉自 2014 年起合併舉行(俗稱「九合一」選舉)，主要目的在於降低社會成本，選舉規模類近於美國期中選舉，但由於國情不同與選舉合併等因素影響，臺灣里長選舉投票率明顯偏高，是否能反映集體效能的程度值得商榷；惟本研究係以資料探勘挖掘潛在影響住宅竊盜之犯罪因素，故對於可能的變項在資料可得情況下，皆先予以納入進行分析，並以各里投票率高低加以衡量居民參與公共事務的意願程度。

¹⁰ 大陸的「居民委員會」出現於 50 年代，為大陸城鎮地區最基層的組織，地位相當於我國的「里」，居民委員會每屆任期 5 年，其經費由政府出資。

(二) 新機會理論變項

在 Felson 和 Clarke 的新機會理論中，日常活動理論強調社會結構的改變而產生犯罪機會，犯罪型態理論強調社區的差異和變化而影響犯罪的機會，理性選擇理論強調個人所處的場所的犯罪機會不同，也會影響其決定犯罪與否，3 種不同層次的觀點均強調機會對於犯罪的重要性；而更早在 1981 年 Cohen 等人(1981) 使用美國犯罪被害調查資料，針對扒竊、傷害罪和住宅竊盜 3 種犯罪類型，將犯罪被害的風險因素歸類為機會暴露(exposure)、犯罪親近性(proximity)、標的物吸引力(target attractiveness)、防衛能力(guardianship)以及犯罪特性(definitional properties of specific crimes)5 項，探討掠奪性犯罪受害風險之間的關係，提出犯罪被害的機會理論(Cohen, Kluegel, & Land, 1981)。基本上，無論是新機會理論或犯罪被害機會理論，除強調犯罪機會外，也提出防衛能力的重要性，而故本研究以此兩大面向提出相關變項。

1、犯罪(被害)機會變項

(1)人口密度

對犯罪地區的人口測量幾乎已成為必要的變項，若干研究認為人口密度較高，犯罪者的隱匿性亦高(Capowich, 2003; Miethe & McDowall, 1993)，故在日常活動理論中有時被視為「有動機的犯罪者」變項；而在 Cohen 等人研究中，則將人口密度歸類在「犯罪親近性」的測量當中，亦即戶數愈多、人口密度愈高，表示彼此距離愈近，遭住宅竊盜被害的可能性也愈高(周德芳，2016)，但亦有研究認為，人口密度越高連帶使得監視能力也同時增強，反而抑制犯罪發生(Andresen, 2016)。然而，在本研究中由於人口數量之資料係以戶籍人口數量為基礎，因此無法將流動人口及活動消費人口一併列入計算。

(2)單獨生活戶數比率

單獨生活戶數比率對於潛在犯罪者而言，代表犯罪機會的成功機率，比例越高犯罪者成高機會越大，Cohen 等人將此歸類於「機會暴露」，強調犯罪標的物的可見性和可接近性。然而，從社會解組理論來看，地區的獨居戶數比可視為地區家庭關係整合程度的反向指標，當單獨生活戶數比越高，代表該地區家庭整合

程度越低，愈有可能產生社會解組，故此變項亦為兩項理論整合之例證。

(3)老化指數

根據 Laub 和 Sampson(2003)的研究指出，中止犯罪可說是一個規則(norm)，即使嚴重、持續性的犯罪者也會停止犯罪，正常人口的犯罪年齡曲線也適用於犯罪人口(許春金，2010)，年齡越高犯罪率越低，相對地，老人對於抑制犯罪力可能也相對降低；因此，老化指數可做為研判被害與加害犯罪趨勢之變項。

(4)老人戶數比率

老人戶數比率，係指僅有 65 歲老人居住之住宅；年老人由於退休與活動力減低關係，在家時間較長，住宅竊盜犯罪應較少有入侵機會，但老人亦可能因為健康等問題，對於住宅入侵警覺力降低，反有較高受害率。

(5)戶量

戶量指每戶平均人口數，通常認為戶內人口數越多，無人在家機率較低，可降低住宅竊盜之風險(游士嫻，2010)。

(6)低度用電戶住宅比率

我國內政部營建署自為配合行政院「整體住宅政策實施方案」中「健全住宅市場目標」內之「加強住宅及不動產相關資訊的建置及發布」執行策略，自 2008 年起利用房屋稅籍資料、臺電用電資料與地籍資料產製低度使用(用電)住宅，將每年 11、12 月平均用電度數低於 60 度的住宅，界定為低度使用(用電)住宅，作為「空屋」的替代統計方式(內政部不動產資訊平臺，2019)；空屋率越高代表無人在家機率較高，相對增加住宅竊盜犯罪者入侵之機會。

(7)房地產價格

房地產價格與住宅竊盜被害的關係，過去的研究結果不一，若干研究顯示房地產價格較高地區與被害的風險成正相關，似符合日常活動理論標的吸引力之假設(Bernasco & Nieuwebeerta, 2005)，但也有些研究顯示相反的結果，有學者研判主要在於貧窮地區因經濟因素，導致住宅防禦能力降低，且破窗效應影響，反而

有更高的住宅竊盜率(Kuo, 2014；余任晴，2019)。

本研究認為，以房地產價格高低，所涉因素甚多，包括地價與建築物屋齡、建材、仲介商哄抬等，難有客觀標準，而我國政府為促進不動產交易資訊透明化，希望能大量蒐集成交資訊，並在保障民眾隱私的前提下，提供大眾查詢，降低資訊不對稱情形，於 2011 年推動不動產實價登錄制度，臺北市則於 2012 年 8 月 1 日起施行(臺北市政府地政局，2012)，制度施行後對於房地產價格之衡量有更為客觀標準，而民間業者為有利房地產交易，近年也紛紛推出許多查詢地區房產資訊的平臺，查詢內容相當細緻，本研究即是利用不動產時價登錄查詢平臺，以里為單位，逐一查詢該地區平均每坪住宅價格做為研究變項。

2、物理防衛能力變項

(1)警力配置

員警人數增加是否可以降低犯罪率，一直是學理和實務上的爭論焦點。國內外許多研究指出，提升警力及見警率來有效減少竊盜犯罪之發生，但是仍有不少研究從實證研究中得出員警人力增加與犯罪減少之間並無顯著關連的結論，主要在於警力與犯罪率之間可能產生「雙向因果」(bi-directional causality)的問題，亦即警力增加固然可能影響犯罪率，但犯罪率的變動也可能是造成警力增加的原因(劉孟奇、張其祿、盧敬植，2000)。但無論就嚇阻理論或日常活動理論而言，警力配置對於犯罪均為抑制因素；本研究使用每萬人警民比率做為測量變項。

(2)監視器密度

隨著科技進步，運用監視器進行犯罪偵防已成為世界各國警政機關必備方法，但裝設監視器是否必然對於各類型犯罪均有遏制效果，仍需進一步檢驗；國內曾有實證研究指出，裝設監視錄影系統對機車、汽車竊盜犯罪數減少程度達統計顯著，但對普通及住宅竊盜犯罪數減少程度未達統計顯著(彭福康，2015)；因此，本研究擬再次驗證監視器對於臺北市住宅竊盜之抑制效果。

(3)路燈密度

路燈是現代都市公共建設重要的一環，不僅讓夜間交通、行人安全多一份保

障，更常被視為是維護居家安寧與防範宵小的重要功臣。也因此犯罪頻繁發生的地點增設或加亮路燈，是治安整治的常見手法；但根據我國相關媒體報導，許多歐美的研究都已證實，路燈與犯罪率的關聯性極低，而臺灣地區的研究與前述也相當一致(臺大新聞E論壇，2017)，惟本研究係以犯罪冷熱區作為比較，故仍將路燈密度列入變項，以探討是否有不同結論。

本研究以臺北市住宅竊盜為自變項(目標變項)，以上各變項為依變項(解釋變項)，以2015至2017年各里平均數據加以統計，資料取得均為政府或民間企業網路公開資料，計算方式如表3-3-1。

表 3-3-1 犯罪區位理論自變項表

變項名稱	單位	計算方式	資料來源	
社會 解 組 理 論	收入中位數	千元	綜合所得總額/納稅單位(中位數)	財政部財政資訊中心
	收入標準差	千元	綜合所得總額/納稅單位(標準差)	財政部財政資訊中心
	低教育程度率	‰	(大專以下教育程度之人口數/總人口數)*1,000	臺北市資料大平臺
	社會人口增加率	‰	(遷入人口數-遷出人口數)/總人口數*1,000	臺北市資料大平臺
	粗離婚率	‰	(離婚登記對數/總人口數)*1,000	臺北市資料大平臺
	社區發展協會比率	‰	(福利據點數/總人口數)*1,000	衛生福利部
	區里活動中心比率	‰	(活動中心/總人口數)*1,000	臺北市資料大平臺
	投票率	%	(2014年村里長選舉投票數/選舉人數)*100	中央選舉委員會
人口密度	人/km ²	(總人口數/面積)*1,000	臺北市資料大平臺	

新 機 會 理 論	單獨生活戶數比	%	$(\text{單獨居住戶數}/\text{戶籍總數}) * 100$	臺北市資料大平臺
	老化指數	%	$(65 \text{ 歲以上人口數}/0-14 \text{ 歲人口數}) * 100$	臺北市資料大平臺
	老人戶數比率	%	$(\text{老人居住戶數}/\text{戶籍總數}) * 100$	臺北市資料大平臺
	戶量	人	總人口數/戶籍總數	臺北市資料大平臺
	低度用電戶住宅率	%	$(\text{低度用電戶}/\text{總戶籍數}) * 100$	內政部不動產資訊平臺
	平均住宅價格	萬元/坪	逐筆查詢各里住宅平均交易價格	永慶房屋官網
	警力配置	人	轄區員警員額數/人口總數*10,000	臺北市資料大平臺
	監視器密度	支/km ²	監視器數/總面積	臺北市資料大平臺
	路燈密度	支/km ²	路燈數/總面積	臺北市資料大平臺

依據表 3-3-1，本研究將變項加入研究概念架構中，如圖 3-3-1。

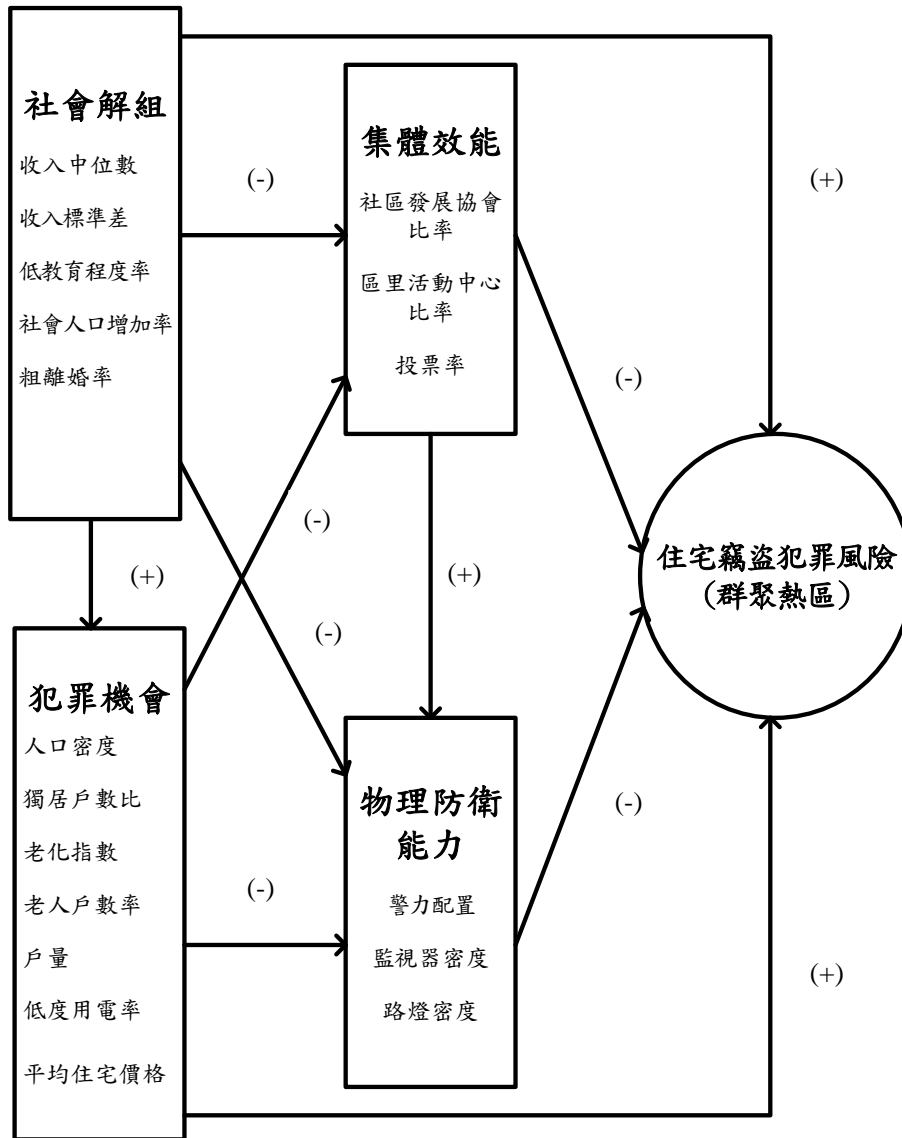


圖 3-3-1 研究概念架構圖(含變項)

二、分析策略

本研究在蒐集與處理資料後，除對於資料進行探索性分析外，主要步驟先以時空掃描統計法對於臺北市住宅竊盜群聚進行空間、時間及時空 3 種不同層次之分析，並將群聚分析結果透過 GIS 予以視覺化。其次，依據掃描結果，住宅竊盜群聚在空間與時空方面可分為冷區與熱區，由於研究資料跨度為 2015 年至 2017 年，進行區位分析時將選取之變項資料加以平均，而平均數值已無法區分時間特徵，故在資料探勘步驟以空間冷熱區做為分組標準，運用不同二元分類的機器學習模型對構成住宅竊盜犯罪群聚冷熱區的因素進行分析。

在此必須說明是，資料探勘主要意義就是從資料訊息中找出脈絡關係；換言之，資料探勘是在「無預設立場」的情況下，找出資料型樣以及資料間關係的一個過程(Hambaba, 1996)；因此，資料探勘與統計分析很大的差異就是在於研究假設與解釋方式不盡然相同。統計分析通常在變項選取後，可根據研究架構與變項建立研究假設，也常會先對分析結果進行一般性預測；但資料探勘則有所不同，如同前述的「無預設立場」，也就是不先建立任何假設，只指定分析演算方法，而往往也無法先行預測最終結果。

本研究將使用5種機器學習模型進行對竊盜犯罪群聚的資料探勘區位分析，由於各模型各有其優缺點，故將對於模型在分類的表現上進行後續評估。在資料探勘分析結束後，最後再以統計方法進行羅吉斯迴歸並與資料探勘比較，以歸納出住宅竊盜冷熱區的重要區位因素。綜合研究流程與概念架構，本研究之分析策略流程如圖 3-3-2。

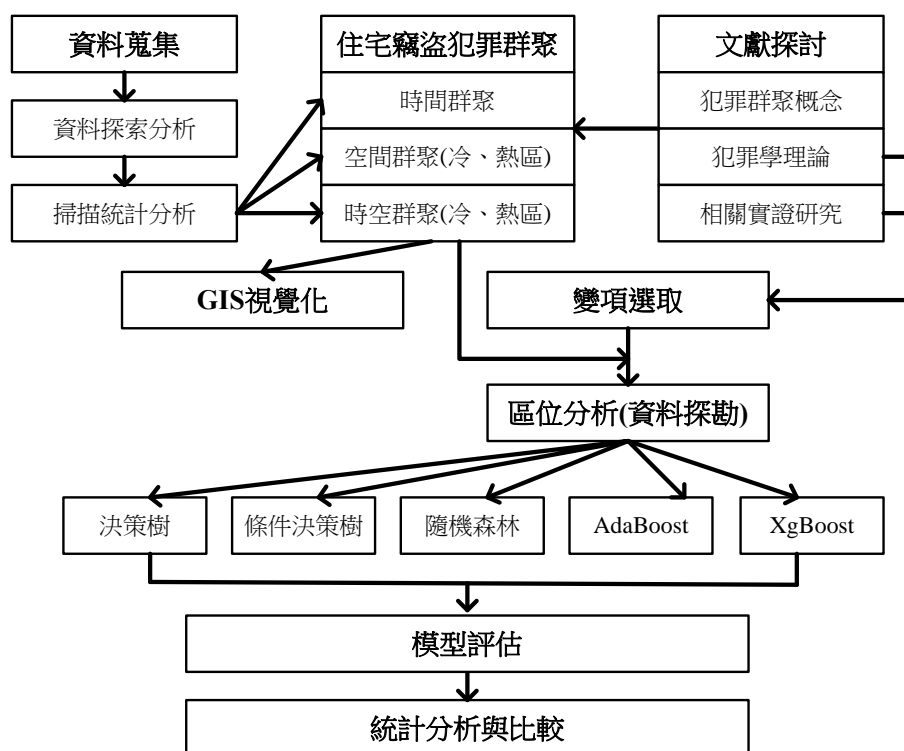


圖 3-3-2 分析策略圖

第四節 研究方法與工具

一、文獻探討法

文獻回顧與探討為研究最為基礎與首要之階段，從文獻中的理論指引到實證研究的分析，對於研究的概念釐清、建立與架構形成皆具有關鍵影響；具體言之，文獻回顧可以幫助研究者深入瞭解研究的背景事實，理論與相關研究的發展狀況，如何選擇適當分析工具進行分析。故文獻回顧可讓研究者瞭解過去、解釋現在及預測未來，無論目的是填補過去研究缺口，或是發掘問題現況並加以解釋，甚至對理論提出挑戰與修正，文獻回顧與探討都是不可或缺的過程。

就本研究而言，即因在文獻回顧過程中，發現過去有關犯罪熱區的研究缺乏將時間同時融入空間的分析方法，故希望能將另一種偵測時空群聚的方法，引用到犯罪學之研究；而進一步審視此類時空群聚方法之相關文獻，似無以資料探勘作為後續分析群聚的區位因素的研究，是以，本研究之發想乃是將為填補過去研究不足，將時空分析結合 GIS 探討住宅竊盜的時空分布群聚情形，並進一步以資料探勘技術探討不同群聚發生的區位特性。

雖然，文獻探討為研究的前期工作，但在研究的進行中，卻必須持續不斷反覆檢視，以考察目前之研究不足之處；此外，也需經常留意是否有更新的相關研究發表，而新的研究對於目前已進行之研究是否有所幫助或因此需調整方向，凡此皆是文獻探討過程中不可忽略的環節。基本上，本研究所涉及的文獻回顧學科領域涉及犯罪學、地理資訊系統、一般統計與時空掃描統計及資料探勘等領域，期能擷取各領域的優勢並加以整合，提出對於住宅竊盜犯罪的研究方法上新思路。

二、時空掃描統計方法

幾十年前國外學者即已針對時空群聚問題進行研究，最早使用時空群聚分析於實證研究的學者為 Knox(1964)對兒童白血病的研究，由於這是一種罕見的疾病，其病因尚不清楚，因此 Knox 假設該疾病中存在傳染因素，並推導出使用疾病發生時間和地點的資料來檢定傳染性的分析方法，亦即假設發病時間和地點為隨機分布，若經由判定拒絕假設則可能為群聚現象，此與目前使用群聚檢定的假設方式相同；發展至今，Knox 時空分析法成為時空分析最廣泛用運用方法之一

(Eckley & Curtin, 2013)。其後，許多學者基於 Knox 之方法，陸續提出改善時空群聚檢定精確度的方法(Baker, 1996; Klauber, 1971; Setsuko Lloyd & Roberts, 1973; Mantel, 1967; Siemiatycki, 1978)，其中 Kulldorff (1998)使用「蒙地卡羅法」(monte carlo method)，透過人口普查區與推估各地區潛在人口 (underlying population) 並結合時間趨勢等，推導出 knox 檢定的無偏誤版本 (unbiased version)，成為現今分析時空群聚接受度很高的新方法(Cheng & Adepeju, 2013; Gao, Guo, Liao, Webb, & Cutter, 2013; Leitner & Helbich, 2011)；其餘較為知名與常用的分析法另有 Baker 的「max test for the knox test」、Jacquez 的「K-NN test」等(Toshiro Tango, 2010)。

Kulldorff(2005)所發展的時空掃描統計分析與 SaTScan 軟體是一種結合不同機率模型與電腦掃描技術，用以檢測事件空間群聚的方法，1997 年發布初版 SaTScan 1.0，在經過二十餘年來模組的擴充與改良之後，2018 年發布最新 SaTScan9.6 版，提供更多樣化的參數設定與分析模式。SaTScan 資料分析型態可分為回顧分析(retrospective analysis)與前瞻分析(prospective analysis)2 種，機率模型可分為離散型掃描統計(discrete scan statistics)與連續型掃描統計(continuous scan statistics)兩大類，其中離散型掃描可使用的機率模型包括「卜瓦松模型」(poisson model)、「伯努利模型」(bernoulli model)、「時空排列模型」(space-time permutation model)、「多項式模型」(multinomial model)、「次序模型」(order model)、「指數模型」(exponential model)及「常態模型」(normal model)7 種，連續型掃描僅「卜瓦松模型」(poisson model)1 種，資料分析型態與機率模型可做不同組合，分別適用於不同的研究目的與資料類型，以協助研究者辨識出群聚的地區及時間，實務上大多用於分析疾病爆發率、死亡率、盛行率、存活率等類型的資料(Kulldorff, Heffernan, Hartman, Assuncao, & Mostashari, 2005)。

(一) 掃描原理

Kulldorff(1999)指出掃描統計的基本要素為掃描的地理區域、區域內隨機事件的機率分配、掃描視窗(Window)的大小與形狀。其掃描程序如下：

1、掃描

SaTScan 的統計方法是透過對時間或空間逐步掃描視窗（代表潛在的群聚

區)，並記錄每個視窗的觀察值和期望值，研究人員可自行定義該視窗之最大的空間範圍和時間區間，直到視窗大小由零達到空間群聚最大值(maximum spatial cluster size)為止，從理論上來說，掃描過程中包含許多位置不同且大小範圍不一之視窗。須注意是，掃描視窗範圍過大，在它覆蓋大部分地理區域之前，可能不再反映掃描內部風險增加的群聚，而是反映掃描視窗外圍風險降低的群聚(Kulldorff, Athas, Feurer, Miller, & Key, 1998)；因此，SaTScan 空間群聚最大預設值限制在視窗面積涵蓋 50% 風險人口(Prates, Kulldorff, & Assunção, 2014)。

2、概似率檢定 (Likelihood Ratio Test)

較早的掃描統計直接將檢驗統計量定義為所有視窗中的最大案例數；然而，在實際操作上，由於無法預知群聚的大小，採用最大案例數存在的問題是大的視窗通常病例數也多，容易造成誤判，故宜採用大小不斷變化的視窗進行掃描；對此，Kulldorff 建構出概似率檢定的方法，此程序乃是估算每個掃描視窗其概似函數值(likelihood function)，藉此比較視窗內外之案發相對風險(relative risk, RR)，出現最大概似函數值的視窗即為第一群聚(most likely cluster)候選，依序為第二群聚等。本研究擬將採用之卜瓦松模型的概似函數的數學式如下所示：(Kulldorff, 2018)

$$\left(\frac{c}{E[c]}\right)^c \left(\frac{C-c}{C-E[c]}\right)^{C-c} I()$$

C：研究區域內的事件總數量；c：視窗內的案件數量觀察值；E [c]：視窗內的案件數量期望值；I ()：指標函數，本研究掃描策略為僅偵測出現群聚的高比率區，則在虛無假說 (H₀：案件為隨機分布) 情況下，當視窗內案件數量觀察值大於期望值時，I () = 1，其它情形則 I () = 0。

3、蒙地卡羅假設檢定(monte carlo hypothesis testing)

蒙地卡羅假設乃是針對掃描到的可能群聚進行顯著性檢定，在虛無假說之條件下 (各地區案件發生的機率一致；若有一地區發生的風險不同，則拒絕虛無假說)，以研究區域內的真實資料為基礎，重複多次模擬產生大量隨機樣本後，進行上述步驟計算隨機樣本的概似函數，再與真實資料計算出之概似函數排列比較進行檢定。例如：經過 999 模擬後，將 999 筆模擬樣本檢定統計量依照數值大小

排列，如真實資料的檢定統計量落在模擬樣本檢定統計量的前 50 筆內，則代表有顯著水準 5% 拒絕虛無假說 ($p < .05$)。最後，再根據「概似率對數」(log likelihood ratio, 簡稱 LLR) 與 P 值來排列風險群聚。

4、相對風險率(relative risk, RR)

計算值為群集內的估計風險除以群集外的估計風險。數學式如下所示 (Kulldorff, 2018)：

$$\frac{c/E[c]}{(C-c)/(E[C]-E[c])} = \frac{c/E[c]}{(C-c)/(C-E[c])}$$

相對風險率可知群聚內的風險率是群聚外風險率的倍數；例如，群聚內的相對風險率數值為 2，表示其風險為群聚外的 2 倍。相對風險率數值大小與群聚區域內的案件數和人口（母體）數有關，在實際的掃描結果中，相對風險率數值大於 1，判定群聚區域為所謂「熱區」，反之則為「冷區」。

(二) 掃描步驟

根據上述掃描原理，SaTScan 掃描統計基本包括 3 個步驟：確定掃描方式，即如何定義掃描視窗的位置和大小；計算統計量，根據輸入資料的類型確定計算統計量的概似函數值，計算每個視窗的統計量，確定最有可能是群聚的視窗（一個或多個）；對群聚視窗的非隨機性假設進行顯著性檢定 (Kulldorff, 2018)。

1、空間掃描

對單純的空間掃描，掃描視窗通常為圓形視窗，如輸入資料是離散資料，則圓形視窗的中心點為輸入資料的各個位置點，視窗的半徑可以根據該中心點與相鄰點的距離來確定，從 0 到某個上限，上限可以是某個距離，也可以視窗內包含的案例數（或對應的人口數等），案例數至少為 2 個；如輸入資料是連續資料，可以利用隨機點或規則網格作為視窗的半徑；相同原理，掃描視窗也可改為橢圓或其它形狀視窗，視研究需求而定。

空間掃描採用二維平面的掃描方式，其原理係利用變化的空間視窗 $[(x,y),r]$ 在研究區域內進行移動，從而探測出群聚存在的空間區域。例如，從以 (x,y) 為中心的區域 ID（行政區代碼等）開始掃描，空間掃描尺度設定為 500 公尺後，開

始進行第一次 $r = 500$ 公尺掃描，第二次 $r = 1,000$ 公尺掃描，第三次 $r = 1,500$ 公尺掃描，以此類推，直到達到事先設置的最大掃描半徑為止；如下圖：

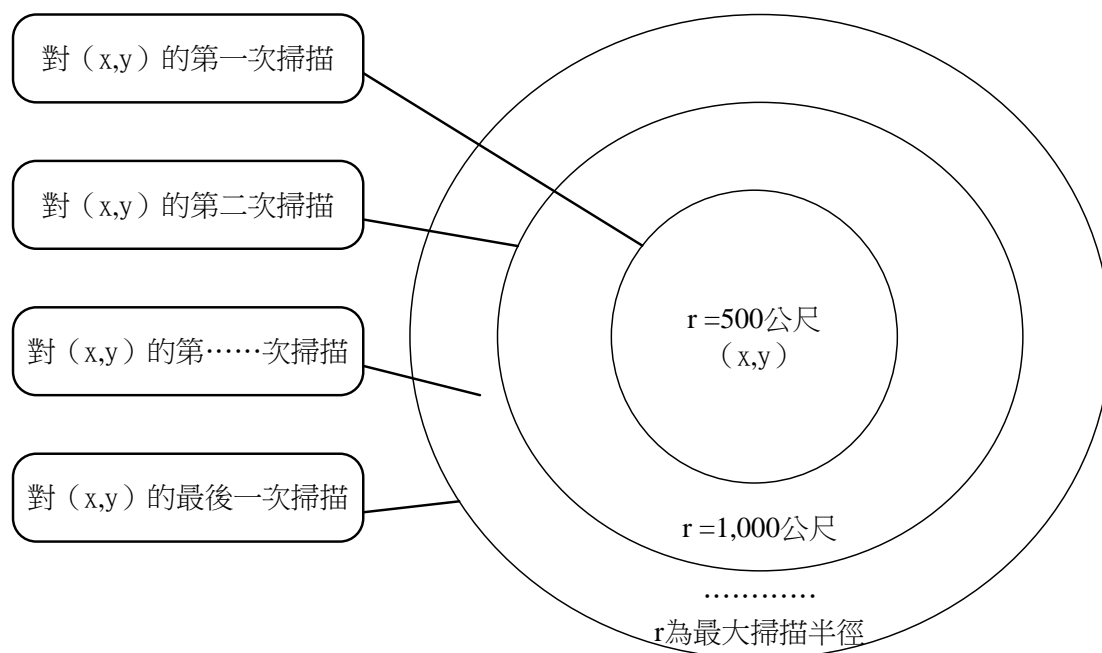


圖 3-4-1 空間掃描視窗變化

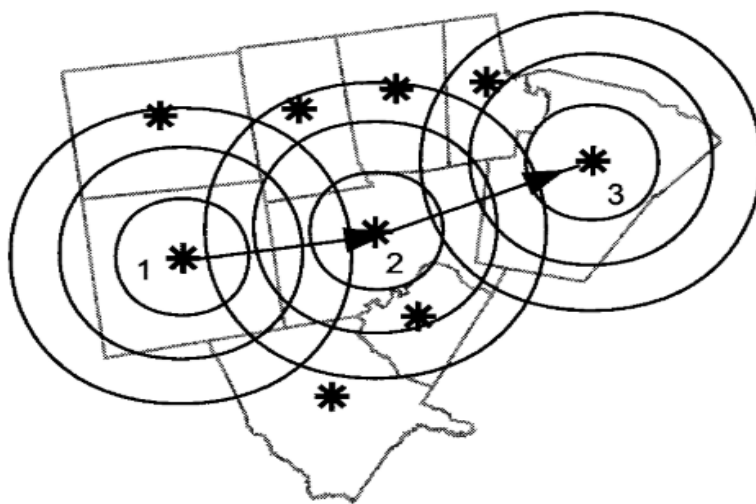


圖 3-4-2 區域中心點的逐步掃描方式

資料來源：Coulston and Riitters(2003)

以上掃描方式關鍵在於 2 個問題：一是空間掃描尺度，二是最大掃描半徑。

其中，空間掃描尺度是指空間掃描視窗中 $[(x,y),r]$ r 的最小值，如圖 3-4-1 空間掃描尺度為 500 公尺；最大掃描半徑是空間掃描視窗中 $[(x,y),r]$ 的最大值，即從某一固定區域點開始掃描時，空間掃描視窗的最大掃描半徑，上述提到以 50% 風險人口為預設值，就是掃描視窗最大半徑，其原理為當視窗以地區中心點進行掃描時，視窗內同時包含全部人口數量，當將風險人口設定為一半時，即是圓形視窗的最大半徑(Kulldorff, 2001)，掃描路徑如圖 3-4-2。

2、時間掃描

在純時間掃描方面，掃描視窗是一個時間間隔。設定一個具有一定時間長度的掃描視窗，該視窗可在時間上移動，視窗的大小和位置均處於動態變化之中，具體來說，時間掃描統計採用一維的掃描統計量概念，利用一個變化的時間視窗 $[t,t+w]$ 在研究時間範圍 $[T_S, T_E]$ 內進行移動，從而偵測出群聚存在的時間段。例如，規定最小掃描尺度是天，從 T_2 時刻點開始掃描，最大掃描時間為 $T_E - T_2 + 1$ 時，第一次掃描 1 天，第二次掃描 2 天，第三次掃描 3 天，以此類推，直到 $t+w=T_E$ 為止。如圖 3-4-3 所示。

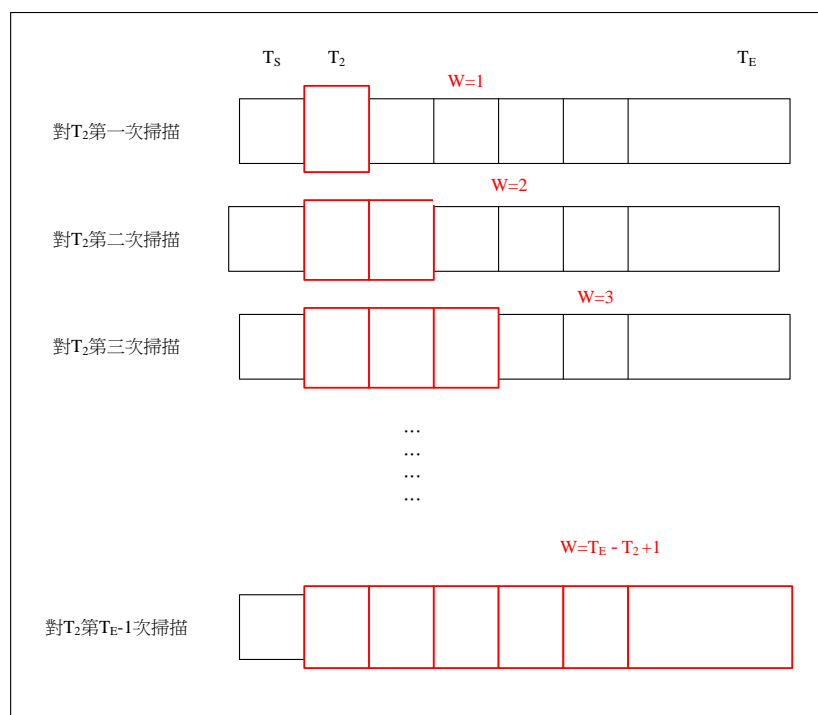


圖 3-4-3 時間掃描方式

與空間掃描相同，對於時間的掃描也必須考慮到時間間隔的設定與最大掃描

時間範圍。其中，時間掃描間隔指的是 $[t, t+w]$ 中， w 的最小值，可設定為天、月、季節及年等，在上例中即是以 1 天為單位；而最大時間掃描長度為 $[t, t+w]$ 中 w 的最大值，即 $T_E - T_2 + 1$ 。

3、時空掃描

結合空間與時間的時空掃描，掃描視窗成為圓柱體或橢圓柱，進行平面與空間移動。首先，隨機在研究區域內選擇一個空間點作為圓柱體掃描視窗底面積的中心；然後，逐漸增加掃描視窗圓柱體的底面半徑及其高度，圓柱體的底面積不斷變化即對應掃描視窗所覆蓋的地理區域變化，圓柱體的高度也不斷變化即對應時間長度的變化，直到達到預先設定的掃描視窗最大的空間和時間上限，圓柱體掃描視窗在研究區範圍內的所有位置都重複相同的掃描過程。具體來說，時空掃描統計係採用三維的掃描統計量，時間和空間構成 X-Y-Z 座標。它利用一個變化的掃描圓柱體 $[(x, y), r, d]$ 在研究區域內進行移動，從而探測出的時空集群存在區域，因為掃描圓柱體暨在時間上變化，又在空間上變化，所以掃描圓柱體的底和高都在不斷的增大。例如，從以 (x, y) 為中心的區域在 T 時開始掃描，首先應該確定時間掃描尺度(1 天)、空間掃描尺度(500 公尺)、以及最大時間長度(2 天)和最大掃描半徑(1,500 公尺)，開始進行多次掃描，以此類推，直到達到事先設置的最大掃描半徑為止。掃描方式如圖 3-4-4，3-4-5 所示：

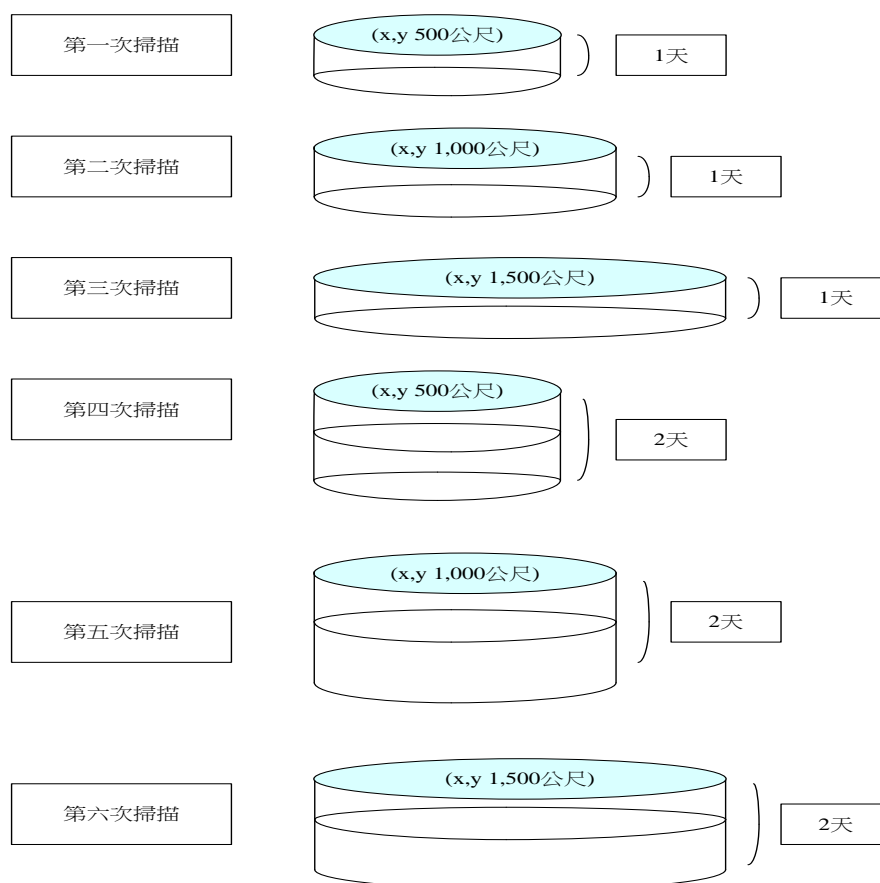


圖 3-4-4 時空掃描方式

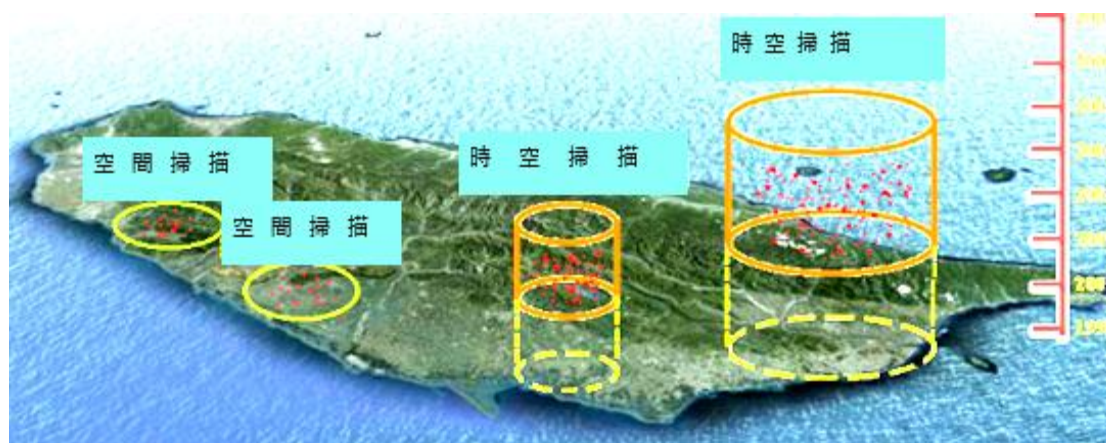


圖 3-4-5 時空掃描示意圖

資料來源：陳薇琳(2012)

(三)參數設定問題

根據 Jin Chen 等人的研究，使用 SaTScan 最大的難題在於掃描範圍的設定，如果使用軟體內建的預設值將風險群體定為 50%，則因掃描範圍過大，通常會將較小的群聚區掩蓋，但如果設定範圍過小，則可能會錯過重要的大型群聚(J. Chen, Roth, Naito, Lengerich, & Maceachren, 2008)，而 DeChello 和 Sheehan 曾建議將掃描範圍定在 25% 或 10% 風險群體進行比較，時間範圍則設為 90%，但設定的原則應視研究對象及目的深入研究(DeChello & Sheehan, 2007)；另有研究專提出各種不同的設定值，期能說明不同設定將導致相異的群聚結果，也將影響對於事件之解釋(Sugumaran, Larson, & Degroote, 2009)。

上述呈現出核心問題為，群聚檢定力受影響的原因在於期望事件發生數的估計，由於期望事件數或期望發生率是由研究區域內的發生事件數及母群體估計，研究區域變動的確對檢定力產生影響，在單一群聚的假設下，選擇較大研究區域通常會有較為顯著的檢定結果，但在實際應用上，研究區域選取太大或太小可能會包含錯誤的訊息或是降低檢定力(蔡丞庭，2011)，但 Kulldorff 掃描統計並非單一群聚，而可能是數十、數百甚或數千個重複群聚。為解決此問題，Kulldorff 等人提出利用基尼係數(gini coefficient)確認基尼係數最佳群聚區，稱之為「基尼群聚」(gini clusters)，而設定在 50% 人口風險掃描範圍，可以在不遺漏各種群聚可能性的情況下，更明確、精簡呈現非重疊群聚(non-overlapping clusters)的研究結果(Junhee Han et al., 2016)。

三、區位分析方法

(一) 區位分析方法概述

區位分析是對所在的區域人文、環境及社會等因素的了解與認知，根據研究的目的不同，有不同層次的分析方式，甚至可以是基於玄學的分析，如古代的「地理師」、「風水師」進行選址，即是區位分析的一種方式。然而，隨著科學的進步，當代的區位分析更講求的是數據上的憑據與精準，涉及的領域也越來越廣，舉凡經濟、社會甚至國防等只要涉及地理空間議題，均可根據不同的目的，進行區位分析，並對區域獨特的資源或優（劣）勢提出意見。

在犯罪學的區位分析方面，主要目的莫過於找出特定區域構成犯罪的成因，

並擬定防治作為；在 GIS 尚未推廣前，常見以統計方式，針對研究區域的社經人文資料進行相關與多元線性迴歸分析(multiple linear regression, MLR)等各項檢定，但此類分析缺乏與空間連結關係，可能會產生較多偏誤，而自 GIS 與犯罪分析結合後，基於對「空間自相關」等空間不穩定(spatial non-stationarity)現象的深入認識，利用空間統計如「空間自相關檢定」(spatial autocorrelation test)、「地理加權迴歸分析」(geographically weighted regression, GWR)等進行區位分析逐漸成為潮流，許多研究結果顯示，空間統計對於犯罪現象似乎有更高的解釋力(彭福康，2015；游柏輝，2011；劉擇昌，2013)。

除以空間統計進行區位分析外，近年採用資料探勘進行區位分析的研究也逐漸增多，包括犯罪學領域(余致廷，2014；林宛宣，2019；榮峻德、孫志鴻、李萬凱，2005；鍾志偉，2010)，一般稱為「空間資料探勘」(spatial data mining)，其演算法與不涉及空間屬性的資料探勘技術大同小異，但更強調由空間面向切入分析，屬於資料探勘的應用領域，主要差別在於分析資料的性質不同；簡言之，空間資料探勘分析的主要對象乃是資料經由 GIS 的各種空間分析模型和空間操作對資料庫中的資料進行再處理後產生的新資訊，例如使用拓撲分析、環域分析、密度分析、距離分析、套疊分析等產生空間資料，但使用空間資料探勘進行區位分析，資料並非侷限於空間屬性，非屬空間屬性資料同樣可以進行綜合分析，故可將「空間資料探勘」描述為專門針對空間或非空間性的資料來找尋有用的空間知識過程(Chawla, Shekhar, Wu, & Ozesmi, 2000)。

目前，常用的「空間資料探勘」大致有 5 種主要技術(Ester, Kriegel, & Sander, 1997；榮峻德等，2005)，分別為：(一)「空間特徵分析」(spatial characterization)：從資料庫中擷取與主題符合的相關資料，進行空間或非空間的描述，類似統計的探索性分析；(二)「空間區別分析」(spatial discriminate analysis)：是指將「目標類」的一般特徵與一個或多個「對比類」的一般特徵做比較，得出用於區分兩者的對比度量。譬如：比較不同地方的坡度和坡向、公用設施的地區差異等空間分析，方便使用者做區別以及加以判讀；(三)「空間分類分析」(spatial classification analysis)：是指對空間物件基於某一特性或空間關係，利用其他的資料按照一定的分類模型進行歸類，屬於監督性分類的一種方式。譬如分析山坡是否會崩塌，空間分類分析就可以利用其他曾經崩塌的相關資料做為基準，進行分類分析，歸納出在符合哪些條件下，可能會有崩塌的危險；(四)空間分群分析(spatial

clustering analysis)：是將空間資料分組，提高組間的異質性，以比較組間差異，譬如：可以利用地震的資料庫對每次地震發生的地點與次數等進行分析，依據資料的特性找出在哪些地區發生地震次數是較高，可能屬於斷層帶，應減少人為開發，以避免有大災情的出現；(五)空間關規則分析(spatial association rules)：指對大量的空間以及非空間資料，找出彼此之間的關連性，並且歸納成規則方式以提供後續進一步分析參考。

各類探勘方法中，與本研究區位分析最具關連性的方式為「空間分群分析」，故將採此分析方式，對臺北市住宅竊盜犯罪的群聚熱區與冷區進行相關因素的區位分析。

(二) 區位謬誤問題與變項篩選

從廣泛 GIS 空間資料分析角度來看，「空間資料探勘」可歸屬於進階的空間資料分析方法，一般 GIS 的資料探索及空間分析方法，在資料探勘過程中可歸屬於資料前處理階段(謝漢欽、汪大雄、鍾智昕，2010)；依此觀點，本研究運用掃描統計對臺北市住宅竊盜進行群聚分析，基本上亦可視為資料探勘的前置階段，而這也要求在後續區位分析時，必須與前面的群聚分析有所連結與一致性。但審視若干以 GIS 進行群聚偵測與區位分析的研究似乎忽略此銜接的環節，而造成所謂的「區位謬誤」(ecological fallacy)；舉例來說，部分研究得出犯罪群聚結果後(局部的研究區域)，在區位分析時卻以全區域進行區位分析，若以本研究為例，研究結果顯示臺北市住宅竊盜犯罪熱區集中在中山區、大同區及萬華區等，在區位分析時卻以臺北市的整體住宅竊盜犯罪率與選取的社經變項進行多項式迴歸(或空間迴歸)等分析，即是陷入「以全概偏」的誤區，此謬誤假設群體中的所有個體都有群體的性質，分析結果將是臺北市住宅竊盜的整體區位因素，而非對特定群聚區域之分析。相對地，某些研究已經注意到從群聚中抽取空間或非空間屬性變項進行分析，並檢定出具統計顯著水準的區位變項，但本研究認為，變項雖是基於理論或參考文獻所選取，但仍具有研究者主觀抉擇的成分，即便是具統計水準的變項，是否就是造成群聚的原因，仍有待更進一步確認。根據以上論點，本研究進行區位分析的方法與流程如下：

1、犯罪群聚分組

本研究為探究何種區位因素造成臺北市住宅竊盜的熱區群聚，首先運用掃描統計將群聚分為「熱區」與「冷區」，由於選取區位變項的資料為3年平均值，沒有時間上的區分，故以空間掃描方式進行分組，以比較兩種不同性質群聚的區位因素差異。

2、區位共變項檢定

掃描統計對群聚偵測時，同時也可對於可能影響群聚的變項進行檢定，即進行群聚偵測時，發現某些地區有明顯住宅竊盜群聚現象，進一步分析發現，這些地區的某些因素（如空屋率、低教育程度率等）高於其他地區，則可假設這些因素（變項）可能導致較高的住宅竊盜發生，從而形成群聚的現象。為驗證犯罪群聚的發生是否與這些因素有關，此時可將這些因素納入共變項分析，以調整誤差來源，若群聚的區域在納入共變項分析後，與未納入前相同，仍呈現統計顯著現象，表示此變項對群聚形成並無影響，非此地區群聚的因素；反之，若納入此變項分析後，原本顯示群聚的區域顯著性消失，則表示此變項與群聚是具有相關性(Kulldorff, 2001；廖興中、徐明莉，2017)。在 SatScan 軟體中，最多可同時納入10個共變項分析，本研究則是每次納入1個共變項，依次檢定以避免變項相互影響。在先前的群聚分析中，本研究並未採用此法對共變項進行檢定，原因在於當時主要旨在於偵測住宅竊盜異常群聚所在位置，不管是什麼區位因素所造成，若當時將共變項納入，則無法對群聚進行有效偵測。以下在逐一檢定各個變項後，若與群聚形成無關，則將予以刪除，不再納入資料探勘的分析模型。

3、重要變項篩選

本研究區位分析之重點，在於建立模型以解釋造成住宅竊盜犯罪群聚發生的可能因素；因此，如何從社會解組與機會理論所選取的諸多解釋變項中進行重要變項的篩選將是影響模型表現的關鍵；篩選變數之優點，除可對模型進行簡化以利於解釋外，也可避免模型過度適配的問題，即在訓練樣本與測試樣本上表現相當優異的模型，但對外部的驗證樣本卻出現較差的預測能力。

對於變項的篩選機制有許多方式，例如，有母數方法中的邏吉斯迴歸可利用向前法(forward selection)等迴歸程序以篩選變項，也有先將所有解釋變數進行ANOVA分析後，再將具有顯著性的變項放入分析模型內(棗厥庸、李永新，2010；

鄧莉雅, 2015)。然而, 本研究擬使用「空間分群分析」的決策樹、條件決策樹、隨機森林及推進法等資料探勘技術屬無母數統計方法, 因無分配上的假設且為非線性模型, 因此篩選變項方法有所不同, 以隨機森林為例, 衡量變項重要性的方法是透過剔除該變項, 並將剔除變項的模型與原模型比較, 差異越大表示變項越重要, 透過 mean decrease gini 指數來衡量變項重要性(Leo Breiman, 2001), 此種篩選變項方法考量到變項間可能存在的交互作用, 故篩選出的變項稱為「相對重要變項」(relative important variables); 同樣地, 推進法中對於重要變項也有篩選機制。

四、資料探勘分析

在區位分析的具體方法方面, 本研究運用資料探勘技術對住宅竊盜犯罪進行區位因素分析, 與若干研究採用統計方法進行分析方式目的相同, 惟資料探勘技術對於資料分布型態、資料性質與變項處理等更具有彈性, 其演算法雖來自於統計概念, 但並不同於統計分析, 且亦無須預先提出假設, 探勘的目的主要在於挖掘各變項之間的關係、分類規則及變項對於研究議題的重要性; 然而, 即便資料探勘與統計方式有所差異, 兩種方式亦可相互配合運用。例如: 運用資料探勘可處理大量變項的優點, 可先行對與研究議題相關或不相關的變項均先予以納入分析, 經分析後可初步篩選出相對重要的變項, 再將其進行統計分析檢定或迴歸, 對於提高統計模型的解釋力將有所助益。

此外, 本研究運用時空掃描統計並結合 GIS, 以辨識住宅竊盜的熱區與冷區, 基本上可將住宅竊盜空間分成兩個不同型態的區位特性, 故在資料探勘的方法運用上適合二元分類的技術分析, 此與統計分析中的二元羅吉斯迴歸(binary logistic regression)原理類似。以下說明本研究使用之資料探勘各項分類技術與模型。

(一) 決策樹

1、原理及優點

決策樹為 Breiman 等人於 1984 年所提出的演算法, 為常用的資料探勘技術, 其方法主要係利用二元分割方式, 將含有許多複雜變項的資料加以歸納(L. Breiman, Friedman, Stone, & Olshen, 1984), 由於其分類的過程與樹木結構相似, 故具備根節點(root node)、葉節點(leaf node)及子節點(child node)3 部分(如圖

3-2-6)；每個內部的節點代表對指定屬性的測試，每一條分支代表測試的結果，葉節點則代表一個類別標籤的值。決策樹的建構關鍵在於每個內部節點的屬性測試，選擇的標準是根據每個屬性對分類所能提供的「資訊增量」(information gain, IG)(Jiawei Han et al., 2011)；演算模式最後藉由產生一連串的「if-then」邏輯條件，讓解釋變項(或稱預測變項, predictors)，能夠預測被解釋變項(或稱目標變項, target variable)，「預測」在此的意涵，就是葉節點不再分割，即代表做出決策之意(make a decision)(何宗武，2016)。

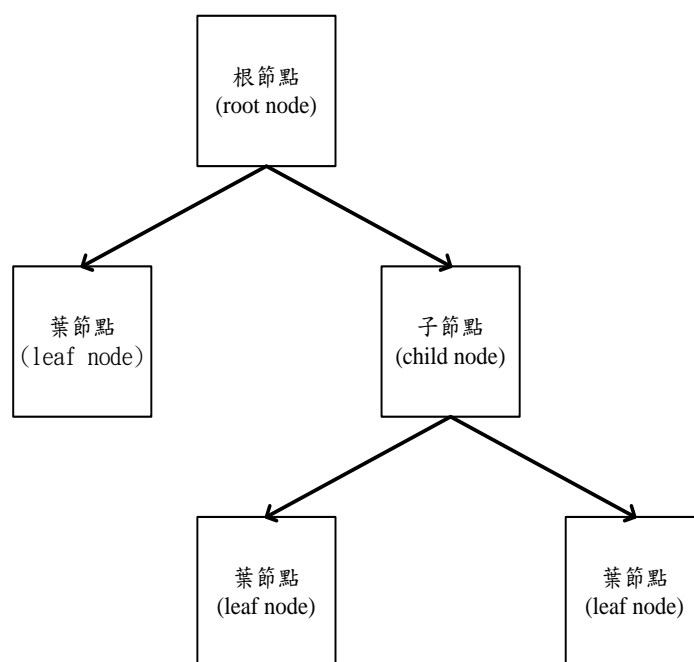


圖 3-4-6 決策樹節點圖

決策樹與使用的線性迴歸模型差異在於，線性迴歸是建立在以方程式為基礎追求最小誤差或最大概似，而決策樹主要演算模式是藉由產生一連串的「if-then」邏輯條件，讓解釋變數能夠預測被解釋變數，且一個解釋變項可在不同的切割階段被重複使用；其主要優點為：

- (1)易於解釋和瞭解；在適當的詮釋後，決策樹所表的的含意和操作意義很容易理解。
- (2)能處理連續性質和類別屬性的資料，其他的技術往往要求數據屬性的一致性；換言之，無須事先將資料轉變為類別資料型態。
- (3)可產生相互獨立的群體，直接和自動提供重要解釋變項的交互作用

情形。

- (4)決策樹在資料分析上不需要特定的資料分配情形(江昇臻、林正昌，2014)。
- (5)樹木形成係利用逐步分類來決定分割的規則，過程中考慮到所有參數的分割可能。
- (6)在統計學中為避免偏誤通常會將資料的離群值予以刪除，但決策樹的演算法並不會受到離群值影響(鄧莉雅，2015)。
- (7)決策樹可以用來建立非線性模型，可被用在迴歸，也可以用在對於遞迴二元變項的分類(recursive binary partitioning)。由於演算法特性，並不會受到「多重共線性」(multicollinearity)的影響¹¹。

然而，決策樹既有優點，當然亦有其缺點，分述如下：

- (1)決策樹關鍵缺點是對於各類樣本數量不一致的數據，資訊增益的結果偏向於具有更多數值的特徵；此外，決策樹對於觀察值數量增減也很敏感，而造成穩定性不夠。
- (2)在訓練資料中很容易生成複雜的樹結構，造成過度擬合(overfitting)¹²。剪枝可以緩解過擬合的負作用，常用方法是限制樹的高度、葉子節點中的最少樣本數量。
- (3)決策樹是基於啟發式(heuristics)演算法所建立，這種演算法不能保證建立全面最優的決策樹，除了過度擬合外，還有變項選擇偏誤的問題。因此，才有後續的條件決策樹(conditional inference tree)、隨機森林(random forest)與推進法決策樹等引入，期望能解決以上所產生

¹¹ 多重共線性 (multicollinearity) 是自變項之間具有很高相互關係的現象，此可能造成統計推論不可靠的情況發生。但是在機器學習中的決策樹、隨機森林等，不會受到共線性的高度多重共線性的影響，在樹的分裂過程中，即使其中兩個變項提供相同的子節點純度，建樹過程只會選擇 1 個節點，故不會降低結果的品質。

多重共線性對於機器學習的實際影響，並不會改變模型的預測能力(至少在訓練資料方面)，但確實會影響的係數估計；但在大多數機器學習應用中，解釋可能不是最重要的目標，模型的預測能力才是必須關注的重點；故對於統計中重視的多重共線性問題，在機器學習領域中討論並不多。

¹² 「過度擬合」是指模型對於範例的過度訓練，導致模型學習到的並非訓練資料的一般特性，反而是訓練資料的局部特性，將造成對測試樣本分類不精確的現象。

的問題。

決策樹全名為分類迴歸樹，其性質如同其名，分析型態可分為兩種類型：分類樹及迴歸樹。分類樹是針對目標變數為離散型的情況，即最後目標是預測樣本的所屬類別；相對地，當目標變數為連續型，則建立迴歸樹(黃文、王正林, 2016)。除分析型態不同外，因演算法的不同，亦有不同的分析方法，目前決策樹的演算法主要大致有 CHAID、CART、C5.0 與 QUEST 演算法；CHAID(chi-square automatic interaction detection)是由 Kass 在 1980 提出，CART(classification and regression tree)是由前述 Breiman 等人 1984 年提出，QUEST(quick、unbiased、efficient statistic tree)是由 Loh 和 Shi 在 1997 提出，而 Quinlan 則分別於 1986、1993 及 1998 年提出 ID3(iterative dichotomiser 3)、C4.5 及改良的 C5.0¹³，3 種演算法(Jiawei Han et al., 2011)。其中，CHAID 及 ID3 只能處理類別變項，其餘各種均可處理連續與類別變項；此外，CART 與 CHAID、ID3、C4.5、C5.0 演算法的最大相異處是其在每 1 個節點上都是採用 2 分法，也就是每次只能夠有兩個子節點，C5.0 等則在每個節點上可以產生不同數量的分枝；本研究所使用的演算法方法為 CART。

2、演算法步驟

決策樹是使用遞迴(recursive)方式來進行二元的切割分類動作，關鍵在於找出分類的臨界值，而用來測量模型不純度(impurity)的最小 Gini 指數就是臨界值，如果 Gini 值中分散的指標程度很高，表示資料中分布許多類別；相反的，如果指標程度越低，則代表單一類別的成分居多，Gini 分類法目標為找到一種分割條件使得純度最高即不純度最低，Gini 係數等於 0 時，資料的不純度最低，在分類的過程中節點(node)將選出能將變異數降低最多的變項(L. Breiman et al., 1984)。通常建構決策樹有以下幾項步驟：

- (1)資料設定：一般資料探勘做預測分析時，會將資料分為兩大部分，一部分為訓練資料組(training set)，用於構建模型，另一部分則是測試資料組(testing set)。但有時候模型的構建過程中也需要檢驗模型之正確性，輔助模型構建，故會將訓練資料再分為兩組，成為訓練資料與驗

¹³ C5.0 是 C4.5 的商業改進版，主要應用於大量資料集合上之分類，並且在執行準確度和消耗記憶體方面進行優化。

證資料(validation set)；驗證資料組用於輔助模型構建，衡量模型的優劣，可以重複使用；而測試資料組則只在模型檢驗時使用，用於評估模型的準確性，避免模型建構生「過度擬合」情形。

(2)決策樹生成：使用訓練資料來建立決策樹，而在每一個內部節點，則依據屬性選擇指標來評估選擇哪個屬性做分支的依據；此又稱節點分割(splitting node)。而分支的準則主要來自於前述的「資訊增量」。所謂「資訊增量」主要是依據資訊理論中的熵(entropy)的計算為基礎，常用以衡量資料離散程度或亂度，利用資料對應類別之分布來當作資訊的純粹(purity)程度指標，並以資訊量計演算法來計算出某個屬性分割前後之差異量，資訊量越接近 1 則表示資訊之凌亂度越高(impure)，反之則純度越高，最後選擇分割屬性的規則是以找出能獲得最大資訊增益之屬性作為其分枝的節點，即該屬性對總資訊量之降低最有效，資訊增量值愈大表示用來分類的資料會表現會更好(鄧莉雅，2015)。

(3)剪枝：使用測試資料進行決策樹修剪，直到所有的新產生節點都是葉節點為止。不過決策樹很容易有「過度擬合」問題，因若沒有對樹的成長做限制，演算法最後就會為每個不同特徵值創建新的分類節點，持續將資料分類，使樹木過大、複雜。因此，為預防過度擬合，常會採取設限及剪枝，包括資料數目不得小於多少才能再產生新節點、要成為葉節點最少需要多少資料、限制樹的高度層數等。

(4)尋求最佳樹：找出所有可能之樹狀構後，即可利用測試樣本來進行錯誤矩陣的交叉驗證(cross-validation)，將測試樣本代入所有可能的樹狀結構得出誤判率(misclassification)，最後挑選最小誤判率的樹狀結構為最佳樹；須說明是，若是樣本太小情況下計算誤判率並沒有太大意義，因為資料一旦有微小變動，結果將出現很大差異(黃文、王正林，2016)。

(二) 條件推論決策樹

傳統決樹分析經常遇到兩個難題：其一，為前述提到的「過度適配」問題；另一問題則為變項選擇的偏誤(selection biases)，尤其有許多類別變項時，

這樣的問題更為嚴重，造成這種問題的關鍵，就在於傳統決策樹在展開樹形時，並沒有使用任何統計方法與檢定，以至於無法對特定變項帶來資訊增量進行判斷。為解決難題，Hothorn 等人(2006)提出「條件推論決策樹」(condition inference tree)以改善樹形的展開，此方法適用類別與連續變項的處理，如：名義(nominal)、尺度(ordinal)、數值(numeric)、設限(censored)、多變量(multivariate)各類資料型態(Hothorn, Hornik, & Zeileis, 2006)。

條件推論決策樹主要原理係依據統計檢定以確定自變項和分割點的選擇。分析過程首先假設所有自變項與依變項均為獨立，再對每一個自變項分別與依變項進行卡方獨立檢定，若該自變項的檢定 P 值小於事先設定的型 1 錯誤時，則將該自變項納入做為分割變項的候選，最後選擇與依變項相關性最強的候選自變項做為分割變項，亦即如果 P 值夠小，就會拒絕分類無效的虛無假設。此外，條件推論決策樹不需要再進行剪枝過程，因為在建立條件推論決策樹的過程中需要設定型 1 錯誤 α ， α 即扮演度量模型複雜程度的角色，一般可透過交叉驗證(cross validation)的方式進行 α 的選取。因此，條件推論決策樹不僅是啟發式，而且是無母數模型(nonparametric models)。下圖為推論決策樹的構成方式。

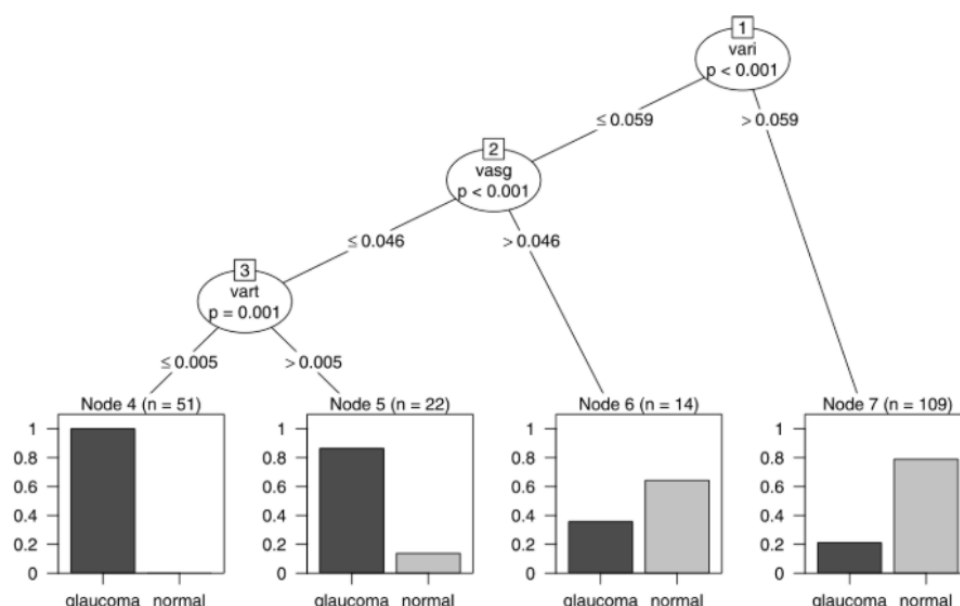


圖 3-4-7 條件推論決策樹圖

資料來源：Hothorn et al. (2006)

(三) 隨機森林

決策樹為單一模型分類法，但當資料變項或資料量大時，就會產生許多可能的分類，導致單一分類結果可靠性降低，且並無任何一種單一分類法的結果是可以解決任何問題；因此，一種增強的取向就是強化式學習法(ensemble learning)，也就是將多種結果綜合起來，強化單一結果。一般常用的強化法有裝袋法(bagging)及推進法(boosting)(何宗武，2016)，其中隨機森林為整合數百棵決策樹所形成，故稱為「森林」。

1、基本原理

機器學習¹⁴領域里，隨機森林是一種進階的分類方法，2001年Breiman進一步把分類樹組合成隨機森林(random forests)，即在變項的使用和資料使用上進行隨機化，產生多個決策樹，再加以整理，係對決策樹演算法的一種改進(Leo Breiman, 2001)。

隨機森林是一個樹狀分類器 $\{h(\chi, \beta_k), k=1, \dots\}$ 的集合，其中 $\{h(\chi, \beta_k)\}$ 是決策樹演算法建構的未剪枝的分類決策樹， χ 為輸入向量， β_k 是獨立同分布的隨機向量，決定單棵樹的生長過程；隨機森林中的每棵決策樹會對輸入資料進行分類，給予一個類別標籤，當聚集所有決策樹的類別標籤後，即可以得到類別標籤的投票(vote)分布，隨後在輸出採用多數投票法，或是取單棵樹輸出結果的平均值。其中多數投票法主要針對是分類模型，平均法則是針對迴歸模型。以分類法的投票方法來說，所謂投票是指每個樹都有 1 票，訓練後可以投 A 分類，也可以投 B 分類，投到哪一類的票數最多的就是隨機森林的分類結果，每個樹都將進行投票。另外，由於各樹都是獨立存在的，相互獨立的樹做出的預測結果涵蓋所有的情況，這些預測結果將會彼此抵消，其中少數優秀的樹的預測結果對於抗資料雜訊¹⁵具有優異的表現，可做出好的預測，即成為隨機森林之分類結果。

¹⁴ 機器學習、資料探勘、與統計分析是由不同的觀點在看待資料的分析技術。其區別在於：

- 1、統計分析：通常為利用數學模型去學習資料，找出一組參數來「描述」資料，目標是找出資料背後的分佈，進而解釋資料間的關係，並以達到最大解釋力為目標。
- 2、機器學習：透過抽象模型學習擬合資料，著重在學習模型如何自我提升以達到最佳化過程，目標是尋求最好的「預測」效果。
- 3、資料探勘：強調如何透過演算法或步驟，目標是找出最符合資料背後的價值，通常會關注資料與分析方法的適配與正確率。

¹⁵ 一般而言，資料雜訊包含錯誤值及離群值。

隨機森林在抽樣方法上是透過拔靴法(boot-strap)¹⁶的重複抽樣，之所以採取抽樣後放回的方式，主要原因在於投票所需；換言之，若採用不重複抽樣，則每棵樹的訓練樣本完全不同，無法產生交集，也就是說每棵樹訓練出來之後得到的評判標準可能都是有很大的差異；而隨機森林最後分類取決於多棵樹(弱分類器)的投票表決，故使用完全不同的訓練集來訓練每棵樹，沒有相同標準是無法得到有效的投票結果。

抽樣的具體作法係從原始訓練樣本集 N 中，取樣後重新放回，隨機重複產生 k 個新樣本，然後根據此新樣本產生 k 個決策樹所組成，每棵樹的建立依賴於一個獨立取出的樣本，森林中的每棵樹具有相同的分布，分類誤差取決於每棵樹的分類能力與它們之間的相關性(黃文、王正林，2016)，直觀上，當每棵樹具有較低錯誤率，成長出較為強健(strength)時，隨機森林整體分類能力就會上升，倘若兩顆樹的相關性同時增加時，其互補的能力就降低，錯誤率也就可能提高。

綜上，隨機森林的概念在機器學習中稱之為「集成學習法」(ensemble learning)¹⁷，亦即單棵決策樹的分類能力可能很小，但在隨機產生大量決策樹集成為森林後，其能力大為增強。而在機器學習集成的過程中，有兩種關係型態，一是各個弱機器學習間有相互依賴關係，此種方法稱之為”boosting”；另一是沒有依賴關係，稱之為”bagging”，隨機森林為獨立的許多決策樹，故使用 bagging 方法。相關圖示如圖 3-4-8

¹⁶ boot-strap 法，亦有稱之為「自助法」，屬於重複抽樣方法，原理係將已有的觀察值當作是母體重複抽樣(與 monte carlo 有真實母體不同)，以求取原先因資料不足而無法探討的資料特性的一種抽樣方法。

¹⁷ 集成學習方法，也稱多分類器系統，它的工作原理是提高單個分類器的性能，透過建立多個分類器並組合這些分類器，最終得到一個性能優於任一基分類器的集成分類器系統。隨機森林全部由決策樹所組成，稱之為「同質」(homogeneous)集成；另有「異質」(heterogeneous)集成，則是由不同性質的學習演算法所組成，例如集成的分類器中包含決策樹與類神經網路(程路文，2017)。

集成學習的理論基礎來自於 Kearns 和 Valiant 提出的基於 PAC (probably approximately correct) 的「可學習性理論」，PAC 定義學習演算法的強弱；所謂弱學習演算法係指辨別錯誤率小於二分之一(即準確率僅比隨機猜測略高的演算法)；強學習演算法則是辨別準確率很高並能在多項式時間內完成的演算法，根據這兩個概念，產生一項重要的結論：強學習與弱學習是等價的；據此，為得到一個優秀的強學習模型，可以將多個簡單的弱學習模型提升(Schapire, 1990)。

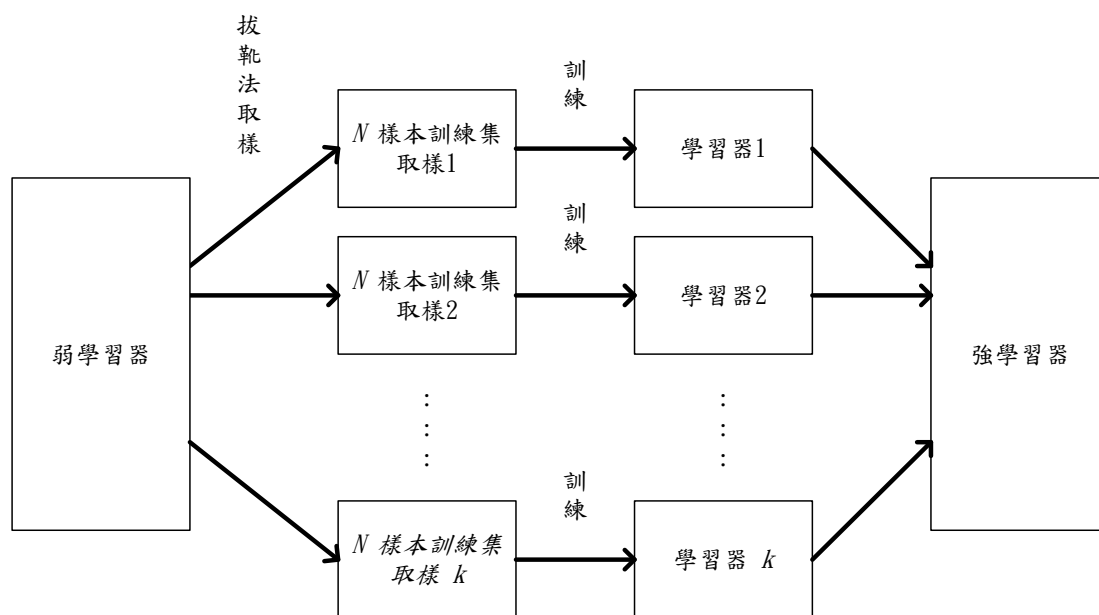


圖 3-4-8 隨機森林 bagging 原理圖

2、隨機森林建構過程

根據圖 3-2-7 隨機森林過程建構如下：

- (1) 從原始訓練集中 (弱學習器) 使用拔靴法隨機取樣後放回取出樣本，共進行 k 次取樣，生成 k 個訓練集。
- (2) 對 k 個訓練集，分別訓練 k 個決策樹模型。
- (3) 對於單個決策樹模型，假設訓練樣本特徵的個數為 M ，那麼每次分裂節點時，每個節點都將隨機選擇 m ($m < M$) 個特徵值，然後運用這 m 個特徵值來確定最佳的分裂點，在決策樹的生成過程中， m 的值保持不變。
- (4) 每棵樹經訓練後，均已學習按此規則繼續分裂，知道該節點的所有訓練樣本都屬於同一類。
- (5) 在決策樹的分裂過程中不需要剪枝。
- (6) 將生成的多棵決策樹組成隨機森林成為強學習器。對於分類問題，按照多棵樹分類器投票 (majority vote) 決定最終分類結果；對於迴歸模型，由多棵樹預測值的平均值決定最終預測結果。

前述提及，在隨機森林的建構過程中，影響其分類效果主要有任意兩棵樹的相關性，相關性越大，錯誤率越大；另每棵樹的分類能力越強，整個森林的錯誤率越低，而其中影響關鍵則在於特徵值的選擇，減小特徵 m 選擇值，樹的相關性和分類能力也會相應的降低；增大 m ，兩者也會隨之增大；所以，如何選擇最佳 m （或者是範圍）的參數將直接影響分類表現。

3、袋外錯誤率問題

為選擇最佳的 m 參數，Breiman(1996)提出「袋外錯誤率」(out-of-bag error, OOB)，用來預測分類的正確性。其原理為，隨機森林在構建每個決策樹的時候，採用的是隨機又放回的抽取，所以對於每棵樹來說，都有一些樣本沒有被抽取到，所以這些樣本成為袋外樣本，此 OOB 大約占全部樣本 37%，而這些未被抽取的樣本沒有參與訓練集模型的擬合，因此可以用來檢測模型的泛化能力與正確性；換言之，使用 OOB 誤差估計可免除對預留測試資料的需要，亦即可以在模型生成過程中取得真實誤差的無偏估計，且不損失訓練數據量。

整體而言，隨機森林的優點如下：

- (1)學習過程很快，在處理大量資料及變數時，效率依舊很好，故當有非常大的訓練資料集和非常多的輸入變項時，隨機森林為相當適當的選擇。
- (2)就像決策樹一樣，幾乎不需要對資料進行預處理，資料不需要標準化，且對異常值具有彈性，不易受到離群值影響。
- (3)如果有許多輸入變項，通常不需要在開始建構模型前進行任何變項篩選，隨機森林的模型構建器能夠針對最有用的變項進行選擇。
- (4)隨機森林其對多元共線性不敏感，結果對缺失資料和非平衡的資料比較穩健。
- (5)由於隨機森林是由每棵獨立樹所組成的模型，建構的過程中不受其他決策樹的影響，因此，模型構建不易造成過度擬合。

(四) 推進法

推進法(boosting)¹⁸是由 Schapire 在 1990 年提出，boosting 是強調使用弱學習 (weak learner) 的演算法，而各個弱學習器之間具有互補的作用，若使用強學習器，則相互之間將產生干擾，影響最後預測或分類結果，唯有彼此都是「弱模型」，才能專注在自己本身的能力發揮，然後再把彼此的成果結合一起，這就是 boosting 的概念。例如，M1 學習器要做到 M2 學習器做不到事情，M3 則要辦到 M1 無完成的任務，在經過多次的疊代(iteration)執行之後，不斷試錯的過程，使輸出的最後結果提高分類的正確率而成為強學習模型(Schapire, 1990)。

boosting 主要係將權重與研究資料中的觀測值相關聯，並增加（增強）那些難以精確分類的觀測值權重，以利於構建模型，並在模型建構後對權重進行修改，再次賦予那些較難分類的觀測值更大的權重，故此類觀察的權重通常會從一個模型到另一個模型上下波動（如圖 3-2-8）。簡單地說 Boosting 是透過將弱分類器的錯誤資料權重提高，然後再訓練新的分類器，這樣新的分類器就會學習到錯誤分類資料(misclassified data)的特性，進而提升分類結果。

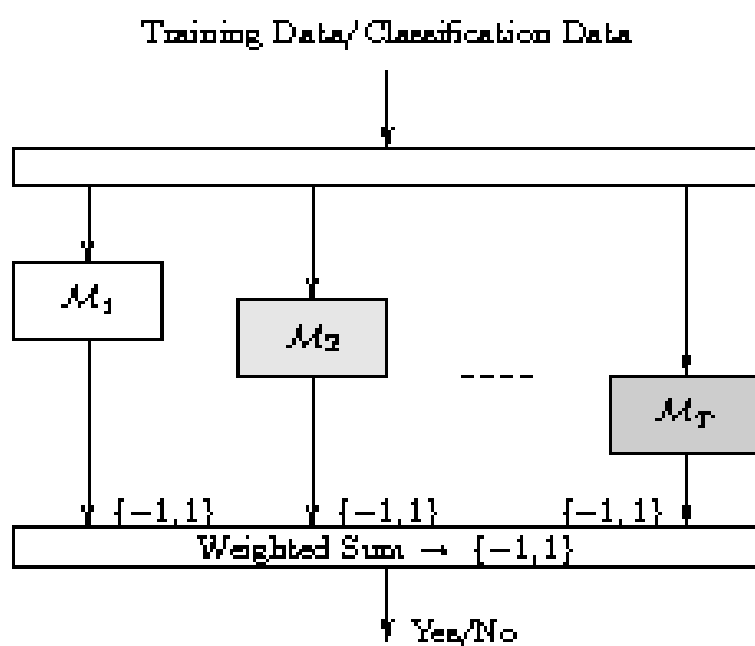


圖 3-4-9 boosting 分類器模型權重波動圖

資料來源：Williams(2011)

¹⁸ 由於推進法有許多的演算方法，boosting 為演算法概念，為避免混淆，均使用其英文名稱。

具體來說，boosting 的過程由以下 4 個步驟組成(如圖 3-2-9)：

- (1)從訓練集 D 中以無放回抽樣方式隨機抽取一個訓練子集 d_1 ，用於弱分類器 M_1 的訓練。
- (2)從訓練集 D 中以無放回抽樣方式隨機抽取一個訓練子集 d_2 ，並將 M_1 中錯誤分類樣本的 50% 加入到訓練集中，訓練得到弱分類器 M_2 。
- (3)從訓練集 D 中抽取 M_1 和 M_2 分類結果不一致的訓練樣本生成訓練樣本集 d_3 ，用 d_3 來訓練第三個弱分類器 M_3 。
- (4)經過多數投票來組合弱分類器 M_1 、 M_2 和 M_3 ，乃至 M_k 。

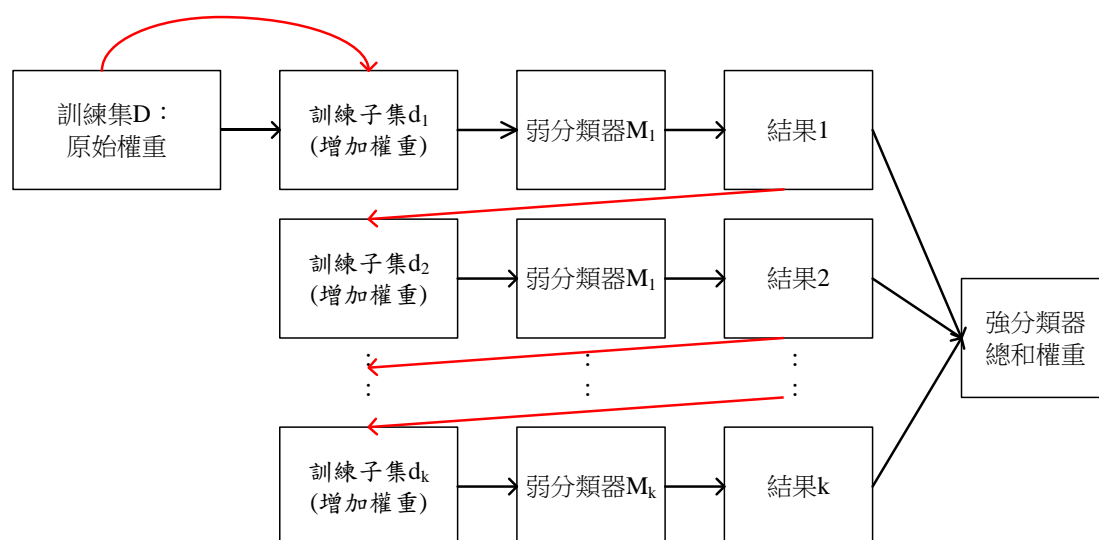


圖 3-4-10 推進法分類器組合圖

對於 boosting 來說，有兩個關鍵問題，一是在如何改變訓練資料的權重；二是如何將多個弱分類器組合成一個強分類器。然而，此兩關鍵問題的難點在於，boosting 要如何預先知道弱分類器識別準確率的下限，並加以判斷；故有一種相當流行且類似傳統推進法的演算法，稱之為 AdaBoost(abbreviation for adaptive boosting，適應性推進演算法)，為 boosting 的變體(Schapire & Singer, 1999)，就是為解決此困難而開發，一般來說效果更優於 boosting 演算法，尤其在資料分布不均勻的情況下更顯其優勢。

AdaBoost 與 bagging 不同處在於，bagging 是對一系列互不相干的分類器進

行投票，但 AdaBoost 的演算法承襲傳統推進法的方式，在依次建構學習器的過程中，同樣根據上一個分類器對各訓練集樣本的預測結果，自行調整在本次分類器建構時，各樣本被抽中的機率。就實際運作來說，如果在上一次的分類中，樣本 i 被錯誤分類，則在下一次的樣本抽取過程中， i 樣本就會賦予更高權重，主要目的就是希望增加 i 樣本被抽到的機會，以進一步加強被正確分類的機率。

對於權重即時調節的過程正是 AdaBoost 演算法最為特殊之處，改良傳統 boosting 的問題，從而更增加精準度和穩定性。總的來說，bagging 和 AdaBoost 的演算法概念可適用於其他分類模型，但大多數仍是運用在決策樹模型，主要原因在於決策樹模型的分類方式可以同時處理數值、類別及次序等各類型變項，且變項的選擇也較容易(黃文、王正林，2016)。

除 AdaBoost 演算法外，近年來在各大科學競賽經常被使用的 XgBoost 也是屬於 boosting 的優化演算法，其演算方法除決策樹等使用 CART 外，¹⁹也支援線性分類器，此時 XgBoost 相當於具有 L1 和 L2 正則化(regularization)²⁰的邏輯迴歸(分類問題)或線性迴歸，頗適合本研究分類方法使用。由於此類演算法相當複雜，在此無法逐一說明，介紹此方法主要是 R 語言已提供 RATTLE 的 XgBoost 演算法套件，本研究也將在區位變項分析的章節中加以使用。

(五) 模型評估

本研究選取以上 5 種目前經常被運用於資料探勘分類的模型與演算法，各有其優缺點；為瞭解各不同模型訓練成效(performance)如何，須有客觀的方法加以評定。然而，各種分類器的成效如何乃是相對的概念，其表現優劣與研究議題、資料特性、樣本大小及變項篩選等均具有關聯，故以某項議題或資料集的表現來評價分類器的優異性，對分類器而言並不公允也不恰當，這也就是經常可以看到在某類議題研究中，若干分類器被認為相當有效，但在其他研究中卻表現不如預期的原因。此外，一般在討論分類的有效性都是基於其成功預測率進行評價，但是這個指標僅是其中之一，並不能斷言分類器好壞；舉例來說：當研究的資料集

¹⁹ XgBoost 的基本分類器是 CART 決策樹，但與前述所提到的 CART 決策樹不同；簡要的說，XgBoost 樹的葉節點輸出值，不是分到該葉節點的所有樣本點的均值(迴歸樹)，而是由一個函數生成的值。

²⁰ 在機器學習中，「正則化」的功能主要在於避免模「過度擬合」問題發生。正則化的思維，就是當過度擬合發生時，可能是因為訓練的假設模型本身就過於複雜，因此嘗試將複雜的假設模型退回至簡單的假設模型。相關解釋可參考臺灣大學林軒田教授的「機器學習基石」講義。<https://zh-tw.coursera.org/learn/ntumlone-mathematicalfoundations>.

有 90 個正值，10 個負值，分類器 M1 將所有資料均判定為正值，則其分類預測率為 90%；而分類器 M2 成功分出 80 個正值，6 個負值，分類預測率僅 86%，在此情況下，尚難認定分類器 M2 不如 M1，此充分表明若僅依分類預測率做為判別指標，則將忽略樣本被錯誤分類的事實。因此，為能較客觀反映分類器的表現程度，本研究將以常用二元分類的「驗證指標」(validation index)做為評定成效的方法。

1、錯誤矩陣(error matrix)

二元分類的模型基本上就分辨資料為「正」和「負」、「有」與「沒有」、「真」與「假」、「是」與「不是」的過程。而錯誤矩陣，或稱「混淆矩陣」(confusion matrix)為資料探勘或是機器學習對於分類方法準確率進行評估的工具，透過將模型預測的數據與測試數據進行對比，「真實情況」(true condition)為來自資料的真實分布情形，「預測結果」(predicted outcome)就是模型預測出來的結果。通常在判定時，以"positive"代表「有」、「真」或是「正」，在本研究中則可設定為「熱區」，反之"negative"為「沒有」、「假」或是「負」，則代表「冷區」。判定結果有下列情況：

- (1)「真陽性」(true positive, TP)：真實情況是「陽性」，模型預測輸出結果為「陽性」的個數。
- (2)「真陰性」(true negative, TN)：真實情況是「陰性」，模型預測輸出結果為「陰性」的個數。
- (3)「偽陽性」(false positive, FP)：真實情況是「陰性」，模型預測輸出結果為「陽性」的個數。統計學將 FP 稱為「型一錯誤」(type I error)，即虛無假設事實上成立，但檢定結果拒絕虛無假設。
- (4)「偽陰性」(false negative, FN)：真實情況是「陽性」，模型預測輸出結果為「陰性」的個數。FN 為統計上所稱的「型二錯誤」(type II error)，即虛無假設事實上不成立，但檢定結果不拒絕虛無假設。

「錯誤矩陣」為相當適合瞭解二元資料的非對稱狀態關係，通常進行模型預測能力評估時多從此方式開始。在此測量方法下，可對真實情況次數與預測輸出結果次數相除得出比率，如定義「真陽性率」為"true positive rate"，反之亦然。

由此衍生之相關概念與指標如下：

- (1)「靈敏性」(sensitivity, SEN):即為「真陽性率」(true positive rate, TPR),又稱為「回溯性」(recall)或「精確性」(precision),此項指標值越高越好。
- (2)「特異性」(specificity, SPE):即為「真陰性率」(true negative rate, TNR),代表正確預測為陰性的命中率。
- (3)「正確性」(accuracy, ACC):為模型整體判斷的正確率,所以有時候也稱為"Overall Accuracy",指標值越高越好。
- (4)「真陽性率」(true positive rate, TPR):預測模型判成陽性,且實際上為陽性的比率,指標值越大越好。
- (5)「真陰性率」(true negative rate, TNR):預測模型判成陰性,且實際上為陰性的比率,指標值越大越好。
- (6)「偽陽性率」(false positive rate, FPR):預測模型判成陽性,但實際上為陰性的比率。
- (7)「偽陰性率」(false negative rate, FNR):預測模型判成陰性,但實際上為陽性比率。
- (8)「陽性預測值」(positive predictive value, PPV):也稱為 Precision,即模型判斷結果呈現陽性且確實為陽性的比率,指標值越高越好。
- (9)「陰性預測值」(negative predictive value, NPV):模型判斷結果呈陰性且實際上也為陰性的比率,指標值越高越好。

以上二元分類常用的指標名稱和計算方式如下(表 3-4-1):

表 3-4-1 二元分類常用的指標名稱和計算方式

		真實情況 (true condition)		
		陽性 (positive)	陰性 (negative)	
預測結果 (predicted outcome)	全部樣本			
	陽性 (positive)	真陽性 (TP)	偽陽性 (FP)	「陽性預測值」(PPV) $\frac{TP}{TP + FP}$
	陰性 (negative)	偽陰性 (FN)	真陰性 (TN)	「陰性預測值」(NPV) $\frac{TN}{FN + TN}$
	正確性 (accuracy) $\frac{TP + TN}{T}$	「真陽性率」 (TPR) $\frac{TP}{TP + FN}$	「偽陽性率」 (FPR) $\frac{FP}{FP + TN}$	
		「偽陰性率」 (FNR) $\frac{FN}{TP + FN}$	「真陰性率」 (TNR) $\frac{TN}{FP + TN}$	

2、ROC 操作特徵曲線

ROC(receiver operating oharacteristic ourve)曲線為可視覺化用來評估二元分類問題的一種技術，最早提出 ROC 的人為 Spackman(1989)；近年來，ROC 圖在資料探勘與機器學習領域的使用日漸增加(Fawcett, 2006)。進一步來說，ROC 是反映「靈敏性」(即真陽性率)和「特異性」(真陰性率)的綜合指標，是用圖示法表達靈敏性和特異性的相互關係，透過設定多個不同的臨界值(threshold)，從而計算出一連串靈敏性和特異性之數值，1 個靈敏性會對應 1 個特異性，將臨界值設定為 1 和 0 時，分別可以得到 ROC 曲線上的(0,0)和(1,1)2 個點，另將所有可能對應點連起來(圖 3-4-11)，得出的就是 ROC 曲線，當臨界值取值越多，ROC 曲線越平滑。最後，以靈敏性為縱座標、「1-特異性」(偽陽性率)為橫座標繪製成曲線。ROC 曲線解讀方式，係以對角線為基準，若是算出來的 ROC 曲線等於對角線的話，代表模型完全沒有鑑別性，相當於隨機猜測，若 ROC 往左上角移動，代表模型敏感度越高(偽陽性率越低)，模型的鑑別力越好。

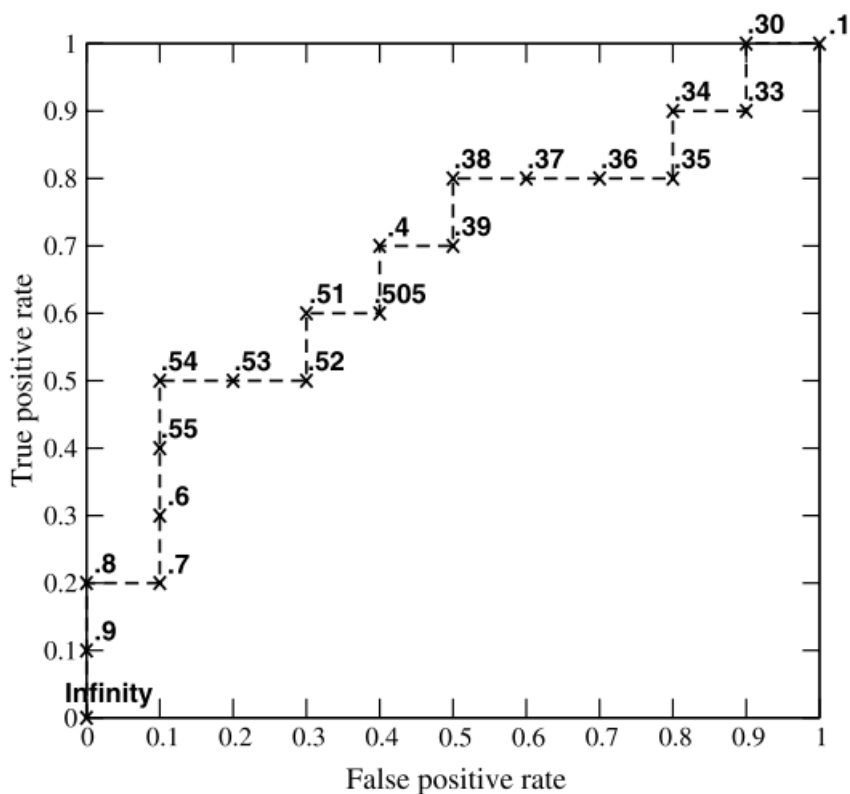


圖 3-4-11 靈敏性與特異性對應值

資料來源：Fawcett(2006)

除觀察 ROC 曲線的圖形之外，也可以利用曲線下的面積(area under curve, AUC)來判別 ROC 曲線的鑑別力，AUC 數值的範圍從 0 到 1，數值愈大愈好（如圖 3-4-13）。以下為 AUC 數值一般的判別規則(Hosmer, Lemeshow, & Cook, 2001)：

- (1) $AUC=0.5$ 無鑑別力(no discrimination)，ROC 為對角線。
- (2) $0.7 \leq AUC \leq 0.8$ 可接受的鑑別力(acceptable discrimination)。
- (3) $0.8 \leq AUC \leq 0.9$ 優良的鑑別力(excellent discrimination)。
- (4) $0.9 \leq AUC \leq 1.0$ 極佳的鑑別力(outstanding discrimination)。

AUC 也可以當作比較不同模型之間比較的指標，AUC 越大代表那個模型越表現較佳(圖 3-4-12)。

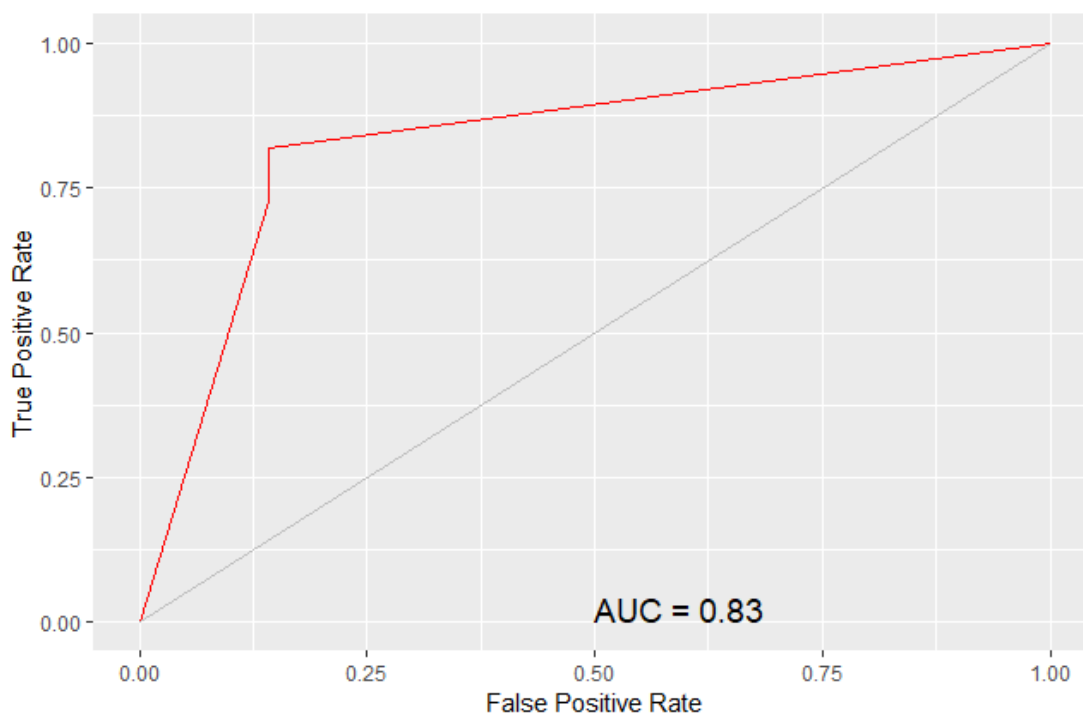


圖 3-4-12 ROC 曲線與 AUC 面積圖

整體而言，本研究使用資料探勘方法，乃是依循資料分析的歸納策略，應用各種機器學習演算法，漸進地從資料中獲取資訊，不預設任何前提假定或假設來認定最後何種結果將會呈現，著重發掘資料中變項之間可能的關連，而不特別強調個別屬性的因果關係，在特性上以探索與機率性多於檢定及因果解釋，此與傳統的推論統計分析，必須有前提條件及統計假定之演繹式方法有所不同，故對於模型建構後進行評估相當重要，避免與真實世界造成太大差異，故上述提到的錯誤矩陣的交叉驗證與 ROC 曲線評估方法，是以模型「樣本外」的預測能力作為評斷標準，而不是以「樣本內」的適配能力判斷最佳模型，此設計可做為調整過度配適問題的依據，提升模型的可信賴程度。

三、研究工具

根據前述之研究策略，本研究所運用之研究工具主要有 GIS 軟體、掃描統計軟體與資料探勘軟體 3 大類。簡介如下：

(一) GIS 軟體

目前常用之 GIS 軟體主要有 ArcGIS、GeoDa、Quantum GIS (簡稱 QGIS)

等，其中 ArcGIS²¹ 是由 ESRI 公司出品的地理資訊系統系列 ArcReader、ArcView、ArcEditor 和 ArcInfo 軟體的總稱，操作簡便，功能強大，惟須支付頗為高額的費用。

GeoDa²² 由 Luc Anselin 和其團隊開發，從 2003 年 2 月 GeoDa 發布第一個版本以來，由於其為免費軟體故頗受歡迎，截止 2017 年 6 月，GeoDa 的用戶數量已經超過 20 萬，包括哈佛，麻省理工、康乃爾等著名大學都在實驗室中安裝並使用 GeoDa 軟體。GeoDa 最新發布的版本是 1.12 版。

QGIS²³ 由 Gary Sherman 於 2002 年開始開發，並於 2004 年成為開源地理空間基金會的一項專案，並由志願者開發團體持續維護，此軟體定期釋出更新和錯誤修正，目前最新版本為 3.8.0 版，且在捐款者贊助下，Quantum GIS 已翻譯為三十餘種語言，廣泛被使用在全世界的學術和專業環境。

由於 GIS 在本研究之功能主要為視覺化介面，本研究選擇使用 QGIS 免費軟體。

(二) SaTScan 掃描統計軟體

SaTScan 是一個實現 Kulldorff 掃描統計的專用軟體，用於空間、時間和時空群聚檢測，其統計方法與軟體運用受到目前公共衛生領域中用於檢測疾病群聚廣泛接受與使用，目前最新版本為 9.6 版。軟體設計主要用途為：

- 1、偵測群聚是否具統計上顯著意義。
- 2、進行前瞻性(prospective)的即時或定時監測，以便及早發現可能的群聚暴發。

由上可知，基於 Kulldorff 掃描統計的 SaTScan，除能偵測時空群聚外，也同時具備預警之功能，此對於攸關人類生命安全的疾病預防而言極具貢獻，故若能將此功能運用於犯罪學領域，及早偵測犯罪活動頻繁週期與熱區，將可對犯罪防治相關作為提供重要協助，此亦是本研究之初衷。

然而，SaTScan 雖然功能強大，但目前尚有許多缺點，以至於在推廣上無法

²¹ ArcGIS 官網 <http://www.arcgis.com/index.html>。

²² GeoDa 官網 <http://geodacenter.github.io/index-cn.html>。

²³ QGIS 官網 <https://qgis.org/en/site/>。

如 GIS 軟體普及，其主要因素為：

- 1、資料輸入要求嚴格且複雜：由於 SaTScan 可進行時空掃描，故資料整理須為同時具備時間與空間之格式，且空間格式除座標之外，另須行政區碼匹配，再加上某時間、地區的發生案例數，在資料整理的過程中即可能讓許多不擅長文書軟體操作（如：Excel 等）的人耗費大量時間，甚至整理後的資料軟體無法讀取；故習於 GIS 的使與者可能會相當不適應，因使用 SaTScan 輸入的每個步驟都需要 1 個單獨的檔案（例如：發生之案例、控制項、人口變項和地理座標均需要單獨的檔案），並且事件的位置座標與屬性是分開存儲；凡此，經常造成學習者不勝其煩而卻步。此外，資料輸入、輸出檔的格式如 ASCII²⁴、dBase²⁵、KML²⁶在運用上也須具備若干資訊專業背景，否則無法讀取或進階利用。
- 2、無法自動產生圖層：SaTScan 軟體統計結果無法自動產生群聚等圖層，必須藉由 GIS 軟體加以轉換，過程頗為繁雜。值得慶幸是，此軟體自 2018 年起，空間資料部分已經可以藉由 HTML²⁷將輸出資料導出到 Google Maps 與 Google Earth，惟須將笛卡爾坐標系統(cartesian coordinate system)，即常用的 X-Y 座標轉換為經緯度，但若須進行精確的群聚分析，仍舊需與 GIS 軟體密切合作，甚至可進一步利用 GIS 內建功能另做其他圖層套疊分析。
- 3、對電腦規格要求較高：由於掃描統計乃針對每一地區進行逐步掃描，若加上時間因素且資料量大時，需耗時相當長時間，故 SaTScan 對於電腦規格要求較高，否則在測試各種設定時，可能出現曠日廢時卻一無所獲的情形。

²⁴ ASCII(american standard code for information interchange)，譯為「美國資訊交換標準程式碼」是基於拉丁字母的一套電腦編碼系統，主要用於顯示現代英語。

²⁵ dBase 是第一個在個人電腦上被廣泛使用的單機版資料庫系統，包括核心資料庫引擎、查詢系統、表單引擎、與腳本程式語言。

²⁶ KML(keyhole markup language)，是基於 XML(extensible markup language，可擴展標記語言)語法標準的一種標記語言(markup language)，主要用來表達地理標記，可應用於 Google 相關軟體中(Google Earth, Google Map 等)，用於顯示地理數據(包括點、線、面、多邊形，多面體以及模型等)。目前許多 GIS 相關企業也開始採用此種格式進行地理數據的交換。

²⁷ HTML(hyper text markup language)，譯為「超文字標記語言」是一種用於建立網頁的標準標記語言。

(三) R 語言

R 語言²⁸由紐西蘭奧克蘭大學的 Ross Ihaka 和 Robert Gentleman 所開發出之免費程式語言與軟體，目前開發的核心團隊是由世界各地不同機構所組成，近年在 Hadley Wickham 與 RStudio 團隊所建構的應用套件以及大數據、資料科學與人工智慧的熱潮加持，廣受統計學家和資料科學家的喜愛(郭耀仁，2018)。

事實上，如果熟悉程式撰寫，使用 R 語言配合相關套件，GIS 與 SaTScan 的功能幾乎均能取代，並可加以視覺化，功能齊全，但畢竟 GIS 係專為視覺化製圖所設計，故 R 語言所繪製之圖層視覺化效果尚不如專業 GIS 軟體。在本研究中，R 語言的運用主要在資料探勘與機器學習方面，使用時機係銜接與前述 2 種軟體分析之後，做為區位分析的工具。

(四) SPSS 統計軟體

SPSS 早期為 Statistical Package for the Social Sciences 的縮寫，原本是在社會科學研究的商業統計軟體，後來功能逐漸加強，加入許多統計分析相關程式套件，逐漸受各行各業使用，產品定義也轉為 Statistical Product and Service Solutions。SPSS 以其試算表的介面及點選式的操作著名，對於部分不擅於編寫程式的使用者來說，SPSS 無疑是相當容易操作的一套軟體。

SPSS 公司於 2009 年被 IBM 公司併購，自 19 版開始軟體正式名稱為 IBM SPSS Statistics，主要強化機器學習演算法程式庫、文字分析、開放程式碼延伸、大數據整合以及應用程式等部分，以配合當前更加多元的研究趨勢。本研究使用版本為第一 8 版，使用時機為有關犯罪資料探索性統計及羅吉斯迴歸分析等。

第五節 研究倫理

一、研究倫理的定義與內涵

根據臺灣學術倫理教育推廣資源中心之定義，所謂「研究倫理」指的是執行研究工作時所必須遵循的行為規範，包含研究者的道德、誠信、社會和研究成本的妥善規劃及運用行為等，皆是研究倫理討論的範疇(朱家嶠，2017)。違反研究

²⁸ R 軟體 官網 <https://www.r-project.org/>。

倫理不但使研究者個人學術誠信、研究成果等遭到質疑，更可能誤導學術研究方向，進而影響到社會與國家群體聲譽，故對於研究倫理進行規範確有其必要性。

在研究倫理的內涵方面，臺灣學術倫理教育推廣資源中心臚列出3項核心基本準則，分別為「涉及人類研究參與者之研究倫理」(human research ethics)、「動物實驗倫理」(animal research and ethics)及「負責任的研究行為」(responsible conduct of research)3項。就本研究而言，由於研究對象與資料來源，為來自政府資料開放平臺，並未涉及人與動物之研究，亦未包含個人基本資料，前兩項準則似較無違反之虞，故「負責任的研究行為」為本研究須特別重視之核心準則。

二、負責任的研究行為

所謂負責任的研究行為是指研究者必須遵守誠實(honesty)、正確(accuracy)、效率(efficiency)及客觀(objectivity)等執行研究的核心價值(朱家嶠，2017)，在實際執行上即是正確蒐集資料、誠實處理分析資料，不可為凸顯研究成果而竄改資料，且對於資料的詮釋必須客觀不帶有價值、偏見，不刻意掩飾負面的研究結果等。

就本研究而言，在蒐集資料方面，數據及圖資來源為政府公開資料，均可自官網下載取得，而住宅竊盜資料以2019年1月自6月，使用者已達約二千五百次，受到各界廣泛使用，故資料之正確性應有其可信度。而參考文獻等資料除自行購買之圖書外，許多為網路查詢或圖書公司所釋出的免費電子書籍等，另部分為由圖書館查詢所得，引用資料儘量以原始資料為主，研究引用皆註明出處，避免侵權行為。

在處理分析資料方面，本研究對於數據之處理將是依循學術準則，例如：研究中如遇原始資料遺漏資訊過多無法以統計方法加以彌補時，或案件無法進行地理編碼等，導致影響分析正確性者，才考慮予以剔除，其餘分析資料即便進行格式轉換，亦務必使與原始資料相符；此外在進行合併資料前先另存新檔，合併後與先前儲存之檔案進行比對，確保儲存格移動後各欄位無錯置情形，以保持資料之正確性。

在分析過程中，由於使用軟體均有其設定邏輯與操作程式，資料輸入後將依照研究目的逐步分析，等待研究結果呈現，不竄改原始犯罪數據來符合自己的預

期研究目標，事實上由於研究所使用之統計與演算法頗為複雜，任意竄改分析資料或結果，更可能導致難以解釋的現象；而對於研究結果詮釋，則將依統計之正確意義詳加說明。

三、不造成研究對象可能受到的傷害

本研究研究對象係以政府公開的臺北市住宅竊盜與相關人文社經資料為主，政府在開放此類資料之前已先行進行初步把關，將有關可能洩漏個體資訊的疑慮先予排除，例如對住宅竊盜精確之地址與時間的模糊化，採取區段數據，研究者只能根據自行定義，以模擬方式建立樣本資料。然而，即便對個別資料已進行隱匿，但經過研究者之研究設計與分析後，研究結果仍有可能造成負面影響。

例如，住宅竊盜開放資料的查詢臺北市並非全球第一，先進城市如紐約、倫敦等，早已提供此項服務。近來開放資料逐漸成為世界各先進城市政策推動趨勢，且唯有將治安資訊公開透明，提供市民完整透明的資訊，才能真正提供市民安心生活的環境。但臺北市政府此項舉措曾遭民眾反彈，深怕公開住宅竊盜的資訊後，將會影響自身居住地點的房價，造成房價下跌。針對此項議題，國內研究顯示短期內，低、中犯罪程度區域對於房價沒有造成顯著的影響，高犯罪程度區域對於房價造成顯著下跌，但長期下公布犯罪地圖，無論在低、中、高犯罪程度對消費者造成影響並不顯著(黃元杰，2017)；因此，本研究中涉及住宅竊盜與房產價值之分析，依該研究結論，應不致對社會造成負面影響。惟本研究地價項僅為諸多變項之一，故對於其他變項衍生之研究結果的詮釋與可能造成民眾疑慮之社會負面影響將會進行風險與利益的審慎評估，以決定研究呈現方式。

此外，就社會解組理論而言，多認為種族的異質性較容易產生社會解組現象；在我國雖然沒有這方面的問題，但近年來隨著外籍配偶的增加，也逐漸成為一項重要議題。由於社會解組理論在本研究的區位分析中為重要理論，為顧及涉及此類議題，故對於有關外籍配偶人數此變項便不予納入。

第四章 臺北市住宅竊盜時空分析

第一節 臺北市住宅竊盜犯罪資料探索分析

本節以 2015 年至 2017 年臺北市住宅竊盜的犯罪資料為基礎，透過探索性資料分析步驟，瞭解不同時間尺度上各類型犯罪案發率之趨勢，接著以推論統計檢定犯罪現象在時間分布上是否存在明顯的差異，屬於一般性統計分析。

在分析方式上，根據所蒐集資料加以分類為年別、月別、星期別、時段別與地區別等變項，進行犯罪資料探索統計分析。本研究假設臺北市在時段與區域範圍各變項方面，住宅竊盜發生數(率)之分布具有顯著差異，虛無假設即為「住宅竊盜犯罪率不因時間或地區而有所不同」，在時間方面使用卡方適合度進行檢定，而在空間方面，由於各行政區人口與住宅數量不同，除以犯罪數呈現外，也將以犯罪率加以比較。使用推論統計的目的在於彌補僅以觀察次數分配但未能證明是否有顯著差異之缺點，以更進一步瞭解臺北市歷年與各區域犯罪現象的時空分布概況。

一、臺北市住宅竊盜時間分布概況

(一) 時間分布

1、年分布

2015 年至 2017 年，臺北市住宅竊盜發生件數，以 2016 年次數最多，為 642 件，占 36.5%，其次為 2015 年 614 件，占 34.9%，2017 年下降至 503 件，占 503 件（如表 4-1-1；圖 4-1-1）。

表 4-1-1 2015-2017 年臺北市住宅竊盜發生數

次數 年度	發生次數	百分比
2015年	614	34.9
2016年	642	36.5
2017年	503	28.6
總和	1,759	100

卡方值=18.434；期望值=586.3；自由度=2；p<.01

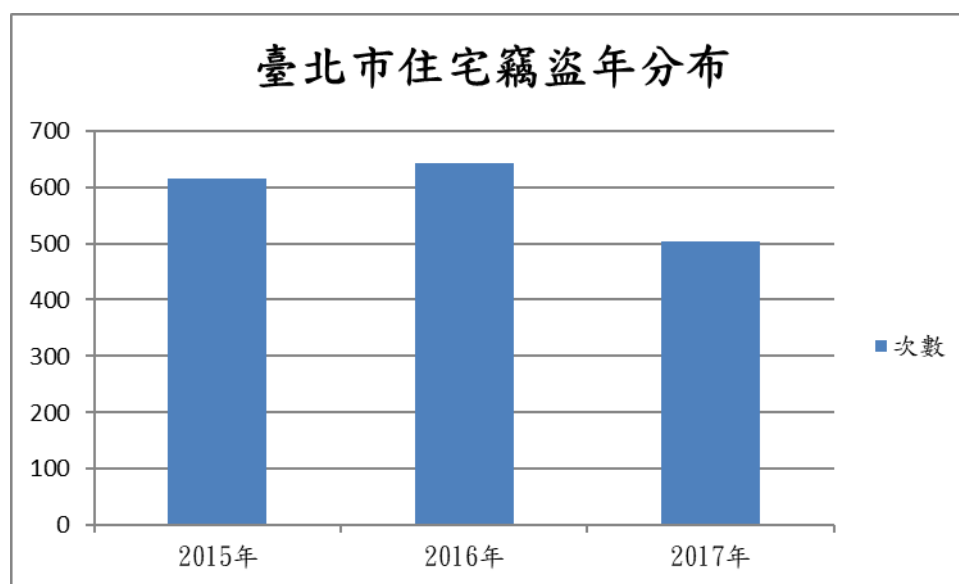


圖 4-1-1 2015-2017 年臺北市住宅竊盜年分布圖

以卡方適合度檢定，檢定統計量顯示臺北市 2015 年至 2017 年住宅竊盜犯罪發生年度的分布上有顯著差異(卡方值=18.434；期望值=586.3；自由度=2；p<.01)。從觀察值與期望值分析，其中 2015 年與 2016 年之觀察值大於期望值，顯示上述年度住宅竊盜犯罪發生率較高；反之，2017 年之發生率較低。

2、月分布

在月分布方面，2015 年至 2017 年各月份住宅竊盜發生次數如下(表 4-1-2；圖 4-1-2)。

表 4-1-2 2015-2017 年臺北市住宅竊盜月發生數

次數 月份	發生次數	百分比
1月	177	10.1
2月	141	8.0
3月	134	7.6
4月	163	9.3
5月	136	7.7
6月	140	8.0
7月	149	8.5
8月	155	8.8
9月	135	7.7
10月	160	9.1
11月	132	7.5
12月	137	7.8
總和	1,759	100

卡方值=15.247；期望值=146.6；自由度=11；p=.171

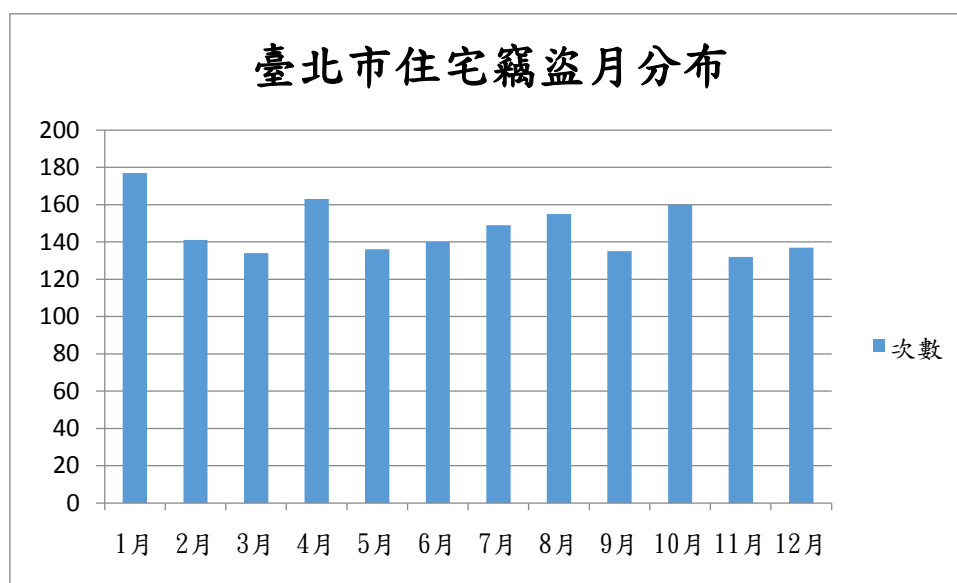


圖 4-1-2 2015-2017 年臺北市住宅竊盜月分布圖

以卡方適合度檢定，檢定統計量顯示臺北市 2015 年至 2017 年住宅竊盜犯罪發生的月分布上無顯著差異(卡方值=15.247;期望值=146.6;自由度=11;p=.171)，但從觀察值與期望值分析，可大致瞭解 1、4、7、8 及 10 月為發生率較高月份，反之其他月份則較低。

3、周分布

在周分布方面，2015年至2017年各星期別住宅竊盜發生次數如下(表4-1-3；圖4-1-3)。

表 4-1-3 2015-2017 年臺北市住宅竊盜星期別發生數

星期別 \ 次數	發生次數	百分比
星期日	227	12.9
星期一	280	15.9
星期二	277	15.7
星期三	234	13.3
星期四	258	14.7
星期五	278	15.8
星期六	205	11.7
總和	1,759	100
卡方值=20.994；期望值=251.3；自由度=6；p<.05		

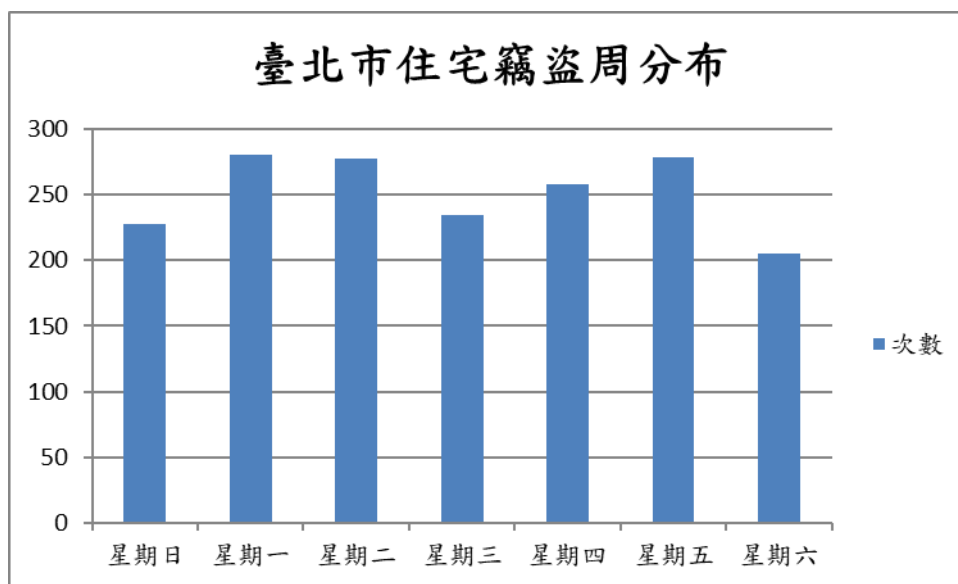


圖 4-1-3 2015-2017 年臺北市住宅竊盜周分布圖

以卡方適合度檢定，檢定統計量顯示臺北市 2015 年至 2017 年住宅竊盜犯罪發生的周分布上有顯著差異(卡方值=20.994;期望值=251.3;自由度=6; $p<.05$)，從觀察值與期望值分析可知，星期一、二、四、五為發生率較高星期別，反之星期六、日及星期三則較低。

4、時段分布

在每日住宅竊盜發生時段方面，前已詳述因住宅竊盜發生時均為無人在家或發現情況下，故報案時間與發生時間將有所落差，連帶影響此變項分析的準確性；惟本研究仍將此部分予以呈現以供參考。

根據原始資料分類方式，係將 1 天 24 小時以每 3 小時加以區分，共分為 8 個時段，各時段竊盜發生次數如下(表 4-1-4；圖 4-1-4)。

表 4-1-4 2015-2017 年臺北市住宅竊盜每日時段發生數

時段別 \ 次數	發生次數	百分比
1時 — 3時	159	9.0
4時 — 6時	134	7.6
7時 — 9時	223	12.7
10時 — 12時	305	17.3
13時 — 15時	245	13.9
16時 — 18時	218	12.4
19時 — 21時	243	13.8
22時 — 24時	232	13.2
總和	1,759	100

卡方值=89.382；期望值=219.9；自由度=7；p<.05

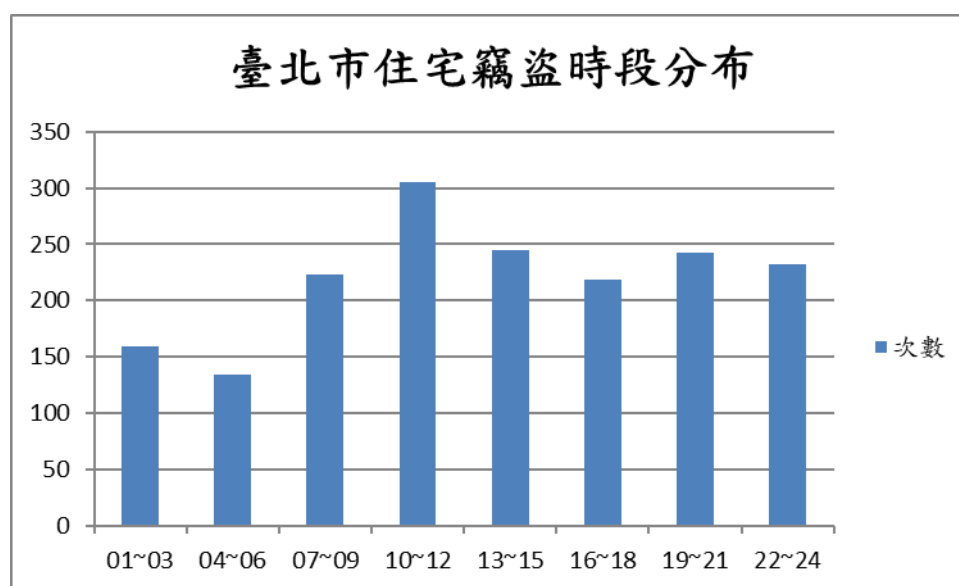


圖 4-1-4 2015-2017 年臺北市住宅竊盜時段分布圖

以卡方適合度檢定，檢定統計量顯示臺北市 2015 年至 2017 年住宅竊盜犯罪發生的時段分布上有顯著差異(卡方值=89.382;期望值=219.9;自由度=7;p<.05)，從觀察值與期望值分析可知，上午 10 時至 12 時及下午 1 時至 3 時為發生率較高的 2 個時段，其次為上午 7 時至 9 時及晚間 7 時至 24 時兩個時段；反之，凌晨

1 時起至上午 6 時之間則為發生率最低之時段。

經上述以年、月、日及時段的時間次數分布與卡方適合度檢定分析結果，臺北市 2015 年至 2017 年間除月分布未達顯著差異外，其餘顯示臺北市的住宅竊盜在時間分布上有所不同。年分布方面，根據警政署統計，臺灣地區的住宅竊盜自 2014 年 5,479 件逐年遞減至 2018 年 3,682 件，5 年來減少 1,797 件，降幅達 32.80%(警政署統計室, 2019)，而臺北市 2015 與 2016 年住宅竊盜數量相差不多，但 2017 年比起 2015 年則下降 139 件，降幅達 22%，基本下降趨勢與全臺一致。

月分方面，分析結果雖無達到顯著水準，但與其他以臺北市住宅竊盜研究相比較均有相符之處，如本研究為 1、4、7、8 及 10 月為發生率較高月份，其他研究則指出 9 月、10 月為最高峰(劉擇昌, 2011)或 1、2、4、5 月為發生率較高月份(邱奕堯, 2013)；如文獻回顧所述，一方面造成月份差異原因可能因為研究時間範圍不同所致，另一方面也凸顯運用次數分布檢定的方式可能有所不足，故本研究積極主張應以時間群聚的概念另行檢定。

周分布方面，星期一、二、四、五為發生率較高星期別，反之星期六、日及星期三則較低，結果與其他研究結果相當類似，表明住宅竊盜易發生於民眾工作家戶監控降低之星期別；而在時間分布上，本研究結果亦與其他研究相當一致，一般工作時段為竊盜高發時間，而非深夜睡眠時間，此印證「日常活動理論」監控能力降低將導致犯罪發生之論述。

二、臺北市住宅竊盜空間分布概況

繪製犯罪地圖係探討犯罪是否具有特殊分布型態之視覺化初探性方法，故對於臺北市住宅竊盜的分析除統計分析外，另將以主題地圖呈現其空間分布情形；以下先以行政區進行資料探索分析，簡要概述臺北市各行政區住宅竊盜犯罪資料概況。

(一) 行政區分布

在行政區方面，2015 至 2017 年各行政區住宅竊盜犯罪數量總和以中山區最多，士林區其次，內湖區第三；反之，最少為大同區、南港區及信義區(如表 4-1-5；圖 4-1-5)。以卡方適合度檢定，檢定統計量顯示臺北市 2015 年至 2017 年住宅竊

盜犯罪發生在行政區別分布上有顯著差異（卡方值=137.075；期望值=146.6；自由度=11； $p < .05$ ），從觀察值與期望值分析可知，北投區、士林區、內湖區、中山區、萬華區、大安區 6 區，有較多的住宅竊盜發生數。

表 4-1-5 2015-2017 年臺北市住宅竊盜各行政區發生數

行政區 \ 年別	2015年 (次)	2016年 (次)	2017年 (次)	總和 (次)	排序
北投區	54	72	48	174	4
士林區	58	76	66	200	2
內湖區	56	68	52	176	3
中山區	91	77	56	224	1
大同區	25	31	29	85	12
松山區	55	42	32	129	8
南港區	18	36	33	87	11
中正區	49	45	41	135	7
信義區	47	37	32	116	10
萬華區	72	51	35	158	5
大安區	43	65	47	155	6
文山區	46	42	32	120	9
總和	614	642	503	1,759	
卡方值=137.075；期望值=146.6；自由度=11； $p < .05$					

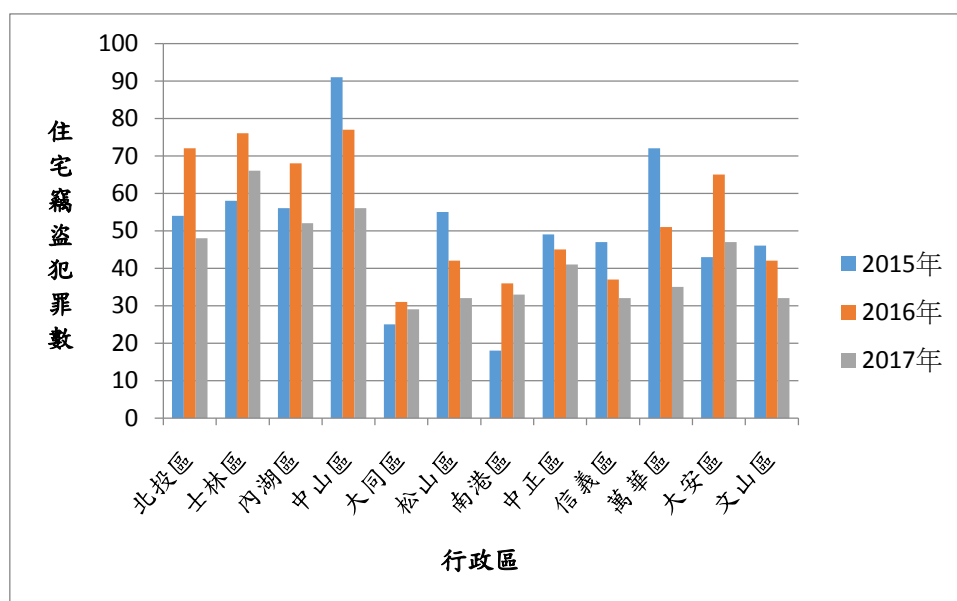


圖 4-1-5 2015-2017 年臺北市各行政區住宅竊盜數

從以上觀察得知，臺北市各行政區在住宅竊盜數有明顯差異，為能進一步瞭解相互間確實差異情形，以下進行單因數變異數分析(analysis of variance, ANOVA)，惟進行此分析之前提必須是自變數為類別變數(categorical variable)，依變數必須是連續變數(continuous variable)；此外，母群體必須是常態分布(Normal Distribution)，樣本須為獨立變項(independent variable)，變異數具同質性(homogeneity of variance)等條件，即各組的樣本不相互影響，且必須取自變異數相等的母體(王保進，2006)；以下針對上述條件逐步進行各項檢定。

1、常態分布檢定

首先，對臺北市各行政區住宅竊盜失竊率分別進行資料常態分布檢定，由於行政區樣本小於 50，故使用 Shapiro-Wilk 統計量來檢定常態性假設(王保進，2006)。檢定結果顯示，Shapiro-Wilk 統計量值為.970，未達.05 顯著水準，無法拒絕虛無假設，表示臺北市各行政區樣本在住宅竊盜數的數量的機率分配符合常態分布假設(如表 4-1-6)。

表 4-1-6 2015-2017 年臺北市各行政區住宅竊盜數常態分布檢定

	Shapiro-Wilk 常態性檢定		
	統計量	自由度	顯著性
數值	.970	12	.912

2、變異數同質性檢定

變異數分析的前提為假設各組內變異數相等，亦稱各組必須具備同質性，以提升變異數分析的品質效果。在變異數同質性檢定結果顯示，臺北市各行政區住宅竊盜數量經 Levene 法的 F 考驗結果，F 值為 1.204，P 值為.336，未達.05 顯著水準，表示各行政區在數量的變異數具有同質性，未違反變異數分析的基本假設(表 4-1-7)。

表 4-1-7 2015-2017 年臺北市住宅竊盜數變異數同質性檢定

Levene 統計量	顯著性
1.204	.336

3、ANOVA 分析

由表 4-1-8 可知，臺北市各行政區在失竊數的變異數分析結果 F 值為 5.003，顯著性達.05 顯著水準，表示各行政區住宅竊盜犯罪數之平均數有明顯差異。由於，事後檢定(post hoc)為檢定當多組樣本平均數有顯著差異時的後續步驟，採取兩兩配對方式，可更詳細觀察各組間顯著差異情形。

表 4-1-8 2015-2017 年臺北市住宅竊盜數量變異數分析

	平方和	自由度	平均平方和	F	顯著性
組間	6697.639	11	608.876	5.003	.000
組內	2920.667	24	121.694		
總和	9618.306	35			

4、事後檢定

由於臺北市各行政區住宅竊盜在數量平均數分析上呈現顯著；因此，以下進行事後檢定的多重比較，以說明各行政區差異(表 4-1-9)。

表 4-1-9 2015-2017 年臺北市各行政區住宅竊盜數多重比較

行政區	平均差異	顯著性	行政區	平均差異	顯著性		
北投區	大同區	29.667	.003	南港區	北投區	-29.000	.004
	南港區	29.000	.004		士林區	-37.667	.000
	信義區	19.333	.042		內湖區	-29.667	.003
			中山區		-45.667	.000	
			萬華區		-23.667	.015	
			大安區		-22.667	.019	
士林區	大同區	38.333	.000		中正區	士林區	-21.667
	松山區	23.667	.015	中山區		-29.667	.003
	南港區	37.667	.000				
	中正區	21.667	.024				
	信義區	28.000	.005				
	文山區	26.667	.007				
內湖區	大同區	30.333	.003	信義區	北投區	-19.333	.042
	南港區	29.667	.003		士林區	-28.000	.005
	信義區	20.000	.036		內湖區	-20.000	.036
	文山區	18.667	.049		中山區	-36.000	.001
中山區	大同區	46.333	.000	萬華區	中山區	-22.000	.022
	松山區	31.667	.002		大同區	24.333	.012
	南港區	45.667	.000		南港區	23.667	.015
	中正區	29.667	.003				
	信義區	36.000	.001				
	萬華區	22.000	.022				
	大安區	23.000	.017				
	文山區	34.667	.001				

大同區	北投區	-29.667	.003	大安區	中山區	-23.000	.017
	士林區	-38.333	.000		大同區	23.333	.016
	內湖區	-30.333	.003		南港區	22.667	.019
	中山區	-46.333	.000				
	萬華區	-24.333	.012				
	大安區	-23.333	.016				
松山區	士林區	-23.667	.015	文山區	士林區	-26.667	.007
	中山區	-31.667	.002		內湖區	-18.667	.049
					中山區	-34.667	.001

由表 4-1-9 各行政區住宅竊盜犯罪數量之平均數相互比較，基本上與表 4-1-5 結果一致，但以事後比較方式更能精確比較各行政區差異之顯著性。

(二) 里分布

由於臺北市有 456 里，表列各里犯罪數量將顯得相當繁雜，難以一目了然，故此部分將另以 GIS 製圖方式呈現；此前，對各里竊盜犯罪資料進行資料描述，以瞭解各里住宅竊盜分布概況。

2015 年至 2017 年臺北市各里住宅竊盜發生次數總和最高為 18 次，最低為 0 次，平均發生次數為 3.86 次，標準差為 2.777，各里累積次數與發生次數的分布統計如下(表 4-1-10；圖 4-1-6)。由圖 4-1-6 觀察，臺北市各里住宅竊盜多數落在 2 至 4 次間，超過 5 次後陡降幅度增大。

以卡方適合度檢定，檢定統計量顯示臺北市 2015 年至 2017 年住宅竊盜犯罪發生在各里別分布上有顯著差異（卡方值=755.735；期望值=4.1；自由度=428； $p < .05$ ），由觀察值與期望值分析可知，共有 148 個里高於期望值，數量約占臺北市 456 個里的三分之一，最高次數為 18，共計 2 次，均是中山區，最低次數為 0，共有 26 個里，散布於各行政區，亦即整體治安較佳地區仍有治安死角，此說明以行政區觀察住宅竊盜僅是趨勢觀察，並未能詳細反映臺北市住宅竊盜分布情形。

表 4-1-10 2015-2017 年臺北市住宅竊盜發生數、里數量分布統計

		里數量	百分比	累積百分比
發生數	0	27	5.9	5.9
	1	58	12.7	18.6
	2	73	16.0	34.6
	3	75	16.4	51.1
	4	75	16.4	67.5
	5	52	11.4	78.9
	6	33	7.2	86.2
	7	17	3.7	89.9
	8	16	3.5	93.4
	9	13	2.9	96.3
	10	7	1.5	97.8
	11	1	.2	98.0
	12	4	.9	98.9
	13	2	.4	99.3
	17	1	.2	99.6
	18	2	.4	100.0
	總和	456	100.0	

平均發生數3.86次；標準差2.777

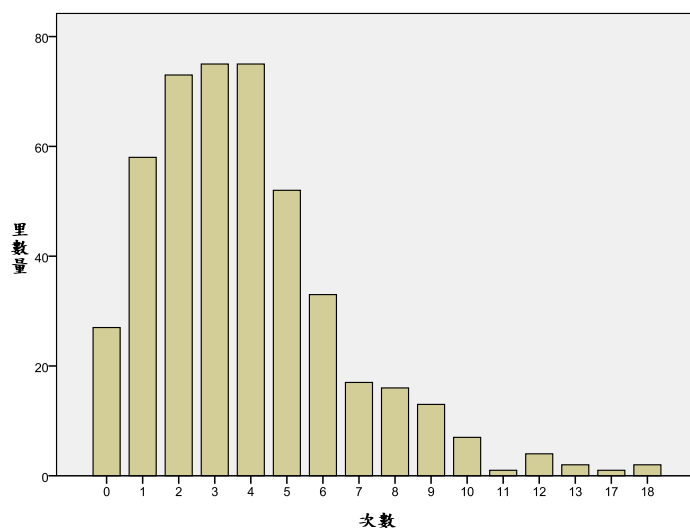


圖 4-1-6 2015-2017 年臺北市各行政區住宅竊盜數

(三) 臺北市各區里住宅竊盜犯罪率

以上各項分析呈現出 2015 年至 2017 年臺北市住宅竊盜數量分布情形；然而，由於各區住宅戶數密度不同，較多住宅的區域相對將有更多被害目標，因此，要評估住宅竊盜風險程度，應由犯罪率加以觀察；本研究以該區總戶數為分母，與住宅竊盜數量做比較，如「該區住宅竊盜戶數」/「該區總戶數」× 10,000=每萬戶失竊率。結果顯示，失竊率最高前 3 者為中山區、萬華區及士林區；反之，失竊率最低為文山區、大安區及信義區（如表 4-1-12；圖 4-1-8）

表 4-1-11 2015-2017 年臺北市住宅竊盜各行政區失竊率

行政區 \ 年別	2015年 (率)	2016年 (率)	2017年 (率)	平均 (率)	排序
北投區	5.64	7.47	4.96	6.02	5
士林區	5.46	7.12	6.17	6.25	3
內湖區	5.26	6.34	4.82	5.47	8
中山區	9.18	7.73	5.59	7.49	1
大同區	4.85	6.01	5.60	5.49	7
松山區	6.85	5.23	3.97	5.35	9
南港區	3.88	7.69	7.01	6.20	4
中正區	4.94	6.92	6.30	5.89	6
信義區	7.20	4.13	3.58	4.75	10
萬華區	9.19	6.49	4.45	6.71	2
大安區	3.57	5.39	3.89	4.29	11
文山區	4.41	4.00	3.03	3.81	12

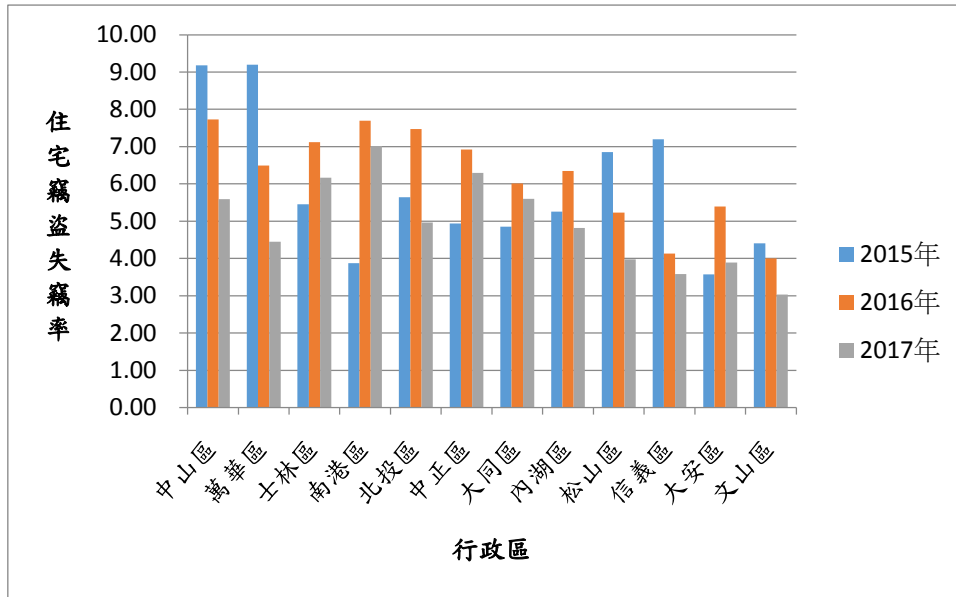


圖 4-1-7 2015-2017 年臺北市各行政區住宅竊盜失竊率

從失竊率的排序來看，中山區在失竊數量上排序第一，同時在失竊率亦為第一，兩項統計均顯示中山區確為臺北市住宅竊盜風險最高區域，此與本區出租套房多、商圈多元，出入複雜可能有相當關係；而觀光客群聚的商圈萬華區則由數量上排序第五，躍升為失竊率排序第二的區域。相對地，原本在數量上排序第三的內湖區，則在失竊率排在第 8 位，反而成為住宅竊盜風險相對較低的區域，而文山區則在竊盜數量與失竊率方面均為最低，可能與此地區為傳統文教區有所關聯。

以里為單位的住宅竊盜失竊率方面，2015 年至 2017 年臺北市 456 里的平均失竊率為每萬戶 5.85 戶，其中有 185 個里超過平均數，分別為中山區 23 個、萬華區 22 個、士林區 22 個、北投區 23 個、中正區 17 個、大同區 9 個、內湖區 15 個、松山區 7 個、大安區 12 個、信義區 14 個、南港區 11 個、文山區 9 個。各里住宅失竊率高於平均數占其行政區比例如表 4-1-13。

表 4-1-12 2015-2017 年臺北市住宅竊盜各里失竊率占比

行政區	總里數	高於平均 (里數量)	百分比 (%)	排序
北投區	42	23	54.8	4
士林區	51	22	43.1	6
內湖區	39	15	38.5	7
中山區	42	23	53.5	5
大同區	25	9	36.0	8
松山區	33	7	21.2	11
南港區	20	11	55.0	2
中正區	31	17	54.8	3
信義區	41	14	31.4	9
萬華區	36	22	61.1	1
大安區	53	12	22.6	10
文山區	43	9	20.9	12

從表 4-1-12 分析，在高於平均數失竊率的里所占比例，以萬華區排第一，南港區居次，中正區第三、北投區第四，而失竊數與失竊率均為第一的中山區反而降為第五，其餘依次為士林區、內湖區、大同區、信義區、大安區，文山區依然為高失竊里數占比最低的行政區。

(四) 臺北市住宅竊盜犯罪數據主題圖

地理界線區域主題圖(thematic mapping of geographic boundary area)或面量圖(choropleth mapping)為最常被運用於展示犯罪資料空間分布的製圖技術，此類主題圖之地理界線可依行政或其他需求而定義，如行政界限(村里界)、警勤區等，犯罪事件在每個地理界線中依據研究單元不同而以整合或加總數據呈現每個單元中之數量，可迅速瞭解那些區域中具有高犯罪事件，並進一步決定下階段的研究方向(劉擇昌、張平吾，2014)。

在上述犯罪資料的統計分析中，我們大致瞭解臺北市各行政區與各里住宅竊盜犯罪數量與失竊率分布情形，以下主題圖即是使用這些統計資料，以呈現臺北

市在不同尺度範圍的住宅竊盜視覺化分布概況。

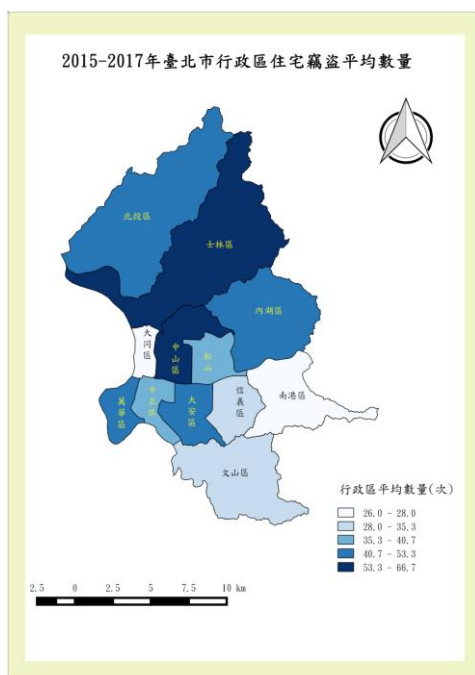


圖 4-1-8 2015-2017 年臺北市各行政區住宅竊盜失竊數主題圖

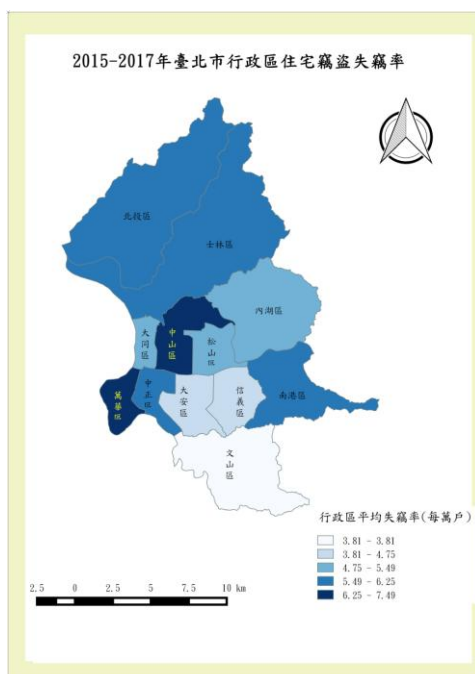


圖 4-1-9 2015-2017 年臺北市各行政區住宅竊盜失竊率主題圖

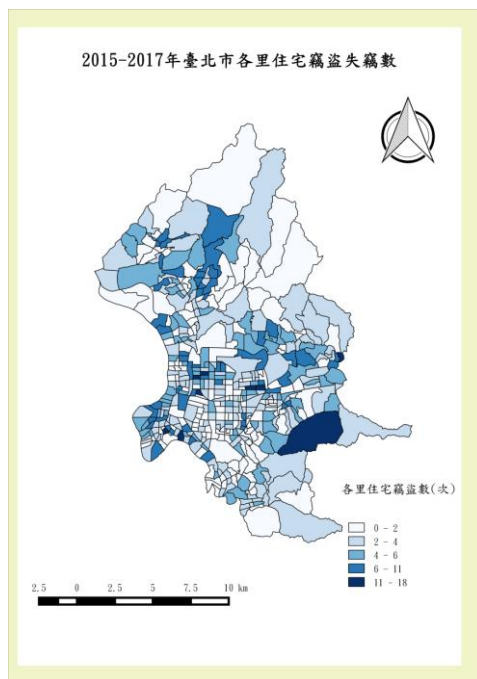


圖 4-1-10 2015-2017 年臺北市各里住宅竊盜失竊數主題圖

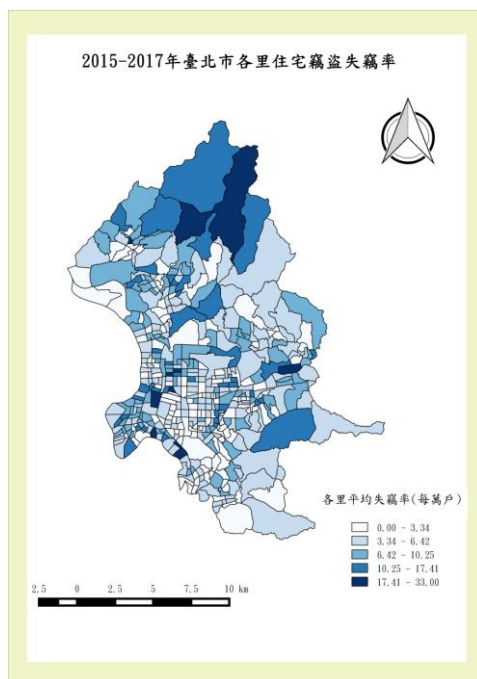


圖 4-1-11 2015-2017 年臺北市各里住宅竊盜失竊率主題圖

(五) 小結

本研究針對臺北市 2015 年至 2017 年住宅竊盜的犯罪資料分別以行政區(里)

不同尺度範圍進行數量與犯罪率的統計分析，最後並以 GIS 繪製相關主題圖，初步分析結果顯示，若以住宅竊盜發生次數進行分析，則在行政區方面中山區、士林區及內湖區為前 3 位，而大同區、南港區及信義區為相對較少後 3 位，此與民眾對於各行政區的治安印象似有落差，但在考量各行政區住宅數量後，改以平均失竊率衡量，則呈現截然不同結果，中山區失竊率依然排序第一，但萬華區攀升至第二位，大安區與文山區則分居失竊率最低的 11、12 位，此結果對於民眾而言，應較符合一般預期，而此說明犯罪標的物的多寡與犯罪數量有重要關聯。而在里的分布方面，臺北市 456 里在住宅竊盜數量上有極大落差，有 27 個里在本研究時間範圍內，並無任何住宅竊盜發生，多數的里均落在 2 次至 4 次之間，最高為 18 次，顯示臺北市各里住宅竊盜犯罪在數量上呈現不均勻分布。

以上以行政區與里不同尺度進行分析，產生不同研究結果的差異，除受住宅數量此變項的影響外，主要的問題在於「可修正區域單元問題」(modifiable areal unit problem, MAUP)；此問題包括兩種層面，一是「區劃效應」(zoning effect)，即不同的劃分方式或比例尺將使相同資料在視覺上產生不同效果；另一為「尺度效應」(scale effect)，即不同的衡量尺度使相同資料產生不同結果(劉澤昌、張平吾，2014)；例如，本研究所使用之資料，分別以區里分析，有相異的解釋，同時在 GIS 的主題圖也清楚看出之間的差異，而以較大面積的行政區主題圖觀察住宅竊盜的分布，並無法區隔行政區內風險差異，容易產生整個行政區都處於受到相同犯罪風險的狀態，從而可能導致錯誤的推論。

整體而言，透過本研究犯罪資料統計分析，可針對研究區域住宅竊盜犯罪概況進行初步瞭解，各項檢定也說明研究資料性質與臺北市各行政區的差異，但對於更小範圍的里單元而言，若進行統計分析，要詳細說明各里間的差異將過於繁瑣，更重要是以統計分析並無法呈現研究區域內住宅竊盜犯罪之確切地點、熱區、時空變遷趨勢等，對擬定犯罪防治對策與工作有助益有限。故下一節中同樣以此資料為基礎，將各年度住宅竊盜犯罪地點予以地理編碼後，並根據每筆資料的發生時間，進行不同尺度的犯罪空間、時間與時空群聚分析。

第二節 竊盜犯罪資料空間探索分析

一、空間統計概述與資料類型

在上節中本研究進行犯罪資料的一般性統計探索分析，對於臺北市各區里的整體住宅竊盜犯罪情形有初步瞭解，在本節中將使用經地理編碼後的犯罪資料對於具有空間定位資訊之屬性資料以視覺化方式加以呈現，並進一步探討空間分布與群聚（犯罪熱區）等相關問題。由於原始資料經地理編碼後與空間產生連結，與一般統計資料性質有所不同，稱為空間統計資料，主要可分為 3 種型態(Cressie, 1993)：

- (一)地理統計資料(geostatistical data)：在地質統計資料中有一個固定的研究區域，在每個位置都有一個觀察到的或未觀察到的隨機變數。例如：該區域可以是美國，隨機變數可以是 2004 年的降雨量。
- (二)格點空間資料 (spatial lattice data)：在研究區域內有一組固定（非隨機）的點集，每個點上定義一個隨機變數；例如，格子可以是美國各縣的中心點，隨機變數可以是該縣患有哮喘兒童的數量。本研究資料為臺北市 456 個里 3 年間住宅竊盜資料，因使用掃描統計，須以各里中心點為基準，隨機變數為各里住宅竊盜數量，故屬於格點空間模型。
- (三)空間點分布型態(spatial point pattern)：觀測的空間位置是隨機的，而這些位置上的觀測若非等於 1，則其本身就是隨機變數；例如，空間格局可以是鳥窩在島上的位置，隨機變數可以是鳥窩中鳥蛋的數量。此模型之隨機機制在於位置本身，其中最基本的空間點過程為「同齊卜瓦松」(homogeneous poisson)點過程，主要假設有以下兩種：1、在空間中任一集合所包含事件之個數為卜瓦松分布，且其分布之平均值與其集合面積（體積）成正比；2、不相交之集合所包含事件之個數呈統計獨立關係(葉茗宏，2013)。值得注意的是，如果資料為空間點分布型態，但條件是對觀察到的隨機點位置進行分析，那麼資料實際上是格點數據，因為空間位置不再是隨機的(Kulldorff, 2006)。

對於空間資料進行統計分析稱為「空間統計」(spatial statistics)，主要思維在於空間中鄰近的資料通常比相離較遠的資料具有較高的相似性，透過位置建立資

料間的統計關係。與一般統計主要差異為：(一)由於空間資料之間並非獨立，而是在 d 維度的空間中具有特殊空間相關性，故不同的空間對於相關性問題將產生不同影響；(二)大多數空間問題僅有一組觀測值（不規則在空間分布），且無重複觀測的資料。因此，空間現象的瞭解與描述較傳統的統計分析更為複雜，尤其是建立在獨立樣本的統計方法，並不適合用來分析空間資料，其與時間序列最大的差異在於空間中並無過去、未來之次序，因而不易透過某種因果關係的描述來建構空間模型(黃信誠，2000)。

空間統計有如一般統計，範圍相當廣泛，其中一個重要問題是檢驗是否在對已知空間不均勻進行調整後，空間分布上是否隨機分布。在本研究中，空間不均勻的測試過程乃是測試臺北市住宅不均勻分布的情況下，住宅竊盜犯罪是否在某一地理區域內隨機分布進行評估。以下首先針對本研究資料地理編碼後，進行空間探索性分析。

二、空間探索分析

空間探索性分析 (exploratory spatial data analysis, 簡稱 ESDA) 之目的在於透過如 GIS 等空間統計工具，探索各種變數的空間分布模式；通常在犯罪學上的空間探索主要係透過製作犯罪主題圖，利用主題圖呈現出某些特定地區在特定議題的集中趨勢，而「空間自相關」(或稱空間相依性)與「空間異質性」為探討空間因素最重要的兩個概念，所謂「空間自相關」表示地區之間的社會現象並非獨立存在，而是與周遭區域透過空間互動，和鄰近空間單元之社會現象有所關連，而形成空間自相關的情形，經濟學的空間外溢效應(spatial spillovers effect)與空間外部性(spatial externality)、社會學的鄰近效應(neighborhood effect)以及流行病學的空間擴散(spatial diffusion)都是屬於「空間自相關」的情形(周孟嫻、紀玉臨、謝雨生，2010)。

對於「空間自相關」的測量方面，簡單地說就是量度「目標」與「鄰近地區」的相關性，通常可透過計算鄰近屬性的相關性與相互距離，進一步測量此地理資料的空間相關指標，該統計值可以反映該現象的正、負或無相關的各種關係，如結果為正空間相關，表示地理屬性資料呈現群聚的分布型態，反之則是離散的分布型態。在 1940 年代末期，澳洲學者 Partick Moran 發展出 Moran' I 的度量指標，利用計算鄰近屬性數值與平均數的差距，以量測空間鄰近相似程度，隨後許多測

量空間鄰近與群聚的指標如 Rober Geary 的 Geary's contiguity ratio(Geary's C)等相繼推出(溫在弘, 2015);目前對於「空間自相關」的檢定主要可分為全域式空間自相關(global spatial autocorrelation)與局部式空間自相關(local spatial autocorrelation)兩大部分,前者以 Moran' I 較常被使用,但 Moran'I 值之量測僅能明瞭屬性相似的單元間是否呈聚集狀態與強度,無法藉由該數值瞭解空間聚集區的分布型態,故在全域分析後,若顯示具有相關性,仍需進行局部的檢定,以發掘地區內空間群聚的分布位置。

有關局部檢定的統計方有許多方式,其中由 Anselin (1995)提出 LISA 統計分析,稱為 LISA 指數(Anselin's LISA index),其主要方法係對 Moran' I 去除不穩定的值,或相似值的取捨進行改良,透過直接計算屬性值的差異比較相鄰地物的屬性值是否相近,做為瞭解局部空間自相關分析或稱區域空間自相關指數,所以可視為局部的 Moran' I(林宜甲、黃柏霖, 2017);除 LISA 指數外,Local G-sraticistic 統計分析方法也具有相當代表性。

綜上,無論是全域或局部檢定,主要目的就是找出區域之間的相似性,也就是所謂的「鄰居」所在,但須注意資料上的限制,通常若以週圍有接壤者判定為鄰近,在地理單元必須是面資料或圖層,否則無法從邊界是否銜接判斷,而如果是點資料,無法判斷是否接壤,一般會用距離判斷,亦即一定距離內的各點都叫做鄰居,如果是面圖層須用距離計算,此時是以面圖層的核心點計算,亦即把面圖層當成點圖層來處理。以下本研究以臺北市住宅竊盜點位資料進行各項空間探索分析,除呈現點位資料的主題圖外,第一要務在於確認是否犯罪點資料在距離上是否具有空間群聚的現象,俾於後續局部群聚分析的相關問題探討。

(一) 點位圖(dot map)

繪製犯罪統計地圖係探討犯罪之特殊分布型態之視覺化初探性方法,較基礎的主題地圖有點位圖、等值線圖及面量圖等。在上一節中,圖 4-1-9 至圖 4-1-12 各主題圖,即是屬於將犯罪資料加總後的面量圖,其基本假設地理現象在各區域內為均勻分布,以不同等級、不同顏色來表示區域間的數值大小,反映出區域間的差異。然而,在現實社會中,犯罪發生地點絕非均勻分布,各地犯罪數量與密度均不相同,此時使用點位圖可用於顯示分布密度,透過此種製圖方式,研究者能觀察地理點位的疏密程度,判讀犯罪現象的空間分布情形,但此類表達方式僅能直觀地認定空間型態,無法準確辨認出群聚位置;因此,必須再進一步運用空

間統計方法，找出具有統計意義上的空間群聚。

圖 4-2-1 為本研究根據 2015 年至 2017 年臺北市住宅竊盜地理編碼資料，以 GIS 繪製的各年度住宅竊盜點位圖，每一點代表犯罪發生地點。從各年度分別加以觀察，可發現儘管各年度在住宅竊盜犯罪數量有多寡區別，但在落點與密度的分布趨勢頗為類似，而從 3 年加總的點位圖觀察，雖然尚未進行群聚的統計檢定，不難發現若干地區的犯罪發生地點呈現十分密集的現象。例如，中山區有較高的住宅竊盜發生數，但在面量資料中只能以顏色深淺表示其數值，在點位圖中即可看出較多的住宅竊盜集中在部分區域，仔細觀察可發現其與士林區鄰近地區，就少有住宅竊盜發生，而位處臺北市北邊的北投區及士林區，愈趨北方罕有住宅竊盜犯罪，主要原因乃是此區域為陽明山國家公園等，並非住宅密集地區。

總之，點位圖係利用圖像的方式呈現密度高低感受，而非呈現密度的數值，雖然直接將點資料繪製於地圖對於直接觀察犯罪概況十分便利，但若資料太過密集時，將可能導致資料點位重疊，此時可以先將原始資料加總至面資料，再用點位圖重新繪製，如此便可同時呈現各地的密度高低，也能調整資料重疊的問題(溫在弘，2015)。

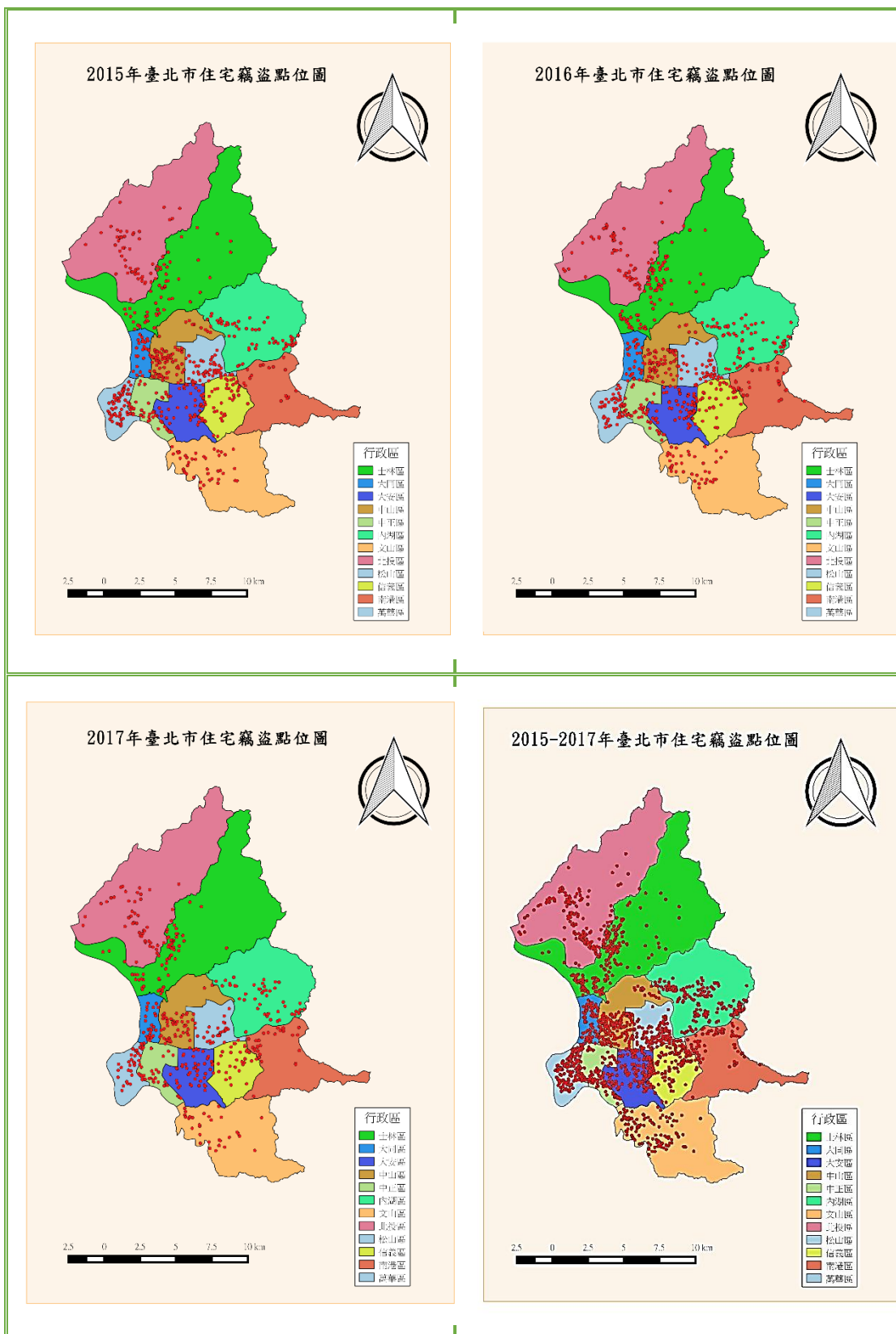


圖 4-2-1 2015-2017 年臺北市住宅竊盜點位圖

(二) 熱區圖(heat map)

熱區圖通常使用在計次的空間分布分析，最適合用來呈現高密度的點資料，它可用於識別點資料中的高度密集區域。圖 4-2-2 左邊的熱區圖係由 2015 年至 2017 年點資料以「最近鄰分析」(nearest neighbor analysis, NNA) 計算座標距離的統計分析結果繪製而成，並非僅是計次的熱區圖，圖中顏色深淺，代表群聚的強度，觀察可發現中山區與萬華區為顏色最深的行政區，故似可初步判斷此兩個行政區最有可能成為群聚強度較高地區。

圖 4-2-2 右邊則是以 Google Map 做為底圖，可進一步識別熱區所在的實際街道位置，對於一般民眾而言，相當易懂，如臺北市公布住宅竊盜熱區圖，即是以類似方式製作。熱區圖與點位圖基本上均是以視覺方式表達犯罪分布型態的初探性方法，若要進一步對於犯罪空間群聚現象進行探討，仍是必須使用局部群聚空間統計分析方法，才能獲致較確切與科學化之結果。

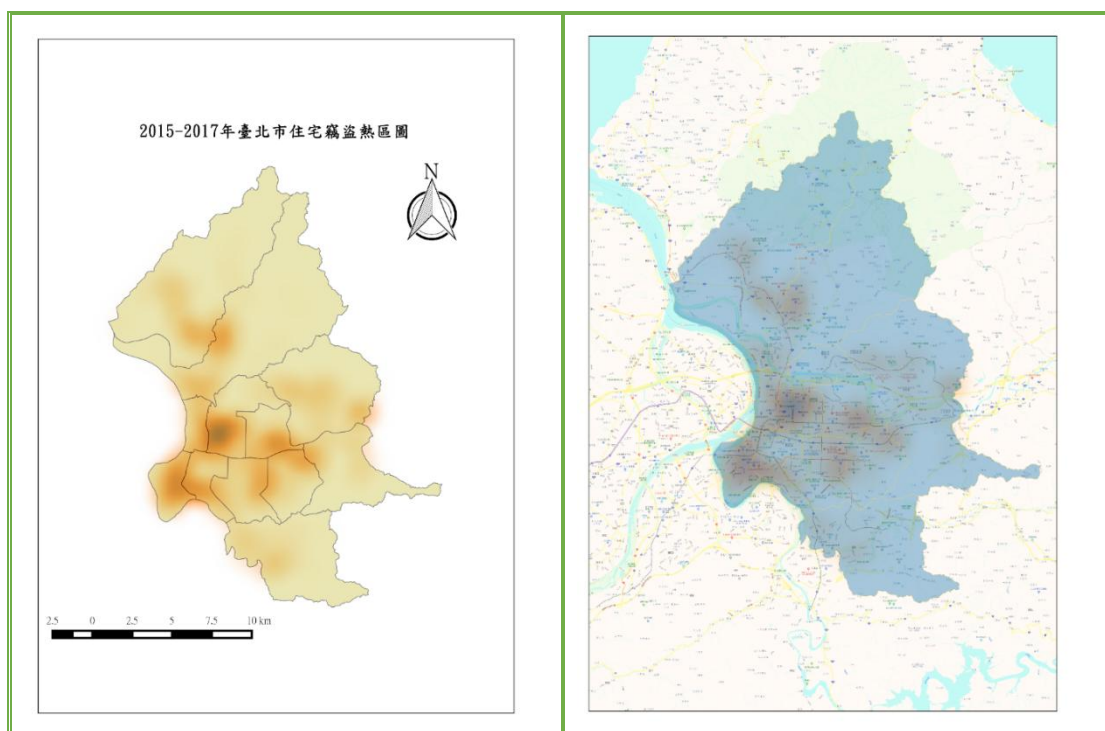


圖 4-2-2 2015-2017 年臺北市住宅竊盜熱區圖

(三) 犯罪點資料全域式群聚分析

以點資料型態的空間聚集分析，目的在於計算資料是否顯著聚集趨勢及其所

在聚集地點；一般而言，係透過統計顯著性檢定評估其空間聚集趨勢是否顯著，其統計檢定的虛無假設為空間的點資料呈現「完全空間隨機(complete spatial randomness, CSR)」，而常用來分析犯罪地點的空間群聚統計方法包括「最近鄰階層群聚」及「核密度推估」等(溫在弘等，2010)。本研究使用 QGIS 軟體，以「最近鄰分析」(nearest neighbor analysis, NNA)針對 2015 年至 2017 年臺北市住宅犯罪點資料，進行全域式群聚分析，其原理係計算一個地物與最近地物的距離，累計所有地物的最近鄰距離取平均值，再與理論平均最近鄰距離相比較，(翁培文、蔡博文，2006)，相較後得到比值稱之為「最近鄰指數」(nearest neighbor index)。

將本研究住宅竊盜犯罪點資料載入 QGIS 分析後，臺北市住宅竊盜全域式空間群聚分布情況如表 4-2-1，表格數據顯示，各年度最鄰近指數均介於 0~1 之間，Z-Score 負值表示有群聚現象、正值表示離散現象，數字越大則現象越明顯；據此說明，在單一年度方面，2016 年臺北市住宅竊盜犯罪全域式群聚現象比起其他年度最為明顯，而從原始資料可知，2016 年住宅竊盜的件數也同時較其他年度為高，也因此說明在固定面積單位中，地物數量越多，在不均勻分布情況下，其產生群聚情況的可能性增加。

表 4-2-1 2015-2017 年臺北市住宅竊盜犯罪全域式群聚分析表

年別	觀測的平均距離 (公尺)	預期的平均距離 (公尺)	最鄰近指數	Z-Score
2015	220.6415	386.8705	0.5703	-17.5723
2016	215.1420	425.9065	0.5051	-20.2382
2017	251.6870	439.8559	0.5722	-15.5497
2015-2017	117.0700	254.9866	0.4591	-36.9475

(四) 小結

從犯罪點資料的最近鄰分析的空間探索分析，可知臺北市住宅竊盜在全域性檢定上具有群聚的現象，配合點位圖及熱區圖同時可觀察出犯罪熱區分布的位置，基本上對於一般居家防護宣導已具實用性；然而，根據前述，由於全域性檢定無法瞭解地區間更詳細的分布型態，對於警政機關而言，若要進一步介入排除並擬定防治策略，資訊仍有不足，故在確認具有全域性群聚後，局部性檢定實屬

必要步驟，才能進一步確認各地區詳細群聚資訊。

目前，對於局部性檢定的方法，如上述以 LISA 較為普遍，在犯罪學領域中被廣為採用，而 Local G-sraticistic 統計分析方法亦可分析地物的高屬性值與低屬性的分布特性，亦即會對每一個地物進行分析，觀察並比較其屬性值與給定距離範圍內鄰近地物的屬性值；換言之，局部檢定的統計方法可進一步區分該地物是否被同樣高或低的物所包圍，並會在各地區分別計算統計值，找出不同顯著程度的冷區與熱區，若以統計值繪製地圖後，即可一目了然觀察出熱區與冷區的群聚位置(溫在弘，2015)，目前常用的 GIS 軟體多內鍵局部性檢定功能，可在讀取資料後直接分析，並繪製地圖。

針對 Moran' I、LISA 與 Local G-sraticistic 等檢定在當前研究中較為常見的現象，本研究認為與 GIS 軟體使用便利性有很大關聯；例如：付費軟體 Arcgis 操作簡單，功能強大，學習門檻較低，很快成為空間分析的主流，而免費軟體如 Geoda、QGIS 等，在操作與功能上雖相較遜色，但因其無須付費，學習者亦日益增多。事實上，對於群聚檢定，學界另有許多專業檢定方法，在本研究文獻回顧中所提到的 Openshaw(1988)、Turnbull(1990)等人以固定圓形窗口掃描，再透過 monte carlo 檢定是否顯著，Besag 和 Newell(1991)以累積至 K 個事件數所需最少的鄰近區塊數作為統計量等，均為優異的局部檢定的方法。而本研究採取的 Kulldorff 等人(1995)的掃描統計方法即是融合 Openshaw 的「地理分析機」與 Turnbull 的(cluster evaluation permutation procedure, CEPP)兩種方法的優點，有極佳的檢定力(蔡丞庭，2011)，缺點是搭配此掃描統計的 SatScan 無論在資料的輸入與輸出都有較高的要求，甚至因時空的運算複雜，對電腦硬體要求也相當高，在 SatScan 的操作手冊中甚至獨列單元，介紹電腦規格與如何節省運算時間的方法，故在目前的群聚研究中，包括犯罪學領域較為少見，最多是運用在流行病學的疾病群聚與防疫工作的偵測或預警，由於醫學領域的研究事涉人類生命安全，故此項方法對於群聚偵測的要求相當嚴謹，也特別注意到群聚的時間性問題，以便能提前發出預警；因此，但此種統計分析方法不但對區域內熱區及冷區的群聚判斷相當敏感，即使是案例數非常稀少，也可進行偵測，同時也依據受風險人口的相對風險性進行群聚的分級，與 Local G-sraticistic 等分析方法在空間局部性檢定有異曲同工之處，但其能同時對時間與空間進行同步分析，則非目前一般 GIS 軟體所能提供，故下一節有關臺北市住宅竊盜犯罪的局部群聚檢定，將以掃描統

計進行。

第三節 空間掃描統計分析

對於犯罪群聚是否發生，很重要的問題在於群聚為偶然爆發，還是由於某些常見、長期的風險因素，如地理、人文或甚或時間因素所造成。對於偶然發生的犯罪群聚(如夜店鬥毆等)，警力立即介入通常可以有效排除，而對於持續性或新興的群聚，則是警政機關必須澈底詳細瞭解，查明危險因素為何，進而擬定具體的犯罪防治政策。然而，如何判定此犯罪群聚為偶然、持續或新興性的群聚呢？

判定犯罪事件發生是否為群聚，有許多不同的方式，在上一節中所提到的局部檢定已提供具體方法，其判斷標準在於假設犯罪為隨機發生，再經由檢定，查驗其顯著水準加以判斷是否具統計意義；另有些非正式科學的方式則是將某群聚區域與更大的區域進行差異性比較，從而認定此群聚是否具有風險，例如將臺北市某區里與全臺北市進行比較，以觀察是否具統計上差異性。本研究認為，任何地理區域都包含一些犯罪發生機率較高的地區，可能也是隨機發生的，對於所蒐集之資料，如果只考慮空間因素，對於某地區有較高數量的犯罪數量經統計檢定後接受（或拒絕）為具統計意義群聚，可能有欠周詳；換言之，本研究認為對於群聚的判定若忽略時間趨勢的因素，僅以當下所蒐集之資料判斷，可能會導致統計上的選擇偏差，無法瞭解整體趨勢，導致警政策略的擬定可能均僅是對隨機事件進行分析，造成無謂資源投入。故除對空間群聚進行檢定外，仍需進一步將時間因素加以考慮，目的除力求尋找較正確的群聚地區外，更重要是希望能發現過去未曾注意到的群聚現象與特性。在本節中，首先對臺北市住宅竊盜犯罪進行空間分析，其次為時間分析，第三部分則是綜合時間與空間交互作用的時空分析。

一、資料輸入與參數設定

以掃描統計進行群聚偵測包含資料輸入、參數設定及判斷群聚是否顯著等階段，以下分別說明：

(一) 資料輸入

本研究使用 SaTScan 9.6 進行時空掃描統計，用於掃描統計分析的資料可以是個體層面(individual-level)資料，也可以是聚合層面(aggregated-level)資料，個

體層面資料記錄每個事件發生的位置、時間、屬性等，聚合層面資料是按單元(如行政單元、人口統計單元、郵遞區號單元、網格單元等)記錄每個單元內事件數(或按時間等分別記錄事件數)(Kulldorff, 2018)。本研究使用分組層面型態資料，包含 3 種個別檔案：

- 1、事件檔案(case file)：檔案提供所有住宅竊盜詳細的資訊，包含各年度每一筆住宅竊盜發生時間、區域代碼，並統計其總數；若進行群聚的共變項分析，在此檔案中可將變項另行加入。
- 2、母群體檔案(population file)：檔案提供研究區域內不同時間的母群體數量，本研究母群體檔案為臺北市各里住宅戶數。在共變項分析中，此檔案也必須輸入與事件檔案相同之變項。
- 3、座標檔案(coordinate file)：檔案主要提供分析地區的座標資訊。本研究的分析檔案為臺北市各里中心點經緯度座標，但橢圓形視窗分析則需要轉換為 X-Y 座標。

(二) 參數設定

資料檔案輸入後，進行參數設定，包括分析模型、掃描視窗大小、蒙地卡羅檢定次數、顯著性 P 值範圍等；說明如下：

- 1、分析模型：本研究根據研究資料特性採用離散型「卜瓦松模型」進行分析。
- 2、掃描視窗：掃描視窗有風險人口比率與掃描半徑兩種選項，可以同時加以設定；若以指定風險人口比率進行掃描，事實上也等於間接指定最大掃描半徑。在進行掃描時，改變視窗的大小實際上是查找鄰近點的過程，最常用的距離是歐氏距離，亦可經由附加檔案，指定為其他測量距離方式；掃描視窗的形狀一般情況下為圓形，第一個視窗即中心點本身，掃描半徑為 0，第二個視窗的半徑是中心點與最鄰近點的距離，第三個視窗的半徑是中心點與次鄰近點的距離，依次類推，直到掃描距離達到設定的最大值。本研究基本上依據 Kulldorff 建議，將掃描視窗定為 50% 的風險人口，掃描的半徑將會覆蓋全區域，並將群聚設定為不重疊的「吉尼群聚」(gini clusters)。

3、蒙地卡羅假設檢定：本研究蒙地卡羅檢定設定為 999 次，顯著性透過蒙特卡羅模擬進行評估，一般而言，在 P 值為.05 的情況下，如果最可能的群聚的模擬 P 值小於或等於.05，則拒絕臺北市住宅竊盜犯罪隨機分布的原假設，判定有群聚發生；惟對於 P 值小於或等於.05 情況下拒絕虛無假設的判定，只是統計上的推論，事實上仍有可能發生誤判的機率，對此統計學家根據不同錯誤情況歸納出型 1 錯誤(Type I Error)與型 2 錯誤(Type II Error)兩種，型 1 錯誤係指當虛無假設為真，但卻遭到拒絕，此類錯誤的嚴重性較低，例如，對臺北市住宅竊盜分析設定虛無假設為犯罪的分布為隨機分布，此時若犯下型 1 錯誤，誤判竊盜犯罪為群聚情形，可能將是造成警政資源的錯誤配置與浪費，但若是犯下型 2 錯誤，則是竊盜群聚發生，司法警政單位卻無法察覺，此對於人民生活保障之危害性更甚於前者，故有學者建議，在空間分析中根據議題的性質，可將顯著水準定在 10%(溫在弘，2015)。

此外，應用掃描統計時，除偵測到最有可能的第一群聚外，也可能探測到其他次要的第二、三等群聚，並且根據概似率的大小對它們進行排序。有時某些次要的群聚與最大可能的群聚幾乎有同樣大的概似率數值，此乃因為對於較大或中等大小的群聚，稍微增大或縮小它的邊界並不會使概似率發生太大變化，而且群聚可能存在重疊的現象；一般而言，這類型的次要群聚所能提供訊息較少，但它們的存在意味著掃描統計的分析結果通常只能指出群聚的約略邊界，無法肯定所有的邊界確切無誤，但若次要的群聚，其邊界跟最大可能的邊界並不重疊，這樣的群聚就比較有意義。在蒙地卡羅假設檢定中，次要群聚的概似率係以最可能群聚的數值相比較以確定其 P 值，但此將導致次要群聚的 P 值估計會較為保守，此亦為可將群聚的統計顯著水準定為 10%的另一原因(Kulldorff et al., 1998; Odoi et al., 2004)。

基於以上，本研究原則仍採.05 的顯著水準，但對於 P 值超出.05 但小於.10 之群聚，也將特別予以觀察與標註，期能發現更多潛藏的可能群聚。在資料輸入與基本參數設定後，以下進行掃描統計的各類群聚分析，有關各種群聚類型的偵測參數調整，將在各類型分析時提出說明。

二、空間掃描

SatScan 掃描軟體在空間掃描的方式上可分為圓形掃描視窗(circular scanning window)與橢圓形掃描視窗(elliptic scanning window)兩種，通常以圓形掃描居多，而橢圓空間掃描統計量在範圍不是很大時表現良好，在某些情況下可以對真實的集群區域做出更好估計；相反地，對於非常大的形狀，橢圓空間掃描因選擇一個狹長的區域，位於此區域的許多集群可能被排除在外，所產生的群聚不見得令人滿意，例如所有的群聚只集中在一個狹長的行政區。但 Kulldorff 曾指出，無論最可能的群聚形狀如何，任何群聚掃描只是偵測出群聚的大致區域，檢測到的群聚的確切邊界仍是不確定，如果計算資源允許，有時可以使用橢圓掃描統計量而不單只是圓形掃描統計量，或許可獲得更好的結果(Kulldorff et al., 2006)，換言之，使用橢圓形的掃描視窗較圓形掃描視窗可能更能判別出異常空間群集，此特性尤其在異常範圍不甚大時更為顯著。

由於本研究之範圍為臺北市，相較許多幅員廣大的掃描統計研究，範圍不算太大，有鑑於 Kulldorff 之建議，本章節的空間掃描將首先以圓形視窗掃描，其次進行橢圓形掃描，比較兩種掃描視窗之差異，期能發現潛藏的群聚區域，並針對檢測的群聚加以歸類，初步劃分出群聚的性質。

(一) 圓形視窗空間掃描

1、2015 年圓形空間掃描結果

在空間掃描方面，本研究以圓形視窗共掃描臺北市 456 個里，10 個群聚區，其中 3 個群聚區達到 $P < .05$ 統計顯著，而第四個群聚區的 P 值為 .081，小於 10%，故可列入參考。第一個最大可能的群聚區(first most likely cluster)幾乎全部集中在中山區，群聚中心點座標在中山區恆安里(25.063737 N, 121.524974 E)，包括中莊里、恆安里、晴光里、集英里、新生里、新福里、聚盛里及萬華區新安里，群聚半徑為 0.74km，對數概似率(LLR)為 15.94，相對風險值(relative risk)為 3.17， $p\text{-value} < 0.01$ ；第二個可能的群聚區包括萬華區及中正區 2 個地區，群聚中心點座標在萬華區榮德里(25.019071 N, 121.491919 E)，群聚半徑為 2.70km，對數概似率為 15.81，相對風險值為 1.99， $p\text{-value} < 0.01$ ，其中萬華區包括 28 個里為最多，其次為中正區 8 個里；第三可能群聚區為松山區安平里，群聚中心點座標則

為此里的中心位置(25.052568 N, 121.567048 E)，群聚半徑為 0km，對數概似率為 9.46，相對風險值為 5.36，p-value < 0.05；第四可能群聚區為士林區與北投區，p-value<.10，群聚中心點座標位於北投區湖田里(25.179213 N, 121.547572 E)，群聚半徑為 5.43km，包含 8 個里，對數概似率為 8.16，相對風險值為 3.75 (表 4-3-1，4-3-2，圖 4-3-1)。

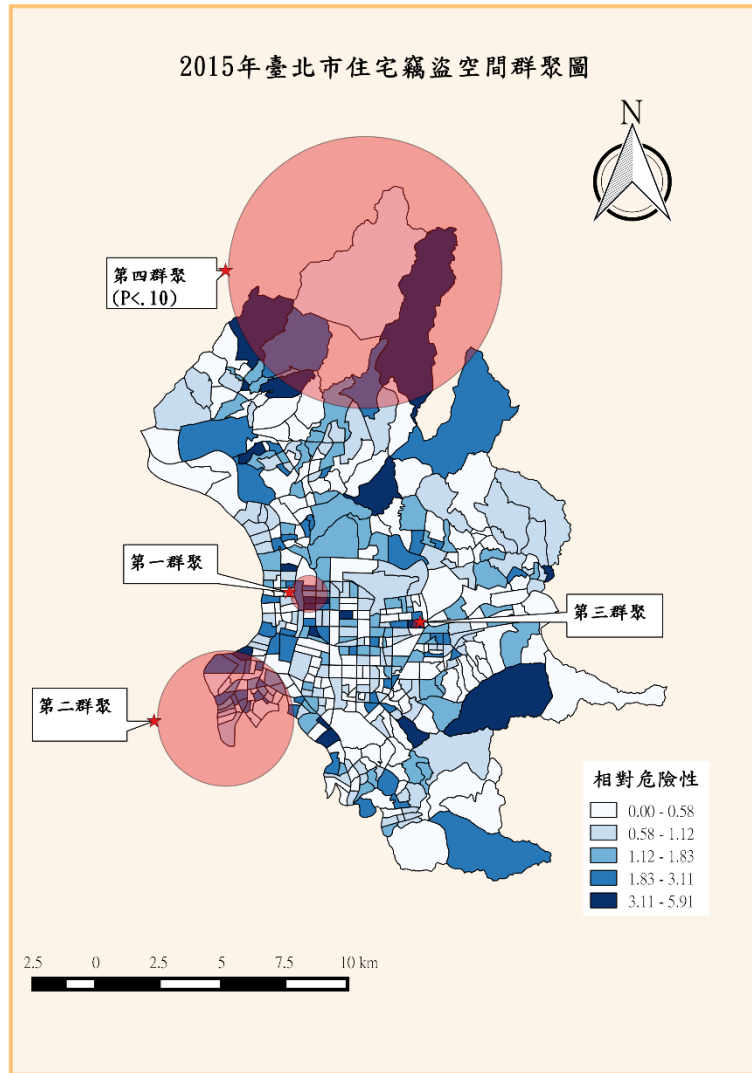
表 4-3-1 2015 年臺北市住宅竊盜圓形空間掃描摘要表

群聚	群聚半徑	期望值	觀察值	LLR	RR	發生率	p-value
1	0.74 km	11.49	35	15.94	3.17	179.3	**
2	2.70 km	50.64	93	15.81	1.99	108.1	**
3	0 km	2.08	11	9.46	5.36	311.0	*
4	5.43 km	3.79	14	8.16	3.75	217.3	0.08

註：**p < 0.01；*p < 0.05。

表 4-3-2 2015 年臺北市住宅竊盜圓形空間群聚區(里)

群聚	區里
1	中山區：中莊里、恆安里、晴光里、集英里、新生里、新安里、新福里、聚盛里、聚葉里
2	萬華區：仁德里、日祥里、日善里、全德里、孝德里、和平里、和德里、忠貞里、忠德里、保德里、凌霄里、頂碩里、富福里、菜園里、華中里、華江里、新和里、新忠里、壽德里、榮德里、福音里、綠堤里、銘德里、糖蔴里、興德里、錦德里、雙園里、騰雲里 中正區：永昌里、忠勤里、南門里、廈安里、網溪里、螢雪里、龍光里 龍興里
3	松山區：安平里
4	士林區：平等里、菁山里、陽明里 北投區：大屯里、林泉里、泉源里、湖山里、湖田里



註：紅色圓形區域為群聚區域範圍。

圖 4-3-1 2015 年臺北市住宅竊空間群聚圖

在圖 4-3-1 中，本研究以各里相對風險數值分級圖做為底圖，在本研究由於使用各里戶數做為母群體數，故各里的相對風險數值即相當於住宅犯罪率，與觀察值除以期望值所得出之數值十分接近，掃描統計主要對於各里母群體分布不均進行調整，考慮所設定風險人口比例後進行群聚範圍選取，故群聚內的整體相對風險性數值高於各里數值，而表 4-3-1 的群聚住宅竊盜發生率則是以每十萬戶為單位，若改為以萬戶為單位，其發生率同樣仍遠高於臺北市各里發生率，故觀察群聚內相對風險數值，可顯示處於此區域內的住宅竊盜風險程度；此外，群聚的排序與對數概似率有關，排序較前的群聚，風險程度不一定是最高；換言之，只

要群聚達到統計顯著後，相對風險值為重要評估指標，警政機關除觀察群聚的排序外，亦應同時參考各群聚之相對風險程度。

此外，由上圖觀察可發現，在群聚範圍中，包含顏色深淺不一，顯示有高低不同相對風險值的里被涵蓋在同一群聚範圍內，而在群聚範圍外，亦有相對風險值較群聚內更高的里，卻未成為群聚，針對此現象，Kulldorff 等人認為，此情形顯示群聚內的確有較高風險值的地區存在，但難免有部分較低風險的地區被整體群聚範圍所覆蓋(Kulldorff, Feuer, Miller, & Freedman, 1997)；此外，SaTScan 圓形掃描視窗在掃描時會將圓形半徑內可能存在的非真實地理單元加入到偵測結果中（如此地區為國家風景區、或湖泊），也將造成此現象發生。而對於高風險地區卻未納入群聚的問題，主要在其鄰近的里均是較低風險值，故在無法選取強度較大的鄰接單元進入群聚候選集合中，所以在蒙地卡羅檢定時，群聚現象並不明顯(董倩楠，2016)，如圖 4-3-1 最右下角文山區指南里，雖顯示出有較高風險值，但在 2015 年群聚掃描中並不具顯著性，即為此原因。

圖 4-3-2 為本研究利用國土資源部的通用電子地圖呈現群聚分析結果，通用電子地圖可用放大地圖方式詳細觀察臺北市住宅竊盜發生的詳細街道地點及周邊景物。由於本研究主要以里為單位，並藉以檢測群聚區域，並將連結各種人文社經等屬性資料進行分析，因上述網路資源之製圖無法辨識群聚中包含哪些里及連結屬性資料，故以下分析仍以臺北市各里的向量里界圖層做為基本底圖。

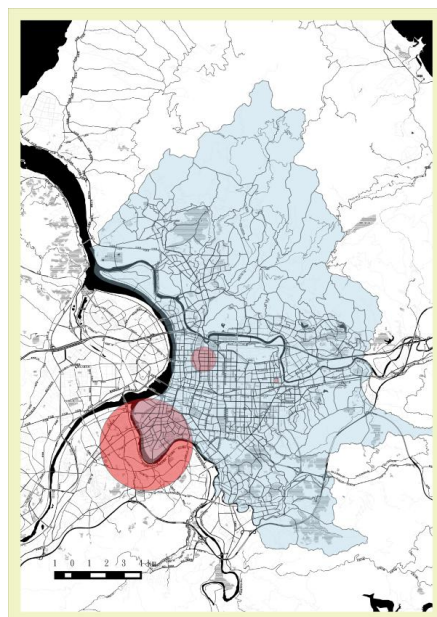


圖 4-3-2 2015 年臺北市住宅竊空間群聚圖

2、2016 年圓形空間掃描結果

在 2016 年空間掃描方面，達到顯著水準之群聚區僅有中山區聚盛里 1 處 (25.059222 N, 121.525213 E)，群聚半徑為 0km，對數概似率為 10.81，相對風險值為 6.99，p-value < 0.01，而第二個群聚區的 P 值為.099，小於 10%，集中在士林區與北投區，群聚中心點座標位於北投區湖田里(25.179213 N, 121.547572 E) 共有 34 個里，群聚半徑為 7.43km，對數概似率為 7.85，相對風險值為 1.70 (表 4-3-3，4-3-4，圖 4-3-3)。

表 4-3-3 2016 年臺北市住宅竊盜圓形空間掃描摘要表

群聚	群聚半徑	期望值	觀察值	LLR	RR	發生率	p-value
1	0 km	1.45	10	10.81	6.99	421.6	**
2	7.43 km	44.40	72	7.85	1.70	99.2	.10

註：**p < 0.01。

表 4-3-4 2016 年臺北市住宅竊盜圓形空間群聚區(里)

群聚	區里
1	中山區：聚盛里
2	士林區：平等里、菁山里、陽明里、三玉里、公館里、天山里、天母里、天玉里、天和里、天壽里、天福里、溪山里、東山里、新安里、北投區：大屯里、林泉里、泉源里、湖山里、湖田里、中心里、中央里、中和里、中庸里、文化里、永和里、永欣里、秀山里、奇岩里、東華里、長安里、清江里、智仁里、開明里、溫泉里、稻香里

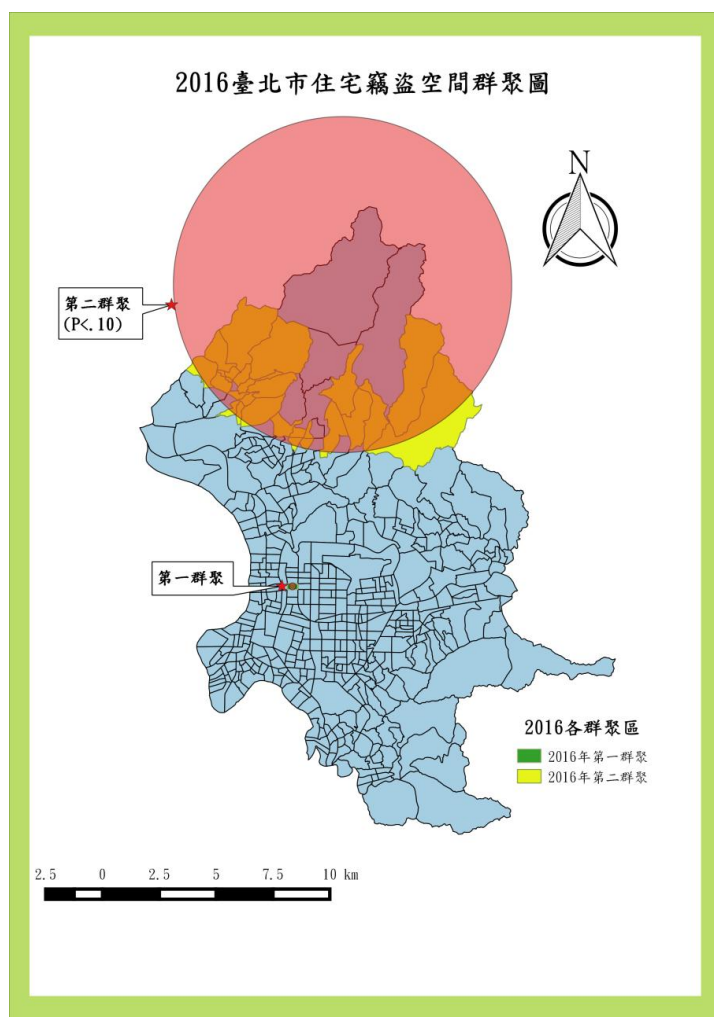


圖 4-3-3 2016 年臺北市住宅竊空間群聚圖

雖然 2016 年比起 2015 年有更多住宅竊盜犯罪數，但群聚檢測之結果，顯著群聚區卻只有中山區 1 個里，若參考上一節的點位圖，大致可觀察出其犯罪點似較 2015 年分散；而若納入 P 值小於 .10 的區域，則落在此範圍的群聚里數量並不是太少，圖 4-3-3 顯示，臺北市住宅竊盜群聚區由較擁擠的市中心位置移向北邊，對此現象過去有若干研究將其稱之為「犯罪轉移」。本研究從群聚角度出發，認為此現象應僅是原本群聚的顯著性消失，不一定是犯罪轉移，且觀察點位圖，犯罪點數量仍不在少數，可能是分布型態屬於狹長型，以圓形掃描產生較大的誤差等原因所造成，此有待橢圓形掃描再次確認。此外，僅以 2016 年與 2015 年相比較，時間過於短暫，若將時間尺度設定在 3 年為 1 單位，進行竊盜數量總數分析，群聚結果將會不同，由此也說明時間的尺度大小與觀察時間長短，對於群聚的形成與判斷具有重要的影響。

3、2017 年圓形空間掃描結果

在 2017 空間掃描方面，出現 6 個群聚區，其中 2 個群聚區達到 $p\text{-value} < .01$ 統計顯著，第三群聚區 $p\text{-value}$ 未達 10% 標準。第一個最大可能的群聚區分布在中山區、中正區、大同區及萬華區，中心點座標在中正區黎明里(25.044538 N, 121.516280 E)，共有 21 個里，群聚半徑為 1.23km，對數概似率為 14.31，相對風險值為 2.52，第二群聚區均在士林區，中心點座標在天祿里(25.113378 N, 121.528557 E)(表 4-3-3，4-3-4，圖 4-3-3)。

表 4-3-5 2017 年臺北市住宅竊盜圓形空間掃描摘要表

群聚	群聚半徑	期望值	觀察值	LLR	RR	發生率	p-value
1	1.23km	19.73	47	14.31	2.52	114.1	**
2	0.45km	4.61	18	11.30	4.1	186.9	**

註：** $p < 0.01$ 。

表 4-3-6 2017 年臺北市住宅竊盜圓形空間群聚區(里)

群聚	區里
1	中山區：正守里、正得里、正義里、民安里 大同區：玉泉里、光能里、建功里、建明里、建泰里、星明里、朝陽里 中正區：光復里、幸福里、東門里、建國里、梅花里、黎明里 萬華區：西門里、新起里、萬壽里、福星里
2	士林區：天壽里、天福里、天祿里、蘭雅里

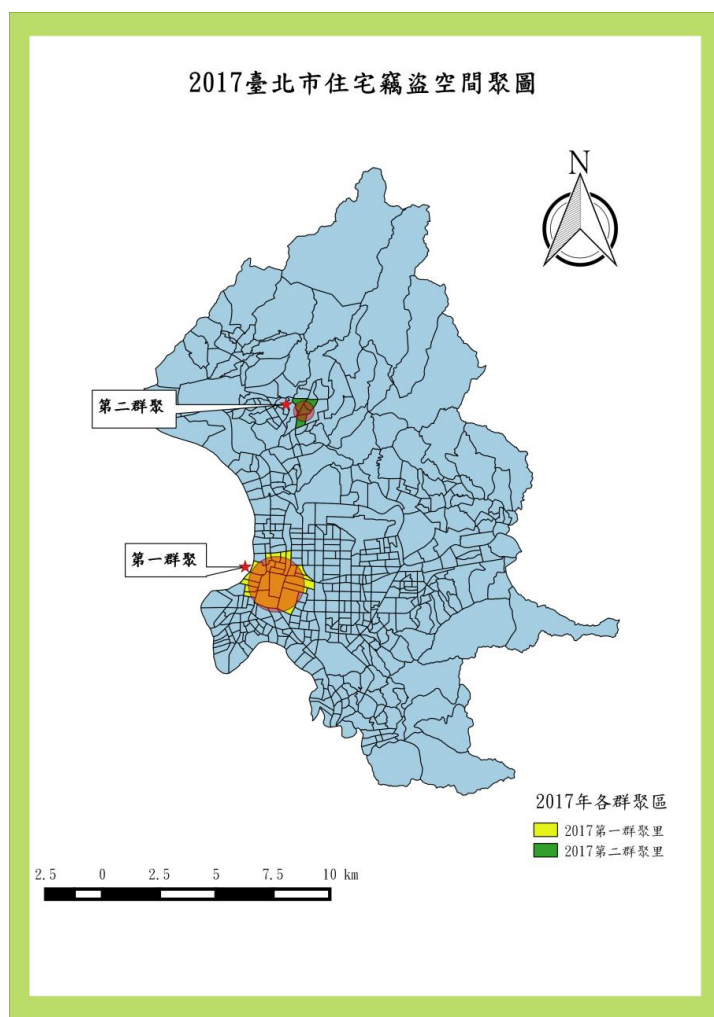


圖 4-3-4 2017 年臺北市住宅竊空間群聚圖

4、2015 年至 2017 年圓形空間掃描結果

以 2015 年至 2017 年 3 年住宅竊盜總數累加掃描結果，出現 87 個可能群聚里，14 個群聚區，其中 4 個群聚區達到 $p\text{-value} < .05$ 統計顯著，第五群聚區則超出 .10。第一個最大可能的群聚區主要集中在中山區，包括中山里、聚盛里和康樂里等 11 個區域及大同區光能里，群聚中心點座標在中山區中山里 (25.056605 N, 121.524944 E)，對數概似率為 20.51， $p\text{-value} < 0.01$ ；第二個可能的群聚區包括萬華區、中正區和大同區 3 個地區，群聚中心點座標在萬華區菜園里 (25.042339 N, 121.499276 E)，對數概似率為 12.06， $p\text{-value} < 0.01$ ，其中萬華區包括 24 個里為最多，其次為中正區 11 個里，大同區則有 5 個里；第三可能群聚區為北投區與士林區，共包括 7 個里，群聚中心點座標在北投區湖田里 (25.179213

N, 121.547572 E)，對數概似率為 10.02， p -value < 0.05 ；第四可能群聚區仍為士林區與北投區，包含 7 個里，群聚中心點座標在士林區天祿里(25.113378 N, 121.528557 E)，對數概似率為 9.62， p -value < 0.05 (表 4-3-5，4-3-6，圖 4-3-4)。

表 4-3-7 2015-2017 年臺北市住宅竊盜圓形空間掃描摘要表

群聚	群聚半徑	期望值	觀察值	LLR	RR	發生率	p-value
1	0.77 km	46.09	95	20.51	2.12	115.4	**
2	2.00 km	140.02	199	12.06	1.47	140.0	**
3	4.96 km	8.21	24	10.02	2.95	163.5	*
4	0.72 km	28.04	54	9.62	1.96	107.8	*

註：** $p < 0.01$ ；* $p < 0.05$ 。

表 4-3-8 2015-2017 年臺北市住宅竊盜圓形空間群聚區(里)

群聚	區里
1	中山區：中山里、聚盛里、康樂里、中原里、聚葉里、民安里、新生里、正得里、正義里、中吉里、中莊里 大同區：光能里
2	萬華區：菜園里、青山里、富民里、西門里、福音里、新起里、萬壽里、仁德里、福星里、富福里、糖蔴里、柳鄉里、頂碩里、雙園里、和平里、綠堤里、新忠里、華江里、日善里、新安里、新和里、忠德里、和德里、全德里 中正區：愛國里、光復里、建國里、廈安里、南門里、忠勤里、黎明里、龍興里、龍福里、龍光里、永昌里 大同區：玉泉里、永樂里、建功里、建明里、朝陽里
3	北投區：湖田里、湖山里、泉源里、大屯里 士林區：菁山里、陽明里、平等里
4	士林區：天祿里、天壽里、天福里、蘭雅里、三玉里、聖山里 北投區：榮華里

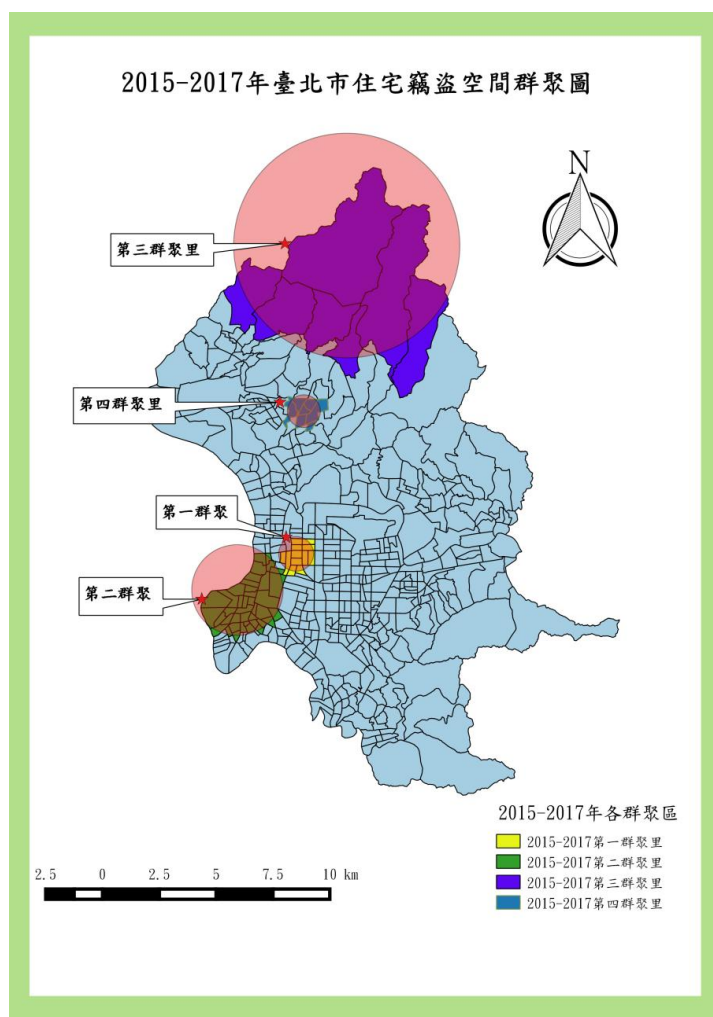


圖 4-3-5 2015-2017 年臺北市住宅竊空間群聚圖

(二) 橢圓形視窗空間掃描

橢圓形空間掃描與圓形掃描的統計原理相似，最大的區別在於如何選擇群聚的問題。橢圓形可以由以下參數加以定義：1、中心點的 x 和 y 坐標；2、形狀(shape)；3、偏心率(eccentricity)；4、角度(angle)；5、大小(Size)，其中橢圓的形狀 S 定義為半長軸(semimajor axis)的長度與半短軸(semiminor axis)的長度之比，即它是橢圓的最長軸與最短軸的比率，絕對數值大於 1 表示長而窄的橢圓，而數值為 1 的橢圓則是特例，通常是圓形。橢圓掃描所偵測地群聚區域形狀 S 與偏心率 e 有關，相關公式為 $S = 1/\sqrt{1 - e^2}$ ，在掃描過程中橢圓的形狀、角度、大小不斷變化，並且引入一個非緊密懲罰函數(penalty function)來約束對數概似率(LLR)的增加，以達到約束群聚區的形狀的目的，其計算公式為 $[4S/(S+1)]^a$ ，懲罰強度

由弱到強依次為 0、1/2、1，a 值越大表示期望偵測到的群聚區形狀越規則，但相對也限制偵測的範圍。

相較圓形掃描統計方法，橢圓形方法能探測到對數概似率值更大、形狀更加不規則的顯著聚集區，有助於發現更多潛藏的群聚，但一個非常長而狹窄的群聚僅包含某一地區，而沒有其他地區，通常這並不符合地理區域群聚的認定；因此，對於每個橢圓形，要考慮一定數量的不同角度，尤其是對於長而窄的橢圓，角度的微小變化將導致一組非常不同的地理中心點(Kulldorff et al., 2006)。以下本研究同樣將風險人口設定為 50%，進行橢圓形空間掃描，懲罰強度設定為弱，主要目的是希望能探測到更多的群聚區域。

1、2015 年橢圓形空間掃描結果

在 2015 橢圓形空間掃描方面，出現 7 個群聚區，其中 4 個群聚區達到 p-value < .01 統計顯著，第五群聚區 p-value 為 0.07。第一個最大可能的群聚區分布在萬華區、中正區、中山區及大同區，中心點座標在萬華區華中里(25.0146 N, 121.493 E)，以萬華區 29 個里占多數，橢圓形 Shape 值為 5.00，角度為 36.00，對數概似率為 20.07，相對風險值為 1.87；第二群聚區原有 12 個群聚里，但 11 個群聚里與第一群聚區重疊，僅有萬華區和平里不同，中心點座標在萬華區日善里(25.0276 N, 121.498 E)，橢圓形 Shape 值為 5.00，角度為 0，對數概似率為 19.21，相對風險值為 3.14；第三群聚區與第一群聚區亦大部分重疊，僅中山區中莊里與朱馥里不同，中心座標在中山區中莊里，(25.0592 N, 121.53 E)橢圓形 Shape 值為 5.00，角度為-72，對數概似率為 19.21，相對風險值為 3.89；第四群聚區中心點座標為中正區螢雪里(25.024 N, 121.515 E)，包含在第一群聚中，Shape 值為 0，角度為 0，對數概似率為 13.47，相對風險值為 1.07；第五群聚區中心點座標為松山區安平里(25.0526 N, 121.567 E)，p-value < .10，Shape 值為 3，角度為 80，對數概似率為 10.74 相對風險值為 4.23 (表 4-3-9，4-3-10，圖 4-3-6)。

表 4-3-9 2015 年臺北市住宅竊盜橢圓形空間掃描摘要表

群聚	Shape值	角度	期望值	觀察值	LLR	RR	發生率	p-value
1	5.00	36.00	90.98	151	20.07	1.87	97.7	**
2	5.00	0	14.36	43	19.21	3.14	176.2	**

3	5.00	-72	8.56	32	19.21	3.89	219.9	**
4	0	0	1.07	10	13.47	9.46	548.9	**
5	3.00	80	3.86	16	10.74	4.23	244.1	0.07

註：**p < 0.01。

表 4-3-10 2015 年臺北市住宅竊盜橢圓形空間群聚區(里)

群聚	區 里
1	<p>萬華區：新和里、新安里、仁德里、日祥里、日善里、忠貞里、忠德里、保德里、凌霄里、頂碩里、富福里、菜園里、新忠里、壽德里、糖蔴里、興德里、青山里、富民里、西門里、萬壽里、福星里、全德里、孝德里、和德里、華中里、榮德里、綠堤里、銘德里、錦德里</p> <p>中正區：永昌里、忠勤里、南門里、廈安里、龍興里、愛國里、黎明里、螢雪里、雙園里、騰雲里、永功里、幸福里、東門里、建國里、梅花里、龍福里</p> <p>中山區：聚盛里、聚葉里、中山里、康樂里、正得里、民安里、正義里、恆安里、晴光里、集英里、新生里、中原里、正守里、新福里、新喜里</p> <p>大同區：光能里、建明里、建泰里、玉泉里、建功里、星明里、朝陽里、雙連里</p>
2	萬華區：和平里
3	中山區：中莊里、朱馥里
5	松山區：安平里、東光里

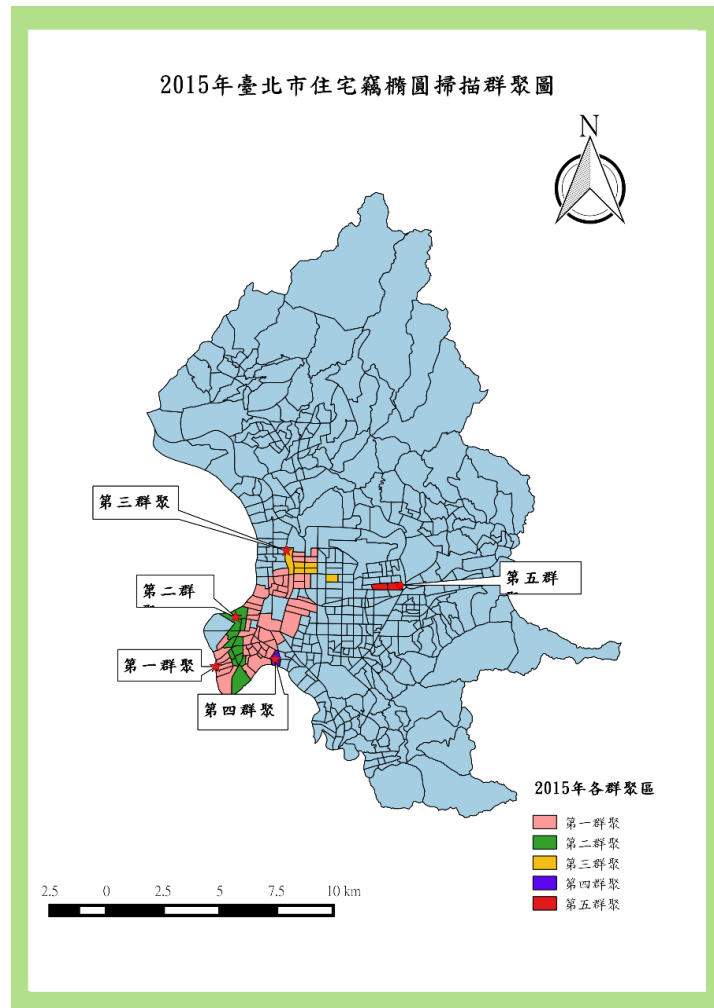


圖 4-3-6 2015 年臺北市住宅竊盜橢圓空間群聚圖

2、2016 年橢圓形空間掃描結果

在 2016 橢圓形空間掃描方面，出現 8 個可能群聚區，其中 2 個群聚區達到 $p\text{-value} < .05$ 統計顯著，第三群聚區 $p\text{-value}$ 為超出 10%。第一個最大可能的群聚區分布在中山區、中正區、萬華區及大同區， $p\text{-value} < .05$ ，中心點座標在中正區黎明里(25.0445N, 121.516E)，此群聚共有 19 個群聚里，以中正區 9 個里占多數，橢圓形 Shape 值為 4.00，角度為 30.00，對數概似率為 14.35，相對風險值為 2.39；第二群聚區在松山區、信義區及大安區，中心點座標在信義區興隆里(25.0393N, 121.56E)，共有 12 個群聚里，以大安區占 6 個里為最多，橢圓形 Shape 值為 4.00，角度為 30.00，對數概似率為 11.57，相對風險值為 2.46。(表 4-3-11，4-3-12，圖 4-3-7)。

表 4-3-11 2016 年臺北市住宅竊盜橢圓形空間掃描摘要表

群聚	Shape值	角度	期望值	觀察值	LLR	RR	發生率	p-value
1	4.00	30.00	22.82	52	14.35	2.39	139.4	*
2	4.00	30.00	16.42	39	11.57	2.46	145.3	0.61

註：* $p < 0.05$ 。

表 4-3-12 2016 年臺北市住宅竊盜橢圓形空間群聚區(里)

群聚	區里
1	中山區：正得里、聚盛里、聚葉里、中山里、康樂里、民安里 中正區：黎明里、永昌里、忠勤里、南門里、龍興里、愛國里、廈安里、光復里、建國里 萬華區：新和里 大同區：建泰里、建明里、光能里
2	松山區：安平里、新聚里 信義區：中興里、正和里、新仁里、興隆里、 大安區：正聲里、光信里、法治里、通化里、通安里、臨江里

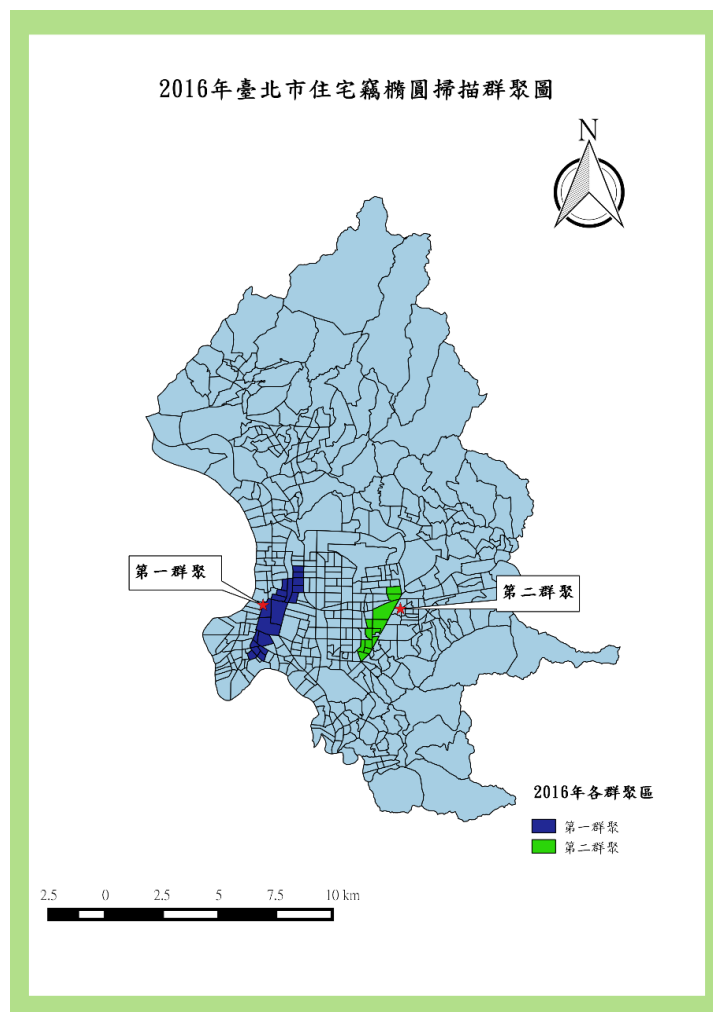


圖 4-3-7 2016 年臺北市住宅竊盜橢圓空間群聚圖

3、2017 年橢圓形空間掃描結果

在 2017 橢圓形空間掃描方面，出現 7 個可能群聚區，其中 3 個群聚區達到 $p\text{-value} < .05$ 統計顯著，第四群聚區 $p\text{-value}$ 為超出 10%。第一個最大可能的群聚區分布在萬華區、中正區、中山區及大同區， $p\text{-value} < .01$ ，中心點座標在中正區黎明里(25.0445N, 121.516E)，此群聚共有 18 個群聚里，以中正區 8 個里占多數，橢圓形 Shape 值為 4.00，角度為 30.00，對數概似率為 14.35，相對風險值為 2.39；第二群聚區在中山區、大安區、松山區及信義區，中心點座標在信義區興隆里(25.0393N, 121.56E)，共有 26 個群聚里，以中山區占 13 個里為最多，橢圓形 Shape 值為 4.00，角度為 30.00，對數概似率為 11.57，相對風險值為 2.46；第三群聚區為士林區與北投區，中心點座標為北投區湖山里(25.1493 N, 121.539 E)，

橢圓形 Shape 值為 5.00，角度為 12，對數概似率為 11.21，相對風險值為 2.48 (表 4-3-13，4-3-14，圖 4-3-8)。

表 4-3-13 2017 年臺北市住宅竊盜橢圓形空間掃描摘要表

群聚	Shape值	角度	期望值	觀察值	LLR	RR	發生率	p-value
1	5.00	84.00	10.24	37	21.52	3.82	173.1	**
2	2.00	-30.00	6.42	39	13.04	3.70	171.5	**

註：**p < 0.01。

表 4-3-14 2017 年臺北市住宅竊盜橢圓形空間群聚區(里)

群聚	區里
1	中山區：正得里、正義里、 中正區：光復里、黎明里、 大同區：建明里、建泰里、玉泉里、建功里、 萬華區：西門里、萬壽里、福星里、
2	士林區：天祿里、三玉里、聖山里、蘭雅里、名山里、岩山里

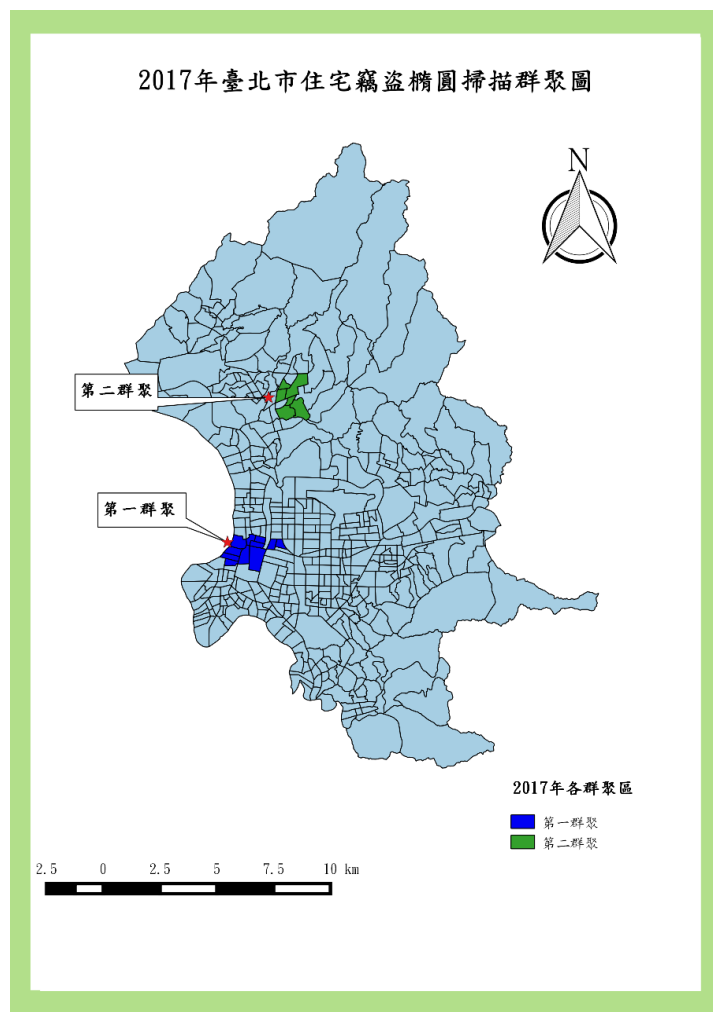


圖 4-3-8 2017 年臺北市住宅竊盜橢圓空間群聚圖

4、2015 年至 2017 年橢圓形空間掃描結果

在 2015 年至 2017 年橢圓形空間掃描方面，出現 7 個可能群聚區，其中 3 個群聚區達到 $p\text{-value} < .05$ 統計顯著，第四群聚區 $p\text{-value}$ 為超出 10%。第一個最大可能的群聚區分布在中山區及中正區，中心點座標在中山區中山里(25.0566 N, 121.525 E)，此群聚共有 15 個群聚里，以中山區 14 個里占多數，橢圓形 Shape 值為 5.00，角度為 24.00，對數概似率為 29.97，相對風險值為 2.20；第二群聚區在北投區與士林區，中心點座標在北投區湖山里(25.1493 N, 121.539 E)，共有 12 個群聚里，以士林區占 10 個里為最多，橢圓形 Shape 值為 5.00，角度為 12.00，對數概似率為 13.86，相對風險值為 1.92；第三群聚區為中正區、萬華區、文山區及松山區，中心點座標為中正區螢雪里(25.024 N, 121.515E)，橢圓形 Shape 值

為 5.00，角度為-48，對數概似率為 13.26，相對風險值為 1.54；值得注意的是，文山區與松山區首次入列，但均僅有 1 個里(表 4-3-15，4-3-16，圖 4-3-9)。

表 4-3-15 2015-2017 年臺北市住宅竊盜橢圓形空間掃描摘要表

群聚	Shape值	角度	期望值	觀察值	LLR	RR	發生率	p-value
1	5.00	24.00	61.71	130	29.97	2.20	117.9	**
2	5.00	12.00	44.29	83	13.86	1.92	104.9	**
3	5.00	-48.00	117.44	175	13.26	1.54	83.4	*

註：** $p < 0.01$ ；* $p < 0.05$ 。

表 4-3-16 2015-2017 年臺北市住宅竊盜橢圓形空間群聚區(里)

群聚	區 里
1	中山區：正得里、正義里、聚盛里、聚葉里、中山里、康樂里、民安里、中莊里、恆安里、晴光里、新生里、新福里、新莊里、新喜里、中原里 中正區：黎明里
2	北投區：湖山里、湖田里 士林區：蘭雅里、天山里、天母里、天和里、天壽里、天祿里、三玉里、聖山里、天福里、天玉里
3	中正區：永昌里、忠勤里、南門里、龍興里、愛國里、網溪里、螢雪里、永功里、文盛里、水源里、板溪里、林興里、河堤里、南福里、頂東里、富水里、螢圃里、廈安里 萬華區：新和里、仁德里、凌霄里、頂碩里、富福里、菜園里、新忠里、糖蔴里、雙園里、騰雲里、青山里、富民里、福音里 文山區：萬年里 松山區：東光里

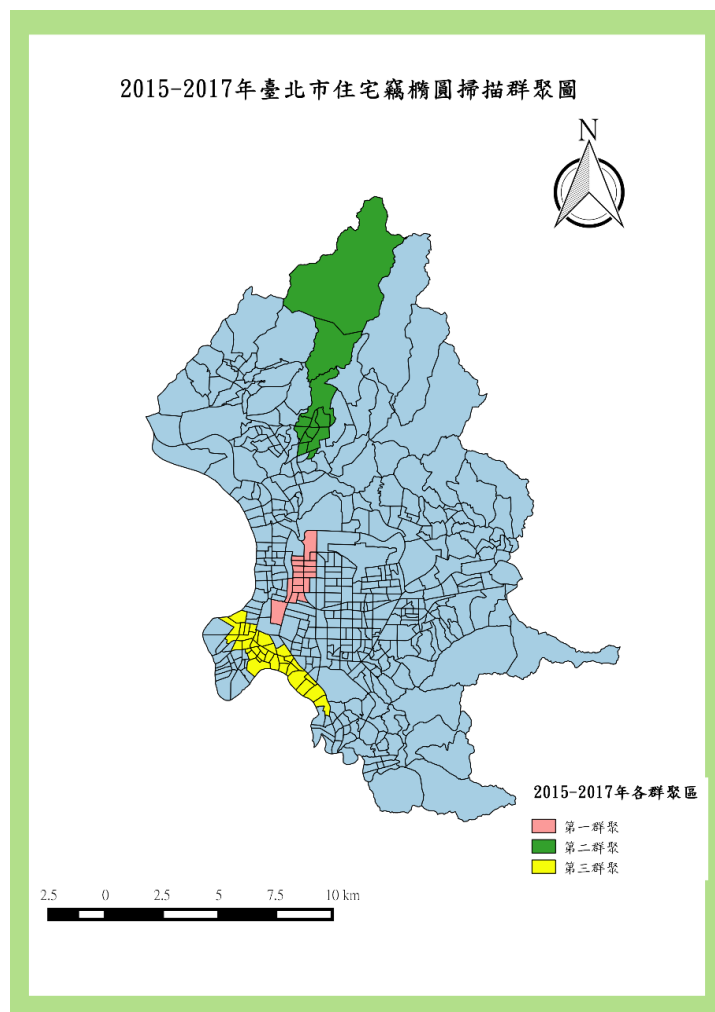


圖 4-3-9 2015-2017 年臺北市住宅竊盜橢圓空間群聚圖

(四) 圓形掃描與橢圓形掃描比較

從以上針對臺北市 2015 年至 2017 年各年度與 3 年總和的圓形、橢圓形視窗掃描的結果觀察，基本上印證 Kulldorff 對於橢圓形比起圓形視窗掃描能得到更大概似率，及對形狀不規則群聚偵測更為敏感的結論，兩種視窗掃描結果差異在 2016 年顯得特別突出，在圓形視窗掃描中本研究若將 P 值設定為 $<.05$ ，僅中山區聚盛里具顯著水準，但使用橢圓形視窗掃描卻可偵測多達 19 個里，由此或可推論 2016 年臺北市住宅竊的群聚形狀可能趨近於狹長型態。以下本研究將圓形與橢圓型視窗掃描結果加以套疊，對兩者在群聚偵測的差異即可一目了然(圖 4-3-10)。

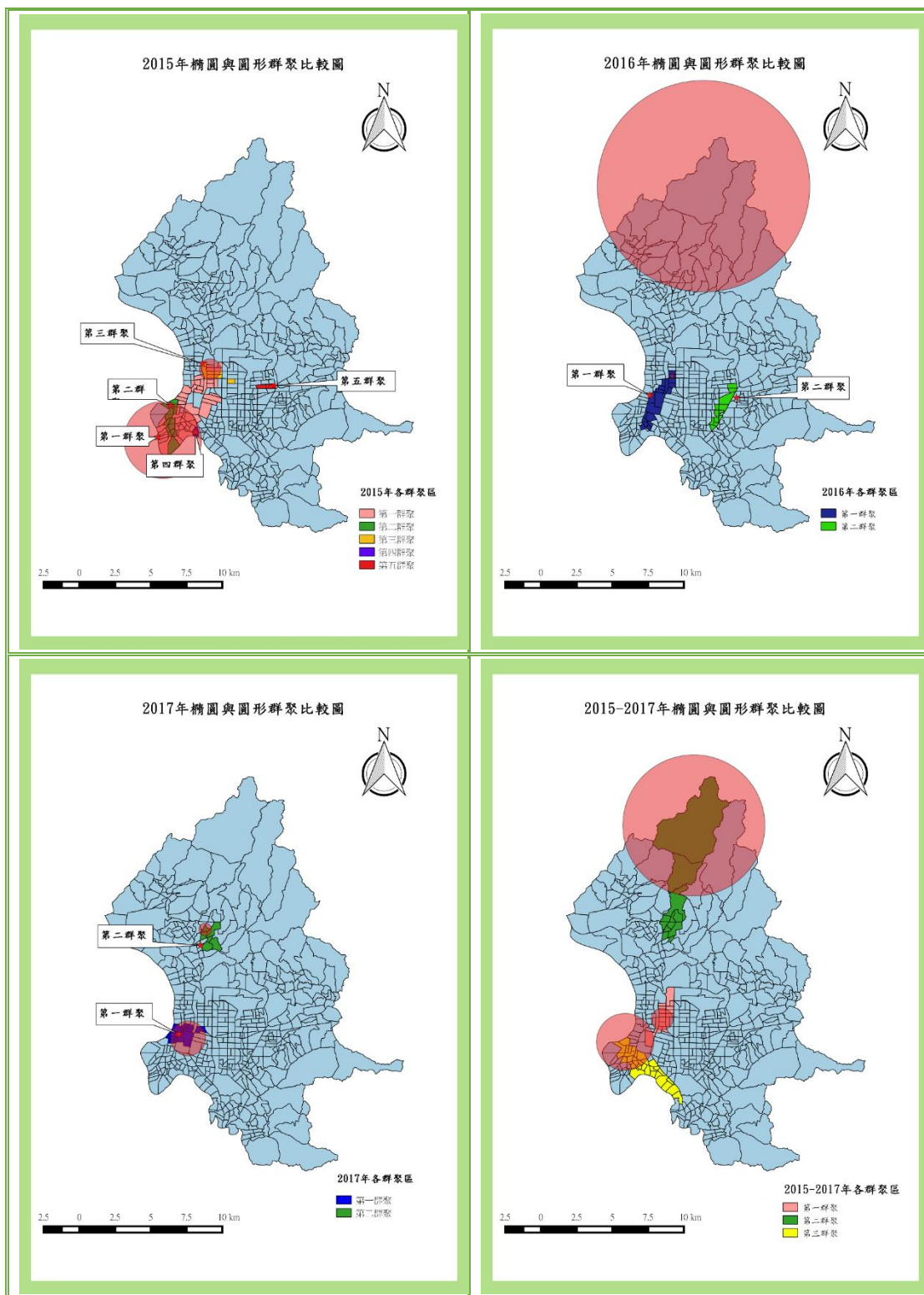


圖 4-3-10 臺北市住宅竊圓形與橢圓比較圖

以上各圖顯示，橢圓形視窗掃描所偵測到的群聚區（里）在各年度均多於圓形視窗掃描，且基本上均能涵蓋圓形視窗掃描的範圍，從前述對型 1 與型 2 錯誤在犯罪防治的意義而言，對民眾的保障宜秉持「料敵從寬」的態度，故橢圓形視窗掃描似乎更符合對於空間犯罪群聚的需求；惟以兩種不同形狀的視窗掃描結果所偵測的群聚里並不全然相同，所以本研究建議應對於除兩者重疊區域應特別予以重視外，對於偵測差異部分，亦不能輕忽，尤其這些群聚里若其相對風險值很高，且 P 值達到小於.05 的顯著水準時，可能需要重新改變掃描視窗的尺度，再次進行確認，例如：原設定 50% 的風險人口，可改為 25% 更小尺度再次掃描，或以掃描半徑進行不同範圍的設定，經過反覆的測試與比對，將可獲知有關群聚的更多訊息。

（五）臺北市住宅竊盜群聚熱點性質的識別與界定

從以上空間掃描結果可初步得知臺北市宅竊盜犯罪群聚的位置與構成群聚的里別，這些在群聚範圍內的各里別，為所謂的竊盜犯罪熱點，雖然根據群聚的相對風險值，大致上可知這些犯罪熱點的風險程度，但對於其性質卻無從得知，此對於犯罪防治策略的制定似仍有不足。

基於對熱點的觀察與認識，通常犯罪熱點的存在往往為周圍環境多方面因素導致的結果，呈現出現象一般有幾種情況：連續多年都是熱點，顯示周圍的環境因素可能沒有發生太多變化，可稱為持續熱點；隨著主要環境因素的改變，某些熱點會逐漸消失，稱為消失熱點；而若干地方又有可能逐漸形成與其他熱點環境相似的環境因素，新的熱點也會逐步形成，可稱為新興熱點；同樣地，亦有一些偶發因素導致某個地方迅速成為熱點，但很快這些因素會自然退卻或受到強制力影響，熱點迅速消失，可稱為偶發熱點；另有某些熱點連續出現多次後消失，間隔若干時間後又重新出現，可稱為間歇熱點。因此，本研究將各年度掃描掃描結果群聚區中的各里別予以歸納，根據其在各年度是否出現的頻率與間隔情形對各熱點進行性質上分類。然而，由於本研究時間範圍僅 3 年，時間跨度可能仍有不足，對於熱點是否將持續發生或消失後再發生等現象，現有歸類並非變動不居，未來若有更多的年度資料，則熱點的性質將會產生改變，例如：持續熱點在往後若干年均未在出現，則將成為消失熱點等。

本研究根據不同時期空間熱點發生的頻率，將熱點概分為持續、間歇、新興、

偶發及消失熱點共 5 類，分類標準如表 4-3-17。

表 4-3-17 臺北市住宅竊盜熱點分類標準表

類別	2015 年	2016 年	2017 年	熱點性質
1	★	★	★	持續
2	★	★		間歇
	★		★	
3	★			消失
4		★		偶發
			★	
5		★	★	新興

- 1、持續熱點：連續 3 年出現在群聚區中。
- 2、間歇熱點：3 年內在群聚中出現 2 次，型態包括 2015、2016 年連續出現，但 2017 年未出現；2015 年出現後，2016 年消失，2017 年再度出現。至於 2016 年及 2017 年連續出現，本研究將另外歸類為新興熱點。
- 3、消失熱點：所謂消失熱點為 2015 年出現後，連續兩年沒有出現在群聚中。
- 4、偶發熱點：2016 年與 2017 年僅出現 1 次。
- 5、新興熱點：即 2016 年與 2017 年連續出現的熱點。

根據表 4-3-17 分類標準表，臺北市圓形與橢圓形掃描住宅竊各里別的犯罪熱點性質如表 4-3-18：

表 4-3-18 臺北市住宅竊盜熱點性質分類

熱點性質	圓形掃描	橢圓形掃描
持續	無	中山區：正得里、黎明里、建泰里、建明里
間歇	士林區：天壽里、天福里、平等里、菁山里、陽明里 中山區：聚盛里、 北投區：湖山里、湖田里、大屯里、林泉里、泉源里 萬華區：綠堤里	大同區：光能里、玉泉里、建功里 中山區：聚葉里、民安里、正義里、聚盛里、中山里、康樂里 中正區：永昌里、忠勤里、南門里、龍興里、廈安里、建國里、愛國里 松山區：安平里 萬華區：新和里、西門里、萬壽里、福星里
消失	中山區：聚葉里、恆安里、晴光里、新生里、新福里、中莊里、集英里 中正區：永昌里、忠勤里、南門里、龍興里、廈安里、螢雪里、網溪里、龍光里 松山區：安平里 萬華區：新和里、仁德里、凌霄里、頂碩里、富福里、菜園里、新忠里、糖蔴里、雙園里、騰雲里、福音里、壽德里、華中里、新安里、日祥里、日善里、忠貞里、忠德里、保德里、興德里、全德里、孝德里、和德里、榮德里、銘德里、錦德里、和平里、華江里	中山區：恆安里、晴光里、新生里、新福里、中莊里、集英里、正守里、新喜里、中原里、朱馥里 中正區：螢雪里、幸福里、東門里、梅花里、永功里、龍福里 松山區：東光里 萬華區：仁德里、凌霄里、頂碩里、富福里、菜園里、新忠里、糖蔴里、雙園里、騰雲里、壽德里、華中里、新安里、日祥里、日善里、忠貞里、忠德里、保德里、興德里、全德里、孝德里、和德里、榮德里、銘德里、錦德里、和平里、綠堤里、青山里、富民里
偶發	士林區：蘭雅里、天山里、天母里、天和里、天祿里、三玉里、天玉里、公館里、東山里、新安里、溪山里 大同區：建泰里、建明里、光能里、玉泉里、建功里、星明里、	大安區：通化里、通安里、臨江里 松山區：新聚里 信義區：中興里、正和里、新仁里、興隆里

	<p>朝陽里</p> <p>中山區：正得里、民安里、正義里、正守里</p> <p>中正區：黎明里、光復里、建國里、幸福里、東門里、梅花里</p> <p>北投區：中心里、中央里、中和里、中庸里、文化里、永和里、永欣里、秀山里、奇岩里、東華里、長安里、清江里、智仁里、開明里、溫泉里、稻香里</p> <p>萬華區：西門里、萬壽里、福星里、新起里</p>	
新興	無	中正區：光復里

從上表中可看出，經過分類後，圓形與橢圓形在熱點的性質有重複部分，但也有相當的不同，特別是在持續、偶發及新興熱點 3 種分類方面；以上述的分類方式，圓形掃描在持續與新興群聚並沒有相關熱點出現，但橢圓形視窗掃描則出現若干熱點訊息，而最大的差異就在於偶發熱點，兩者的結果完全不相同，圓形視窗掃描所顯示的熱點為平均犯罪率較高地區，也是民眾一般認為區域環境較混雜的地區，但橢圓形視窗掃描的熱點完全集中在臺北市所謂高級商業或住宅地段區域。本研究初步認為造成差異的原因至少以下幾點：

- 1、掃描視窗不同所造成的差異。
- 2、分類標準造成差異。例如，將 2017 年出現之熱點均歸為新興群聚，而非有些歸入偶發群聚。
- 3、以年度進行分別統計，時間尺度是否適當。
- 4、研究時間僅有 3 年，時間跨度可能不足。
- 5、以年度進行區隔，但實質上仍是純粹空間的範疇，無法進一步了解空間在時間趨勢中的動態變化。

綜上，本研究認為除主觀劃分標準及掃描視窗不同所造的差異外，未能將時間因素與空間因素兩者加以融合為主要原因；換言之，以上對於熱點分類標準係以各年度發生在空間的次數為基準，無法考慮熱點群聚時間的長短，若從時間觀

點而言，群聚時間 1 年與 1 個月相比，則前者對後者而言，可稱為持續性熱點，而新興熱點若出現在年初短暫時間，但之後幾乎全年未再發生，此熱點亦可視為偶發熱點。

進一步來說，如果在一個時間段內進行純空間分析，通常很難發現群聚為長期存在（持續群聚）或最近才出現的群聚（新興群聚），簡單解決這個問題的一種方法即是如以上本研究以分年劃分方式的純空間分析；然而，合適的年數通常是未知的，如果納入考慮的年份太少，可能沒有足夠的資訊來檢測已經存在相當長一段時間的低至中等水準的風險群聚；相反地，如果考慮的年份太多，新興的高風險群聚可能被過去各類群聚所稀釋或覆蓋，這也是傳統 GIS 在進行時空分析時，僅將不同時期的空間變化進行比較經常出現的問題與難題。由此可知，對於群聚熱點的分析，以空間模式分析因為受限於時間的資訊，可能造成選擇性的判斷偏誤，解決方案即是使用時空掃描統計分析。故以下將從時間與時空兩個面向，對臺北市住宅竊盜群聚與熱點進行更進一步分析。

第四節 時空掃描統計分析

一、時間掃描分析

臺北市住宅竊盜的時間掃描分析是單純從時間維度對案件的發生進行分析，目的在偵測案件發生的熱點時段。具體來說，指的是案件的發生是否特別集中在某一段時間內，從群聚的觀點而言，亦可說是偵測案件發生時間分布是否發生群聚的情形。分析方式是採用整個時間週期的平均值作為期望值，主要偵測在整個研究時間跨度內案件發生較高的時段，特別是季節性、週期性的波動，利用離散的卜瓦松模型計算每個年份的熱點時間，假定每個視窗的期望值是視窗的時間長度與總時間長度之比乘上總事件數(王占宏，2013)，其餘 LLR、LR 與顯著性水準等檢定方式與空間掃描檢定方法原理相同。

在參數設定方面，本研究設定最大時間群聚大小為總時間長度的 50%，與空間掃描設定 50% 風險人口意義類似，最小時間群聚大小可以指定為特定的天數、月或年，而為減少運算時間，亦可以將案件時間聚合(time aggregation)為時間間隔，這樣做的另一個原因也是為調整週期性的時間趨勢。例如，當使用 1 年的時間間隔時，運用參數設定方式，分析將根據案件數的季節變化自動調整，當使用

7天的時間間隔時，分析將根據工作日(星期別)的影響自動調整(Kulldorff, 2018)，通常在醫學防疫的領域中，會根據疾病潛伏期，做為時間間隔的週期，以偵測流行病是否爆發後形成群聚，並採取各項防疫措施，但在犯罪偵防方面卻很難有客觀標準對時間間隔加以設定；因此，本研究將先分別以年、月及日進行掃描，時間間隔均設定為1單位，總時間長度為3年，根據基本掃描結果再進行更多調整與掃描模型變換。

(一)「年」掃描分析

表 4-4-1 臺北市住宅竊盜「年」掃描摘要表

群聚時間	期望值	觀察值	LLR	RR	p-value
2016/1/1 ~ 2016/12/31	587.40	642	3.75	1.15	*

註：* $p < 0.05$ 。

表 4-4-1 顯示，以3個年度進行掃描，臺北市住宅竊盜年分布的時間群聚在2016年，主要與該年度有最多犯罪數量有關，時間群聚達到顯著水準。

(二)「月」掃描分析

在月掃描分析方面，首先將針對各年度進行逐月掃描，其次再以3年時間總和進行掃描。

1、各年度「月」掃描

表 4-4-2 各年度「月」掃描摘要表

年度	群聚時間	期望值	觀察值	LLR	RR	p-value
2015	2015/12/1 ~ 2015/12/31	49.75	70	3.77	1.42	0.07
2016	2016/1/1 ~ 2016/11/30	537.65	596	4.47	1.16	*
2017	2017/5/1 ~ 2017/5/31	49.75	56	0.39	1.13	0.91

註：* $p < 0.05$ 。

2、2015 年至 2017 年「月」掃描

表 4-4-3 2015-2017 年「月」掃描摘要表

群聚時間	期望值	觀察值	LLR	RR	p-value
2015/7/1 ~ 2016/11/30	832.96	925	9.64	1.23	**

註：** $p < 0.01$ 。

掃描結果顯示，若以分年的月掃描，僅有 2016 年出現時間群聚，時段與年掃描相當；而以全部 3 年期間總和掃描，臺北市住宅竊盜最集中的月份為 2015 年 7 月 1 日至 2016 年 11 月 30 日，時間長達約 1 年半時間，亦即在此段期間中，住宅竊盜案件較多，時間也較為密集(圖 4-4-1)

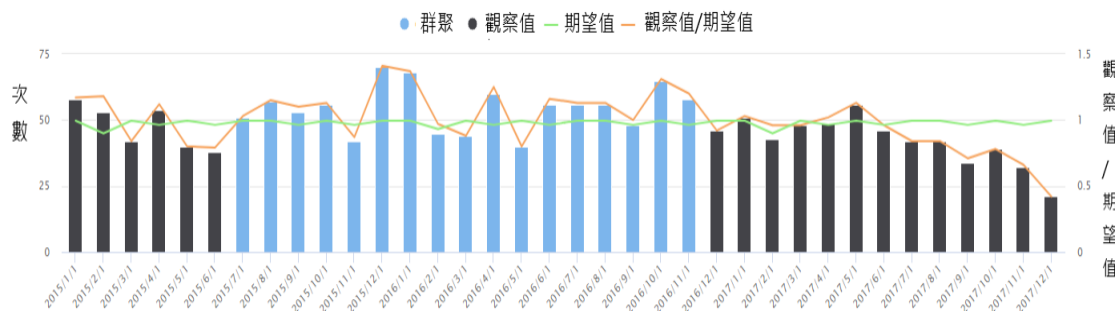


圖 4-4-1 2015-2017 年「月」掃描群聚月份圖

(三)「日」掃描分析

1、各年度「日」掃描

表 4-4-4 各年度「日」掃描摘要表

年度	群聚時間	期望值	觀察值	LLR	RR	p-value
2015	2015/12/19 ~ 2015/12/23	8.02	23	9.31	2.89	*
2016	2016/4/24 ~ 2016/4/27	6.42	17	6.01	2.66	0.26
2017	2017/1/1 ~ 2017/1/1	1.60	7	4.92	4.38	0.59

註：* $p < 0.05$ 。

各年度逐日掃描結果顯示，僅有 2015 年達到顯著水準，其餘兩年案件發生時間分別為 4 月與 1 月元旦假期，但並不具統計顯著水準，顯示在時間發生的分

布為隨機，原因可能是由於局部地區的案件數增加尚不足以使整個地區的曲線呈現上升趨勢，故單純以時間偵測方式，發現群聚的時間就會延後，此時若較多案件發生時間出現跨年度情形，使用分年掃描，可能會造成原本群聚的時間被切割，此與地理邊緣界線的問題相同，更無法檢測到顯著群聚情形；因此，建議對於總體時間的掃描時間跨度不宜太短。此外，犯罪的形成經常是時間積累的結果，就社會解組理論而言，某地區的犯罪發生通常是長時間所形成的結構因素；是以，較長的時間觀察對於某地區犯罪趨勢變化才能有更深入的理解。

2、2015 年至 2017 年「日」掃描

表 4-4-5 2015-2017 年「日」掃描摘要表

群聚時間	期望值	觀察值	LLR	RR	p-value
2015/8/10 ~ 2017/1/1	820.60	920	11.25	1.25	**

註：**p < 0.01。

日掃描結果顯示，全部 3 年期間內，臺北市住宅竊盜最集中的時段為 2015 年 8 月 10 日至 2017 年 1 月 1 日，與月掃描相比較，在各項數據相當接近，不同是群聚時間往後延 1 個月，此結果說明無論以月掃描或日掃描，竊盜犯罪集中均為自 2015 下半年至 2016 年底期間。由下圖 4-4-2 可看出，在群聚區間內，觀察值與期望值的比值與前後區間相比呈現更高的趨勢。

綜合年、月及日 3 種掃描的結果，可初步推論在 2015 年至 2017 年間，住宅竊盜犯罪時間群聚很重要是受到 2016 年有較高犯罪數量影響，且此影響的發生是從 2015 年下半年逐漸產生的積累效應，到 2016 年底達到高峰，所以若將年度加以切割後，時間群聚的顯著現象便不再顯著；而自 2016 年達到高峰期後，2017 年無論在各別年度或整體的檢測中均沒有出現群聚情況；從圖 4-1-1 可看出時間群聚的整體趨勢從 2015 年逐漸往上攀升，2016 年處於最高峰，2017 年後逐漸下降的趨勢。

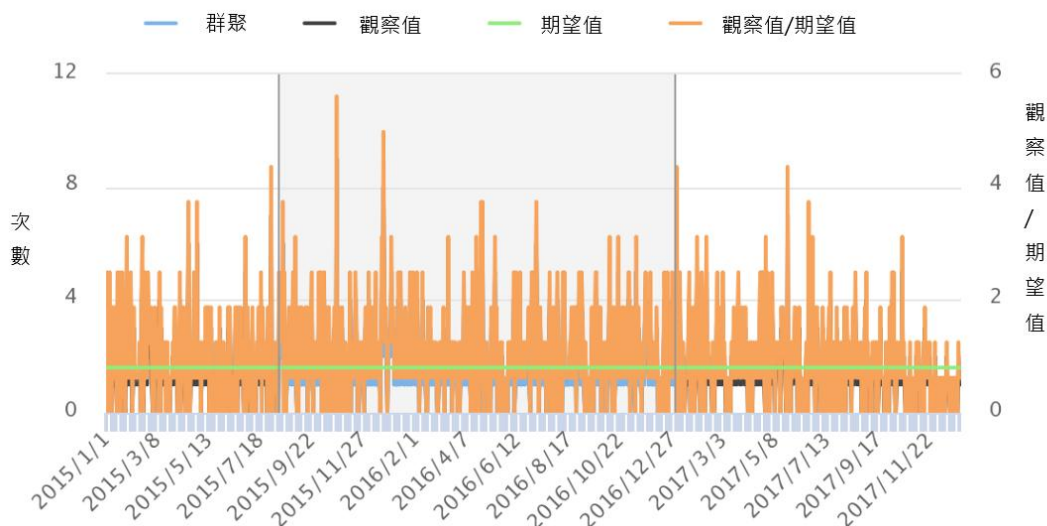


圖 4-4-2 2015-2017 年「日」掃描群聚月份圖

(四)「季節性」掃描分析

時間數列在概念上可拆解成 4 個項目：趨勢(trend)、循環(cycle)、季節性(seasonal)、不規則項(irregular)，季節性是反映時間序列在不同年份的相同季節所呈現之週期性變化(黃月盈, 2013)，進行分析時資料數量通常需要至少 3 年各月或各季的統計資料，故以下使用季節性分析以了解臺北市住宅竊盜發生的時間週期，掃描方式分為月掃描與週掃描。

表 4-4-6 臺北市住宅竊盜季節性「月」掃描摘要表

群聚時間	期望值	觀察值	LLR	RR	p-value
12/1 ~ 1/31	248.76	293	4.37	1.21	*

註：* $p < 0.05$ 。

表 4-4-7 臺北市住宅竊盜季節性「日」掃描摘要表

群聚時間	期望值	觀察值	LLR	RR	p-value
12/29 ~ 1/4	4.81	16	8.06	3.34	**

註：** $p < 0.01$ 。

對臺北市住宅竊盜的季節性掃描結果，月掃描時間群聚集中在歲末年終與跨年之際，期間約 2 個月；就臺灣社會的生活型態而言，此時工作通常較為忙碌，

同時各類活動也較多，如：跨年活動、尾牙等，同時居家也正為即將到來的農曆春節進行各項採買，由於戶外活動較多，居家時間減少，故從機會理論或日常活動理論來看，此時居家防護力減弱，提供犯罪更多的機會；相對地，歲末年終之際，通常有更多的財物需求，也增加犯罪者動機。

從週期性的日掃描觀察，結果顯示較為精確的群聚時間為 12 月 29 日至 1 月 4 日，時間恰為我國每年元旦連續假期，通常民眾多會利用此假期全家外出旅遊，此提供住宅竊盜犯罪絕佳的犯罪機會。

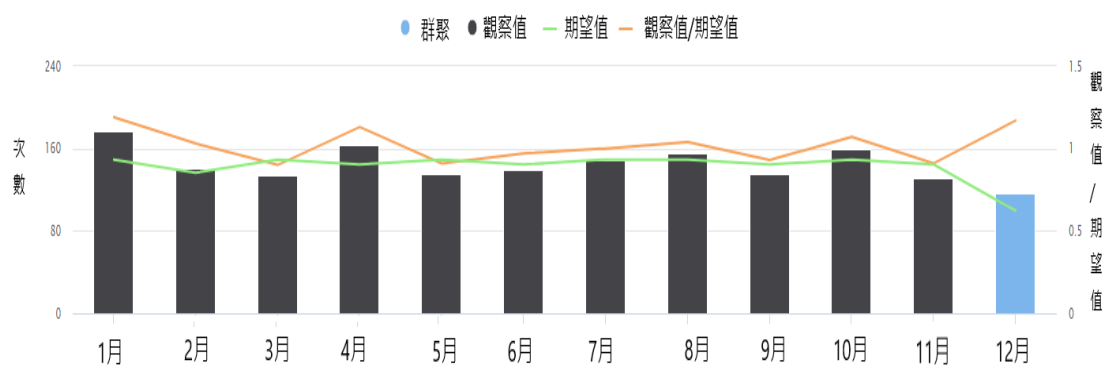


圖 4-4-3 2015-2017 年「日」掃描群聚月份圖

二、時空掃描分析

犯罪熱點時空分布會隨著時間不同、地區差異而發生變化。源於經濟學理論的「二八法則」指出，幾乎 80% 的事件只和 20% 的人或事件有關，就環境犯罪學而言，此法則指的是絕大部分犯罪發生在很小一部分地區或時段內，且僅與少數犯罪行為者有關(Clarke & Eck, 2003)，此意味著犯罪分布時間及熱點形成並非無章可循。在本章前面部分，本研究分別以空間與時間掃描，對 2015 年至 2017 年臺北市住宅竊盜進行熱點與熱時的偵測，對於各年度的犯罪空間群聚與熱點有許多具體的結果，同時在時間掃描分析方面，也得知 3 年間犯罪發生集中時間，且從時間序列分析中更清楚了解臺北市住宅竊盜發生的週期性。然而，雖然我們知道群聚之地點與時間，但我們對於犯罪熱點發生的確切時間並不能確定，亦即犯罪時間與空間兩者同時出現的情況為何仍無法知曉；對此問題，時空掃描統計可有效對於空間熱區與時間集中兩種問題加以解決，這也是 Kulldorff 掃描統計與一般 GIS 群聚偵測方法很大不同之處。

就時空掃描統計範疇而言，針對時間與空間的複雜關係，Kulldorff 在 1998 年首先提出「回顧性」時空掃描統計方法，2001 年和 2005 年又分別提出「前瞻性」時空掃描統計方法和「時空重排」掃描統計方法，這幾種方法都是針對時間、時空兩種資料進行掃描統計，但側重點有所不同。「回顧性」掃描主要是使用歷史資料對於時空某現象的出現是否具有群聚或隨機性進行檢測，本質上乃是對於過去某段時間與地區時空群聚情形進行觀察與分析；相對於「回顧性」掃描，「前瞻性」掃描則是致力於檢測仍處於「活動」(alive)狀態的群集，這意味著在所蒐集資料中最後一段時間內仍然存在具有較高風險的群聚，在統計過程中會針對許多可能的時間長度進行調整，由於不斷重複對時間週期進行多次測試，若將資料隨時更新，即可構建出對於群聚的定期監視系統，不論群聚是過去或近期所形成，只要有異常群聚現象，均可立即發出預警，此對於司法警政機關而言甚具有實用性。

「時空重排」掃描統計則是僅使用案例發生數量、地點及時間進行估計，其基本假設是研究區域內的各個小區域母群體增長速度一致，故此方法的優點是不需要額外的母群體資料，並且自動調整案件資料中的母群體密度和季節變化，使用時機通常是對於母群體不確定情況下，例如：使用人口資料做為母群體時，由於無法取得詳細人口資料，或某地區多數人口沒有登記戶籍等，官方的統計數據可能造成較大的偏差，此時運用「時空重排」可忽略母群體不確定性所產生的影響(Kulldorff, 2001)，而「前瞻性」掃描進行預警時，由於母體資料(如戶籍人口、住宅數量)可能無法隨之立即更新，亦可考慮搭配「時空重排」掃描統計同時分析。

值得注意的是，「時空重排」掃描統計雖有優點，但是缺點在於若偏離其基本假設，探測到的群聚既有可能是案件數量增加，也不排除是由於不同時間、不同地區母群體分布的改變造成；換言之，若某些地區母群體增長速度高於其它地區，則「時空重排」掃描可能發出錯誤警訊。因此，如果研究時間較長，各區域母群體增長速度不一致，則不宜使用此方法，由於本研究具備各年度詳細住宅統計數，且各年度數值變化並不大，故無使用「時空重排」掃描統計之必要。以下分別使用「回顧性」與「前瞻性」時空掃描，針對臺北市住宅犯罪資料進行分析。

(一)「回顧性」時空掃描

1、掃描原理

「回顧性」時空掃描統計的原理與空間掃描統計類似，主要係將平面上的掃描統計拓展到時空維度，即為圓形或橢圓形掃描視窗增加時間維度，掃描視窗變為圓柱或橢圓柱形，掃描視窗的底面覆蓋空間域上的聚集區，掃描視窗的高度代表時間的長度，時空掃描統計透過位置和大小變化的視窗沿著時間軸和空間軸移動，完成對時空區域內的資料的掃描統計。掃描過程中，同樣以區域中心點座標為掃描視窗底面中心，柱的底面半徑和高度不斷變化，且分別對應視窗覆蓋的空間區域和時間段的變化，計算不同位置和大小視窗內的 LLR 值，並進行 999 次蒙地卡羅模擬檢定。

2、參數設定

在參數設定方面，由於增加時間因素並考量使用不同模型與電腦計算資源，因此不進行尺度較小的日分析，全部採用月尺度分析，相關設定如下：

- (1)最大時間範圍設定為研究時間的 50%，此設定可自由調整，但最高上限不能超過 90%。
- (2)空間最大掃描範圍以設定母群體 50% 做為掃描半徑。
- (3)時間間隔的設定設為「月」，聚集單位為 1。
- (4)在空間掃描方面，本研究曾分別以圓形與橢圓形視窗進行掃描比較，目的在凸顯兩者掃描方式的差異；在本單元中，研究重點將聚焦於時空的關係，考量時空計算量相當龐大，若以橢圓形進行，電腦資源恐有不足，為顧及可操作性，故僅以圓形視窗掃描進行測試，不再重複使用橢圓形視窗掃描進行比較。
- (5)群聚掃描報告採非重疊群聚(non-overlapping clusters)，即第二群聚的結果報告不會與第一群聚相重疊，可減少判斷重疊區域的繁複程序及誤判。

有鑑於時間群聚分析結果顯示分年度掃描存有將時間切割後分析的侷限，故以下掃描將以 2015 年至 2017 年資料進行整體趨勢之觀察。

3、分析結果

在「回顧性」時空掃描方面，結果呈現 57 個可能群聚里，14 個群聚區，其中 6 個群聚區達到統計顯著。第一個最大可能的群聚區為中山區，共有 15 個群聚里，對數概似率 24.13， $p\text{-value} < 0.01$ ，群聚時間為 2015 年 5 月 1 日至 2016 年 3 月 31 日；第二個可能的群聚區為中正區螢雪里，對數概似率為 23.85， $p\text{-value} < 0.01$ ，群聚時間為 2015 年 10 月 1 日至 2015 年 10 月 31 日；第三可能群聚區為松山區安平里，對數概似率為 19.41， $p\text{-value} < 0.01$ ，群聚時間為 2015 年 5 月 1 日至 2016 年 4 月 30 日；第四可能群聚包括中正區、萬華區與北投區，包含 10 個里，對數概似率為 17.45， $p\text{-value} < 0.01$ ，群聚時間為 2017 年 3 月 1 日至 2017 年 5 月 31 日；第五群聚區為北投區與士林區，共 3 個里，對數概似率為 15.79， $p\text{-value} < 0.05$ ，群聚時間為 2016 年 7 月 1 日至 2016 年 7 月 31 日；第六群聚區為萬華區與中正區，對數概似率為 14.73， $p\text{-value} < 0.05$ (表 4-4-8, 表 4-4-9, 圖 4-4-4)。

表 4-4-8 2015-2017 年臺北市住宅竊盜圓形時空掃描摘要表

群聚	群聚時間	群聚半徑	期望值	觀察值	LLR	RR	p-value
1	2015/5/1 ~ 2016/3/31	0.86 km	17.24	53	24.13	3.14	**
2	2015/10/1 ~ 2015/10/31	0 km	0.09	7	23.85	81.19	**
3	2015/5/1 ~ 2016/4/30	0 km	1.99	16	19.41	8.11	**
4	2017/3/1 ~ 2017/5/31	0.81 km	2.62	17	17.45	6.53	**
5	2016/7/1 ~ 2016/7/31	3.43 km	0.08	5	15.79	63.01	*
6	2016/4/1 ~ 2016/4/30	0.54 km	0.50	8	14.73	16.13	*

註：** $p < 0.01$ ；* $p < 0.05$ 。

表 4-4-9 2015-2017 年臺北市住宅竊盜時空群聚區(里)

群聚	區(里)
1	中山區：中原里、中吉里、中山里、康樂里、中庄里、聚盛里、正義里、中央里、興亞里、新生里、民安里、正得里、聚葉里、正守里、松江里
2	中正區：螢雪里
3	松山區：安平里
4	中正區：光復里、黎明里、建國里 萬華區：福星里、萬壽里、西門里、新起里 大同區：玉泉里、建明里、建功里
5	北投區：湖田里、湖山里 士林區：菁山里
6	萬華區：騰雲里、凌霄里、忠貞里、新和里 中正區：永昌里

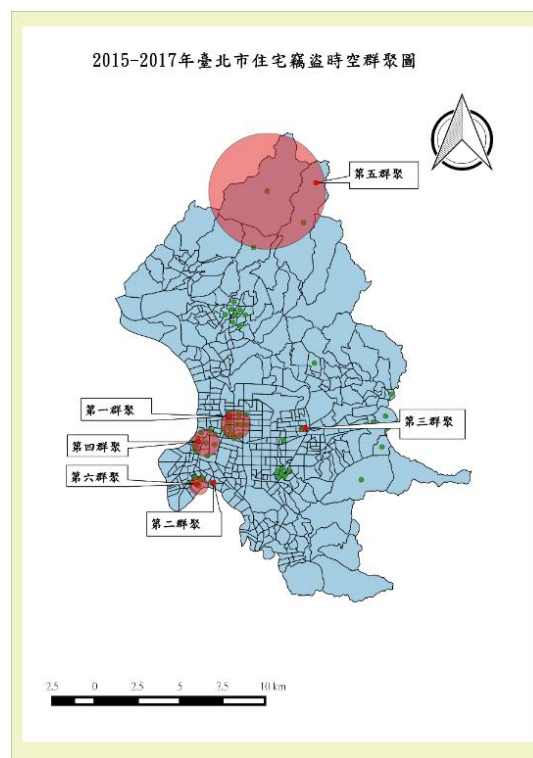


圖 4-4-4 2015-2017 年時空掃描群聚圖

根據上表進一步分析；首先，比較同為圓形視窗時空掃描與空間掃描發現，就數量而言，純空間掃描偵測的群聚里數量較時空掃描為多，這主要是因為時空掃描必須考量時間與空間同時群聚兩項因素的結果；具體來說，空間群聚的熱點若未同時涵蓋在所設定時間範圍內群聚，將會加以排除，此結果說明若僅以空間群聚熱點對犯罪群聚進行研判，可能造成許多警政資源的不當配置，尤其是在警力缺乏與勤務繁多情況下，徒增工作負擔；是以，對於提升警政效能，熱點與熱時合併的時空群聚熱點無疑是擬定肅竊方案與執行計畫的更加選擇。

其次，就各群聚區的各项數據觀察，第一群聚區群聚時間將近 1 年（2015 年 5 月 1 日至 2016 年 3 月 31 日），全部集中在中山區，其中正得里在前面單元中本研究將其分類為持續性熱點，其相對風險率為 3.14，在所有群聚中較低，可稱為長期性的中低度危險熱區。第二群聚區為中正區瑩雪里，群聚時間集中在 2015 年 10 月，相對風險率高達 81.19，為所有群聚最高，查其原始資料，竊盜發生日期全部在 10 月 23 日（星期五）同一地點，此後便未再發生，建築物形式為大樓，很顯然此為連續竊盜的偶發性群聚。第三群聚區松山區安平里，群聚時間雖為 2015 年 5 月 1 日至 2016 年 4 月 30 日，但分成兩個區間，分別為 2015 年 5 月間發生 10 件及 2016 年 1 月至 4 月間 6 件竊盜案，可說是典型的間歇性群聚熱點。第四群聚區集中在中正、萬華及大同區交界，群聚半徑不到 1 公里，群聚時間為 2017 年 3 月 1 日至 2017 年 5 月 31 日共 2 個月，相對風險率 6.53 在各群聚中並不突出，似乎為偶發群聚，但若與發生在 2016 年 4 月間的第六群聚區合併觀察，可歸類在間歇性群聚熱點類型。值得注意的是，位處臺北市北部的第五群聚北投與士林區，群聚時間為 2016 年分布在 7 月，共有 5 件住宅竊盜案件，但是相對風險率卻高達 63.01，主要原因是此群聚區外的地區屬於「地廣人稀」，且群聚時間較為密集，但此後並未在有群聚發生，亦可歸類為偶發性群聚熱點。

由以上對時空群聚的分析可知，對於群聚發生是否偶發或長期性等，若是以主觀人為的方式對空間進行時間區隔，將產生許多偏誤，而加上群聚時間後，群聚的性質即可一目了然。惟本研究仍再次強調，群聚的性質仍會隨著時間的積累產生改變，研究時間範圍越長，資料越多，最後形成所謂的「大數據」，此對於群聚的真實性質才能有更深入認識。

(二)「時間趨勢中的空間變化」掃描分析

「時間趨勢的空間變化」分析(spatial variation in temporal trends analysis, SVTT) 為流行疾病監測的重要工具，對於制定預防疾病的計劃有相當大幫助，並可藉以衡量預防工作的進展、影響和效力，此方法旨在檢測在時間趨勢中具有異常的空間區域(Moraga & Kulldorff, 2013)。就本研究而言，此分析方法除可檢測 2015 年至 2017 年臺北市住宅竊盜犯罪率的整體趨勢變化外，重要是偵測時間趨勢在不同的地理區域中的差異，亦即有異常趨勢的犯罪群聚區域。

SVTT 統計亦屬於「回顧性」掃描方式，其原理係以總體趨勢為基礎條件，進而估計各地區趨勢的差異，其視窗掃描方式與時空掃描方式相同，而趨勢估計的方式則是透過使用卜瓦松迴歸(poisson regression)進行，以時間為自變項，案例數為依變項，隨著時間變化的母群體為偏移量(offset)，根據趨勢估計以調整各別地區和時間的期望值，由於趨勢估計的不同，視窗內部和外部的調整也將不同，最後使用新的期望值找出所有視窗的 LLR(Ramasamy & Janarthanan, 2019)。

從分析結果顯示，2015 年至 2017 年臺北市住宅竊盜犯罪的時間總體趨勢年下降率為 9.05%，共偵測 8 個趨勢異常的群聚，其中 2 個群聚區達到 P 值小於 .05 的顯著水準。第一個最大可能的群聚區中心座標為大同區朝陽里(25.054752 N, 121.512726 E)，共有 7 個群聚里，全部集中在大同區，內部時間趨勢每年增長 217.30%，外部時間趨勢每年下降 10.26%，對數概似率 8.81，相對風險率為 0.92，p-value=0.05；第二個可能的群聚區中心座標為南港區中南里(25.046181 N, 121.607006 E)，包含南港區 7 個里，內湖區 4 個里，信義區 1 個里，群聚內部時間趨勢每年增長 54.91%，群聚外部時間趨勢每年下降 11.64%，對數概似率 8.80，相對風險率為 1.09，p-value = 0.05(表 4-4-10，表 4-4-11)。

SVTT 掃描統計的分析結果與時空群聚相比較，除大同區建功里相同外，其餘各群聚區熱點均無重複，其被列為異常群聚主要原因在於這些群聚區在臺北市總體犯罪率下降趨勢中卻呈現「逆勢上揚」的異常趨勢，其特點為相對風險性並非很高，即便未達統計顯著水準之其他群聚亦是如此，故在一般的時空群聚分析聚焦於較高相對風險率的標準下很難被偵測出來，甚至此類群聚的相對風險率經常低於 1，與周遭地區相比尚屬於「冷區」。但這些趨勢異常的群聚卻是警政單位在未來特別須加以注意可能爆發成新的熱點，若以群聚性質區分，此類群聚在

某種程度上可視之為「新興熱點」。從表 4-4-10 可看出，大同區老舊社區居多，群聚偵測結果相對風險率雖然仍低於區域外，但群聚區內年增長率異常高，似有犯罪轉移與逐漸集中現象發生。南港與內湖區在空間群聚的偵測中從未出現熱區，但在此卻成為趨勢異常的區域，可能與此兩個行政區為臺北市高科技產業聚集地，有較多單身或年輕人開始入住，白天工作外出，導致監控較低，至於是否有其他因素造成其不降反升的犯罪趨勢，相當值得進一步關注與分析。

表 4-4-10 異常趨勢之群聚摘要表

群聚	內部時間趨勢	外部時間趨勢	期望值	觀察值	LLR	RR	p-value
1	217.30% (年增長率)	10.26% (年下降率)	23.99	22	8.81	0.92	*
2	54.91% (年增長率)	11.64% (年下降率)	84.76	92	8.80	1.09	*

註：* $p < 0.05$ 。

表 4-4-11 異常趨勢之群聚區(里)

群聚	區(里)
1	大同區：朝陽里、建功里、永樂里、星明里、延平里、大有里、建泰里
2	南港區：中南里、中研里、新光里、東明里、南港里、東新里、新富里、重陽里、仁福里、成福里、三重里、西新里、聯成里、百福里、鴻福里、合成里、萬福里 內湖區：蘆洲里、石潭里、五分里、南湖里 信義區：中行里

(三)「前瞻性」時空掃描分析

1、「前瞻性」掃描原理

所謂「前瞻性」分析，乃是對現存的群聚進行偵測。用數學符號表示，假定 $[T_1, T_2]$ 為分析的時間區間， s 和 t 是掃描視窗的時間起點和終點，則所有掃描視窗均應滿足 $T_1 \leq s \leq t \leq T_2$ ；故以「前瞻性」掃描出的結果若未特別設定時間範圍，則偵測到的群聚時間是從資料研究時間最後期限向前推算，如本研究若以此方式

掃描，在時間群聚的報告上將如「2017/05/01~2017/12/31」、「2017/05/02~2017/12/31」此格式。

「前瞻性」掃描偵測群聚的數據解釋與「回顧性」掃描相似，包括可能群聚的排序、群聚半徑、群聚時間、LLR、LR 及 P 值等；其中，P 值可以計算「無效預警發生率」(null occurrence rate, NOR)，其定義為 $NOR=1/P$ 天，「無效預警發生率」是指假設不存在群聚性，出現錯誤預警信號的時間間隔；例如，P 值為 0.05，意即每 20 天可能出現 1 次錯誤的預警信號，若 P 值為 0.01，則是每 100 天可能出現 1 個錯誤預警，由此可知 P 值越小，NOR 值越大，群聚程度越高(殷菲、李曉松、馮子建、馬家奇,2007)，在 SatScan 報表中，可查看”recurrence interval”天數，即可知道發出錯誤信號的間隔時間。

2、「前瞻性」掃描實務運用概況

建立犯罪案件預警系統是有效監測犯罪發展情勢的重要措施，早期預警使司法警政機關得以在可能的最早的時間採取對策，對突發或日漸惡化的犯罪進行預防和控制；近年，國內已有許多研究陸續提出對犯罪預測的方法，但多是採用單純時間聚集性方法，例如，使用時間序列分析來預測犯罪之趨勢(李良益,2014；傅鈺婷,2018)，本研究此前也運用時間序列掃描方式進行季節性週期預測，此分析方法對於犯罪預測有其實用性，但也存有許多侷限，主要原因之一在於犯罪的發生與地點有重大關聯，而時間分析方式並未考慮犯罪發生的空間，在實務的運用方面顯有不足。其次，時間序列模型都會遇到的問題，就是長期預測不準確的困難，此乃因時間序列模型是假設未來與過去的情況不變下所做出的預測與預警；因此，隨著時間逐漸增加，環境的變化程度相對愈大，例如，當法律對於某些犯罪處罰做出修改或政府實施重大打擊犯罪方案後，即可能改變犯罪發生的頻率與型態等，這些都會影響預測的準確性；是以，時間序列模型用於犯罪分析是否能夠做為長期的預測與預警仍是值得商榷。

國外對早期預警的研究多數開始於 2001 年之後，紐約市、紐澤西州、華盛頓、鹽湖城、西雅圖及匹茲堡等地先後建立早期預警系統的試驗(Lober et al.,2002)，對各類早期預警方法的實用性和有效性進行驗證，有鑑於純時間群聚性方法的侷限性，國外學者多致力於時空群聚監測方法的研究，其中 Kulldorff 等提出的時空掃描統計已廣被使用。目前，我國基層警政單位均會定期將犯罪案件

匯總陳報，資料屬於聚合(aggregated)性質，所陳報數據基本上已按時間和地區進行分類，如本研究使用臺北市政府住宅竊盜犯罪資料即具備此格式，配合社會經濟資料庫各里住宅戶數統計資料，頗適合時空群聚掃描的分析方法。

現階段在國內使用掃描統計進行預警的研究多為醫學方面，對於犯罪監測及預警的研究尚付諸闕如，過往的研究內容僅是止於「回顧性」分析，主要目的為測試犯罪在地理區域的空間和時空上隨機分布情形，有關時間序列及時間趨勢中的空間變並無討論，而對於未來趨勢的研判乃是觀察「回顧性」分析的結果後進行推測，通常這樣的做法並非毫無根據，也相當具有參考價值，但受限於分析方法，此類推測與預警仍有性質上不同，故本單元將以 Kulldorff 的「前瞻性」時空掃描方式，模擬如何以官方資料進行定期犯罪群聚的監測與預警，以提供司法警政機關參酌。

3、即時監測與預警模式建構

即時監測與預警指的是對於關注現象不斷進行觀察並紀錄結果，所得資料可隨時提供查詢，並根據所設定的臨界值等，對於異常狀況發出警訊，俾進一步採取相應措施。建置即時監測系統非常仰賴科技的支援，除電腦軟硬體設施外，也需要完備的倉儲資料庫以儲存與提取資料；目前，即時監測系統已經應用在各行各業領域，較常見與民眾日常生活相關包括：交通路況、氣象、土石流、環保空汙及疫情等監測。

在犯罪偵防方面，近年因 AI 技術進步，利用影像辨識（如人臉辨識）進行犯罪防治成為主流，相關研究與硬體設備日新月異；而在傳統數據資料領域中，最為所知是針對洗錢犯罪所建置的監測系統，此系統是對於可疑金融交易發出警訊，以利於偵查人員採取防制作為，無論是影像辨識與洗錢防制監測系統等，其主要功能在於補足以人工方式無法對大量資訊處理的缺點。由於電腦對於複雜資訊處理較人工方法識別具有更大優勢，故對於犯罪群聚的監測應切實可行且有其必要性。

基於以上，本研究認為「前瞻性」時空掃描分析早已運用於醫學領域中疫情群聚監測多年，雖然犯罪與疫情群聚有許多現象並不能等同，例如：疫情有傳染或所謂的「爆發」(breakout)現象，但此並不影響統計方法對於群聚偵測的原理，差異在於對於結果的詮釋不同。以下運用「前瞻性」時空掃描分析，模擬對住宅

竊盜犯罪群聚的監測及預警模式。

(1) 參數設定

「前瞻性」時空掃描分析的主要參數設定與「回顧性」掃描基本上類似，包括最大群聚的時間長度和空間掃描半徑均設定為 50%；群聚掃描的時間間隔設定，設定為 1 日。

(2) 模擬資料範圍與工作流程

完備的監測預警掃描工作原則應是對資料庫所有相關資料進行逐日分析。以本研究資料為例，2015 年至 2017 年共有 1,096 天（2016 年為閏年），分析時段可以自由選取，如以 2017 年開始進行全年監測，從 2017 年 1 月 1 日起，逐日分析並發布預警即需重複 365 次，每次掃描為選取發布日向前納入全部累積資料進行掃描。

執行預警掃描時，無論時間間隔為何，基本上每次分析均需經過以下流程：
A. 對資料進行預處理，選擇某特定時間段的住宅竊盜案件資料，製作案件數量檔案，若採人工分析，則每日分析均需添加新資料後的獨立檔案；B. 根據選擇時段，調整母群體資料檔案，考量母群體可能變動過大，此時亦可採用「時空重排」掃描方法測試；C. 利用 SPSS 或 Excel 讀取掃描資訊，並對其進行編輯後另存為 shp、dbf 及 kml 等檔案以供 GIS 系統讀取繪圖。預警系統架構圖如下：

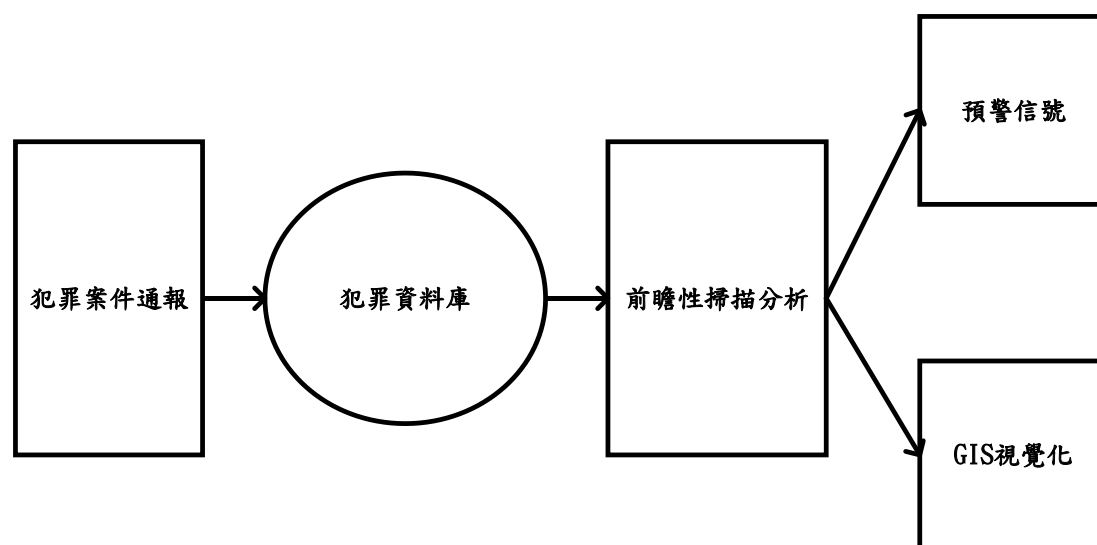


圖 4-4-5 監測預警系統架構圖

上述流程以人工操作過程繁瑣，工作量極大，故此類逐日預警通報系統必須建置自動化掃描功能，並連接數據庫，且最好有許多電腦同時運作，以達到最高效率。本研究限於資訊設備，將逐日發布警報簡化為以月底發布，選取 2017 年 9、10、11 月 3 個月底日為發布日期，期限以發布日期前 6 個月為資料測試範圍，共發布 3 次監測結果，最後再與 2017 年「回顧性」時空掃描進行對照比較。

4、監測與預警分析結果

(1)2017 年 9 月 30 日預警發布結果

以 2017 年 9 月 30 日進行「前瞻性」時空掃描，共掃描到 9 個群聚區，其中僅第一群聚達 P 值小於.05 顯著水準，群聚時間為 7 月 2 日起到 9 月 30 日止，期望值為 0.78，觀察值 8，LLR 為 11.53，相對危險率為 10.58，「無效預警發生率」為間隔 38 日；群聚區中心座標在士林區天福里(25.116077 N, 121.531345 E)，群聚熱點全部集中在士林區，分別為天福里、天祿里、三玉里及天壽里。

表 4-4-12 2017 年 9 月 30 日預警發布結果

群聚	群聚時間	群聚半徑	期望值	觀察值	LLR	RR	p-value
1	2017/7/2 ~ 2017/9/30	0.48 km	0.78	8	11.53	10.58	*

註：* $p < 0.05$ 。

表 4-4-13 2017 年 9 月 30 日預警發布群聚區(里)

群聚	區(里)
1	士林區：天福里、天祿里、三玉里、天壽里

(2)2017 年 10 月 31 日預警發布結果

2017 年 10 月 31 日預警掃描結果，共掃描到 7 個群聚區，第一、二群聚 P 值達統計顯著水準，第一群聚區中心點位置為士林區天祿里(25.113378 N, 121.528557 E)，群聚時間為 8 月 18 日起到 10 月 31 日止，期望值為 0.65，觀察值 8，LLR 為 12.82，相對危險率為 12.64，「無效預警發生率」為間隔 83 天，群聚熱點集中在士林區的 4 個里，但與 9 月 30 日發布結果有些許變動。第二群聚

區中心座標在南港區中南里(25.046181 N, 121.607006 E)，群聚時間為 10 月 3 日起到 10 月 31 日止，期望值為 1.83，觀察值 12，LLR 為 12.60，相對危險率為 6.83，「無效預警發生率」為間隔 77 天，群聚熱點分部在南港區、內湖區及信義區，以南港區占 19 個里數量較多。

表 4-4-14 2017 年 10 月 31 日預警發布結果

群聚	群聚時間	群聚半徑	期望值	觀察值	LLR	RR	p-value
1	2017/8/18 ~ 2017/10/31	0.45 km	0.65	8	12.82	12.64	*
2	2017/10/3 ~ 2017/10/31	2.80 km	1.83	12	12.60	6.83	*

註：* $p < 0.05$ 。

表 4-4-15 2017 年 10 月 31 日預警發布群聚區(里)

群聚	區(里)
1	士林區：天福里、天祿里、天壽里、蘭雅里
2	南港區：中南里、中研里、新光里、東明里、南港里、東新里、新富里、重陽里、仁福里、成福里、三重里、西新里、聯成里、百福里、鴻福里、合成里、萬福里、九如里、玉成里 內湖區：蘆洲里、石潭里、五分里、內湖里、葫洲里、週美里 信義區：中行里、大道里、大仁里、中坡里、永春里、永吉里

(3)2017 年 11 月 30 日預警發布結果

2017 年 11 月 30 日預警掃描結果，共掃描到 10 個群聚區，第一、第二群聚達 P 值達統計顯著水準，第三群聚區 p-value 為 0.092，第一群聚區中心座標為南港區中南里(25.046181 N, 121.607006 E)，群聚時間為 10 月 3 日起到 11 月 30 日，期望值為 3.61，觀察值 16，LLR 為 11.78，相對危險率為 4.69，「無效預警發生率」為間隔 48 天。第二群聚區中心座標為士林區聖山里(25.107410 N, 121.531447 E)，群聚時間為 10 月 5 日起到 11 月 30 日，期望值為 0.36，觀察值 6，LLR 為 11.25，相對危險率為 16.89，「無效預警發生率」為間隔 38 天。第三群聚區僅有松山區東光里，群聚時間為 11 月 15 日起到 11 月 30 日，期望值為 0.04，觀察值

3，LLR 為 9.97，相對危險率為 74.98，「無效預警發生率」為間隔 11 天。

表 4-4-16 2017 年 11 月 30 日預警發布結果

群聚	群聚時間	群聚半徑	期望值	觀察值	LLR	RR	p-value
1	201710/3 ~ 2017/11/30	2.80km	3.61	16	11.78	4.69	*
2	2017/10/5 ~ 2017/11/30	0.48km	0.36	6	11.25	16.89	*
3	2017/11/15 ~ 2017/11/30	0km	0.041	3	9.97	74.98	0.092

註：* $p < 0.05$ 。

表 4-4-17 2017 年 11 月 30 日預警發布群聚區(里)

群聚	區(里)
1	南港區：中南里、中研里、新光里、東明里、南港里、東新里、新富里、重陽里、仁福里、成福里、三重里、西新里、聯成里、百福里、鴻福里、合成里、萬福里、九如里、玉成里、 內湖區：蘆洲里、石潭里、五分里、內湖里、葫洲里、週美里 信義區：中行里、大道里、大仁里、中坡里、永春里、永吉里
2	士林區：聖山里、忠誠里、蘭雅里
3	松山區：東光里

(4)2017 年全年「前瞻性」掃描結果

以上對 2017 年 9 月至 11 月的「前瞻性」預警掃描結束後，12 月份為本研究資料最後月份，故以全年度進行掃描，結果共掃描到 7 個群聚區，第一、第二群聚 P 值達統計顯著水準。第一群聚區中心座標為士林區三五里(25.113448 N, 121.534388 E)，群聚時間為 7 月 7 日起到 12 月 31 日，期望值為 2.55，觀察值 15，LLR 為 14.31，相對危險率為 6.04，「無效預警發生率」為間隔 230 天。第二群聚區中心座標為南港區中南里(25.046181 N, 121.607006 E)，群聚時間為 7 月 6 日起到 12 月 31 日，期望值為 11.03，觀察值 31，LLR 為 12.48，相對危險率為 2.93，「無效預警發生率」為間隔 59 天。

表 4-4-18 2017 年「前瞻性」時空掃描結果

群聚	群聚時間	群聚半徑	期望值	觀察值	LLR	RR	p-value
1	2017/7/7 ~ 2017/12/31	0.86km	2.55	15	14.31	6.04	**
2	2017/7/6 ~ 2017/12/31	2.80km	11.03	31	12.48	2.93	*

註：** $p < 0.01$ ；* $p < 0.05$ 。

表 4-4-19 2017 年「前瞻性」時空掃描群聚區(里)

群聚	區(里)
1	士林區：三玉里、天福里、天祿里、聖山里、天山里、蘭雅里、天壽里
2	南港區：中南里、中研里、新光里、東明里、南港里、東新里、新富里、重陽里、仁福里、成福里、三重里、西新里、聯成里、百福里、鴻福里、合成里、萬福里、九如里、玉成里 內湖區：蘆洲里、石潭里、五分里、南湖里、葫洲里、週美里 信義區：中行里、大道里、大仁里、中坡里、永春里、永吉里

(5) 「前瞻性」與「回顧性」掃描比較

從 2017 年的 9 月起至 11 月的預警發布不難發現，群聚的地區幾乎落在士林區、南港區、內湖區與信義區，因掃描時間範圍設定在警報發布前 6 個月，故群聚預警時間出現在下半年為可預期之情形，但全年度的「前瞻性」掃描結果顯示，無論地區與時間的群聚現象與預警極為相似，同樣集中在這些區域及 7 月起的下半年時段，值得注意的是這些預警的區域熱點與先前純空間、時空掃描顯示的結果有很大不同，反倒與「時間趨勢中的空間變化」掃描分析的結果較為雷同，原本較為「冷門」的南港區、內湖區及信義區再度被列為群聚，此一現象是否因 2017 年犯罪發生較集中在這些區域及時間所造成？為此，本研究使用「回顧性」掃描對 2017 年全年進行時空群聚分析加以驗證說明。

2017 年「回顧性」掃描結果，共掃描到 9 個群聚區，其中第一、第二及第三群聚 P 值達統計顯著水準。第一群聚區中心座標為萬華區福星里(25.047280 N, 121.506451 E)，群聚熱點分布在萬華區、大同區及中正區，群聚時間為 4 月 17 日至 5 月 21 日，期望值為 0.98，觀察值 14，LLR 為 24.37，相對危險率為 14.66。

第二群聚區中心座標為中正區網溪里(25.022230 N, 121.518545 E)，群聚區域全為中正區，時間為5月22日至5月28日，期望值為0.064，觀察值5，LLR為16.91，相對危險率為79.31。第三群聚區中心座標為中山區行政里(25.022230 N, 121.518545 E)，群聚區域全部在中山區，時間為1月23日至6月4日，期望值為7.90，觀察值27，LLR為14.44，相對危險率為3.55（如表4-4-20）。

從「回顧性」時空掃描結果可知，2017年臺北市住宅竊盜犯罪的群聚區域與2015年至2017年3年間的整體時空掃描並沒有很大差異，依舊是大同區、中山區、萬華區及中正區等，這更凸顯出「前瞻性」掃描在功能上的特殊之處。進而言之，「前瞻性」掃描主要目的在於「預警」、「提醒」，即所謂的「防微杜漸」，此掃描方法在流行病疫情預警上經常是對於局部出現但尚未構成嚴重群聚的地區提出警告，以免疫情擴散；雖然犯罪預防的性質與防疫有所不同，但「前瞻性」掃描同樣可對於臺北市過去住宅竊盜的發展情勢提出預警，此前在異常趨勢的分析中曾顯示南港、內湖及信義區相對「冷區」等有不尋常的群聚現象，但該分析並無法判斷群聚時間，而「前瞻性」掃描不但再次確認這些不尋常群聚區域，也進一步提供群聚的時間，此對於司法警政機關的犯罪預防實務工作具有重大實質意義。

表 4-4-20 2017 年「回顧性」時空掃描結果

群聚	群聚時間	群聚半徑	期望值	觀察值	LLR	RR	p-value
1	2017/4/17 ~ 2017/5/21	1.04km	0.98	14	24.37	14.66	**
2	2017/5/22~ 2017/5/28	0.36km	0.064	5	16.91	79.31	**
3	2017/1/23~ 2017/6/4	1.11 km	7.90	27	14.44	3.55	*

註：** $p < 0.01$ ；* $p < 0.05$ 。

表 4-4-21 2017 年「回顧性」時空掃描群聚區(里)

群聚	區(里)
1	萬華區：福星里、萬壽里、西門里、新起里 大同區：玉泉里、永樂里、建功里、菜園里 中正區：光復里、黎明里
2	中正區：網溪里、螢圃里、河堤里
3	中山區：行政里、行孝里、新喜里、行仁里、新福里、松江里、新生里、 江寧里、下埤里、中庄里、江山里、新庄里、恆安里、晴光里、 朱馥里、聚葉里、中央里、中吉里

第五章 臺北市住宅竊盜區位分析

在上一章中，本研究運用掃描統計對臺北市住宅竊盜進行空間、時間及時空群聚探索的面向分析，最後並根據資料模擬建構預警發布模型，對於警政機關的肅竊、防竊提供具體的犯罪時空群聚訊息；在本章中將運用資料探勘的分析方式，嘗試從諸多人文社經與空間屬性因素中篩選出臺北市住宅竊盜群聚的重要區位因素，提供犯罪群聚形成的更多資訊，以強化對於住宅竊盜防治的策略擬定的作為。

第一節 區位分析資料前置處理

進行區位分析前，有 2 項資料前置處理工作；首先，對臺北市住宅竊盜犯罪群聚進行分組，以利於資料探勘的二元分類分析；其次，將所選取的各變項加入空間群聚偵測的共變項分析，以排除與群聚無相關的因素。

一、臺北市住宅竊盜犯罪群聚熱區與冷區分布

本研究以 2015 年至 2017 年臺北市住宅竊盜犯罪全部資料進行風險群聚分析，相對風險率高於 1 為所謂「熱區」，反之則為「冷區」。最大掃描視窗設定為風險母群體的 50%。掃描結果共偵測到 17 個群聚區，其中 P 值達到統計水準.05 的群聚共有 7 個群聚區(熱區有 4 個，計有 61 個里，冷區 3 個，計有 173 個里)。在冷區方面，雖有 3 個群聚區但第三、第四群聚區與第一群聚重疊；而原本對熱區偵測時共有 66 個里，但在冷熱區合併偵測後減為 61 個里，主要原因為冷熱區的交界處不免有重疊問題，故以 LLR 與 P 值較顯著者為優先考量，部分里被併入到冷區範圍。各冷熱區群聚情形如表 5-1-1；5-1-2。

表 5-1-1 2015-2017 年臺北市住宅竊盜冷熱群聚區摘要表

群聚	群聚半徑	期望值	觀察值	LLR	RR	發生率	p-value
1	5.03 km	661.06	502	31.89	0.66	42.5	**
2	0.77 km	46.09	95	20.51	2.21	115.4	**
3	1.71 km	191.42	114	20.22	0.57	33.3	**
4	3.43km	222.05	146	16.69	0.63	36.8	**
5	2.00 km	140.02	199	12.06	1.47	79.5	**
6	4.96 km	8.21	24	10.02	2.95	163.6	*
7	0.72 km	28.04	54	9.62	1.96	107.8	*

註：**p < 0.01；*p < 0.05。灰色欄位為熱區群聚。

表 5-1-2 2015-2017 年臺北市住宅竊冷熱群聚區(里)

群聚	區 里
冷區	中山區：力行里、朱崙里、朱園里、埤頭里
	大安區：大學里、仁愛里、仁慈里、古風里、古莊里、正聲里、民炤里、民輝里、永康里、光明里、光武里、光信里、全安里、住安里、車層里、和安里、昌隆里、法治里、臥龍里、芳和里、虎嘯里、建安里、建倫里、通化里、通安里、敦安里、敦煌里、華聲里、新龍里、群英里、群賢里、義安里、義村里、誠安里、福住里、德安里、黎元里、黎孝里、黎和里、學府里、錦安里、錦泰里、錦華里、龍生里、龍安里、龍坡里、龍門里、龍泉里、龍陣里、龍淵里、龍雲里、龍圖里、臨江里
	中正區：三愛里、文北里、文盛里、文祥里、水源里、永功里、永昌里、幸市里、幸福里、東門里、板溪里、林興里、河堤里、南門里、南福里、梅花里、頂東里、富水里、新營里、網溪里、螢圃里、螢雪里、龍光里、龍福里、龍興里
	文山區：木柵里、木新里、老泉里、忠順里、明義里、明興里、政大里、博嘉里、景仁里、景行里、景東里、景美里、景華里、景慶里、華興里、順興里、萬年里、萬有里、萬和里、萬芳里、萬美里、萬盛里、萬祥里、萬隆里、萬興里、試院里、樟文里、樟林里、

	<p>樟新里、樟腳里、樟樹里、興光里、興安里、興邦里、興旺里、興昌里、興家里、興泰里、興得里、興業里、興福里、興豐里</p> <p>松山區：中正里、中崙里、吉仁里、吉祥里、美仁里、復建里、復盛里、復勢里、復源里、敦化里、福成里</p> <p>信義區：三張里、三犁里、大仁里、中行里、中坡里、中興里、五全里、六合里、六藝里、四維里、正和里、安康里、西村里、松友里、松光里、松隆里、長春里、泰和里、國業里、富台里、惠安里、敦厚里、景勤里、景新里、景聯里、新仁里、嘉興里、廣居里、黎平里、黎安里、黎忠里、黎順里、興隆里、興雅里、雙和里</p> <p>萬華區：日祥里、凌霄里、騰雲里</p>
熱區	<p>士林區：三玉里、天祿里、天壽里、天福里、平等里、菁山里、陽明里、聖山里、蘭雅里</p> <p>大同區：永樂里、玉泉里、光能里、建功里、建明里、朝陽里</p> <p>中山區：中山里、中吉里、中庄里、中原里、正得里、正義里、民安里、康樂里、新生里、聚盛里、聚葉里</p> <p>中正區：光復里、忠勤里、建國里、廈安里、愛國里、黎明里</p> <p>北投區：大屯里、泉源里、湖山里、湖田里、榮華里</p> <p>萬華區：仁德里、日善里、全德里、西門里、和平里、和德里、忠德里、青山里、柳鄉里、頂碩里、富民里、富福里、菜園里、華江里、新安里、新和里、新忠里、新起里、萬壽里、福星里、福音里、綠堤里、糖蔴里、雙園里</p>

二、區位共變項資料敘述統計

本研究依社會解組理論與新機會理論共選取 18 個變項，資料來源主要取自政府公開資料平臺，以里為單位，部分資料並未使用 GIS 處理，此類資料性質為非空間屬性，如學歷、收入中位數、男女比率等；而若干變項則是經由 GIS 運算之數值，如人口、路燈及監視器密度等，已具有空間物理特性，可歸為空間屬性資料；有關各變項資料敘述統計表如下(表 5-1-3)。

表 5-1-3 臺北市住宅竊盜區位變項敘述統計表

變項名稱	最小值	最大值	平均數	標準差
低教育程度率	0.22	0.68	0.40	0.10
社區發展協會比率	0.00	1.12	0.15	0.16
區里活動中心比率	0.00	0.99	0.18	0.13
投票率	53.83	83.18	70.55	3.43
人口密度	62.00	112964.00	29974.16	20040.50
社會人口增加率	-90.94	53.10	-5.99	9.91
粗離婚率	0.59	4.52	2.06	0.52
戶量	1.73	3.05	2.56	0.22
老化指數	37.05	335.52	133.58	46.89
收入中位數	540.00	1269.00	747.36	122.17
收入標準差	82.63	2888.13	222.23	223.28
地價	20.80	94.80	60.03	15.72
老人戶數比率	3.58	233.88	33.51	22.52
低度用電戶住宅率	1.71	29.38	7.09	3.52
單獨生活戶數比	19.97	59.74	34.98	6.98
路燈密度	51.00	2372.00	995.94	452.03
監視器密度	0.00	410.00	72.94	57.02
警力配置	0.00	39.53	8.90	4.11

三、群聚共變項檢定

表 5-1-3 所列變項多為過去文獻中常用來解釋住宅竊盜犯罪之區位因素，其中不少變項已被證實與住宅竊盜具有高度相關，但變項的解釋力在不同時空中經常會有不同的效果，就如同日常活動理論的標的物也是隨著時空不斷產生改變。因此，本單元目的即是將這些變項納入臺北市住宅竊盜的掃描分析中，以檢測這些變項是否與本研究的群聚具有相關性，若無相關性則將予以刪除，可避免造成後續資料探勘不必要及可能錯誤的分析結果(有關群聚共變項分析的說明請參閱第三章第三節研究方法)。

本研究群聚共變項掃描方法為空間熱區掃描，掃描半徑設定為風險母群體

50%，掃描結果定在 P 值.05，除置入案件數、住宅數及座標檔案外，另增加共變項的檔案進行掃描。原本在不加入共變項時，判定群聚顯著性即以是否小於.05 做為標準，故若加入共變項再次進行群聚偵測，P 值仍維持.05 以下，則表示加入此變項對群聚並沒有產生影響，判定此因素與群聚形成不具相關性；反之，群聚顯著性消失，則選取此變項進入下階段資料探勘分析，此類似共變數同質性檢定原理。

表 5-1-4 檢定結果顯示，共有 3 個變項 P 值小於.05，分別為「社區發展協會比率」、「區里活動中心比率」及「戶量」，表示與本研究住宅竊盜群聚較無相關性，而其餘 15 個變項在重新掃描後，群聚顯著性消失，這也意味著這些變項與本研究之住宅竊盜犯罪群聚有一定程度的相關性，故將以此 15 個變項進行區位因素分析。

表 5-1-4 臺北市住宅竊盜群聚共變項檢定

變項名稱	p-value	變項名稱	p-value
低教育程度率	1.00	收入中位數	0.96
社區發展協會比率	0.0026	收入標準差	1.00
區里活動中心比率	0.0065	地價	0.96
投票率	1.00	老人戶數比率	1.00
人口密度	1.00	低度用電戶住宅率	1.00
社會人口增加率	1.00	單獨生活戶數比	1.00
粗離婚率	0.91	路燈密度	1.00
戶量	0.0012	監視器密度	0.079
老化指數	1.00	警力配置	1.00

註：P 值小於.05 表示此變項與群聚無相關性。

(一)變項共線性診斷

本研究所採取的決策樹等資料探勘分類分析，基本上並不會受到變項共線性太大影響，但為了解以上所選取變項之間是否存有高度相關現象，仍對以上 15 個變項進行共線性診斷(collinearity diagnostics)。

一般而言，當自變數間的相關達到 0.85 以上即具有明顯的共線性威脅，常

用的判定方式則是透過由容忍值 (tolerance)或變異數膨脹因素(variance inflation factor; VIF)。經 VIF 診斷，本研究所選取之變項 VIF 值均低於 10，顯示並無嚴重共線性現象(Midi, Sarkar, & Rana, 2013)，數值如表 5-1-5。

表 5-1-5 區位變項共線性檢定

變項名稱	VIF	變項名稱	VIF
人口密度	2.818	低度用電戶住宅率	2.031
社會人口增加率	1.054	單獨生活戶數比	3.614
粗離婚率	1.818	路燈密度	1.993
老化指數	2.036	監視器密度	2.262
收入中位數	5.523	警力配置	1.864
收入標準差	1.143	低教育程度率	5.857
平均住宅價格	2.727	投票率	1.729
老人戶數比率	1.348		

第二節 臺北市住宅竊盜區位變項分析

經過以上資料前置處理後，本階段以臺北市住宅竊盜群聚為目標變項，資料計有 234 筆，分為冷區(173 筆)與熱區(61 筆)，並投入群聚共變項初步篩選的 15 個變項，據此建立資料探勘的二元分類與預測模式，以發掘形成不同性質群聚的重要區位因素；主要分析工具為建立在 R 語言的 RATTLE 套件，其優點為可快速完成變項選取、建構模型和評估，無須反覆修改程式，但對模型進行微調與檢查時，仍以返回到 R 的環境中進行較為適當；以下分別以決策樹、條件決策樹、隨機森林、推進法及支援向量機進行分類分析。

一、決策樹分析

(一)樣本與參數設定

本研究在決策樹演算法過程中，主要以 rpart 套件建立分類及相關遞迴劃分(recursive partitioning approach) cart 的演算法，初步將樣本分割為 3 部分，其中 70% 為訓練樣本(training samples)用來建立模型，共有 163 個訓練樣本；15% 為驗證樣本(validation samples)用來檢測當訓練樣本訓練出多個模型後，使用各個模型對

驗證樣本數據進行預測，並記錄模型準確率，以找出效果最佳的模型；剩餘 15% 為測試樣本(testing samples)，功能在於透過訓練樣本和驗證樣本得出最優模型後，使用此部分資料進行模型預測，用來衡量該最優模型的效能和分類能力。

在其他參數設定方面，RATTLE 有幾項可調整的設定：1、「min split」表示每個節點中所含樣本數的最小值，預設值為 20；2、「min bucket」則表示每個葉節點中所含樣本數的最小值，預設取 Min Split 設定 1/3 的四捨五入值，故數值為 7；3、「complexity」指複雜度參數，預設值為 0.01，表示在建構模型的過程中僅保留可以使得模型擬合程度提升 0.01 及以上的節點，該參數的作用在於可以修剪對模型貢獻度不大的分支，以加強演算的效率；4、「max depth」預設值為 30，此項功能可控制樹成長的高度，即是設定節點層次的最大值，其中根節點的高度為 0，依此類推(黃文、王正林，2016)；本研究首先以基本設定進行分析。

(二)分析結果

以 rpart 函數及基本設定建立的決策樹分析，共得出 4 條葉節點規則(如圖 5-2-1，表 5-2-1)。

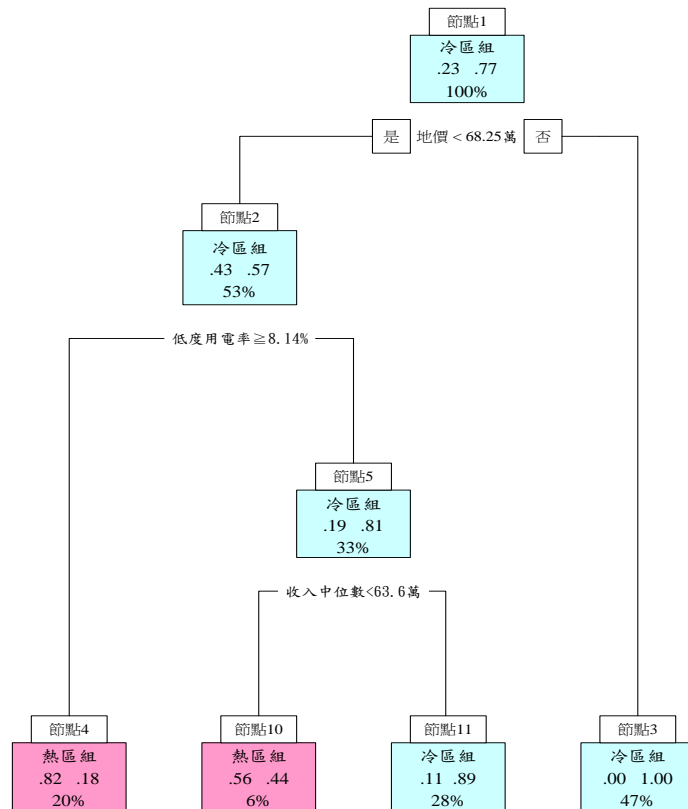


圖 5-2-1 臺北市住宅竊盜冷熱區決策樹分類規則圖

表 5-2-1 臺北市住宅竊盜冷熱區決策樹分類規則表

規則	葉部結點	分類條件
1	3	IF 每坪地價 ≥ 68.25 萬 THEN 1.類別為 冷區組 2.本組中：冷區組機率 100%；熱區組機率 0% 3.占訓練樣本數 47%
2	11	IF 每坪地價 < 68.25 萬 AND 低度用電率 $< 8.14\%$ AND 收入中位數 ≥ 63.6 萬 THEN 1.類別為 冷區組 2.本組中：冷區組機率 89%；熱區組機率 11% 3.占訓練樣本數 28%
3	10	IF 每坪地價 < 68.25 萬 AND 低度用電率 $< 8.14\%$ AND 收入中位數 < 63.6 萬 THEN 1.類別為 熱區組 2.本組中：冷區組機率 44%；熱區組機率 56% 3.占訓練樣本數 6%
4	4	IF 每坪地價 < 68.25 萬 AND 低度用電率 $\geq 8.14\%$ THEN 1.類別為 熱區組 2.本組中：冷區組機率 18%；熱區組機率 82% 3.占訓練樣本數 20%

從圖 5-2-1 中可看出，在建構決策樹的過程中使用變項有「地價」、「低度用電率」及「收入中位數」3 項，階層的高低顯示這 3 種變項對決策樹分支的重要性；表 5-2-1 則為臺北市住宅竊盜冷熱區變項分類樹詳細分析結果，表中所呈現是以「IF」～「THEN」條件式語句方式來說明分類規則和最終分類結果；此種規則表達方式凸顯的是變項的依賴關係，與因果關係解釋不同，故採用機率方式說明此種情況所發生的可能性。

根據以上決策樹分析結果顯示，對臺北市住宅竊盜犯罪群聚冷熱區具最大影響的區位變項為「地價」，其臨界值在每坪 68.25 萬元，大於或等於此數值，無須考慮其他變項，立即判定為犯罪群聚冷區，且機率為 100%；第二冷區地價每坪雖低於 68.25 萬元，但低度用電率少於 8.14%(較少空屋率)，收入中位數高於或等於 63.6 萬元，則仍可能是冷區，機率為 0.89。在熱區方面，地價每坪低於 68.25 萬元，低度用電率少於 8.14%，但收入中位數低於 63.6 萬元，則有 56% 成為熱區；而每坪地價低於 68.25 萬元，但只要低度用電率高於或等於 8.14%(較多空屋率)，無須考慮收入，則判定為熱區，機率為 82%。

綜上，傳統決策樹對臺北市住宅竊盜犯罪的區位變項分析，結果顯示地價較高、空屋率較低及較高的收入的地區，不易成為犯罪群聚的熱區，此符合社會解組理論的高收入低犯罪的論述，進而解釋具有高收入者通常也居住在高地價區域，而空屋率較高則易形成熱區，此即日常活動理論所指監控力減低所造成，同時空屋率高也代表犯罪標的可見性增加。

(三)模型調整

以上所使用為訓練樣本與模型基本設定，並無法知曉此模型的鑑別力如何，是否需進一步調整以達更優化的模型；為此，先以 ROC 操作特徵曲線進行評估，其餘模型之績效綜合評估將在全部分析後一併比較說明。

對於模型調整有許多方式，常見基本的兩種方式為：第一，重新分割訓練樣本與測試樣本比例；第二，對各輸入參數進行修改。無論使用哪一種方式或同時進行調整，相關組合變化繁多，必須不斷的測試才能得出最佳結果，惟經不斷調整仍無法達到可接受的鑑別力，可能必須回到資料蒐集階段，重新對變項進行處理，或透過給予權重改變樣本各類別的比例進行調整(黃文、王正林，2016)。基於過去許多研究指出，使用相同演算法時，當訓練資料愈多，常能獲得較高正確率及穩定性，故以下本研究進行樣本重新分割，提高訓練樣本至 80%，共有 187 個訓練樣本，其餘分別各占 10%。經 ROC 評估調整後結果如表 5-2-2。

表 5-2-2 ROC 操作特徵曲線評估表

分割比例	訓練樣本	驗證樣本	測試樣本
70/15/15	0.9419	0.7973	0.9147
80/10/10	0.9783	0.9107	0.9219

表 5-2-2 顯示，當訓練樣本由 70% 提高為 80% 後，各項樣本測試的正確率均有所提高，特別在驗證資料方面，調整前正確率為 0.7973，僅達「可接受」的 $0.7 \leq AUC \leq 0.8$ 標準，但在訓練樣本及測試樣本的表現，數值達到 0.9419，顯示模型在訓練時似有「過度適配」的現象；當提高訓練資料至 80% 後，驗證樣本正確性提升，達到 $0.9 \leq AUC \leq 1.0$ 極佳的鑑別力標準，3 項樣本集測試的數值也相當平均，表示重新分割樣本可優化本研究決策樹之模型，同時調整後的決策樹模型在分類規則也與先前有所不同，由原來 4 條規則增加為 8 條規則，變項也增加「單獨生活戶數比」1 項(圖 5-2-2，表 5-2-3)。

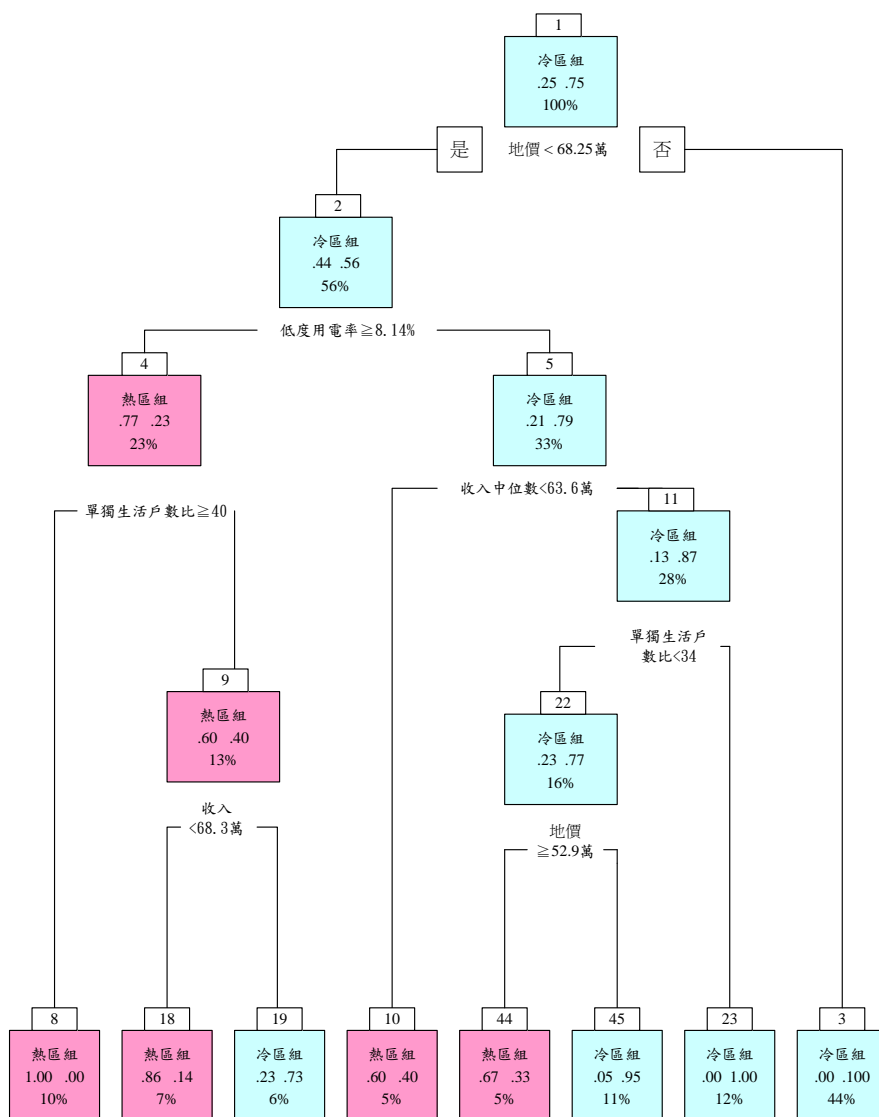


圖 5-2-2 調整後決策樹分類規則圖

表 5-2-3 調整後決策樹分類規則表

規則	葉部結點	分類條件
1	3	IF 每坪地價 ≥ 68.25 萬 THEN 1.類別為冷區組 2.本組中：冷區組機率 100%；熱區組機率 0% 3.占訓練樣本數 44%
	23	IF 每坪地價 < 68.25 萬 AND 低度用電率 $< 8.14\%$ AND 收入中位數 ≥ 63.6 萬

		<p>AND 單獨生活戶數比 ≥ 33.69</p> <p>THEN 1.類別為冷區組</p> <p>2.本組中：冷區組機率 100%；熱區組機率 0%</p> <p>3.占訓練樣本數 12%</p>
3	45	<p>IF 每坪地價 < 68.25 萬</p> <p>AND 低度用電率 $< 8.14\%$</p> <p>AND 收入中位數 ≥ 63.6 萬</p> <p>AND 單獨生活戶數比 < 33.69</p> <p>AND 每坪地價 < 52.9 萬</p> <p>THEN 1.類別為冷區組</p> <p>2.本組中：冷區組機率 95%；熱區組機率 5%</p> <p>3.占訓練樣本數 11%</p>
4	19	<p>IF 每坪地價 < 68.25 萬</p> <p>AND 低度用電率 $\geq 8.14\%$</p> <p>AND 單獨生活戶數比 < 40.28</p> <p>AND 收入中位數 ≥ 63.6 萬</p> <p>THEN 1.類別為冷區組</p> <p>2.本組中：冷區組機率 73%；熱區組機率 27%</p> <p>3.占訓練樣本數 6%</p>
5	10	<p>IF 每坪地價 < 68.25 萬</p> <p>AND 低度用電率 $< 8.14\%$</p> <p>AND 收入中位數 < 63.6 萬</p> <p>THEN 1.類別為熱區組</p> <p>2.本組中：冷區組機率 40%；熱區組機率 60%</p> <p>3.占訓練樣本數 5%</p>
6	44	<p>IF 每坪地價 < 68.25 萬</p> <p>AND 低度用電率 $< 8.14\%$</p> <p>AND 收入中位數 ≥ 63.6 萬</p> <p>AND 單獨生活戶數比 < 33.69</p> <p>AND 每坪地價 ≥ 52.9 萬</p> <p>THEN 1.類別為熱區組</p> <p>2.本組中：冷區組機率 33%；熱區組機率 67%</p> <p>3.占訓練樣本數 5%</p>
7	18	<p>IF 每坪地價 < 68.25 萬</p> <p>AND 低度用電率 $\geq 8.14\%$</p> <p>AND 單獨生活戶數比 < 40.28</p> <p>AND 收入中位數 < 68.3 萬</p> <p>THEN 1.類別為熱區組</p>

		2.本組中：冷區組機率 14%；熱區組機率 86% 3.占訓練樣本數 7%
8	8	IF 每坪地價 < 68.25 萬 AND 低度用電率 \geq 8.14% AND 單獨生活戶數比 \geq 40.28 THEN 1.類別為 熱區組 2.本組中：冷區組機率 0%；熱區組機率 100% 3.占訓練樣本數 10%

從以上圖表可知，「單獨生活戶數比」比率越高，形成住宅竊盜群聚熱區的機會越大，此驗證日常生活理論中監控力降低易導致犯罪發生的情形。此外，「地價」此變項在決策樹的葉節點分類中重複被使用，再次顯示此變項的重要性，但並非形成冷熱區的絕對因素，從各項規則可看出，冷熱區的區位因素是各變項相互影響的結果，沒有單一因素可以決定環境的特性，這也說明構成犯罪環境的因素的複雜性，同時也對犯罪防治形成巨大的挑戰。

二、條件推論決策樹

為避免傳統決策樹分析「過度適配」問題，Hothorn 等人提出「條件推論決策樹」試圖解決，其主要方法係運用統計檢定以確定自變項和分割點的選擇，在 R 語言中主要方式為使用「ctree」套件進行模型建構。以下本研究同樣使用 80/10/10 的分割比例進行分析，俾利於與傳統決策樹進行效能比較。

分析結果顯示，條件推論決策樹透過統計檢定，先將「大專以下教育程度率」分為 2 組，再依照「低度用電率」進行分群。圖 5-3-3 可看出節點 5「大專以下教育程度率」高於 0.434 立即分到熱區，計有 43 個樣本，其餘部分則是以「低度用電率」大於 11.32 為熱區，小(等)於 11.32 為冷區；換言之，條件推論決策樹認為「大專以下教育程度率」與「低度用電率」為形成住宅竊盜冷熱區最重要的 2 個變項。

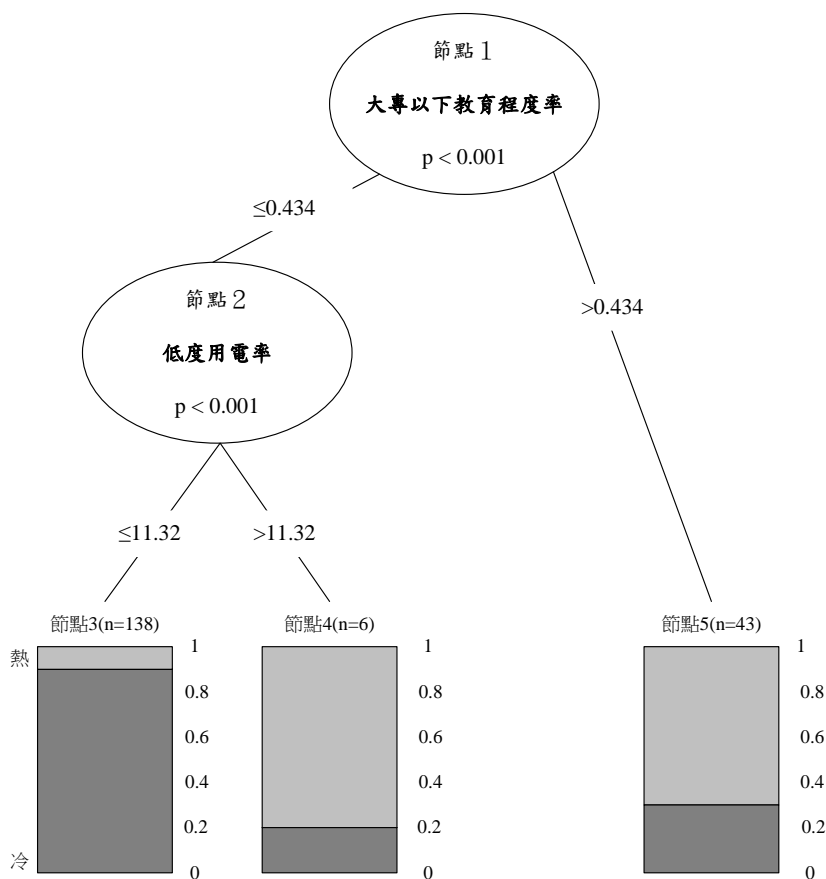


圖 5-2-3 條件推論決策樹分類規則圖

三、隨機森林

屬於集成學習的隨機森林，幾乎不需要對資料進行預處理，也不需要對資料進行標準化，對異常值也具有彈性，比起決策樹表現更多的穩定性；例如：在決策樹中，若刪除少許的訓練樣本，分析結果可能產生很大的改變，但在隨機森林中幾乎不受到任何影響，主要原因在於其演算法對於「噪音」（即與目標變項關係不大的解釋變項）具有相當的「穩健性」（robust to noise）。

此外，隨機森林可以很好地處理代表性不足的分類；例如：若本研究冷區與熱區的樣本比例相差甚多(如冷區占 95%)，在決策樹中可能就會偏向於較多樣本的特徵，有時需要進行比例調整，但隨機森林可以透過建立每棵樹的最大深度(即不修剪單個決策樹)，以得到偏差較小的模型，此時每棵單獨的樹都可能產生「過度適配」的現象，但是藉由不同的變項輪流進行測試，匯集許多測試數據，最後以投票的方式達到模型的最佳組合(Williams, 2011)。

(一) 參數設定

在參數設定上，隨機森林演算法通常會構建 100 到 500 棵樹，每個樹所做的決策投票都是等值，最終決定將是組成樹的大多數的決定；例如，如果隨機森林中的 100 棵樹中有 51 棵判定明天會下雨，那麼就會做出此預判。本研究預設樹木數量為 500 棵，分析結果將是投票的結果。

除設定樹木的數量外，也需設定分支節點的變項數量，在分支時將只先考慮此選取的變項，這些變項是隨機選擇，且在建構模型的過程中會輪流不斷更換變項的組合，本研究設定變項數量為 4。

(二) 分析結果

由於隨機森林為 500 棵樹所組成，無法逐一對於單一樹結果進行描述，故觀察重點在於「相對重要變項」的排序與預測正確率，以下分別說明。

1、重要變項

隨機森林對於解釋變項重要性的選擇(或者稱對目標變項的影響力)，其方法是統計上使用的「重排檢定」(permutation test)。其基本思路是不考慮解釋變項間的相關性，而只單獨對解釋變數給予分數，也就是重要性；其方法為使用訓練樣本建構模型後，檢查模型分析驗證樣本的準確性，接著單獨將訓練樣本中的某個變項的資料重新排列，並以此資料重新建構模型，比較此新舊模型分析驗證資料的準確性，若重新排列前後的準確度差很多，則表示此變項是很重要的變項；由於隨機森林對變項重要性的計算時需重新建構樹群，換言之，在本研究中計算 1 個變項的重要性就必須重新建構 500 棵決策樹，相當耗費時間，故通常使用具無偏誤的袋外樣本(OOB)進行計算 (陳時仲，2015)。

表 5-2-4 列出每個輸入變項與 4 個欄位的數值，根據重要性的準確性來排序，較高的值表示該變項相對重要。表中各數值欄位代表意義為：(1)「熱區」某變項數值高代表此變項對於熱區的解釋具有重要性；(2)「冷區」某變項數值高代表此變項對於冷區的解釋具有重要性；(3)「MeanDecreaseAccuracy」是指當變項被排除時，準確度下降多少所計算之數值，該值越大表示該變項的重要性越大；(4)「MeanDecreaseGini」是指當變項當時被用來分割一個節點時，吉尼雜質的減

少的數值，該值越大表示該變項的重要性越大(圖 5-2-4)。

表 5-2-4 隨機森林相對重要變項排序表

變項名稱	熱區	冷區	MeanDecreaseAccuracy	MeanDecreaseGini
平均住宅價格	17.61	14.69	19.15	5.56
低度用電戶住宅率	17.29	10.03	18.34	5.32
低教育程度率	14.33	9.43	15.87	5.96
收入中位數	12.93	10.59	15.21	5.73
單獨生活戶數比	11.15	6.73	11.99	3.64
老化指數	6.89	4.54	7.38	2.22
警力配置	4.08	6.48	7.31	2.35
路燈密度	1.77	6.98	6.63	1.63
粗離婚率	0.42	4.75	4.14	2.07
收入標準差	1.01	4.05	3.76	1.89
人口密度	0.07	3.61	2.89	1.65
老人戶數比	1.6	2.58	2.85	1.6
投票率	-0.8	3.66	2.64	1.81
監視器密度	-0.62	3.3	2.49	1.37
社會人口增加率	-3.77	2.07	-0.29	1.27

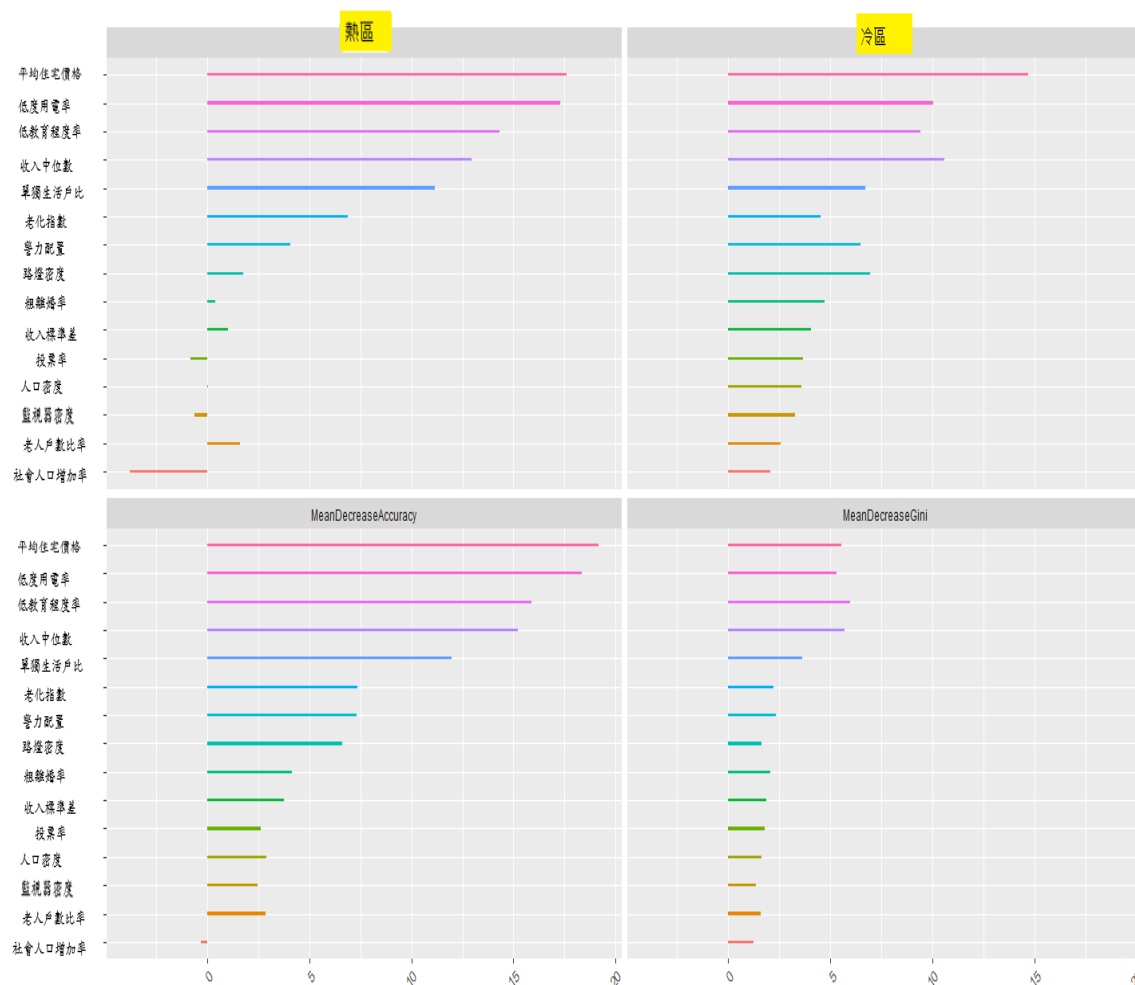


圖 5-2-4 隨機森林相對重要變項排序圖

表 5-2-4 顯示，使用「MeanDecreaseAccuracy」及「MeanDecreaseGini」兩種總體指標衡量，「平均住宅價格」、「低度用電戶住宅率」、「較低教育程度率」、「收入中位數」及「單獨生活戶數比」為相對重要變項的前 5 位，與傳統決策樹與條件推論決策樹所篩選出的重要變項一致。而在熱區與冷區的各別指標方面，此 5 種變項依然對於冷熱區的形成均具有重要解釋意義，而較特別的是在熱區中的「投票率」、「監視器密度」及「社會人口增加率」3 個變項出現負值，表示這些變項對於熱區的解釋並不重要，但不一定是負相關。

2、分類正確性評估

預測隨機森林模型正確性以袋外樣本做為驗證資料時，其計算方法是每棵生成的決策樹分析各自的袋外的資料，並計算錯誤率，稱為袋外錯誤率(out of bag error rate)，再將所有決策樹的錯誤率加總平均即可得到整體隨機森林的錯誤率，

反向解釋即為其正確率。分析結果顯示，本研究的袋外錯誤率為 11.76%，也就是準確度為 89.24%，模型表現相當不錯；在 ROC 特徵操作曲線的 AUC 值為 0.922(如圖 5-2-5)。

此外，亦可以錯誤矩陣(error matrix)對隨機森林訓練樣本中的表現進行評估，在 187 個訓練樣本上，熱區組正確預測有 31 次、錯誤 7 次，總體分類錯誤率為 0.326；冷區組正確預測 134 次、錯誤 7 次，總體分類錯誤率為 0.050，訓練資料對於冷區的正確性顯然遠高於熱區(表 5-2-5)，造成此差距之原因主要在於本研究熱區與冷區樣本數不平衡所致；對此問題，隨機森林可以對樣本比例進行調整。本研究將樣本權重比例調整至各 30 再進行測試，結果冷熱區在錯誤率的差別顯然縮小，熱區錯誤率降低為 0.196，冷區略升高為 0.128(表 5-2-6)，雖然袋外錯誤率略為升高為 14.44%，但 AUC 值提升至 0.932(圖 5-2-6)；值得注意的是，隨著樣本比例改變，重要變項在排序上也有所變動，但「平均住宅價格」、「低度用電戶住宅率」、「較低教育程度率」及「收入中位數」並沒有發生變動，顯然這 4 個變項對於冷熱區的影響力穩定性相當高 (表 5-2-7)。

表 5-2-5 隨機森林訓練樣本分類錯誤率評估表

	熱區	冷區	總體分類錯誤率
熱區	31	15	0.326
冷區	7	134	0.050

表 5-2-6 調整後隨機森林訓練樣本分類錯誤率評估表

	熱區	冷區	總體分類錯誤率
熱區	37	9	0.196
冷區	18	123	0.128

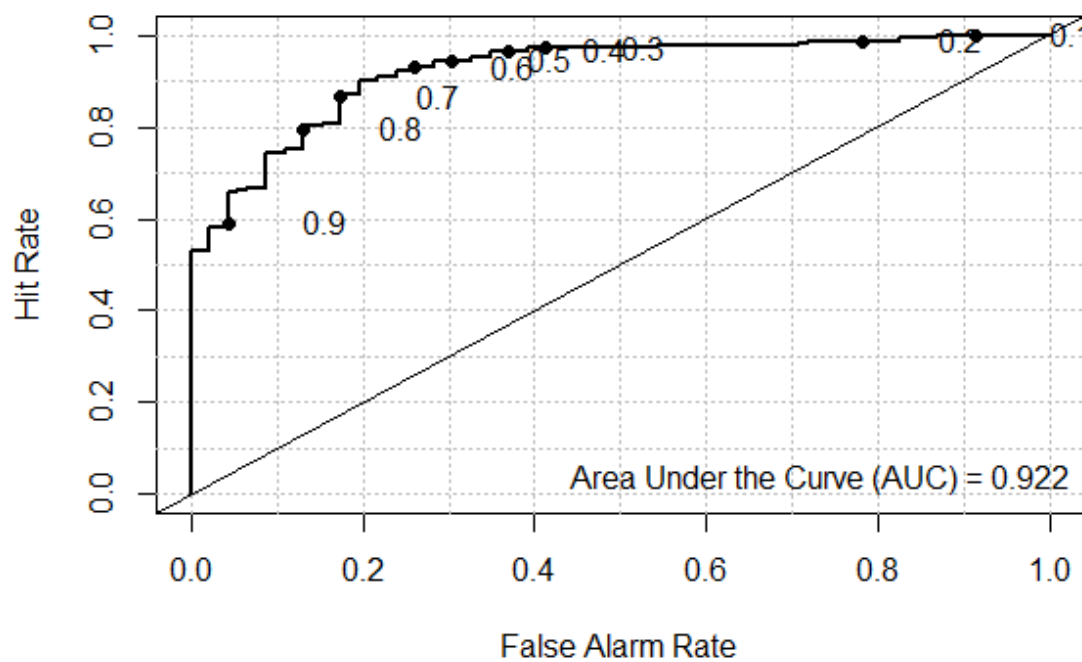


圖 5-2-5 調整前隨機森林袋外樣本 ROC 圖

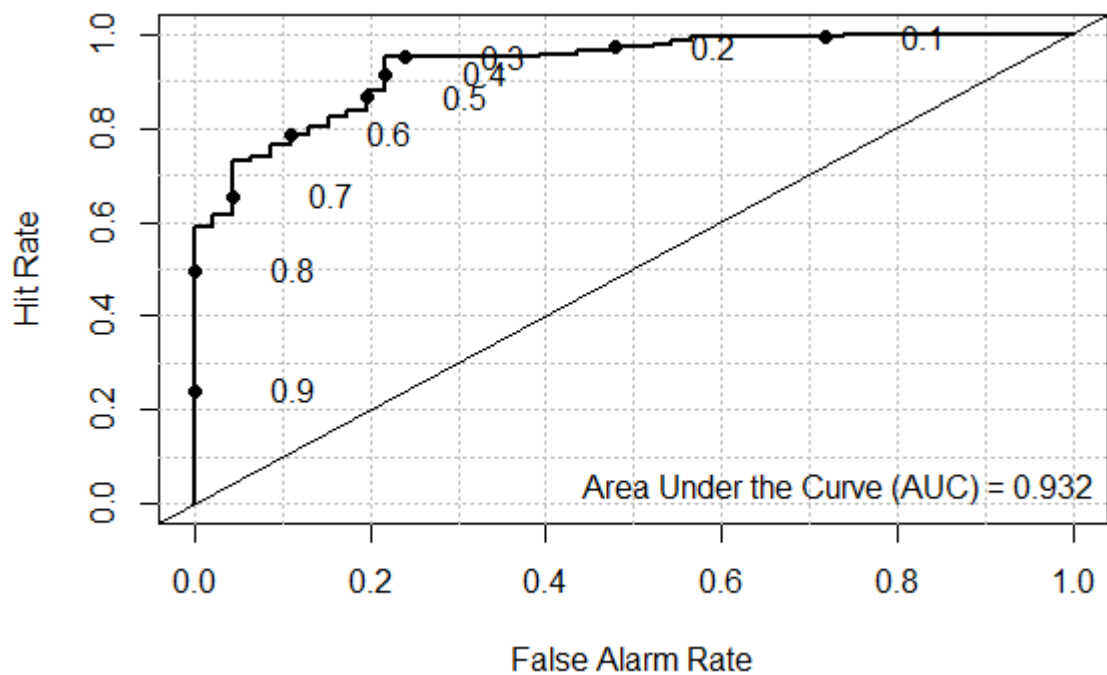


圖 5-2-6 調整後隨機森林袋外樣本 ROC 圖

表 5-2-7 調整後隨機森林相對重要變項排序表

變項名稱	熱區	冷區	MeanDecreaseAccuracy	MeanDecreaseGini
平均住宅價格	20.37	14.66	20.38	5.47
低度用電戶住宅率	13.67	9.28	13.88	3.79
低教育程度率	13.46	7.1	11.83	4.07
收入中位數	13.91	5.74	11.21	3.53
警力配置	6.77	6.07	7.95	1.65
路燈密度	2.51	6.3	6.8	0.94
收入標準差	2.5	5.89	6.28	1.21
老化指數	3.91	4.71	5.87	1.37
單獨生活戶數比	8.93	2.09	5.60	1.86
粗離婚率	1.9	4.07	4.52	1.44
老人戶數比	0.67	3.67	3.62	1.10
監視器密度	0.42	3.53	3.57	0.98
人口密度	0.45	3.54	3.54	0.95
投票率	-1.17	1.21	0.83	0.92
社會人口增加率	-3.72	0.9	-0.45	0.75

觀察隨機森林調整前後整體表現，兩者差距不大，較大的差異在於對冷熱區的正確預測能力，但對於本研究而言，探求住宅犯罪熱區遠比了解冷區的變項重要，故調整後的隨機森林模型更貼近本研究之目的。

3、建樹的規模

前述提到隨機森林建構模型時通常以 100 至 500 棵樹木為範圍，但究竟該選取多少數量較為合適？最好的方法是不斷嘗試，並比較各項數值，理論上應能產生最佳的數據，但此方法缺乏效率。在 RATTLE 中提供「error plots」的功能，以圖形表示樹木的數量與錯誤率的關係(圖 5-2-8)。圖中看出，在 100 棵樹以前，無論是冷、熱區或袋外樣本錯誤率的變化相當大，直到 400 棵樹後，熱區錯誤率十分平穩，冷區與袋外樣本也不再升高，考慮到袋外錯誤率在接近 500 棵樹時仍有下降，故本研究選擇 500 棵樹建構模型應是合適，而實際以各數量加以驗證，

均獲證實²⁹。

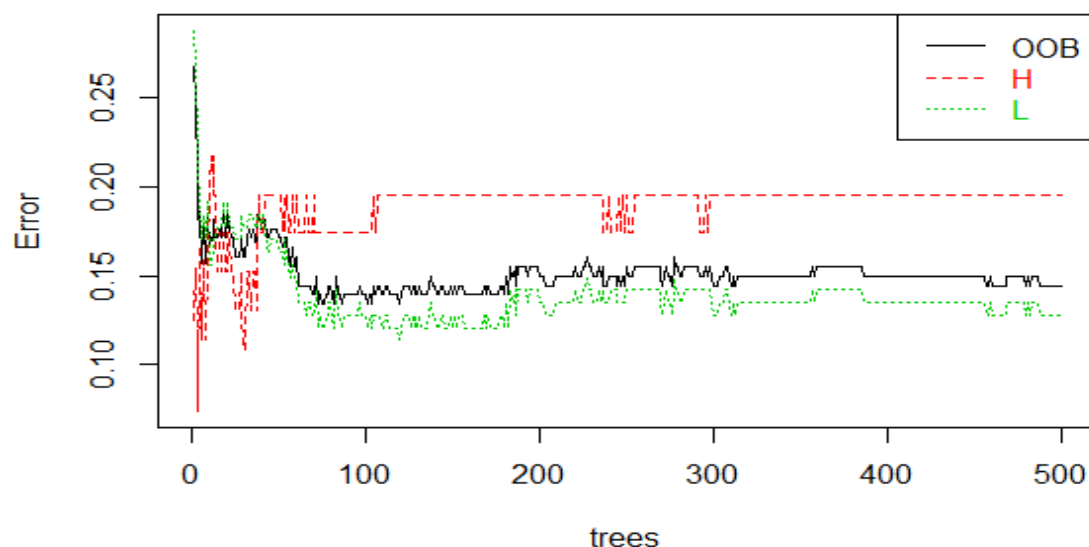


圖 5-2-7 隨機森林建樹量與錯誤率分析圖

四、AdaBoost 推進法

決策樹的改良模型除採取 bagging 策略的隨機森林外，另一為採用 boosting 策略的 AdaBoost 決策樹。boosting 對於訓練樣本內的每個觀察值均給予權重，每次疊代後，對於分類錯誤的資料加大權重給予更多的關注，藉此不斷增加分類的準確率，其抽樣方法是以取後不放回(sample without replacement)為基準。本研究在「推進法」決策樹首先使用“ada”套件建構模型，演算法為「適應性推進」演算法(abbreviation for adaptive boosting, AdaBoost)，以下簡稱 AdaBoost。

(一)參數設定

在參數設定方面：1、樣本數的最小值(min split)設定為 20；2、樹的最大深度(max depth)設為 30；3、複雜度(complexity)設為 0.01；4、樹木數量為 50，疊代(iteration) 次數與樹木量相同為 50(Williams, 2011)。

(二)分析結果

與隨機森林相同，AdaBoost 為多棵樹所組成，無法逐一對於單一樹結果進

²⁹ 300 及 400 棵樹的 AUC 值均為 0.929；而增加至 600 棵樹，AUC 仍與 500 棵樹時相同為 0.932。

行描述，觀察重點同樣在於「相對重要變項」的排序與預測正確率，以下分別說明。

1、相對重要變項

AdaBoost 對於變項的衡量同樣是相對度量，因此變項排列的順序比實際得出的數值更重要。首先，以上述設定進行測試，初步分析結果建議將樹木量下調為 48 棵即可(圖 5-2-8)；因此，樹木量與疊代均調整至 48 進行再測試，訓練錯誤率由 0.032 下降到 0.027，從圖 5-3-8 可看出，隨著樹木(疊代)數量越高，錯誤率不斷下降，到 48 次疊代時下降到最低點，同時繪製相對重要變項如圖 5-2-9。

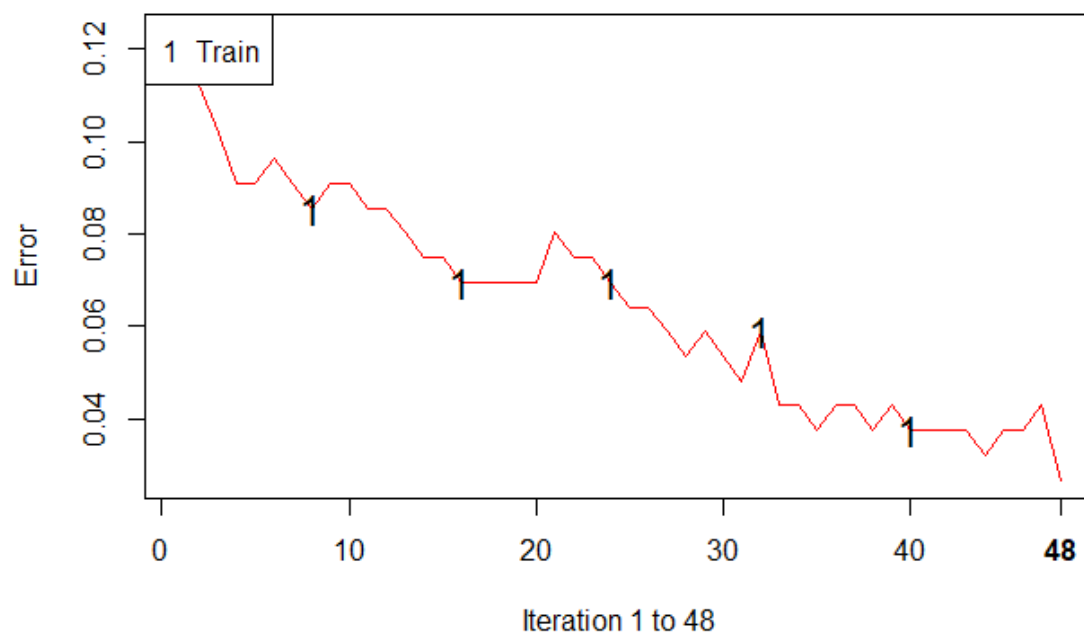


圖 5-2-8 AdaBoost 推進法疊代次數與錯誤率相關圖

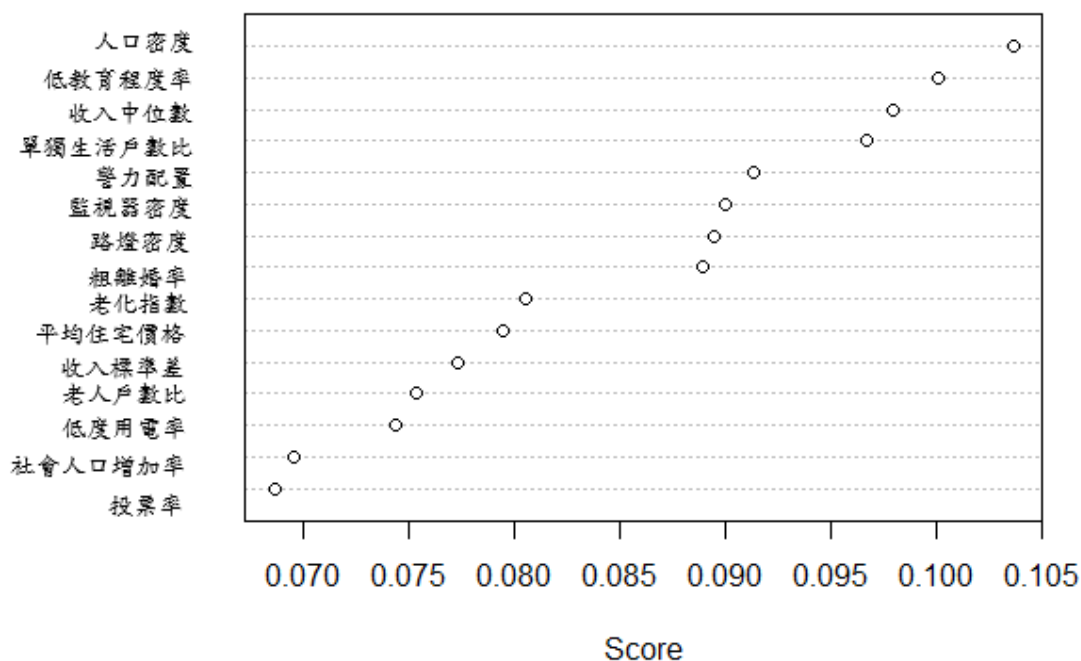


圖 5-2-9 AdaBoost 推進法相對重要變項圖

從圖 5-2-9 看出，以 AdaBoost 所得出的相對變項與決策樹、隨機森林在前 4 項的變項比較，在「低教育程度率」、「收入中位數」及「單獨生活戶數比」相同，最大差異在第一項的「人口密度」。至於「平均住宅價格」及「低度用電戶住宅率」則下降許多，而「社會人口增加率」與「投票率」再度位列最後，顯示此 2 變項在住宅竊盜的區位因素中可能確實不具重要性。

2、分類正確性評估

AdaBoost 模型分類正確性的評估，與隨機森林同樣使用袋外樣本做為驗證資料，分析結果顯示，此模型的袋外錯誤率為 7%，也就是準確度為 93%，模型表現十分優異；混淆矩陣在 187 個訓練樣本上，熱區組正確預測有 42 次、錯誤 4 次；冷區組正確預測 140 次、錯誤 1 次，總共有 182 個樣本正確預測，總體分類錯誤率為 0.027。

五、XgBoost 法

XgBoost(extreme gradient boosting)同樣是一種基於決策樹的集成機器學習演算法，適用於分類和迴歸問題，優點是速度快、準確率高、能處理大規模數據，對於過度擬合問題也有很好的處理能力；在 RATTLE 中建構模型時使用"xgb"套

件。

(一)參數設定

在參數設定方面，本研究採用基本設定：1、樹的最大深度(max depth)為 6；2、學習率(learning rate)為 0.3；3、疊代(iteration) 次數設為 50；4、執行緒(threads) 為 2^{30} ；5、目的(objective)選擇「binary:logistic」類型。

(二)分析結果

1、相對重要變項

表 5-2-8 的「gain」欄位，是指某個變項在樹的分支過程中帶來準確性提高的增益(貢獻度)，此數值大小決定變項重要性排序；「cover」是指與變項相關的觀察值的相對數量；「frequency」是百分比數值，表示特定變項在模型樹中出現的相對次數比例。

從表 5-2-8 與圖 5-2-10 可清楚看出，在相對重要變項方面，排序在前的前 3 項為「低教育程度率」、「低度用電戶住宅率」及「平均住宅價格」，與隨機森林的前 3 項變項完全相同，排序略有改變而已，「老化指數」雖排第四項，但 Gain 的數值與前 3 項已差距頗多，從圖形中可觀察各變項的重要程度，其中「投票率」仍是相對不重要的變項。

³⁰ 在程式語言中，Thread 代表從某個起始點開始(例如 main)，到目前為止所有函數的呼叫路徑，以及這些呼叫路徑上所用到的區域變項。

表 5-2-8 XgBoost 相對重要變項排序表

變項名稱	gain	cover	frequency
低教育程度率	0.245560	0.089869	0.035242
低度用電戶住宅率	0.202690	0.168699	0.127753
平均住宅價格	0.175033	0.248566	0.171806
老化指數	0.066850	0.067413	0.096916
收入標準差	0.060376	0.041566	0.066079
老人戶數比	0.048482	0.061666	0.088106
單獨生活戶數比	0.048080	0.086806	0.092511
收入中位數	0.045697	0.058041	0.061674
警力配置	0.026625	0.040987	0.044053
路燈密度	0.018096	0.033195	0.057269
人口密度	0.017630	0.027591	0.017621
社會人口增加率	0.012071	0.015873	0.039648
監視器密度	0.011967	0.025400	0.044053
粗離婚率	0.011453	0.009164	0.017621
投票率	0.009395	0.025163	0.039648

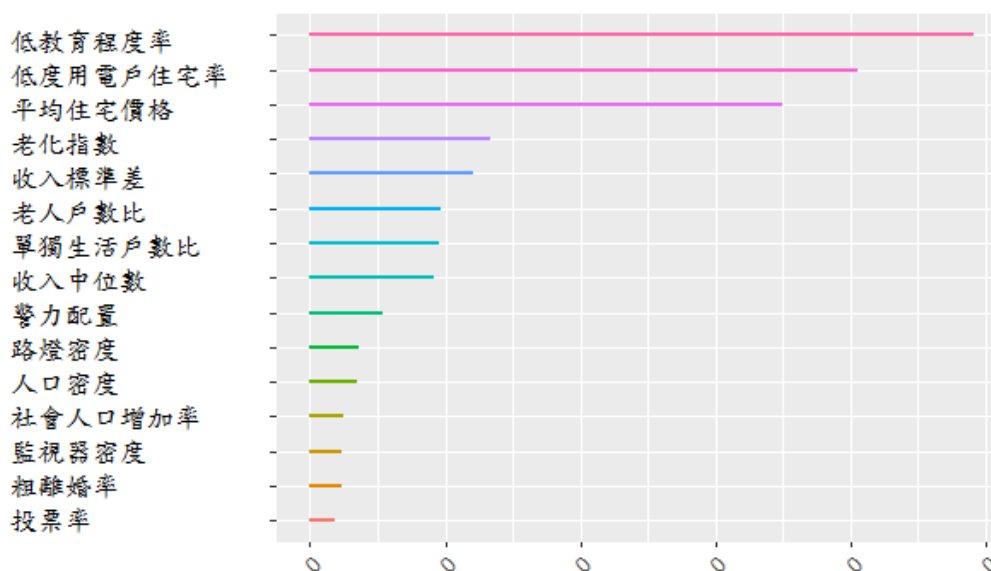


圖 5-2-10 XgBoost 相對重要變項圖

2、分類正確性評估

圖 5-2-11 顯示，隨著疊代次數越多，錯誤率也越低，大約在第 9 次時候，訓練錯誤率為最低點，隨後增加更多次的疊代也無法對模型效能有所提升。將疊代次數調整為 10 進行驗證，表 5-2-9 顯示，在第 9 次及其之後，訓練錯誤均為 0，觀察相對重要變項，前 3 項仍無改變，僅有最後幾項排名略有變化，對後續研究並無影響，故原始設定為可接受範圍。

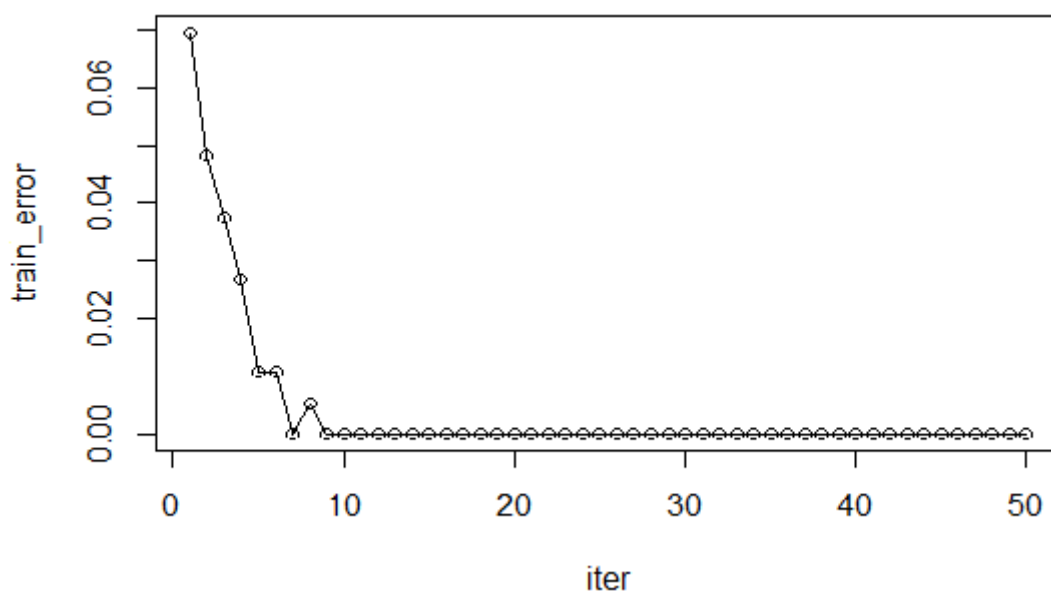


圖 5-2-11 XgBoost 疊代次數與訓練錯誤率關係圖

表 5-2-9 XgBoost 疊代次數與訓練錯誤率關係表

疊代次數	訓練錯誤率
1	0.069519
2	0.048128
3	0.037433
4	0.026738
5	0.010695
6	0.010695
7	0.000000
8	0.005348
9	0.000000
10	0.000000

以上本研究針對臺北市住宅竊盜犯罪冷熱區變項進行多種與決策樹有關的各類模型分析，主要目的在於找出造成犯罪區域特性的區位因素；從各種模型的分析中大致得出若干重要變項對於住宅竊盜群聚有相對重要的影響。由於各類模型對於重要變項的分析結果不一，故在下一節中除將對各模型的優劣進行綜合評比外，也將對各變項做綜整與篩選，進一步精簡模型。

第三節 模型評比與變項篩選

一、模型評比

在資料探勘或機器學習的領域中，一般情況下，取得資料數據後很難立即構建出完美的模型，在做法上經常是運用許多模型並不斷測試，探索模型和數據的擬合程度，最後比較哪個模型更優於其他模型，並探討其優缺點，從而獲得較精確結果。在本單元中，將使用錯誤矩陣的交叉驗證方式與 ROC 曲線(AUC 值)兩種方式，評估上一節中各模型在訓練樣本中對於住宅竊盜犯罪冷熱區變項的分類效能，並根據各模型分析結果，篩選出重要區位變項，以精簡模型的解釋。

(一)錯誤矩陣評比

錯誤矩陣(error matrix)，適合使用在目標變項為二元分類性質，主要在預測資料集中每個觀察結果，錯誤率越低表示分類的效能越佳；作法上係將預測結果與實際觀測結果進行比較。以下先以本研究決策模型分析結果為例進行計算說明。

1、計算方式

錯誤矩陣計算方式主要以預測結果和實際結果的次數進行比較，同時也可變換成百分比的比較方式。表 5-3-1 顯示為本研究決策樹在驗證資料中的實際次數(counts)，表 5-3-2 則是百分比(proportions)的形式。

表 5-3-1 傳統決策樹模型錯誤矩陣次數表

實際	預測		
	熱區	冷區	錯誤
熱區	6	1	14.3
冷區	0	16	0.0

從表 5-3-1 的次數來看，矩陣的主對角線 6 次，指的是正確預測熱區的「真陽性」(true positive)，16 指正確預測冷區次數的「真陰性」(true negative)。矩陣總數為 23 次，故表 5-3-2 的比率矩陣計算如下：

表 5-3-2 傳統決策樹模型錯誤矩陣百分比表

實際	預測		
	熱區	冷區	錯誤率
熱區	$26.1 = \frac{6}{23}$	$4.3 = \frac{1}{23}$	$14.3 = 4.3 / (26.1 + 4.3)$
冷區	$0.0 = \frac{0}{23}$	$69.6 = \frac{16}{23}$	$0.0 = 0.0 / (0.0 + 69.3)$
整體錯誤率：4.3% 類別平均錯誤率：7.15%			

(1)模型整體錯誤率(overall error)

整體錯誤率與真陽(陰)性率正好相反，為離對角線次數加總除總次數：
 $1 / (6 + 1 + 0 + 16) = \underline{4.3\%}$ 。

(2)類別平均錯誤率(averaged class error)

熱區與冷區錯誤率之平均值。決策樹對熱區的預測錯誤率為 14.3%，冷區的預測錯誤率為 0%，故類別平均錯誤率為： $(14.3 + 0.0) / 2 = \underline{7.15\%}$ 。從各別類預測來看，此模型對於冷區的預測的正確率明顯高於熱區。

2、各模型錯誤矩陣評估

根據以上計算方式，本研究其他模型的錯誤矩陣(百分比表)如下(表 5-3-3，

表 5-3-4，表 5-3-5，表 5-3-6)。

(1)條件推論決策樹

表 5-3-3 條件推論決策樹模型錯誤矩陣百分比表

實際	預測		
	熱區	冷區	錯誤
熱區	17.4%	13.0%	42.9%
冷區	8.7%	60.9%	12.5%
整體錯誤率：21.7% 類別平均錯誤率：27.7%			

(2)隨機森林

表 5-3-4 隨機森林模型錯誤矩陣百分比表

實際	預測		
	熱區	冷區	錯誤
熱區	21.7%	8.7%	28.6%
冷區	4.3%	65.2%	6.2%
整體錯誤率：13.1% 類別平均錯誤率：17.4%			

(3)AdaBoost

表 5-3-5 AdaBoost 模型錯誤矩陣百分比表

實際	預測		
	熱區	冷區	錯誤
熱區	21.7%	8.7%	28.6%
冷區	0.0%	69.6%	0%
整體錯誤率：8.7% 類別平均錯誤率：14.3%			

(4)XgBoost

表 5-3-6 XgBoost 模型錯誤矩陣百分比表

實際	預測		
	熱區	冷區	錯誤
熱區	26.1%	4.3%	14.3%
冷區	0.0%	69.6%	0%
整體錯誤率 4.3% 類別平均錯誤率：7.15%			

以上各表顯示，各模型對於熱區的預測能力並不如冷區，本研究認為，主要在於熱區原本在樣本數就偏少，而經過樣本分割後，用於驗證的樣本數僅 23 個，只要誤判 1 個樣本，錯誤率就會提高甚多；因此，熱區的錯誤率明顯較冷區高出許多。

(二)ROC 曲線指標評比

ROC 操作特徵曲線圖是反映「靈敏性」(即真陽性率)和「特异性」(真陰性率)的綜合指標，其 AUC 值是由靈敏性 (x 軸) 與 1-特异性 (y 軸) 所畫出曲線下的面積，經常用來判定模型分類能力的好壞，AUC 值愈大表示模型的分類能力愈好。以下以驗證資料進行各模型 AUC 值評比，而 ROC 操作特徵曲線圖將在與羅吉斯迴歸比較後一併繪製。表 5-3-7 各模型 AUC 值：XgBoost > AdaBoost > 隨機森林 > 傳統決策樹 > 條件推論決策樹，顯示在本研究中以 XgBoost 效果最佳，條件推論決策樹不甚理想，其餘各模型亦均達到極佳鑑別力標準。

表 5-3-7 各模型 AUC 值比較

模型名稱	AUC 值	模型分類能力
傳統決策樹	0.9107	極佳的鑑別力
條件推論決策樹	0.7321	可接受的鑑別力
隨機森林	0.9688	極佳的鑑別力
AdaBoost	0.9732	極佳的鑑別力
XgBoost	0.9821	極佳的鑑別力

(三)模型在各類樣本績效評估比較

本研究在建構模型時將樣本分為訓練、驗證與測試樣本 3 部分，通常以驗證樣本的正確率進行評估；換言之，有些模型在訓練時表現非常優異，但在樣本外的表現並不如預期，造成因素除樣本分配比例、參數設定及過度適配等問題外，模型演算法差異也是影響關鍵，理想的情況下，好的模型除要有較高的預測能力外，也應同時具備穩定性，而所謂的穩定性即為無論在樣本內訓練或樣本外測試的各階段均有一致的表現，上述對於各模型的評估主要是基於訓練樣本；因此，以下對各模型在 3 類樣本中錯誤矩陣與 AUC 數值進行相互比較，以對各模型的預測能力與穩定性有更多了解，。

表 5-3-8 各類樣本集錯誤矩陣評估比較

模型名稱	訓練樣本	驗證樣本	測試樣本
傳統決策樹	6.9%	4.3%	4.1%
條件推論決策樹	5.9%	21.7%	16.7%
隨機森林	5.3%	13.1%	4.1%
AdaBoost	2.6%	8.7%	8.3%
XgBoost	0%	4.3%	4.1%

表 5-3-8 可看出，在本研究各樣本中表現最優異的模型為 XgBoost，在訓練樣本中即可達到完全正確分類，而在驗證與測試過程亦表現良好的穩定性與泛化能力，相差不到 5%；反觀，條件推論決策樹添加嚴謹統計檢定嘗試改良決策樹的作法，在本研究並未有太多優異表現，其驗證樣本與訓練樣本差異高達約 16%，其次隨機森林差異約 8%較 AdaBoost 的 6%為高；而傳統決策樹反而在此評估表現較隨機森林更好。

表 5-3-9 各類樣本集 AUC 值評估比較

模型名稱	訓練樣本	驗證樣本	測試樣本
傳統決策樹	0.9783	0.9107	0.9219
條件推論決策樹	0.8276	0.7321	0.7617
隨機森林	0.9998	0.9688	0.9688
AdaBoost	0.9963	0.9732	0.9531
XgBoost	1.0000	0.9821	0.9609

表 5-3-9 可看出，各模型在 AUC 值表現最優異的依然為 XgBoost，而隨機森林在此評比中則超越傳統決策樹，與 AdaBoost 表現不相上下，條件推論決策樹仍為殿後。綜合兩種不同評估方式，強化(集成)學習的 XgBoost、AdaBoost 及隨機森林模型，整體表現更優於單棵樹木的傳統與條件推論決策樹，這也說明強化演算方法在資料探勘與機器學習領域逐漸成為主流的原因。

二、變項篩選

本研究區位分析重點，在於建立資料探勘模型，運用解釋變項說明造成臺北市住宅竊盜冷熱區的重要因素；因此，解釋變項之適當選取將是影響模型表現的重要關鍵。在第一階段的變項選取中，本研究參考過去文獻並依據犯罪學理論，初步選取 18 個變項；隨後在第二階段以群聚共變項分析進行篩選，共排除與群聚無相關的 3 個變項；剩餘 15 個變項納入資料探勘模型後，在各模型中均獲得相對重要的變項。因此，在本單元中將對各模型中的相對重要變項進行歸納與篩選，主要目的在於讓模型更為精簡，俾於判讀與實用性；雖然，根據統計學理論，較多的變項可以提升模型的適配性³¹，得到更具一般性的解釋結果，篩選變項後可能導致模型的正確率下降(裘厥庸、李永新，2010)，但若能將正確性控制在優良的鑑別力($0.8 \leq AUC \leq 0.9$)範圍，所損失若干精確度仍為值得。以下說明本研究第三階段篩選變項的步驟與標準。

(一)變項篩選步驟與標準

本研究對各模型相對重要變項，篩選步驟與標準說明如下：

³¹ 在資料探勘或機器學習中，較多變項也有可能導致「過度適配」的問題。

- 1、第一步驟：由於演算方式不同，不同模型的相對重要變項數值難以相互比較，必須個別衡量。具體作法上，傳統決策樹中選取重要子(葉)節點的變項，計有 4 個；條件推論決策樹從根節點開始，選取達到統計顯著水準的變項，共 2 個；隨機森林在則選取相對重要變項中數值明顯較高的變項，計有 4 個；AdaBoost 的相對重要變項分為 4 群，選取第一群中的 4 個變項；XgBoost 第三個變項後，增益數值下降許多，故選取前 3 個變項。各模型選取之變項如表 5-3-10 所示。
- 2、第二步驟：將各模型篩選之變項，全部投入到各模型進行訓練與驗證。
- 3、第三步驟：若篩選變項後之模型正確性低於優良的鑑別力範圍，則仿效迴歸分析的向前法(forward selection)，逐漸增加模型中次重要的解釋變項，以改善模型鑑別能力；反之，若首次驗證即達到優良鑑別能力範圍，則不再增加變項，以免徒增模型複雜程度。

表 5-3-10 各模型相對重要變項篩選結果

模型名稱	相對重要變項
傳統決策樹	平均住宅價格、低度用電戶住宅率、收入中位數、單獨生活戶數比
條件推論決策樹	低教育程度率、低度用電戶住宅率
隨機森林	平均住宅價格、低度用電戶住宅率、較低教育程度率、收入中位數、單獨生活戶數比
AdaBoost	低教育程度率、收入中位數、單獨生活戶數比、人口密度
XgBoost	低教育程度率、低度用電戶住宅率、平均住宅價格

(二)變項篩選後各模型 AUC 值評估

經變項篩選後，本研究首先投入「平均住宅價格」、「低度用電戶住宅率」、「收入中位數」、「單獨生活戶數比」、「低教育程度率」及「人口密度」6 個變項，初步驗證各模型的鑑別力並無太大改變，惟值得注意的是 AdaBoost 曾經出現的「人口密度」變項，再次分析時不再出現，故予以刪除後重新投入 5 個變項進行驗證，各模型 AUC 驗證結果觀察，以 5 個變項投入後，傳統決策樹完全不受影響，改變最大的為條件推論決策樹，各類樣本的 AUC 值均有所提升；隨機

森林與 AdaBoost 在驗證樣本上也有所上升，顯示模型鑑別力更佳；XgBoost 除訓練樣本依然全部正確外，其餘樣本集數值略有下降，不過依然達到相當優良的鑑別力。整體而言，經由變項選取程序的各模型，在降低模型複雜度的同時並無出現過度模型適配(擬合)問題的發生，模型的預測能力亦有所提升，最重要是模型精簡後，使研究結果更具實務上意義(表 5-3-11)。

表 5-3-11 變項篩選後各模型 AUC 值評估

模型名稱	訓練樣本	驗證樣本	測試樣本
傳統決策樹	0.9783	0.9107	0.9219
條件推論決策樹	0.8603*	0.7634*	0.8008*
隨機森林	0.9985	0.9732*	0.9609
AdaBoost	0.9903	0.9911*	0.9609*
XgBoost	1.0000	0.9643	0.9531

註：*表示篩選後 AUC 數值增加。

第四節 羅吉斯迴歸統計分析

在統計學領域中，線性迴歸模型是用來預測連續型變項，羅吉斯迴歸(logistic regression 或 logit regression)則是建立類別目標變項與解釋變項之間的關係，其與線性迴歸主要不同之處在於：第一，典型的羅吉斯迴歸模式為 S 曲線，與線性迴歸呈現直線的態樣不同，其機率值位於 0 與 1 之間³²，且解釋變項對目標變項的影響是以指數的方式變動，故不需要常態分配的假設即可應用；其次，羅吉斯迴歸的各項參數是透過最大概似法(maximum likelihood)進行估計，與線性迴歸以最小平方法(ordinary least square estimation) 估計參數不同(劉盈君，2008)。長期以來，羅吉斯迴歸模型在統計的運用上已極為普遍，而自從大數據分析受到重視後，其雖然並非分類器，但由於羅吉斯迴歸對於變項分類的統計特性，也成為資料探勘與機器學習領域中相當受到歡迎的分類模型；由於羅吉斯迴歸模型在檢定與解釋與前述決策樹等資料探勘有較多的差異，故特以此節予以探討。

本節中使用的羅吉斯迴歸的統計分類方式與決策樹運用資訊理論

³² 羅吉斯迴歸模式呈現 S 曲線，主要是使用乙狀函數 (Sigmoid Function，或稱為 Logistic Function)，此函數的 Y 值介於 0 至 1，符合機率值範圍。

(information theory) 計算屬性的資訊增益量，最後選取資訊量最大的屬性為節點，反覆進行後建立樹狀結構的分支演算法並不相同，為比較兩種分類方式效果，故重新投入 15 個變項，隨機抽取樣本 80%，首先進行整體模型分析，再以「向後選取法」(backward selection) 進行變項篩選，此方法是逐一剔除重要性較低的解釋變項，直到當再度剔除任何一個解釋變項時，模型會損失過多的解釋力，此時即停止篩選，刪除變項以 $\ln(\text{likelihood ratio})$ 統計量作為檢定標準。

一、羅吉斯迴歸整體模型分析

本研究羅吉斯模型使用 SPSS 進行分析，以住宅竊盜群聚類型為目標變項，分成熱區及冷區兩類，投入的 15 個變項先經標準化(data normalization method) 處理，各項分析結果如下：

(一)Cox & Snell Rsquare 值與 Nagelkerke R square 值檢定

在 Cox & Snell Rsquare 值與 Nagelkerke R square 值檢定方面，Cox & Snell R 平方為 0.497，表示變項組合解釋分類關聯強度為 49.7%，Nagelkerke R 平方為 0.710(表 5-4-1)，則表示全體變項與執行適當性之關聯強度約為 71.0%，關聯性頗高(Hair, Black, Babin, & Anderson, 2013)。

表 5-4-1 Cox & Snell Rsquare 與 Nagelkerke R square 適配度檢定

-2 對數概似	Cox & Snell R 平方	Nagelkerke R 平方
116.994	.497	.710

(二)Omnibus 檢定

羅吉斯迴歸模型所包含的解釋變項必須對目標變項有顯著的解釋能力，才能進行有意義的解釋，由 SPSS 輸出之模式係數 Omnibus 檢定表中的卡方值，可檢驗是否所有迴歸係數皆等於零之虛無假設。檢定結果顯示，卡方值為 128.330，自由度為 15，顯著性 $p = 0.000 < 0.05$ ，達統計顯著，故拒絕虛無假設，表示本模式所選取的解釋變項能有效的聯合影響目標變項。

(三) Hosmer-Lemeshow (HL) 指標

HL 指標可用來檢定羅吉斯迴歸模型的適配程度³³，模型適配較好是由於觀測分類和預測分類的差異較小(Hair et al., 2013)。其基礎運作方式，是以風險十分位數將觀察值予以分組，並比較每個十分位數中的觀察機率與期望機率，利用 HL 指標與卡方分布進行比較，若卡方檢定未達顯著則表示模型適配良好(劉盈君，2008)。經檢定，HL 指標卡方值為 8.081，自由度為 8， $p=0.426 > 0.05$ ，未達統計顯著，故無法拒絕模型適配資料良好之假設(表 5-4-2)，表示整體模型適配程度良好。

表 5-4-2 整體模型 HL 適配度檢定

分組	冷區		熱區		總數
	觀察次數	期望	觀察次數	期望	
1	0	.302	19	18.698	19
2	2	2.935	17	16.065	19
3	10	8.562	9	10.438	19
4	17	14.718	2	4.282	19
5	16	16.761	3	2.239	19
6	16	17.757	3	1.243	19
7	19	18.397	0	.603	19
8	18	18.704	1	.296	19
9	19	18.891	0	.109	19
10	16	15.972	0	.028	16
Hosmer 和 Lemeshow 檢定					
卡方值		自由度		P值	
8.081		8		.426	

³³ HL 指標遵循 Pearson 卡方分布的關係，檢定結果受樣本數影響，通常樣本數建議值應大於 400(本研究冷熱區樣本為 234)(葉柏村、張光宗，2019)。

(四)模型係數顯著性檢定

羅吉斯迴歸通常使用 Wald 統計量對迴歸係數進行顯著性統計檢定，也就是檢定係數等於零的虛無假設是否成立，若拒絕虛無假設即表示解釋變項對目標變項的發生可能具有影響。表 5-5-3 顯示模式中有 3 個變項具顯著性，分別為平均住宅價格、低度用電戶住宅率及路燈密度。羅吉斯迴歸係數可以視為解釋變項的變化導致目標變項的變化情況。羅吉斯迴歸模式的係數如果為正值且統計顯著，意味在控制其他解釋變項的條件下，勝算值(OR 值)隨對應解釋變項的增加而增加³⁴；反之，顯著的負係數代表對數勝算隨對應解釋變項的增加而減少；如果係數的統計性不顯著，說明該解釋變項的作用在統計上與零無異。

表 5-4-3 的 3 個顯著變項表示，隨著低度用電住宅率及路燈密度係數負值越大，臺北市住宅竊盜犯罪冷區機率減少，熱區機率增加($\text{Exp}(B) < 1$)，而平均住宅價格正值升高(正值)，形成冷區機率將增加，熱區機率減少($\text{Exp}(B) > 1$)；換言之，空屋率越高(低度用電住宅率)、路燈密度較高及房價越低之地區有有較高住宅竊盜熱區群聚的機率，此部分變項與資料探勘結果相符。

³⁴ 羅吉斯迴歸勝算值(Odds)公式： $\ln \frac{p}{1-p} = f(x) = \beta_0 + \beta_1 X + \beta_2 X + \dots + \beta_k X_k$

表 5-4-3 整體模型變項統計檢定表

變項	B 之估計值	Wals	顯著性	Exp(B)
低教育程度率	-.525	.688	.407	.592
投票率	-.311	.563	.453	.733
人口密度	-.741	2.207	.137	.477
社會人口增加率	-.252	.399	.528	.777
粗離婚率	-.097	.080	.777	.908
老化指數	-.049	.014	.905	.952
收入中位數	.522	.510	.475	1.685
收入標準差	-.338	2.956	.086	.713
平均住宅價格	1.505	8.729	.003**	4.502
老人戶數比率	-.141	.123	.726	.869
低度用電戶住宅率	-1.167	5.667	.017*	.311
單獨生活戶數比	.063	.014	.906	1.065
路燈密度	-1.011	5.516	.019*	.364
監視器密度	.658	2.928	.087	1.931
警力配置	-1.004	3.124	.077	.366
常數	1.566	20.924	.000	4.788

註：**<.01，*<.05。

二、變項篩選與模型正確性評估

(一)變項篩選

本研究採用向後法進行變項篩選，以 LR 統計量作為刪除變數之檢定標準(設定加入標準： $\alpha = 0.05$ 、刪除標準： $\alpha = 0.1$)，共刪除不符合檢定標準之 10 個變項；分別為：單獨生活戶數比(第二步驟刪除)、老化指數(第三步驟刪除)、粗離婚率(第四步驟刪除)、老人戶數比率(第五步驟刪除)、社會人口增加率(第六步驟刪除)、收入中位數(第七步驟刪除)、投票率(第八步驟刪除)、收入標準差(第九步

驟刪除)、人口密度(第十步驟刪除)及監視器密度(第十一步驟刪除);最終納入模型之變項如表 5-4-4,較表 5-4-3 增加低教育程度率與警力配置 2 個變項,其中警力配置亦為資料探勘分析中未曾出現之變項;此 2 變項表示熱區比起冷區低教育程度率較高及較多的警力配置。

表 5-4-4 變項篩選統計檢定表

變項	B 之估計值	Wals	顯著性	Exp(B)
低教育程度率	-.990	9.102	.003	.372
平均住宅價格	1.239	9.839	.002	3.453
低度用電戶住宅率	-.988	7.891	.005	.372
路燈密度	-.895	8.875	.003	.409
警力配置	-.651	2.705	.100	.522
常數	1.486	26.142	.000	4.420

(二)模型正確性評估

在羅吉斯迴歸模型正確性評估方面,表 5-4-5 顯示變項篩選前對於冷區錯誤預測為 5.2%,熱區錯誤預測為 18.5%,運用前述章節計算方式得出整體錯誤率為 9.0%,平均類別錯誤率為 11.85%;換言之,總體準確度達 91%,模型的分類效能十分優異。

變項篩選後在錯誤矩陣正確性評估方面,表 5-4-6 顯示對於冷區錯誤預測為 4.5%,熱區錯誤預測為 19.2%,整體錯誤率為 8.5%,平均類別錯誤率為 11.85%,分類成效與未篩選前差異不大,整體錯誤率反略為下降 0.5%,重要是在不損失預測正確性情況下,得到對於影響冷熱區區位的其他可能變項資訊。

表 5-4-5 變項篩選前模型錯誤矩陣表

	預測次數		
觀察次數	熱區	冷區	錯誤
熱區	44(23.5%)	10(5.3%)	18.5%
冷區	7(3.7%)	126(67.4%)	5.2%
整體錯誤率 9.0%			
類別平均錯誤率 11.85%			

表 5-4-6 變項篩選後模型錯誤矩陣表

	預測值		
觀察值	熱區	冷區	錯誤
熱區	42(22.5%)	10(5.3%)	19.2%
冷區	6(3.2%)	127(67.9%)	4.5%
整體錯誤率 8.5%			
類別平均錯誤率 11.85%			

三、羅吉斯迴歸與決策樹等模型績效比較

羅吉斯迴歸分析結果顯示，在重要變項方面計有「低教育程度率」、「平均住宅價格」及「低度用電戶住宅率」與前面決策樹等相同，但另 2 個變項「路燈密度」及「警力配置」則是在之前各種模型中並非被列為特別重要之變項，而經常出現的「收入中位數」在此卻並未被列入，此凸顯統計方式與資料探勘的分類演算法對於資料的分類標準確有相當不同，以下將對資料探勘與羅吉斯迴歸進行模型評估比較。

(一)各類樣本集錯誤矩陣評估比較

本研究在建構資料探勘模型時將樣本分為訓練、驗證與測試樣本 3 部分，其中訓練樣本為 80%，但使用 SPSS 進行羅吉斯迴歸分析時，僅以隨機方式抽取相同數量樣本進行統計測試，並未建立驗證與測試樣本，此種隨機抽樣方式，每次計算出的結果均略有差異；因此，為比較羅吉斯迴歸與前面各資料探勘模型在錯誤矩陣與 AUC 值差異，此部分評估將使用 RATTLE 進行，俾能有一致的標準。

表 5-4-7 各類樣本集錯誤矩陣評估比較

模型名稱	訓練樣本	驗證樣本	測試樣本
傳統決策樹	6.9%	4.3%	4.1%
條件推論決策樹	5.9%	21.7%	16.7%
隨機森林	5.3%	13.1%	4.1%
AdaBoost	2.6%	8.7%	8.3%
XgBoost	0%	4.3%	4.1%
羅吉斯迴歸	8.5%	4.3%	12.5%

表 5-4-7 顯示，羅吉斯迴歸在訓練樣本的錯誤率與使用 SPSS 之數據大致相當，而在驗證樣本方面，表現十分突出，與資料探勘相比毫不遜色，測試樣本方較不理想，但整體而言仍優於條件推論決策樹。

(二)各類樣本集 AUC 值評估比較

表 5-4-8 與圖 5-4-1 顯示，羅吉斯迴歸無論在樣本內(訓練樣本)或樣本外(驗證樣本、測試樣本)的 AUC 表現均不亞於資料探勘方法，且數值當平均，表現出相當穩定性，這也是統計方法模型的優點；總體來說，羅吉斯迴歸較單棵決策樹的演算法表現更好，而多棵決策樹的強化學習則優於羅吉斯迴歸，不過差距並不大，在評定標準上同樣擁有極佳的鑑別力。

表 5-4-8 各類樣本集錯誤矩陣評估比較

模型名稱	訓練樣本	驗證樣本	測試樣本
傳統決策樹	0.9783	0.9107	0.9219
條件推論決策樹	0.8276	0.7321	0.7617
隨機森林	0.9998	0.9688	0.9688
AdaBoost	0.9963	0.9732	0.9531
XgBoost	1.0000	0.9821	0.9609
羅吉斯迴歸	0.9342	0.9286	0.9453

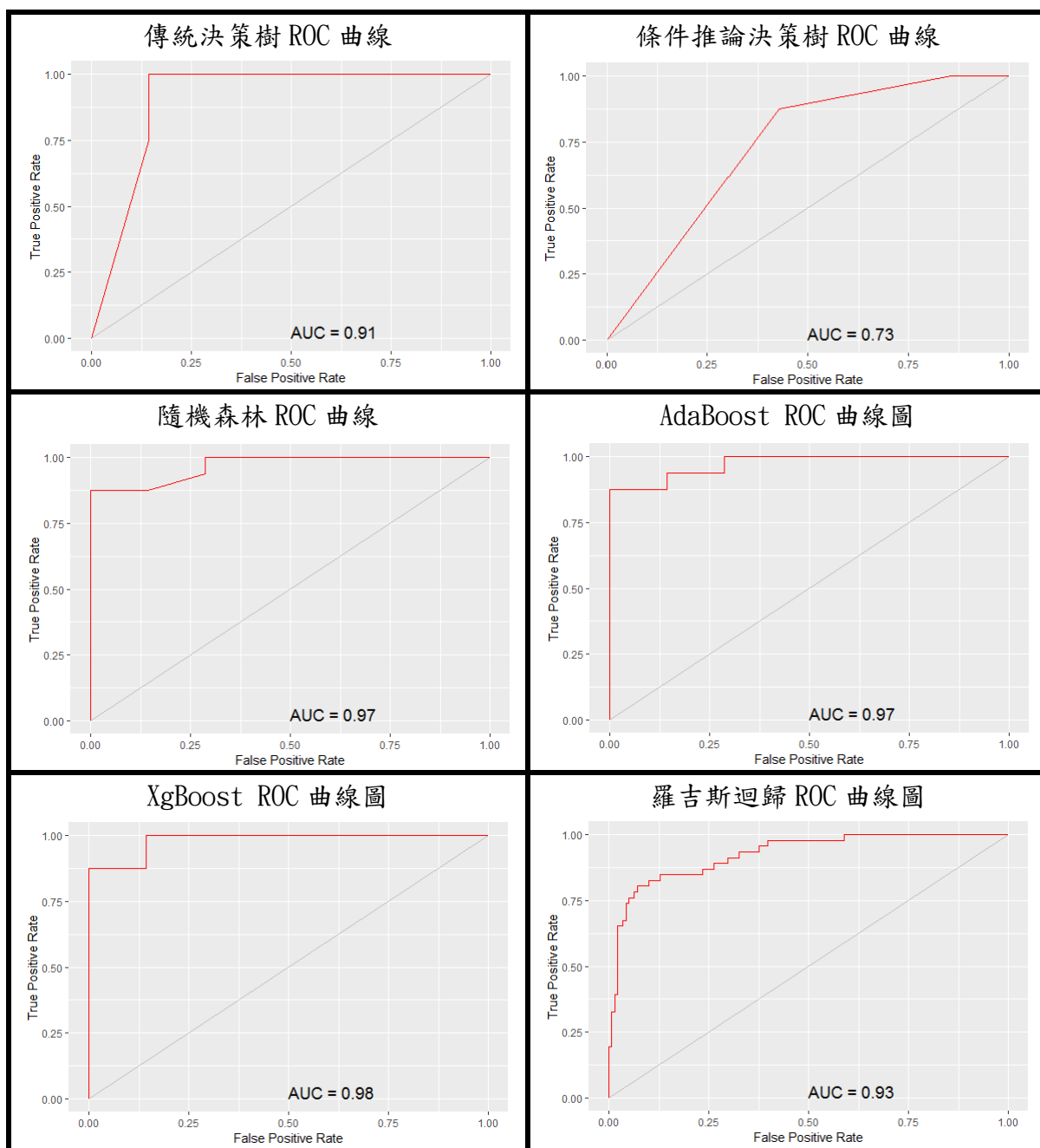


圖 5-4-1 各模型 ROC 曲線圖

針對以上分析與評比，本研究歸納羅吉斯迴歸與資料探勘的差異如下：

1、模型展示的差異

與傳統的迴歸模型相似，羅吉斯迴歸模可以看到非常明確的參數估計值、P 值等重要指標，估計誤差並列出迴歸方程式；但在資料探勘的方法中，較重視微

調參數，以達到模型最佳效果；故除傳統(條件推論)決策樹外，很難看到模型的展示，甚至於其運算方式也如同「黑箱子」(black box)，難以列出方程式並詳細解釋。簡言之，羅吉斯迴歸統計對於現象的解釋較具優勢，而資料探勘模型並不關注研究假設與資料分布型態，著重在發掘未知及模型的預測能力，兩者各有所長，具有互補的作用。

2、模型效能評估問題

在其他許多研究中，進行模型評估時羅吉斯迴歸經常不會是最佳效果，事實上這個結論並不是很可靠，造成此結果原因與羅吉斯迴歸若進行隨機抽樣，每次抽樣後的統計結果，均略有差異，很難立即統整出最佳模型；但在資料探勘的機器學習中，藉由不斷試錯的過程與比對，模型可以迅速優化，但也有可能產生過度學習的現象。因此，在條件允許下，進行資料探勘後，最好能以適當統計方式再次分析，相互印證，較能獲致更客觀的結果。

第五節 區位變項理論意涵與實務意義

在第三章中，本研究基於社會解組理論與新機會理論，建構研究概念架構，並選取 18 個區位變項進行驗證；經本章上述各節的資料探勘模型分析與重要變項篩選後，共得出低教育程度率、平均住宅價格、低度用電戶住宅率、收入中位數、單獨生活戶數比、路燈密度、警力配置及人口密度 8 個變項，其中低教育程度率與收入中位數為在本研究中屬社會解組理論範疇，而其他變項則歸為新機會理論部分。雖然，從前述的決策樹等的條件式分類規則與羅吉斯迴歸分析中，對於這些變項在群聚冷熱區的構成與相互關係已有初步了解，但尚未能確認各變項間的關聯強度；因此，以下對此 8 個變項進行相關矩陣分析，以進一步了解各變項之相關程度，並配合本研究群聚分析結果，說明其在社會解組理論與新機會理論之意涵及實務上意義。

一、區位變項相關矩陣分析

兩個連續變項的關聯情形，除用散布圖的方式來表達，亦可以統計的運算檢驗模式來進行精確的分析，即是建立一個用以描述相關情形的量數，即相關係數；本研究的區位變項均屬於連續變項，故採「皮爾森相關係數」(pearson correlation coefficient)進行分析。從表 5-5-1 的相關矩陣分析結果得知，各區位變項相關情

形如下：

(一) 社會解組理論變項

「低教育程度率」與「收入中位數」呈現高度負相關 ($r = -.856, p < .01$)³⁵。

(二) 新機會理論變項

- 1、「人口密度」與「平均住宅價格」為低度正相關 ($r = .297, p < .01$)。
- 2、「人口密度」與「低度用電戶住宅率」為低度負相關 ($r = -.291, p < .01$)。
- 3、「人口密度」與「單獨生活戶數比」為低度負相關 ($r = -.159, p < .05$)。
- 4、「人口密度」與「路燈密度」為中度正相關 ($r = .544, p < .01$)。
- 5、「人口密度」與「警力配置」為低度負相關 ($r = -.240, p < .01$)。
- 6、「平均住宅價格」與「低度用電戶住宅率」為低度負相關 ($r = -.269, p < .01$)。
- 7、「平均住宅價格」與「單獨生活戶數比」未達統計顯著。
- 8、「平均住宅價格」與「路燈密度」為低度正相關 ($r = .356, p < .01$)。
- 9、「平均住宅價格」與「警力配置」未達統計顯著。
- 10、「低度用電戶住宅率」與「單獨生活戶數比」為低度正相關 ($r = .344, p < .01$)。
- 11、「低度用電戶住宅率」與「路燈密度」未達統計顯著。
- 12、「低度用電戶住宅率」與「警力配置」為中度正相關 ($r = .572, p < .01$)。
- 13、「單獨生活戶數比」與「路燈密度」為低度正相關 ($r = .288, p < .01$)。
- 14、「單獨生活戶數比」與「警力配置」為中度正相關 ($r = .422, p < .01$)。
- 15、「路燈密度」與「警力配置」未達統計顯著。

(三) 社會解組與新機會理論變項

- 1、「低教育程度率」與「人口密度」為低度負相關 ($r = -.180, p < .01$)。
- 2、「低教育程度率」與「平均住宅價格」為中度負相關 ($r = -.697, p < .01$)。
- 3、「低教育程度率」與「低度用電戶住宅率」為低度正相關 ($r = .323, p < .01$)。
- 4、「低教育程度率」與「單獨生活戶數比」為低度正相關 ($r = .333, p < .01$)。
- 5、「低教育程度率」與「路燈密度」為低度負相關 ($r = -.180, p < .01$)。

³⁵ 相關係數範圍 (絕對值) 顯示變項關聯程度, 通常 1.00 稱為「完全相關」(perfect correlated); 0.70 至 0.99 為「高度相關」(highly correlated); 0.40 至 0.69 為「中度相關」(moderately correlated); 0.10 至 0.39 為「低度相關」(modestly correlated); 0.10 以下為「微弱相關」(weakly correlated)(王保進, 2006)。

- 6、「低教育程度率」與「警力配置」為低度正相關 ($r=.173, p<.01$)。
- 7、「收入中位數」與「人口密度」未達統計顯著。
- 8、「收入中位數」與「平均住宅價格」為中度正相關 ($r=.588, p<.01$)。
- 9、「收入中位數」與「低度用電戶住宅率」為低度負相關 ($r=-.180, p<.01$)。
- 10、「收入中位數」與「單獨生活戶數比」為中度負相關 ($r=-.431, p<.01$)。
- 11、「收入中位數」與「路燈密度」未達統計顯著。
- 12、「收入中位數」與「警力配置」為低度負相關 ($r=-.136, p<.05$)。

從各變項的相關矩陣結果觀察，無論是社會解組變項或新機會理論變項，兩類變項在各別與交互間均存在不同程度的顯著相關性，除說明兩項理論可藉由變項整合的可行性外，也顯示若干變項同時具備兩項理論的屬性。是以，以下將從理論與實務兩層面，探討本研究所篩選出重要區位變項之意義，並驗證本研究第三章提出之研究概念架構假設。

表 5-5-1 區位變項相關矩陣

	低教育程度率	人口密度	收入中位數	平均住宅價格	低度用電戶住宅率	單獨生活戶數比	路燈密度	警力配置
低教育程度率	1							
人口密度	-.180**	1						
收入中位數	-.856**	.062	1					
平均住宅價格	-.679**	.297**	.588**	1				
低度用電戶住宅率	.323**	-.291**	-.173**	-.269**	1			
單獨生活戶數比	.333**	-.159*	-.431**	-.056	.344**	1		
路燈密度	-.180**	.544**	.049	.356**	-.066	.228**	1	
警力配置	.173**	-.240**	-.136*	-.005	.572**	.422**	.079	1

**在顯著水準為0.01時(雙尾)，相關顯著。*在顯著水準為0.05時(雙尾)，相關顯著。

二、區位變項之理論解釋及意涵

本研究進行區位分析的方式為將臺北市住宅竊盜犯罪群聚分為冷、熱區兩組，掃描結果顯示，犯罪熱區主要集中在中山、士林、大同、萬華、北投區，冷區主要為大安、信義及文山區，故進行區位變項解釋時，僅以冷、熱區稱之，不再重述各區概況。以下從社會解組理論與新機會理論說明本研究住宅竊盜犯罪重要區位變項之理論意涵與實務意義，並同時驗證本研究前述所提出之概念假設。

(一) 社會解組理論方面

1、教育程度與收入較低之地區易形成住宅竊盜犯罪熱區

在社會解組理論方面，本研究原選取 8 個變項，抽出其中 3 個列為「集體效能」抑制作用的概念，經群聚共變項分析、資料探勘及羅吉斯迴歸分析後，最後篩選之變項為「低教育程度率」與「收入中位數」兩變項，列為「集體效能」之「社區發展協會比率」、「區里活動中心比率」及「投票率」前 2 個變項在共變數分析時即予以汰除，而「投票率」則是相對不重要之變項；具體來說，本研究結果顯示，臺北市住宅竊盜犯罪群聚與「低教育程度率」與「收入中位數」2 變項最具關聯性，犯罪群聚的熱區中有較高「低教育程度率」與較低「收入中位數」，而冷區則恰好相反，此結果再次支持社會解組理論主張教育程度較低與收入較低的地區易造成社會解組現象，從而有較高的犯罪情形。值得注意的是，在共線性檢定中已確認各變項均無共線性情況下，相關矩陣分析顯示此 2 個變項呈現高度負相關，驗證過去研究中得出教育程度高低對於收入確有明顯影響的結論(黃映翎，2008)，而收入多寡在過去多項研究中被認為與各類財產型犯罪具有重要關聯；本研究架構概念假設「社會解組與住宅竊盜犯罪風險為正向關係，易形成住宅竊盜熱區」獲得支持。

2、臺北都會型態之集體效能有待進一步研究

在第二章文獻回顧中，本研究探討有關抑制社會解組主要可分為兩大主張，其一為社會連結模式，主張人際關係聯繫可提供個人歸屬感，加上緊密的人際互動，使得個人的行為得以被有效的控制，而不致於偏離行為規範(Bursik & Webb，1982)；另一則主張近代社會中，由於人際關係疏離，所以社會連結模式的概念已不再適用，基於社會凝聚與信任感的社區集體效能才能有效抑制犯罪的發生(Sampson et al., 1997)。

在本研究的概念架構中以社區發展協會比率、區里活動中心比率及投票率 3 個變項測試集體效能對於犯罪是否具有抑制作用，結果如上所述，均無法有效被選取，故集體效能對於犯罪有抑制效果的推論在本研究並未獲得支持；探究相關原因，一方面可能由於變項選取的指標性不足，亦有可能是國外社區的生活型態與臺北市的都會生活型態有相當大差異所造成，換言之，由於居住環境型態的差異，都市同心圓的居住分布模式仍存在於西方社會中，故西方集體效能的影響易鑲嵌於社區居民整體的網絡關係中，但在臺灣都會地區尤其是臺北市，由於住商

混雜，少有純粹的社區型態，即便將較大規模建案視為社區，各住戶間因工作忙碌，互動亦少。因此，以里為單位所組成的組織，並非整體自發性地關心地方事務，集體效能的產生可能僅是由少部分熱心民眾或具官方身分人士所帶領下形成的表象(Zhang et al., 2007)；故以社區發展協會比率、區里活動中心比率此具有半官方性質所成立之社團組織做為衡量集體效能指標似乎並不恰當，未來研究需另覓其他指標進行更多驗證；故本研究有關集體效能概念與其他概念之關係無法獲得支持。

(二) 新機會理論方面

在新機會理論方面，本研究原始投入 10 個變項，最後得出 6 個重要變項與住宅竊盜群聚具有關聯，其中人口密度在 AdaBoost 模型出現 1 次，後來暫予以刪除，警力配置則是在羅吉斯迴歸中因調高閾值而被選入為重要變項，但為求較廣泛多元的解釋，在此均列入討論。經由相關分析後，各變項解釋如下：

1、較高房價及收入有助物理防衛能力提升以遏制住宅竊盜犯罪

過去若干研究顯示房地產價格較高地區與被害的風險成正相關，似符合日常活動理論標的吸引力之假設(Bernasco & Nieuwbeerta, 2005)，但也有些研究顯示相反的結果(Kuo, 2014；余任晴, 2019)。在本研究中，從犯罪熱區的群聚掃描中即可發現大安、信義等住宅價格較高地區並非群聚的高風險區，而區位變項的分析明確顯示房價較高地區多為冷區範圍，且具很高的顯著性；若進一步從「平均住宅價格」與「收入中位數」的相關分析來看，兩者呈現中度正相關，據此或可解釋為房價較高地區之住宅由於屋主經濟能力較佳，為保護身家安全，花費在防竊的設施也較多，且此地區之住宅建築物以華廈居多，一般均設有保全人員，門禁管制嚴格，竊盜犯罪者難有機會進入，此與日常活動理論的監控能力加強，阻斷犯罪之機會相符合，說明監控強化對於防止犯罪的發生具有良好效果，而此兩變項的相關分析也顯示社會解組與犯罪機會兩者具有正向關係；故本研究「社會解組與犯罪機會為正向關係」獲得支持，而雖然本研究未能將住宅保全列為物理防衛變項，但從上述推論亦可間接支持「物理防衛能力與住宅竊盜犯罪風險為負向關係，較不易形成住宅竊盜犯罪熱區」的假設。

2、空屋率高導致領域感與監控能力降低易形成住宅竊盜犯罪熱區

在本研究中，「低度用電戶住宅比率」代表是空屋率意義，無論是資料探勘或羅吉斯迴歸，此變項均不斷被列出為重要變項，分析顯示空屋率較高地區成為住宅竊盜熱區機會大增，就新機會理論而言，空屋率代表住宅因常無人居住，造成監控力降低，增加犯罪機會，過去相關研究也得出住宅周圍為無人居住之空屋，由於缺乏領域感而容易遭竊的結論(劉擇昌、鄧煌發、張智雄，2013)。

此外在相關分析顯示，低度用電住戶比率與人口密度、收入中位數及住宅價格等均呈現負相關；換言之，空屋率越高則住宅價格、收入及人口密度也越低，代表空屋率較高地區也有可能是社會解組較嚴重地區，此支持本研究所提出「犯罪機會與住宅竊盜犯罪風險為正向關係，易形成住宅竊盜熱區」及「社會解組與犯罪機會有正向關係」之假設。

3、單獨生活戶數比越高監控力越低易增加住宅竊盜犯罪機會

單獨生活戶數比率對於潛在犯罪者而言，代表犯罪機會的成功機率，強調犯罪標的物的可見性和可接近性。在本研究中將此變項列在新機會理論中，但同時強調，從社會解組理論來看，當單獨生活戶數比越高，代表該地區家庭整合程度越低，愈有可能產生社會解組。

在住宅竊盜犯罪冷熱區分析結果顯示，單獨生活戶數比在熱區有較高的情形，另在相關分析中則顯示，此變項同時與「低度用電戶住宅率」呈現正相關，說明單獨生活戶數比與空屋率相同都會造成住宅缺乏監控，而導致竊盜犯罪發生機會增高，再次印證新機會理論之主張。而從社會解組理論變項來看，此變項與「收入中位數」呈現中度正相關，與「低教育程度率」呈現低度負相關，一方面可解釋可能是因為收入較低無法扶養他人，另一方面也可能意味著喪偶或離婚等；但無論原因為何，此與前述所謂單獨生活戶數比越高，代表該地區家庭整合程度越低，愈有可能產生社會解組的推論頗為一致，此驗證本研究「犯罪機會與住宅竊盜犯罪風險為正向關係，易形成住宅竊盜熱區」與「社會解組與犯罪機會為正向關係」之假設。

4、臺北市警力配置呈現雙向因果現象

根據日常活動理論理論，警察的出現有助加強監控力，阻絕犯罪者動機，減少犯罪機會，但過去相關研究指出，警力配置與犯罪率之間可能產生「雙向

因果」的關係，即增加警力配置與見警率固然對於抑制犯罪有所幫助，理論上應是對於降低犯罪率有正面影響，但實際的情況卻常是因區域治安情況不佳，故有增加警力之必要。本研究分析結果顯示，住宅竊盜犯罪熱區呈現較多警力配置之現象，同時相關分析也得出，警力配置與新機會理論的「低度用電戶住宅率」、「單獨生活戶數比」呈現中度正相關，與「人口密度」為低度負相關，此代表目前我國對於警力配置著重在犯罪機會較高之地區，而並非僅以人口密度為主要考量。此外，若再從社會解組理論變項觀察，警力配置與「低教育程度率」為低度正相關，與「收入中位數」為低度負相關，同樣可推論社會解組現象較嚴重地區有更多警力配置；此項結果與本研究「社會解組與物理防衛能力為負向關係」及「犯罪機會與物理防衛能力為負向關係」2項假設相反。

5、監視器及路燈的防竊效果尚難一概而論

監視器與路燈向被視為對犯罪的發生有重要的監控抑制效果；在本研究中警方監視器的裝設密度對於住宅竊盜冷熱區形成並非重要影響因素，原因可能是監視器裝設地點多在路口，重點在於監控交通或其他類型犯罪，防竊效果有限，而目前私人對於居家防護多是採取自行裝設監視器，此部分數據尚無法取得，故僅以官方資料驗證恐產生較多誤差。

路燈密度在國內外的研究中咸認為與犯罪率的關聯性不大，然而在本研究中，住宅竊盜熱區卻有更高的路燈密度，除與警力配相同可能存有「雙向因果」的問題外，通常路燈密度較高地區也是人口較密集地區，表示有更多的民眾居住，犯罪的標的物增加，使得路燈較密集區域成反為住宅竊盜風險較高地區；此外住宅竊盜多於白天民眾外出期間發生，故路燈對於該區域住宅竊盜犯罪影響不大，且竊賊對於標的物大多會透過許多方式交叉比對以確定是否無人在家，路燈對於竊賊之侵入動機之遏制效果有限。而從相關分析結果看出，「路燈密度」與「人口密度」呈現中度正相關，與社會解組變項「收入中位數」未達統計顯著，顯示路燈照明的設置仍係基於民眾使用需要為主。

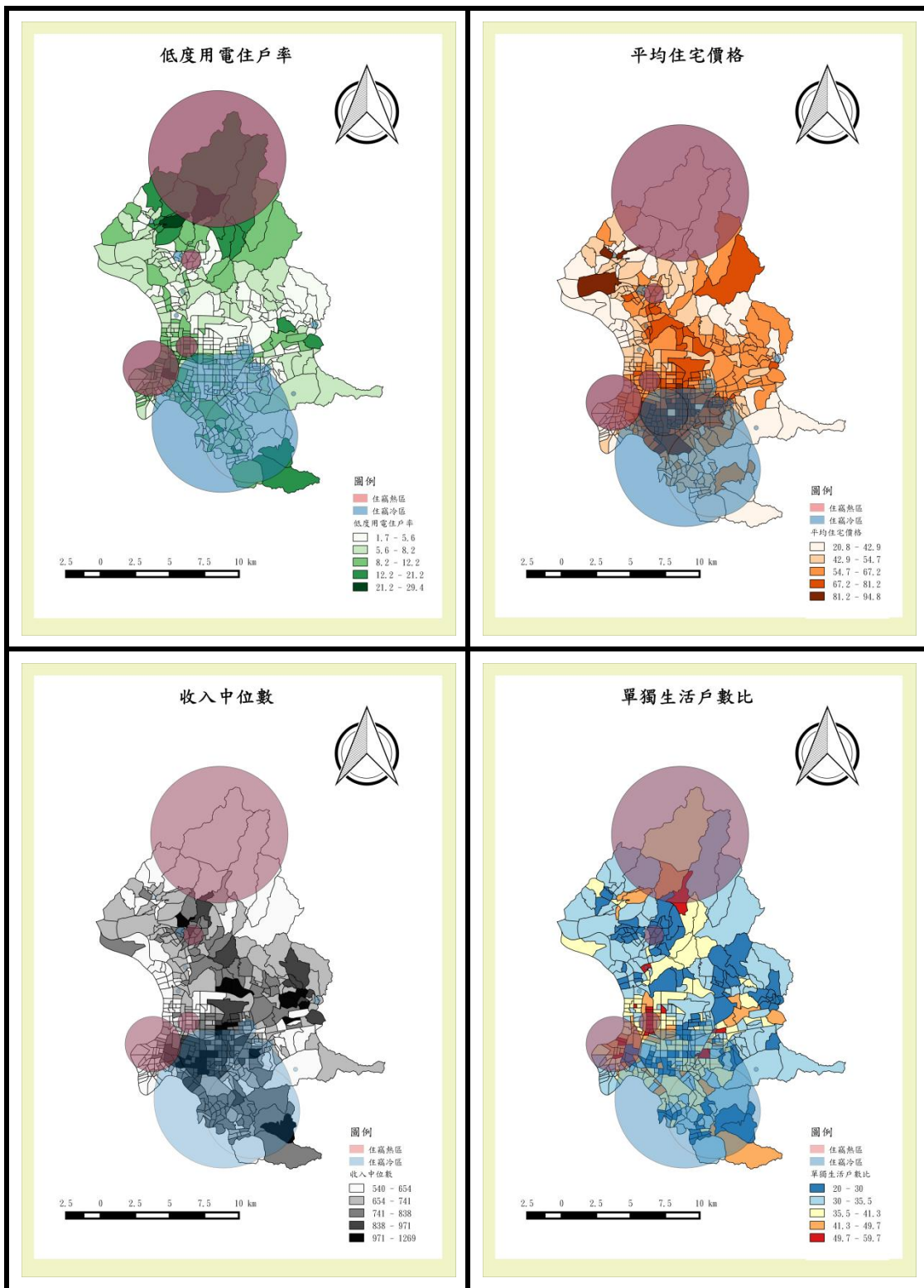
由於本研究無法取得私人監視器正確數據，而對於路燈密度之驗證與其他研究並非一致，故有關監視器與路燈防竊效果尚難一概而論；本研究「物理防衛能力與住宅竊盜犯罪風險為負向關係，較不易形成住宅竊盜犯罪熱區」之假設未能獲得支持。

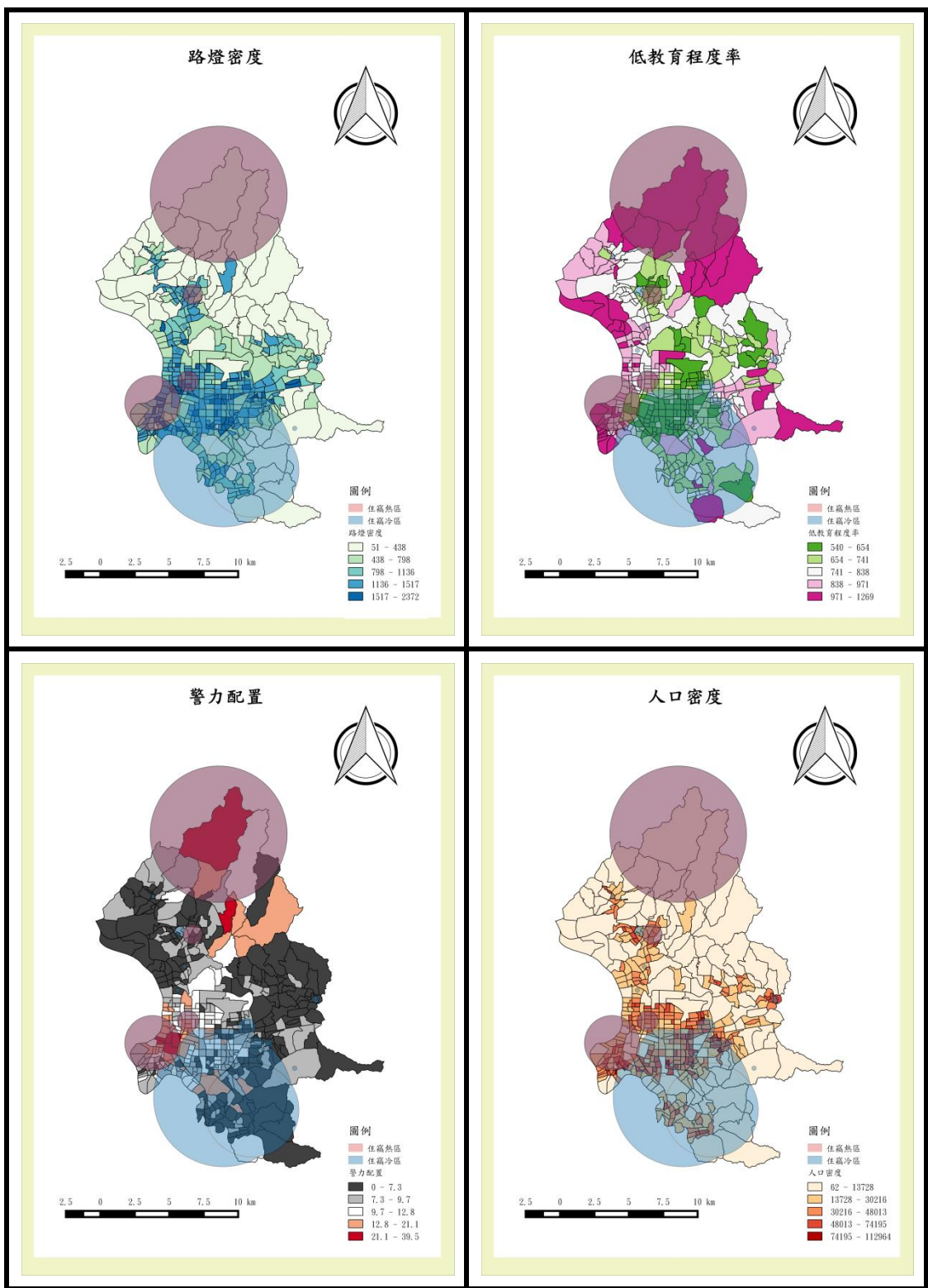
(三) 社會解組與新機會理論方面

從相關統計觀察，社會解組理論的 2 個變項與新機會理論 6 個變項間，除「收入中位數」與「人口密度」、「收入中位數」與「路燈密度」未達統計顯著外，其餘均有若干不同程度的正負相關；具體來看，低教育程度率越高地區平均住宅價格越低($r=-.180, p<.01$)、空屋率及單獨生活戶數比越高($r=.323, p<.01$; $r=.333, p<.01$)；而從決策樹的規則分類觀察，如果平均住宅價格較低、單獨生活戶數比較高、低度用電住戶比較高且收入較低則為住宅竊盜犯罪熱區；此項分析更直接指出社會解組與犯罪機會間具有直接正向關係，與前述之假設驗證相當一致。

總結而言，本研究運用掃描統計與地理資訊系統將臺北市住宅竊盜犯罪群聚分為冷熱區，並依循社會解組理論與新機會理論的指引，選取 18 個區位變項進行分析，藉由多種資料探勘模型與羅吉斯迴歸測試與驗證，得出與住宅竊盜犯罪群聚的重要變項後再佐以相關分析觀察變項間關聯強度。研究結果顯示，在社會解組理論方面，社經地位變項較人口流動率、集體意識等其他指標更能解釋住宅竊盜犯罪群聚的原因；而新機會理論諸多變項除再度證明犯罪標的物的可見性、可接近性及監控力降低將導致較高的犯罪機會外，經由其與社會解組變項的相關分析，與決策樹的條件式推論，可發現兩項理論之變項有相當程度之關聯，呈現社會解組較高地區較易有犯罪機會發生的現象。

此外，較特別是警力配置、路燈密度等抑制犯罪的監控機制，在本研究出現雙向因果的現象，顯示司法警政機關仍傾向以加強監控來控制犯罪的治安策略，至於成效如何，尚待未來更多研究持續評估。最後，本研究以視覺化方式呈現區位變項的分析結果，將冷熱區的覆蓋範圍與各重要變項加以套疊，相信能更清楚觀察造成臺北市住宅竊盜犯罪群聚的風險因素，也同時完成整合掃描統計、資料探勘技術與地理資訊系統的研究目的。





註：紅色圓形區域為熱區，藍色為冷區。

圖 5-5-1 群聚冷熱區之區位變項圖

第六章 研究結論與建議

第一節 研究結論

犯罪時空群聚與區位分析為整合多項技術的綜合性研究領域，也是國內外研究的焦點領域之一，如何客觀、準確的探測犯罪的時間、空間以及時空熱區，進而發掘造成犯罪熱區的因素，長久以來為學界努力的方向。本研究結合傳統統計、時空掃描統計、地理資訊系統與資料探勘技術，以臺北市 2015 年至 2017 年住宅竊盜犯罪資料為實證研究對象，逐步進行各類主題分析，除對犯罪資料進行探索分析外，並藉由時空群聚分析，確認犯罪時空群聚態樣，從而建構住宅竊盜犯罪的預警模型；最後運用資料探勘與統計方式對造成犯罪群聚之因素進行區位分析。有關研究結果重點摘整如下。

一、資料探索性統計分析方面

本研究對於臺北市住宅竊盜時空分布的探索性分析主要使用統計方式進行檢定，根據所蒐集資料加以分類為年別、月別、星期別、時段別與地區別等。在時間分析方面主要使用卡方適合度檢定犯罪時間分布是否差異，此方式可以彌補頻率分析僅能顯示犯罪發生時間最多次數，而不能證明時間分布是否差異的問題；而對於空間分布的差異分析，除卡方適合度檢定外，另加入單因子變異數分析，以進一步了解臺北市各行政區之間的差異情形。

(一)住宅竊盜犯罪易發生於非假日，時段因報案落差無法確認

時間次數分布與卡方適合度檢定分析結果，臺北市 2015 年至 2017 年間除月分布未達顯著差異外，其餘顯示臺北市的住宅竊盜在時間分布上確有所不同。在年分布方面，以 2016 年發生數最多；月分布方面，以 1、4、7、8 及 10 月為發生較多月份；周分布方面，星期一、二、四、五為發生率較高星期別，反之星期六、日及星期三則較低，顯示住宅竊盜易發生於外出工作日家戶監控相對較低之星期別；在每日發生時段的分布方面，上午 10 時至 12 時及下午 1 時至 3 時為發生率較高的 2 個時段，其次為上午 7 時至 9 時及晚間 7 時至 24 時兩個時段；反之凌晨 1 時起至上午 6 時之間則為發生率最低之時段，同樣顯示民眾工作時段易發生住宅竊盜犯罪，惟在時段的分析方面，因多數民眾發覺受竊後再行報案與真實發生時間有所落差，故此部分恐有較多偏誤。

(二)傳統住商混合區有較高住宅竊盜數與風險

在空間分布的差異方面，臺北市各行政區卡方適合度檢定結果，北投區、士林區、內湖區、中山區、萬華區、大安區 6 區，有較多的住宅竊盜發生數，達統計顯著水準；進一步以單因子變異數分析，除顯示各行政區住宅竊盜犯罪數之平均數有明顯差異外，多重比較的結果更可清楚觀察各行政區的具體差異大小。然而，由於各區住宅戶數密度不同；因此，評估住宅竊盜風險程度，宜由犯罪率加以觀察，本研究將住宅竊盜失竊數與地區住宅戶數相除得出住宅失竊率，失竊率最高前 3 者為中山區、萬華區及士林區；反之，失竊率最低為文山區、大安區及信義區。以里為單位的住宅竊盜失竊率方面，臺北市 456 里的平均失竊率為每萬戶 5.85 戶，其中有 185 個里超過平均數，其中仍以中山區 23 個、萬華區 22 個、士林區 22 個里分居前 3 位，但以百分比來看，萬華區排第一，南港區居次，中正區第三。

從行政區與里不同尺度進行分析的結果顯示，受到「尺度效應」(scale effect)影響，不同的衡量尺度將使相同資料產生不同分析結果，通常較小尺度的分析能獲得更詳細的資訊；此外，對於犯罪風險評估的空間分析不能僅以地區的犯罪發生數量做為基準，必須同時考量人口或住宅等標的物的密度，才能獲致準確的結果；然而，無論是以不同空間尺度或考量標的物的密度，分析結果均顯示臺北市傳統的住商混合區如中山區、萬華區及士林區仍有較高的住宅竊盜風險，初步推論與這些地區的商業活動性質、人員往來複雜、住宅普遍老舊等有相當關聯，也凸顯這些地區可能存有較高社會解組的現象。

(三) 2016 年住宅竊盜犯罪群聚最為明顯且集中在傳統住商混合區

住宅竊盜犯罪是否群聚與發生次數不必然有直接關聯，群聚的判定主要在於事件發生是否呈現隨機分布，並經統計檢定後方能進行推論。在統計的探索分析中，雖顯示中山區及萬華區等有較高的住宅竊盜犯罪發生數(率)，但無法直接推論這些傳統住商混合區具有犯罪群聚現象；因此，本研究在 GIS 空間探索分析部分，使用經地理編碼後的犯罪資料對於具有空間定位資訊之屬性資料以視覺化方式加以呈現，逐步將研究重點導入群聚(犯罪熱區)等相關問題。

首先，本研究以點位圖方式呈現臺北市住宅竊盜發生地點，並藉由此資料進行熱區圖分析，從視覺化呈現結果觀察，發現中山區與萬華區為顏色最深的行政

區，故初步判斷此兩個行政區最有可能成為群聚強度較高地區。隨後，本研究再以「最近鄰分析」進行臺北市住宅竊盜犯罪的全域式群聚檢定；數據顯示，2015年至2017年間，均出現群聚現象，其中以2016年比起其他年度更為明顯，但通常進行全域式群聚檢定的目的只是初步確認有無群聚現象，並無法詳細指出哪些地區為群聚熱區，但從GIS的探索及全域性分析，似乎顯示臺北市傳統住商混合地區可能具有明顯住宅竊盜犯罪群聚情形，且以2016年最為嚴重。

二、掃描統計分析方面

在局部性群聚檢定方面，本研究使用掃描統計進行處理，主要原因在於此檢定方式不但可處理空間問題，也能同時將時間因素納入分析，此對時間與空間相互作用的關係可更深入了解。

本研究在掃描統計方面共分為空間掃描、時間掃描與時空掃描3部分，而在時空掃描分析又分為回顧性與前瞻性分析2類，重要發現如次：

(一) 空間掃描分析結果

1、橢圓形視窗對小區域及不規則群聚掃描有更好偵測能力

在空間掃描分析方面，本研究使用圓形與橢圓形兩種視窗進行掃描。基於3年總體資料掃描結果顯示，圓形視窗共掃描出87個可能群聚里，14個群聚區，其中4個群聚區達到本研究設定的統計顯著($p\text{-value} < .10$)，群聚區的排序係根據「對數概似率」大小依序排列，第一個最大可能的群聚區主要集中在中山區，第二個可能的群聚區包括萬華區、中正區和大同區3個地區，第三可能群聚區為北投區與士林區，第四可能群聚區仍為士林區與北投區。橢圓形空間掃描方面，出現7個可能群聚區，其中3個群聚區達到統計顯著，第一個最大可能的群聚區分布在中山區及中正區，第二群聚區在北投區與士林區，第三群聚區為中正區、萬華區、文山區及松山區。

從橢圓形視窗掃描的結果觀察，本研究基本上印證Kulldorff對於橢圓形掃描在範圍不是太大的地區，比起圓形視窗掃描能得到更大概似率，及對形狀不規則群聚偵測更為敏感的結論，有助於發現更多潛藏的群聚。

2、不宜以群聚發生次數多寡做為群聚性質分類標準

根據圓形與橢圓形掃描結果，本研究嘗試以各里在3年間被偵測出群聚熱點的年份與次數，設定各項條件，歸納持續熱點、間歇熱點、消失熱點、偶發熱點及新興熱點5種類型。分類結果顯示，圓形與橢圓形在熱點的性質有相當的不同，特別是在持續、偶發及新興熱點3種分類方面；本研究認為，除掃描視窗不同與主觀性劃分標準所造的差異外，未能將時間因素與空間因素兩者加以融合亦為原因之一；因此，後續的時空掃描群聚分析再次對此差異進行更多檢視與說明。

(二) 時間掃描分析結果

1、臺北市住宅竊盜犯罪群聚時間占研究總時間二分之一

本研究在時間掃描方面分別以「年」、「月」、「日」與「季節性」進行不同時間型態的分析。首先，以3個年度進行總體掃描，臺北市住宅竊盜年分布的時間群聚在2016年；其次，對於各年度的逐月掃描結果顯示，2015年住宅竊盜群聚時間為12月初開始，但此群聚延續至2016年的11月才中斷，2017年則集中在5月份；最後，以全部3年期間掃描，臺北市住宅竊盜群聚月份為2015年7月1日至2016年11月30日，時間長達約1年半時間。在逐日掃描方面，僅有2015年達到顯著水準，時間群聚在2015年12月19日至2015年12月23日，與月掃描結果相符，但縮小尺度可更看出精確的群聚日期。

2、歲末年終與跨年假期為住宅竊盜犯罪高發生時段

季節性是反映時間序列在不同年份的相同季節所呈現之週期性變化。本研究在「季節性」掃描方面，首先以月為尺度，掃描結果顯示時間群聚集中在歲末年終與跨年之際，期間約2個月，就臺灣社會的生活型態而言，此時工作通常較為忙碌，同時各類活動也較多，從日常活動理論來看，此時代表監控能力降低與標的物的增加，導致有更多的住宅竊盜犯罪選擇在此時期進行；若以日掃描為尺度，則群聚時間為12月29日至1月4日，時間恰為我國每年元旦連續假期，通常民眾多會利用此假期全家外出旅遊，此提供住宅竊盜絕佳的犯罪機會。在過去的實證研究中，由於研究時間與地區不同，故對於住宅竊盜犯罪發生的確切月份並沒有一致結論，本研究的季節性掃描分析可有效解決此一問題，並進而將群聚的時間縮小至確定之日期。

(三) 時空掃描分析結果

時空掃描分析方面，本研究分別進行「回顧性」與「前瞻性」掃描，其中「回顧性」掃描尚包括「時間趨勢中的空間變化」掃描分析，而「前瞻性」掃描主要用於建構住宅竊盜預警模型。

1、時空掃描可明確辨識不同群聚發生時空與群聚性質

在「回顧性」掃描部分，本研究以整體3年時間進行時空掃描，結果呈現第一個最大可能的群聚區仍為中山區，第二個可能的群聚區為中正區螢雪里，第三可能群聚區為松山區安平里，第四可能群聚包括中正區、萬華區與北投區，第五群聚區為北投區與士林區，第六群聚區為萬華區與中正區。值得注意的是，若以純空間進行分析，無論是本研究的掃描統計或以GIS分析，僅能發現群聚程度較高之區域，如中山區；但時空掃描則可進一步發現同一地區在不同時間重複發生群聚，也可看出在相同時間範圍內在不同地區發生群聚，此更有助肅竊方案的規劃與警力布署。

與純空間掃描相比較，時空掃描在偵測群聚熱點的總數量雖較少，但能夠偵測空間掃描未能偵測之地點並顯示出群聚的時間，在警力缺乏與勤務繁多情況下，熱點與熱時合併的時空群聚無疑是擬定肅竊方案與執行計畫的更加選擇。

此外，由時空群聚的分析可知，對於群聚發生是否偶發或長期性等，若是以人為主觀的方式對空間進行時間區隔，將產生許多偏誤，而加上群聚時間後，群聚的性質即可一目了然，故時空掃描對於前述持續熱點、間歇熱點、消失熱點、偶發熱點及新興熱點5種類型有更清楚之詮釋，可有效辨識各類型熱點的性質。

2、SVTT分析具「防微杜漸」的犯罪預防作用

「回顧性」掃描中的「時間趨勢的空間變化」(SVTT)分析，旨在檢測在時間趨勢中具有異常現象的空間區域，此為掃描統計非常特殊的偵測功能，一般GIS並不特別關心此類群聚現象。掃描結果顯示，除大同區的建功里外，先前掃描中未被偵測的群聚熱點，如南港區、內湖區及信義區的若干里，在SVTT分析中被列為異常地區，主要原因在於這些群聚區在臺北市總體犯罪率下降趨勢中卻呈現「逆勢上揚」的異常趨勢，某些地區從犯罪風險來看尚非屬於住宅竊盜犯罪熱區，此為警政單位須特別加以注意可能爆發的新興熱點。

3、建置犯罪群聚預警模型有利於犯罪預防「超前部署」

「前瞻性」掃描乃是對現存的群聚進行偵測，以預測未來發展趨勢，有如統計學的犯罪預測，其掃描方式是從資料研究時間最後期限向前推算，並以「無效預警發生率」判定群聚機率高低。在本研究中，嘗試運用「前瞻性」掃描原理模擬建置即時監測與預警模型；預警掃描結果顯示，從 2017 年的 7 月起，群聚的地區落在士林區、南港區、內湖區與信義區，與「回顧性」的 SVTT 掃描結果頗為類似，但增加群聚時間的資訊，此凸顯「前瞻性」掃描的預警特性，對於犯罪預防超前部署實質功能。

三、臺北市住宅竊盜區位分析

在第三章中本研究根據社會解組理論與新機會理論，由各類資料庫中初步選取 18 個與住宅竊盜相關之變項，區位分析之目的即在確認哪些變項對於臺北市住宅竊盜的群聚具有影響；使用方法包括掃描統計、資料探勘方法、羅吉斯迴歸及相關矩陣分析，過程中除篩選出重要區位變項外，也同時對資料探勘與羅吉斯迴歸的模型效能進行評估。研究發現如下：

(一) 與群聚無關之變項應盡量予以排除避免影響研究正確性

掃描統計在本研究的區位分析中主要用途為進行住宅竊盜群聚冷熱區分組與共變項檢定。在冷熱區分組方面，偵測到 4 個熱區，計有 61 個里，3 個冷區，計有 173 個里，全部共計 234 個里，將冷熱區分組結果做為環境區位分析的分類組別。而在共變項分析方面，原本 18 個變項經檢定後，其中做為社會解組理論集體效能機制的「社區發展協會比率」、「區里活動中心比率」兩變項，及屬新機會理論的犯罪機會「戶量」與本研究住宅竊盜群聚無相關性，應予以排除，以免此類變項影響研究結果與推論，故區位分析以 15 個變項進行，並於分析前進行共線性診斷；診斷結果顯示，本研究所選取之變項 VIF 值均低於 10，共線性現象並不嚴重。

(二) 強化學習法具有較優異之模型效能

本研究資料探勘模型共有「決策樹」、「條件決策樹」、「隨機森林」、「AdaBoost」及「XgBoost」5 種，均屬於決策樹族模型，其中「決策樹」、「條件決策樹」屬

於單一樹模型，而後 3 種屬於多樹所組合的強化(集成)學習方法。各模型經調整優化後，篩選之重要區位變項為：「平均住宅價格」、「低度用電戶住宅率」、「收入中位數」、「單獨生活戶數比」、「低教育程度率」及「人口密度」6 個變項，其中「人口密度」僅出現 1 次，其餘變項重複出現頻率較高。

在模型表現方面，本研究以「錯誤矩陣」與「ROC 曲線」的 AUC 值進行評估，並分別以「訓練樣本」、「驗證樣本」及「測試樣本」的數值進行觀察。錯誤矩陣的評估顯示，表現最優異的模型為 XgBoost，在訓練樣本中即可達到完全正確分類，而在驗證與測試過程亦表現良好的穩定性與泛化能力；反觀，條件推論決策樹在此項評估中表現並不佳，其驗證樣本與訓練樣本差異最大。而在 AUC 值表現方面，XgBoost 依然表現最為優異，而隨機森林在此評比中則超越傳統決策樹，與 AdaBoost 表現相當，此 4 種模型均達到極佳的鑑別力的標準，而條件推論決策樹僅達可接受的鑑別力的範圍。

綜合兩種不同評估方式，強化(集成)學習的 XgBoost、AdaBoost 及隨機森林模型，整體表現更優於單棵樹木的傳統與條件推論決策樹，這也說明強化演算法在資料探勘與機器學習領域逐漸成為主流的原因。

(三) 羅吉斯迴歸易於解釋且模型表現優異

在羅吉斯迴歸分析方面，本研究重新投入經「標準化」處理的 15 個變項，進行檢定，結果顯示所選取的解釋變項能有效的影響目標變項且整體模型適配程度良好。另由模型係數顯著性檢定顯示，空屋率越高(低度用電住宅率)、路燈密度較高及房價越低之地區有較高住宅竊盜熱區的機率。本研究再以向後法進行變項篩選，得出低教育程度率、平均住宅價格、低度用電戶住宅率、路燈密度及警力配置為達到顯著的重要變項；因此，羅吉斯迴歸最終對於臺北市住宅竊盜犯罪熱區與冷區變項解釋為：較高的低教育程度率、低度用電戶住宅率、路燈密度、警力配置及較低的平均住宅價格成為犯罪熱區的機率較高。

由於羅吉斯迴歸分類方式與決策樹運用資訊理論計算屬性的資訊增益量的分支演算法並不相同，故對於變項的篩選略有差異。為進一步評估此兩大類模型的效能，本研究再次以錯誤矩陣與 AUC 值進行評估；結果顯示，在兩種不同的評估方式中，羅吉斯迴歸表現均不亞於資料探勘的方法；總體來說，羅吉斯迴歸較單棵決策樹的演算法表現更好，而多棵決策樹的強化學習則優於羅吉斯迴歸，

不過差距並不大，在評定標準上同樣擁有極佳的鑑別力。本研究認為，由於羅吉斯迴歸統計可列方程式，並對變項進行檢定，此對於整體現象的解釋較具優勢，而資料探勘模型並不關注研究假設與資料分布型態，著重在發掘未知與模型準確性，兩者各有所長，具有互補的作用，故在條件允許下，進行資料探勘後，最好能以適當統計方式再次分析，相互印證，較能獲致更客觀的結果。

(四) 具社會解組特性地區提供更多的住宅竊盜犯罪機會

社會解組理論假設犯罪是貧困、人口流動及社會階層較低等地區特性的產物，新機會理論則強調合適標的物、方法或工具、情境等變化對於犯罪型態與數量的影響，兩項理論在許多研究中常被做為區位分析的理論基礎，並依據理論指引選取變項進行區位分析。

從本研究的決策樹的分類規則觀察，收入較低地區，也就是社會解組特性較明顯地區成為住宅竊盜群聚熱區的機率較高，而這些地區若同時出現空屋率較高、單獨生活戶數比率較高等易增加犯罪機會的情況，則幾乎成為熱區的機率為100%。此外，藉由相關分析中也顯示此兩種理論之變項具有密切的關連，從而可推論具社會解組特性的地區提供更多的住宅竊盜犯罪機會，此說明兩項理論具有互補且可頗適合整合的可行性，也是相當適用於犯罪區位分析的重要理論。

根據以上，第二章之相關實證研究結果與本研究比較分析如下（表 6-1-1）

表 6-1-1 住宅竊盜犯罪相關實證研究與本研究結果比較分析表

研究內容	實證研究結果	本研究結果	異同
犯罪時間分布分析	1、臺北市住宅竊盜犯罪高發生時段為8時至23時(彭福康，2015)。 2、臺北市住宅竊盜發生日分布以星期一至星	1、住宅竊盜高發時段分別為上午10時至12時間，下午1時至3時間；凌晨1至6時為發生最低時段。 2、日分布以星期	1、本研究結果與實證研究在日分布結論相似，均印證住宅竊盜易發生於民眾平日工作外出。 2、在時間分布方

	<p>期五較高(劉擇昌, 2011; 邱奕堯, 2013)。</p> <p>3、臺北市住宅竊盜發生日分布以1至5月為高發月份(邱奕堯, 2013; 彭福康, 2015)。</p>	<p>一、二、四、五最易發生。</p> <p>3、月分布次數以1、4、7、8及10月為最多次數, 但未達統計顯著。</p>	<p>面, 因竊盜報案時間可能有所落差, 本研究僅列為參考。</p> <p>3、本研究使用卡方適合度檢定, 部分實證研究係以次數統計。</p>
犯罪空間群聚分析	<p>1、2008年至2014年, 臺北市住宅竊盜群聚主要集中在信義區之中心位置、松山區與大安區交界處及涵蓋萬華區、大同區、中正區之交界範圍(彭福康, 2015)。</p> <p>2、2010年, 臺北市住宅竊盜犯罪群聚主要集中在大安區、中正區、大同區及萬華區全境幾乎均在群聚區內, 而中山區、與松山區也有</p>	<p>1、本研究以圓形視窗進行掃描, 結果顯示2015至2017年臺北市空間群聚主要集中在中山區、萬華區、中正區及大同區。</p> <p>2、本研究另以橢圓形視窗掃描, 除中山區、中正區及萬華區外, 北投區、士林區、文山區及松山區部分里亦偵測到群聚現象。</p>	<p>1、由於研究時間範圍不同, 臺北市住宅竊盜犯罪群聚區域有所不同, 其中大安區在本研究時間範圍已非熱區。</p> <p>2、本研究使用方法為掃描統計法, 與前述2013年研究方法相同, 但另以橢圓形掃描進行測試。</p> <p>3、2015年之研究係使用核密度分析, 較難判定確切群聚地點; 而掃描統計</p>

	相當比例納入 (邱奕堯， 2013)。		可藉由地區編 碼辨識精確區 域。
犯罪時空群聚分析	<p>1、2010年臺北市住宅竊盜犯罪群聚包括中山區、中山區、中正區、信義區、松山區及大同區全境。群聚持續的時間長度由長漸短，上半年度的群聚時間往往長達半年，但下半年出現的群聚皆僅持續2個月(邱奕堯，2013)。</p> <p>2、上海市住宅竊盜犯罪時空群聚月份多在11月、12月和1月，異常的持續時間較短，一般在1到3天之間，惟全市幾被熱區覆蓋(王占宏，2013)。</p>	<p>1、2015至2017年臺北市住宅竊盜犯罪群聚時間長達一半時間。</p> <p>2、以不分年度的季節性掃描分析，歲末年終與跨年假期為住宅竊盜高發生時期。</p> <p>3、「回顧性」時空掃描顯示，最大可能群聚區為中山區，其餘如中正區、萬華區及北投區仍是次要群聚區域。</p> <p>4、「時間趨勢的空間變化」掃描分析，偵測出南港區、內湖區及信義區的若干里，犯罪情形並不嚴重，但有異常群聚現象。</p>	<p>1、在時空掃描方面，本研究掃描之群聚地點與前述研究大同小異，但藉由群聚時間長短及是否間隔等特徵，對群聚性質進行歸納與分類。</p> <p>2、除時空掃描外，本研究另使用季節性掃描，可排除不同年度犯罪資料分布相異的問題，確認歲末年終為住宅竊盜犯罪集中之時間。</p> <p>3、本研究使用「時間趨勢的空間變化」掃描分析，可發現犯罪冷區發生異常群聚現象，有利於警方先期預</p>

		5、「前瞻性」時空掃描顯示，群聚的地區落在士林區、南港區、內湖區與信義區。	防。 4、「前瞻性」時空掃描分析可用於建置犯罪預警模型，對於犯罪預防具有「防微杜漸」的實質功能。
區位因素分析	<p>1、在社會解組理論方面，實證研究結果，貧窮地區、單親家庭較多、失業率較高、經濟收入較低、離婚率較高、低教育人口比率較高、人口流動性頻繁等因素較易發生住宅竊盜犯罪。</p> <p>2、在機會理論或日常活動理論方面，人口密度、獨居比率、男女性別比、老人人口率、監視器密度等對於住宅竊盜犯罪均有影響，但研</p>	<p>1、本研究在社會解組理論得出收入較低及低教育人口比率較高，易成為住宅竊盜犯罪熱區。</p> <p>2、在機會理論方面，本研究結果為，平均住宅價格較低、低度用電戶比率(空屋率)較高、獨居比率高，易成為住宅竊盜犯罪熱區，而熱區中有更多警力配置。</p>	<p>1、實證研究區位分析方法多採用相關及直線迴歸分析，以犯罪率為依變項，區位因素為自變項，得出研究結果。</p> <p>2、本研究先以群聚共變項進行變項篩選後，再以資料探勘二元分類技術，依據群聚冷區與熱區為基準，進行變項分類，篩選相對重要變項。</p> <p>3、本研究根據決策樹分類規則與相關分析，顯</p>

	<p>究結果並非一致(Martin, 2002; 邱奕堯, 2013; 周德芳, 2016; 林宛宣, 2019; 陳惠貞, 2009; 彭福康, 2015; 賴致瑜, 2006)。</p>		<p>示社會解組特性較明顯地區，似有更多住宅竊盜犯罪機會，此符合研究架構之假設，而相關實證驗就並未就兩理論間之關係有進一步驗證及論述。</p>
--	---	--	---

第二節 研究建議

一、實務建議

本研究基於「提升警政效能並減輕警力的負擔」之研究動機，在第一章提出3項研究目的—「有效辨識犯罪時空熱區」、「了解犯罪熱區之區位因素」及「提供被害風險資訊俾於勤務規劃」；依據研究目的與成果，提出下列實務建議，以供司法警政機關參酌。

(一)針對不同熱區特性研擬具特色警務規劃並進行 SVTT 評估

由於本研究以里做為研究單位，運用掃描統計的方式，除可立即辨識犯罪熱區發生的里別外，更可明確觀察群聚之時段，以此判斷發生之月份、季節，特別是藉由犯罪群聚持續時間的長短可以研判此熱區為持續或偶發性等性質。例如，中山區在時空掃描中不但在空間上有很強的犯罪群聚現象（第一群聚），群聚時間也最久，故本區即是屬於持續性犯罪群聚區域；而中正區雖然也有相當強的犯罪空間群聚，但群聚時間較短且分散，表示此群聚現象為間歇性群聚，其中螢雪里發生住宅竊盜之月份集中在某月份，且之後未再發生，可判斷為偶發性群聚。因此，對於高度持續性的犯罪群聚可能並非加派警力或增加預算等可以收立竿見影效果，或許更需要其他政府機關介入，例如從市容整治、行業規範等進行區域

治理，改變區位特性，才能達到長治久安之效果，警務的規劃朝向保持穩定控制、逐步降低的長期防治策略。

對於間歇性與偶發性群聚，可參考 Goldstein(1979)的「問題導向警政」(problem-oriented policing strategy) SARA 模式³⁶，對於住宅竊盜群聚的區域，除檢視轄區警力配置是否足夠或監視器、警民連線等監控設備能否正常運作等外，另配合運用本研究時空掃描結果，詳細比對群聚地點與時間，將警政資源(如警力數)部署在這些特定地區，並提醒該地區里長或巡守組織於特定時間加強自我防護，如此應是較為經濟且有效的做法。

以上無論是針對任何型態的住宅竊盜犯罪所進行的防治策略與具體作為，本研究建議應佐以「時間趨勢的空間變化」(SVTT)分析，藉由定期或行動方案前後時間的掃描分析，觀察並比較地區的整體犯罪趨勢的波動及異常群聚變化，以評估實施前後成效，隨時進行策略與方案調整。

(二)善用警政資料以建構住宅竊盜犯罪預警系統

長久以來，我國犯罪防治雖強調打擊與預防犯罪相結合，既治標亦兼治本，但在實踐上仍是偏於「重打擊、輕預防」的取向，地區性警察機關因勤務繁多，資源有限，或囿于傳統思維方式，在犯罪預防工作仍以描述性統計、文獻分析、理論思辨等做為主導方法，成效有限。然而，考察目前許多先進國家，在科技支援與數據累積情況下，對於犯罪預防多採大數據分析進行，將大量的司法警政資料整合加工，透過預測犯罪趨勢、推測犯罪模式及確定犯罪熱點等分析方法，採取適時適地的警務作為，例如：加強巡邏、增設監控等對高犯罪率發生之環境進行防控，均獲得不錯之成效，其中即時監測系統與預警系統更成為犯罪預防的重要利器。

所謂即時監測與預警指的是對於關注現象不斷進行觀察並紀錄結果，並根據所設定的臨界值等，對於異常狀況發出警訊，俾進一步採取相應措施，當前已運用在交通路況、氣象、土石流、環保空汙及疫情等監測。目前警政署「警政統計

³⁶ SARA 模式為：

- 1、檢視問題 (Scanning)：掃描資料，以確認日常處理事件的方法。
- 2、分析問題 (Analysis)：針對問題詳細分析。
- 3、回應問題 (Response)：執行干預，讓問題減少發生。
- 4、評估成效 (Assessment)：重複評估干預的影響。

通報」固定每周發布，資料彙整自各地警察機關，但目前呈現方式僅止於描述性統計，尚未能發揮數據資料累積的應有價值；因此，在本研究中以官方發布之資料，運用「前瞻性」掃描分析，模擬建構犯罪「即時監測與預警」模型，並建議分析資料可即時提供轄區警方參考，俾於先期進行犯罪預防布署作為。

由於「即時監測與預警」模型所偵測的犯罪群聚與一般犯罪群聚所偵測的結果相當不同，所提供訊息屬於「防微杜漸」，故警察機關可根據轄區狀況進行研判，自行衡量介入時機與方式，在避免犯罪情勢擴大前提下，以合理的警政資源便可達到消弭犯罪於無形的目標。

(三)根據最犯罪群聚最大概率與被害風險數據合理配置警政資源

警政資源包括人力、物力與財力等，為相當複雜的問題，如何合理配置更是艱鉅的工作，在此難以深入探討。僅就本研究的住宅竊盜犯罪而言，在犯罪熱區的分析中，除提供熱區地點與時間外，另有「對數概似率」(LLR)與「相對危險率」(RR)兩項數據指標，前者指最可能發生犯罪地區，後者指風險最高地區，故各轄區警政機關可根據此兩項指標，給予不同權重加以計算，從而訂定治安目標，以科學計算方式投入適當資源，可減少警政資源的不當配置。

(四)根據不同區位因素採因地制宜之防治策略

對於導致住宅竊盜犯罪的因素有許多研究面向，部分實證研究聚焦於犯罪人、犯罪前因論或巨觀之社會結構、社會經濟因子，而提出如拯救貧窮、提升教育水準、降低失業率等必須耗費大量財政與人力且超出警政機關職權之建議，通常這些結論多來自於社會解組理論；對此，本研究認為從巨觀的社會經濟因素探討地區住宅竊盜成因並非一無可取，此對於警政機關了解地區特性仍有一定幫助，從而在擬定防治策略時也才能「因地制宜」；例如，低收入與低教育程度地區有較高住宅竊盜率，警方則可從人口組成進一步了解是否地區中有更多的犯罪再犯比例；此外，此地區居民守望相助與防竊意識通常較為低落，故可預期強調警民合作、社區集體意識的防治策略成效將不會有太多效果，仍必須回到新機會理論範疇，針對路燈、監視器密度、單獨生活戶數比、低度用電戶住宅率及平均住宅價格的高低等因素，從建構安全之防衛空間及加強情境犯罪預防著手，以阻絕犯罪之機會。

二、後續研究建議

本研究以臺北市 2015 年至 2017 的住宅竊盜為分析樣本，依據犯罪學理論蒐集各類區位變項，並採用各種不同分析方式與工具進行分析，最終雖得出若干研究結果，但影響犯罪發生的因素相當複雜，且不同研究方法與工具所獲得之結果也不盡相同；因此，以下就本研究未能操作與顧及的層面，提出若干建議供未來後續研究參考。

(一)宜以不同尺度的時空環境進行分析

從空間尺度來看，本研究以臺北市為總體研究範圍，對熱區分析以里為主要分析單位，在犯罪空間尺度可歸類為大尺度的實質環境空間分析；因此，對於犯罪熱區中更細微的街道、商圈及住宅等地方(place)特性未能詳加探討，無法提供進一步資訊。相對地，臺北市為典型都會型態，無論是人口分布、居家工作型態與其他縣市有相當不同，若能以全臺地區或選擇其他農業縣市做為比較，所得出之結論應更為全面。此外，由於雙北緊鄰且交通便利，若以臺北市為分析樣本，最好同時納入新北市，才能有更完整的犯罪群聚觀察。

在時間跨度方面，本研究因資料限制僅有 3 年資料，若能以更長期的資料(如 10 年以上)，或能從熱區在時間趨勢變化中，進一步探討都市發展和犯罪發生之間的關係與影響。

(二)加強區位分析的多元化

本研究之區位分析資料以政府公開資料為主，大部分屬於非空間屬性資料，用以解釋空間現象不免造成偏差與侷限性；事實上，許多非空間屬性資料均可藉由 GIS 功能予以加工後變成具有空間屬性之資料，例如人口密度即是最簡單的空間屬性資料；此外，也可藉由加入自然環境、土地使用分區圖、公共建設、特殊地標、路網圖及警政機關位置等，透過多種圖層整合技術或以距離矩陣等計算遠近關係，均可獲得各種空間屬性的區位變項。雖然，窮盡再多的區位變項仍無法對犯罪的複雜因子做出最完整的解釋，但相信區位分析的多元化更有助於掌握犯罪區位的全貌。

除了增加區位變項的來源外，根據新機會理論與許多研究指出，許多不同類

型的犯罪是相互關聯，一個犯罪會衍生出另外一個犯罪機會；例如某地區毒品犯罪盛行，若吸食者因缺錢購買毒品，可能進行竊盜或搶奪，造成該地區同樣具有更高的竊盜犯罪；因此，若有不同類型的犯罪資料，可考慮合併分析，也是拓展區位分析多元化的有效方法。

(三)運用跨領域與不同分析工具進行研究

自 GIS 技術興起後，透過 GIS 科技針對犯罪相關之空間或非空間資料整合的研究日益增多；在實務運用方面 CompStat 的資訊統計警務模式，即是一套結合電腦統計和 GIS，讓警方可更精準打擊犯罪的治安利器；因此，如何運用跨領域技術於犯罪分析與偵查實務已是當前各司法警察機關積極推動的目標。

本研究使用流行病學偵測疾病群聚的掃描統計，在犯罪領域尚少有運用，但研究結果顯示，此項技術可以將時間與空間同時結合，解決 GIS 系統難以處理的時空群聚問題，而使用之介面仍為 GIS，基本上可視為 GIS 的強化版本，值得向有志於 GIS 之研究者推薦，並運用於其他類型犯罪之研究，或許能獲得更多過去未曾發現的新知識。此外，目前對於犯罪熱點的分析工具可謂玲郎滿目、各有所長，即便是本研究所使用的時空掃描軟體 SatScan 也有許多的競爭軟體(如 FleXScan、Splancs 等)，各種分析工具得出之結果不盡相同，建議可使用多種不同工具分析，結果相互參證、互補不足，更增加研究之可信度。

第三節 研究限制

一、資料蒐集之限制

本研究原始資料為臺北市政府自 2015 年起始公布之住宅竊盜犯罪資料，但此資料為保護個資，地址部分採取模糊處理，透過地理資訊圖資雲服務圖臺(TGOS)進行批次定位前，處理程序略顯繁複，且以門牌號碼中間值進行定位難免造成偏差；定位後雖有 96% 定位率，符合 Ratcliffe 指出犯罪案件的定位率 85% 標準，但仍約 50 筆資料無法納入分析。針對此問題，桃園市政府之作法為將地址先行轉換為座標，避免個資外洩問題，但缺少地址部分，改用轄區分局為範圍。故本研究認為，若能將兩者資料格式互補，對於住宅竊盜研究將更為便利，研究結果更易為轄區警方運用。

二、硬體設備有待提升

時空掃描統計分析的運算過程非常耗時並耗費電腦資源，礙於個人硬體設備不足，以至於有關橢圓形時空群聚及時空重排逐日等部分分析難以進行，無法進一步豐富論文內容。此外，本研究模擬建置的監測預警模型，一方面礙於個人能力無法完成整體硬體系統建構，另一方面以現有之硬體恐亦無法進行即時掃描分析；進行此類研究必須注意電腦設備是否足以勝任。

三、增加質性訪談與實地勘察以彌補研究

偵辦住宅竊盜為基層警察之職掌之一，日積月累之工作歷練對於易發生竊盜之犯罪環境與接觸犯罪者之經驗，遠比一般民眾與研究人員更貼近犯罪原貌，而受竊家戶的被害經驗亦是了解犯罪手法與居民意識之重要參考。本研究以行政區里單位，主要對犯罪熱區與區位因素進行分析，卻難以對竊盜犯罪者之動機、手法、住宅防衛設計及被害者經驗有更深入了解與剖析，故未來相關研究宜增加與相關人員之質性訪談，始能更深刻瞭解犯罪環境之特性及研究結果應如何轉化為對警方與民眾更具實質幫助之建議。

此外，從破窗理論而言，住宅竊盜犯罪之原因除人文社經等區位因素外，住宅與民眾實際互動所營造的環境意象也是造成地區是否成為住宅竊盜高風險的重要因素；就此而言，實地勘察並配合質性訪談更有助彌補巨觀研究之不足。

第四節 結語

社會解組理論假設犯罪是貧困、人口流動及社會階層較低等地區特性的產物，新機會理論則強調合適標的物、方法或工具、情境等變化對於犯罪型態與數量的影響；本研究的實證結果發現此兩種理論的變項除對於犯罪群聚均具有解釋力外，藉由相關分析對概念假設的驗證也顯示此兩種理論之變項具有密切的關連，從而推論具社會解組特性的地區較易有更多的住宅竊盜犯罪機會，此說明兩項理論可具體整合、相互為用的可行性，也是相當適用於犯罪區位分析的重要理論。

須特別指出，新機會理論所強調的犯罪「機會」，在本研究的時空意義方面，指的就是犯罪容易發生的某些時段與地點；故無論是從日常活動理論的3項要素或環境犯罪學的「犯罪鐵三角」理論來看，3項要素出現都與時間有重要關係，

即犯罪的發生必須是「時間正好」(time coincides)。過去許多以 GIS 對住宅竊盜的時空分析研究，對於時間分析的一種方式是將時間獨立於 GIS 之外，以統計加以進行，如同本研究在第四章第一節的做法，雖可得到犯罪時間分布情形，卻無法與空間直接產生連結；另一種分析方式則是第四章第四節，以分年度的方式展示空間的群聚變化，但基本上此方式仍屬於純空間的形式。為發現空間群聚發生的「正好」時間，本研究使用時空掃描統計分析，明確指出空間與時間群聚關係，並從群聚時間的長短與發生頻率，進一步對群聚的性質做出判斷，此為本研究特色之一。

此外，有別於以統計方式進行區位分析，本研究嘗試以資料探勘技術對影響犯罪區位的變項進行分類，分類的標準乃是基於掃描統計對群聚偵測的結果，除使研究能具有一貫性，並可避免造成區位誤謬的錯誤，隨後再以統計方式相互驗證比較，最後篩選出重要區位變項，並進行變項間的相關強度分析，力求篩選之變項具有代表性與解釋力。

總體而言，本研究引用跨領域技術對犯罪議題進行分析，主要目的在於探索各類研究方法運用在犯罪分析的可行性，由於國內外似尚無類似研究可供參考，僅能分別從各領域的相關技術中汲取養分，嘗試各種分析方式並加以結合，思慮欠周在所難免，但仍衷心期盼此研究對國內住宅竊盜犯罪問題分析能有拋磚引玉之貢獻。

參考文獻

中文部分

- 內政部不動產資訊平臺。低度使用(用電)住宅、新建餘屋(待售)住宅。取自
<https://pip.moi.gov.tw/V3/E/SCRE0104.aspx>
- 內政部統計處。中華民國 107 年各縣市內政統計指標。取自
https://www.moi.gov.tw/files/site_node_file/8714/107%E5%B9%B4%E5%90%84%E7%B8%A3%E5%B8%82%E5%85%A7%E6%94%BF%E7%B5%B1%E8%A8%88%E6%8C%87%E6%A8%99.pdf.
- 內政部統計處。107 年第 9 週內政統計通報。取自
https://www.moi.gov.tw/chi/chi_site/stat/news_detail.aspx?sn=13553
- 內政部警政署內政資料開放平臺。犯罪資料統計數據。取自
<https://data.moi.gov.tw/MoiOD/Data/DataDetail.aspx?oid=6D9C7F00-3E4C-4FC7-BEDB-28D70FF96FEE>
- 內政部警政署全球資訊網。警政統計年報電子書(2015 年-2018 年)。取自
<https://www.npa.gov.tw/NPAGip/wSite/lp?ctNode=12898&CtUnit=2662&BaseDSD=7&mp=1>
- 內政部警政署全球資訊網。警政統計名詞定義。取自
https://www.npa.gov.tw/NPAGip/wSite/lp?ctNode=12599&xq_xCat=02&nowPage=3&pagesize=15
- 王占宏 (2013)。**基於掃描統計方法的上海犯罪時空熱點分析** (未出版博士論文)。華東師範大學，上海市。
- 王保進 (2006)。**英文視窗版 SPSS 與行為科學研究(第三版)**。臺北市：心理出版社。
- 王翔正(2014)。**侵入住宅竊盜被害影響因素及被害後反應之研究** (未出版碩士論文)。中央警察大學，桃園縣。
- 王麗雲 (2005)。理論在量化研究中的地位。**教育資料與研究雙月刊**，64，139-148。
- 朱家嶠 (2017)。「與研究相關之倫理定義」。臺灣學術倫理教育推廣資源中心。取自網址：https://ethics.moe.edu.tw/files/demo/demo_u01/p02.html
- 朱群芳 (2015)。**臺灣社會中的犯罪控制：比較社區中的社會連結與集體效能之**

- 效應。取自 file:///D:/endnote%20library/E10249r.pdf
- 江羿臻、林正昌 (2014)。應用決策樹探討中學生學習成就的相關因素。**教育心理學報**，45(3)，303-327。
- 江振亨 (2002)。犯罪是理性選擇?-理性選擇理論的實証與面臨之挑戰。**犯罪學會會訊**，3(4)，13-19。
- 行政院主計總處(2018)。國情統計通報 (第一 40 號)。取自 <https://www.stat.gov.tw/public/Data/87261643BYSYDTB1.pdf>
- 何宗武 (2016)。**R 資料採礦與數據分析—以 GUI 套件 Rattle 結合程式語言實作**。臺北市：基峰資訊公司。
- 余任晴 (2019)。**從社會解組與機會觀點看住宅竊盜的發生：集體效能的作用** (未出版碩士論文)。國立臺北大學：新北市。
- 余致廷 (2014)。**運用地理資訊系統與資料探勘技術於基層診所選址分析與研究—以臺北市為例**(未出版碩士論文)。國立中央大學，桃園縣。
- 李良益(2014)。**臺灣地區犯罪趨勢之研究—時間數列模型之應用**(未出版碩士論文)。國立臺北大學，新北市。
- 李波 (2011)。論多層面犯罪理論整合模型。**犯罪研究**，4，24-31。
- 李珀宗 (2005)。**社區犯罪基圖在警察機關防制住宅竊盜犯罪之應用—以臺北市松山區為例** (未出版碩士論文)。中央警察大學，桃園縣。
- 周孟嫻、紀玉臨、謝雨生 (2010)。臺灣自殺率具空間群聚嗎？模仿效應或結構效應。**人口學刊**，41，1-65。
- 周愷嫻 (2017)。全球犯罪率為何同步下降。**刑事政策與犯罪研究論文集**，20，1-13。
- 周愷嫻、曹立群(2007)。**犯罪學理論及其實證**。臺北：五南書局。
- 周德芳 (2016)。**基隆市住宅竊盜犯罪空間與區位分析** (未出版碩士論文)。國立臺北大學，新北市。
- 孟維德 (2001)。犯罪熱點的實證分析—警察局勤務指揮中心報案紀錄的分析與運用。**犯罪學期刊**，8，27-63。
- 林宛宣 (2019)。**應用資料探勘於臺北市住宅竊盜環境特性之關聯研究** (未出版碩士論文)。國立臺灣科技大學，臺北市。
- 林宜甲、黃柏霖(2017)。利用空間性局部指標 (LISA) 分析高雄市 C-Bike 站點與空間相關性初探，**經營管理學刊**，12，125-141。

- 林建隆 (2010)。刑案隱性鏈結關聯模式之研究—以臺北市搶奪與住宅竊盜案為例 (未出版碩士論文)。中央警察大學，桃園縣。
- 林進發 (2005)。臺中市搶奪犯罪熱點之空間分析(未出版碩士論文)。國立彰化師範大學，彰化縣。
- 林維真 (2017)。以時間與空間觀點探討住宅竊盜的發生 (未出版碩士論文)。國立清華大學，新竹市。
- 邱奕堯 (2013)。臺北市犯罪現象之時空分析(未出版碩士論文)。國立臺灣大學，臺北市。
- 邱靖方 (2008)。臺灣地區家戶特性與區域特性對於住宅竊盜被害風險之影響 (未出版碩士論文)。國立臺北大學，新北市。
- 邱豐光 (2008)。常業住宅竊盜犯罪歷程之研究 (未出版碩士論文)。國立臺北大學，新北市。
- 洪百亮 (2009)。常業住宅竊盜犯罪標的空間搜尋模式之研究(未出版碩士論文)。國立臺北大學，新北市。
- 唐建波、鄧敏、劉啟亮 (2013)。時空事件聚類分析方法研究。地理信息世界，20(1)，38-45。
- 徐晨剛(2017)。應用資料探勘技術建構竊盜犯罪預測模型(未出版碩士論文)。中央警察大學，桃園市。
- 殷菲、李曉松、馮子建、馬家奇(2007)。前瞻性時空掃描統計量在傳染病早期預警中的應用。中華預防醫學雜誌，41(增刊)，155-158。
- 翁培文、蔡博文 (2006)。空間離散指標：舊觀念、新公式。台灣地理資訊學刊，4，1-12。
- 翁慈宗(2009)。資料探勘的發展與挑戰。科學發展，442，，32-39。
- 袁帥(2017年2月4日)。壹讀。從5大行業領域看大數據場景應用當前進展。取自 website: <https://kknews.cc/zh-tw/tech/v94qaky.html>
- 張平吾、黃富源(2012)。受害者學新論。桃園縣：臺灣警察學會。
- 張平吾、蔡德輝 (2005)。住宅竊盜重複被害特性與防制策略之研究。政府研究資訊系統(計畫系統編號 PG9403-0341)。取自 <https://www.grb.gov.tw/search/planDetail?id=1071473&docId=203159>
- 張淑貞(2008)。街頭搶奪犯罪之空間與時間群聚性研究—以臺中市西屯區為例(未出版博士論文)，逢甲大學，臺中市。

- 張淑貞、李素馨(2012)。都市街頭搶奪犯罪熱點分析：日常活動理論之觀點。**都市與計劃**，**39(1)**，71-94。
- 張璋倫 (2016)。科技橘報。犯罪能精準被預測？洛杉磯警局用大數據分析降低36%犯罪率。取自
<https://buzzorange.com/techorange/2016/09/21/use-data-analysis-predict-crime/>
- 曹立群、周素嫻 (2007)。犯罪學理論及其實證。臺北市：五南圖書出版股份有限公司。
- 許春金 (2009)。人本犯罪學(增訂二版)。臺北市：三民書局。
- 許春金 (2010)。犯罪學(修訂五版)。臺北市：三民書局。
- 許春金、陳玉書、蔡田木、洪千涵、白鎮福 (2015)。102 年犯罪狀況及其研析。**刑事政策與犯罪研究論文集**，**18**，1-31。
- 許華孚、吳吉裕(2015)。大數據發展趨勢以及在犯罪防治領域之應用。**刑事政策與犯罪研究論文集**，**18**，341-375。
- 許福生 (2007)。刑事政策學(修訂二版)。臺北市：元照出版公司。
- 郭耀仁 (2018)。**DataInPoint**。輕鬆學習 R 語言：起步走。取自
<https://medium.com/datainpoint/r-essentials-quick-start-8226b33514f9>
- 陳陸毅 (2008)。臺中市住宅竊盜空間分布之研究 (未出版碩士論文)。逢甲大學都市計畫所，臺中市。
- 陳惠貞 (2008)。高雄市住宅竊盜犯罪地圖繪製與時空分析(未出版碩士論文)。國立高雄師範大學，高雄市。
- 陳瑞基 (2010)。連續住宅竊盜犯作案手法與空間行為模式剖析(未出版博士論文)。中央警察大學，桃園縣。
- 陳榮昌、林育臣(2003)。群聚演算法及群聚參數的分析與探討。**朝陽學報**，**8(1)**，327-353。
- 陳薇琳(2012)。臺灣登革熱疫情之時空群集研究 (未出版碩士論文)。國立陽明大學，臺北市。
- 黃文、王正林 (2016)。利用 R 語言打通大數據的經脈。臺北：佳魁資訊。
- 傅鈺婷 (2018)。臺灣的毒品犯罪趨勢分析-時間序列模型之應用(未出版碩士論文)，國立臺北大學，新北市。
- 彭福康(2015)。監視錄影系統與竊盜犯罪率之時空分析 (未出版碩士論文)。中央

- 警察大學，桃園市。
- 裘厥庸、李永新 (2010)。公司危機預測：計量模型與變數選取。期貨與選擇權學刊，3(1)，1-26。
- 游士嫻 (2010)。罪惡不能侵擾之蒂：臺灣低犯罪率鄉鎮市區影響因素之研究 (未出版碩士論文)。國立臺北大學，新北市。
- 游柏輝 (2011)。以時空加權迴歸模型探索地域犯罪機制的時空變異性(未出版碩士論文)。國立臺灣大學，臺北市。
- 湯巧君 (2007)。利用犯罪地圖分析臺中市 2005 年搶奪案之時空差異(未出版碩士論文)。國立高雄師範大學，高雄市。
- 程路文 (2017)。基於集成學習的空污與呼吸系統疾病發病預測模型(未出版碩士論文)。元智大學，桃園縣。
- 黃元杰 (2017)。犯罪地圖政策公佈對房價之探討—以臺北市為例(未出版碩士論文)。國立臺灣大學，臺北市。
- 黃月盈 (2013)。臺灣景氣指標季節調整方法之研析。經濟研究，14，1-37。
- 黃俊能 (2008)。從國外經驗談犯罪製圖與犯罪防治。取自網址：
<http://cnhuang.cpu.edu.tw/html/gis/files/971117.pdf>
- 黃信誠 (2000)。空間統計簡介。自然科學簡訊，12(3)，101-104。
- 黃映翎 (2008)。臺北市士林區指標型犯罪空間分 (未出版碩士論文)。東吳大學，臺北市。
- 黃瀾嬪 (2010)。社區犯罪基圖對於警察派出所預防住宅竊盜犯罪之研究—以臺北市中正第一分局為例(未出版碩士論文)。華梵大學，新北市。
- 楊武德 (2006)。住宅竊盜被害特性、影響及反應之研究 (未出版碩士論文)。中央警察大學，桃園縣。
- 溫在弘 (2015)。空間分析方法與應用。臺北市：雙葉書廊。
- 溫在弘、劉擇昌、林民浩 (2010)。犯罪地圖繪製與熱區分析方法及其應用：以 1998—2007 年臺北市住宅竊盜犯罪為例。地理研究，52，43-63。
- 葉柏村、張光宗 (2019)。山崩潛感模型的逐步建構與決策分析—以陳有蘭集水區為例。中華水土保持學報，50(2)，49-62。
- 葉茗宏 (2013)。身心障礙人力資源需求之空間異質性因素探討-以雙北地區為例 (未出版碩士論文)。國立臺北大學，新北市。
- 董倩楠 (2016)。基於 Kulldorff 掃描統計量的聚類方法研究及應用(未出版碩士

- 論文)。中國石油大學，山東省。
- 廖興中、徐明莉 (2017)。臺灣地方政府貪腐現象之時空掃描。公共行政學報，53，1-23。
- 榮峻德、孫志鴻、李萬凱 (2005)。普查資料之空間資料挖掘研究—以臺北市便利商店區位分析為例。台灣地理資訊學刊，2，45-56。
- 臺大新聞 E 論壇 (2017)。「路燈可以降低犯罪率？資料顯示兩者關聯不大」。取自網址：<https://castnet.nctu.edu.tw/ntuereporting/article/12194?issueID=674>
- 臺北市政府地政局 (2012)。「不動產實價登錄 101 年 8 月 1 日上路」。取自網址：https://land.gov.taipei/News_Content.aspx?n=0ABE9F8A3E5B75C2&sms=72544237BBE4C5F6&s=5641CBBD75AF0755
- 劉孟奇、張其祿、盧敬植 (2000)。警力增加能導致竊盜犯罪率降低嗎？公共行政學報，34，1-27。
- 劉盈君 (2008)。大專院校學生死亡交通事故類型與影響因子分析之研究 (未出版碩士論文)。國立交通大學，新竹市。
- 劉擇昌 (2011)。住宅竊盜犯罪熱區空間分析與環境特性之研究--以臺北市大安區為例(未出版博士論文)。中央警察大學，桃園縣。
- 劉擇昌 (2013)。臺灣地區都市化因子對竊盜犯罪影響及其區域變異：全域性與區域性迴歸分析。犯罪與刑事司法研究，21，135-174。
- 劉擇昌、張平吾 (2014)。地理資訊系統與執法。臺北市：三民書局。
- 劉擇昌、鄧煌發、張智雄 (2013)。臺北市大安區住宅竊盜犯罪環境差異之實證研究。犯罪與刑事司法研究，20，61-104。
- 蔡丞庭 (2011)。焦點檢定方法比較(未出版碩士論文)。國立政治大學，臺北市。
- 蔡德輝、楊士隆 (2012)。犯罪學。臺北：五南圖書公司。
- 蔡德輝、楊士隆 (2019)。犯罪學(增訂八版)。臺北：五南圖書公司。
- 鄧莉雅 (2015)。決策樹分析與羅吉斯迴歸於資料探勘的整合運用：以人事資料與民眾健康影響因素之探討為例(未出版碩士論文)。國立臺灣師範大學，臺北市。
- 賴致瑜 (2006)。臺北市住宅竊盜犯罪地圖製作與犯罪區位分析 (未出版碩士論文)。國立臺灣大學，臺北市。
- 謝漢欽、汪大雄、鍾智昕 (2010)。談空間資料倉儲及資料探勘的觀念與技術。林業研究專訊，17(2)，61-68。

鍾孝宇 (2017)。巨量資料與隱私權—個人資料保護機制的再思考(未出版碩士論文)。國立政治大學，臺北市。

鍾志偉 (2010)。以地理資訊系統結合資料探勘技術從事郵局設點分析 (未出版碩士論文)。國立政治大學，臺北市。

外文部分

- Aamodt, G., Samuelsen, S. O., & Skrondal, A. (2006). A simulation study of three methods for detecting disease clusters. *International Journal of Health Geographics*, 5(1), 15. doi:10.1186/1476-072X-5-15
- Akers, R. L., & Sellers, C. S. (2013). *Criminological Theories: Introduction, Evaluation, and Application*: Oxford University Press.
- Alm, S. E. (1997). On the Distributions of Scan Statistics of a Two-Dimensional Poisson Process. *Advances in Applied Probability*, 29(1), 1-18. doi:10.2307/1427857
- Alm, S. E. (1998). Approximation and Simulation of the Distributions of Scan Statistics for Poisson Processes in Higher Dimensions. *Extremes*, 1(1), 111-126. doi:10.1023/a:1009965918058
- Andresen, M. A. (2006). A spatial analysis of crime in Vancouver, British Columbia: a synthesis of social disorganization and routine activity theory. *Canadian Geographer-Geographe Canadien*, 50 (4),487-502.
- Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis*, 27(2), 93-115.
- Awal, M. A., Rabbi, J., & Rana, I. (2017). Using Data Mining Technique to Analyze Crime of Bangladesh. 6(4), 489-494.
- Baker, R. D. (1996). Testing for space-time clusters of unknown size. *Journal of Applied Statistics*, 23(5), 543-554.
- Bellamy, L. C. (1996). Situational crime prevention and convenience store robbery. *Security Journal*, 7(1), 41-52.
- Benefiel, R. (2014, 01/22). Theoretical Integration. *Encyclopedia of Criminology and Criminal Justice*. Retrieved from https://www.researchgate.net/publication/282778316_Theoretical_Integration
- Bernasco, W., & Nieuwbeerta, P. (2005). How Do Residential Burglars Select Target Areas? A New Approach to the Analysis of Criminal Location Choice. *British Journal of Criminology*, 45, 296 - 315 (2005), 45. doi:10.1093/bjc/azh070
- Besag, J., & Newell, J. (1991). The detection of clusters in rare diseases. *Journal of the Royal Statistical Society, Series A*, 154(pt1), 143-155.
- Besag, J., & Newell, J. (1991). The Detection of Clusters in Rare Diseases. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 154(1), 143-155. doi:10.2307/2982708
- Bhunja, G. S., & Shit, P. K. (2019). *Geospatial Analysis of Public Health*. Cham, Switzerland: Springer.

- Braithwaite, J. (1989). *Crime, Shame and Reintegration*. Cambridge: Cambridge University Press.
- Brantingham, P. J., & Brantingham, P. L. (1991). *Environmental Criminology*: Waveland Press.
- Breiman, L. (2001). randomforest. *Machine learning*, 45(1), 5-32.
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and Regression Trees*: Taylor & Francis.
- Burgess, R. E. P. E. W., Park, R. E., Burgess, E. W., Janowitz, M., McKenzie, R. D., & McKenzie, R. D. (1967). *The City*: University of Chicago Press.
- Bursik, R. J., & Webb, J. (1982). Community change and patterns of delinquency. *American Journal of Sociology*, 88(1), 24-42. doi:10.1086/227632
- Capowich, G. E. (2003). The Conditioning Effects Of Neighborhood Ecology On Burglary Victimization. *Criminal Justice and Behavior*, 30(1), 39-61. doi:10.1177/0093854802239162
- Ceccato, V., Haining, R., & Signoretta, P. (2002). Exploring Offence Statistics in Stockholm City Using Spatial Analysis Tools. *Annals of the Association of American Geographers*, 92(1), 29-51. doi:10.1111/1467-8306.00278
- Chawla, S., Shekhar, S., Wu, W. L., & Ozesmi, U. (2000). Extending Data Mining for Spatial Applications: A Case Study in Predicting Nest Locations. In S. Minnesota Univ Minneapolis Minneapolis United (Ed.).
- Chen, J., Roth, R. E., Naito, A. T., Lengerich, E. J., & Maceachren, A. M. (2008). Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of U.S. cervical cancer mortality. *Int J Health Geogr*, 7, 57. doi:10.1186/1476-072X-7-57
- Chen, P., Yuan, H., & Li, D. (2013). Space-time analysis of burglary in Beijing. *Security Journal*, 26(1), 1-15. doi:10.1057/sj.2011.4
- Cheng, T., & Adepeju, M. (2013). *Detecting emerging space-time crime patterns by prospective STSS*. Paper presented at the In Proceedings of the 12th International Conference on GeoComputation.
- Clarke, R. V., & Cornish, D. B. (1985). Modeling Offenders' Decisions: A Framework for Research and Policy. *Crime and Justice*, 6, 147-185.
- Clarke, R. V., & Eck, J. E. (2005). Crime Analysis for Problem Solvers in 60 Small Steps. Retrieved from <https://www.semanticscholar.org/paper/Crime-Analysis-for-Problem-Solvers-in-60-Small-Clarke-Eck/be59dec7e23070bf8857745e737d9f4c90aef5e>
- Clarke, R. V. G., & Eck, J. (2003). *Become a problem-solving crime analyst: in 55 small steps*. London: Jill Dando Institute of Crime Science.
- Cohen, L. E., Felson, M., & Land, K. C. (1980). Property Crime Rates in the United

- States: A Macrodynamic Analysis, 1947-1977; With Ex Ante Forecasts for the Mid-1980s. *American Journal of Sociology*, 86(1), 90-118.
doi:10.1086/227204
- Cohen, L. E., Kluegel, J. R., & Land, K. C. (1981). Social Inequality and Predatory Criminal Victimization: An Exposition and Test of a Formal Theory. *American Sociological Review*, 46(5), 505-524. doi:10.2307/2094935
- Coulston, J., & Riitters, K. (2003). Geographic Analysis of Forest Health Indicators Using Spatial Scan Statistics. *Environmental Management*, 31(6), 764-773.
doi:10.1007/s00267-002-0023-9
- Cressie, N. A. C. (1993). *Statistics for spatial data* (Rev. ed. ed.). New York: John Wiley.
- DeChello, L. M., & Sheehan, T. J. (2007). Spatial analysis of colorectal cancer incidence and proportion of late-stage in Massachusetts residents: 1995-1998. *International Journal of Health Geographics*, 6, 20-20.
doi:10.1186/1476-072X-6-20
- E. Cohen, L., & Felson, M. (1979). *Social Change and Crime Rate Trends: A Routine Activity Approach* (Vol. 44).
- Eck, J. E., Chainey, S., Cameron, J. G., Leitner, M., & Wilson, R. E. (2005). Mapping Crime: Understanding Hot Spots.
- Eckley, D. C., & Curtin, K. M. (2013). Evaluating the spatiotemporal clustering of traffic incidents. *Computers, Environment and Urban Systems*, 37, 70-81.
- Ester, M., Kriegel, H.-P., & Sander, J. (1997, 1997//). *Spatial data mining: A database approach*. Paper presented at the Advances in Spatial Databases, Berlin, Heidelberg.
- Fawcett, T. (2006). Introduction to ROC analysis. *Pattern Recognition Letters*, 27, 861-874. doi:10.1016/j.patrec.2005.10.010
- Felson, M., Clarke, R. V. G., Great Britain. Home Office. Research, D., Policing, S. D., & Unit, R. C. (1998). *Opportunity makes the thief: practical theory for crime prevention*: Home Office, Policing and Reducing Crime Unit, Research, Development and Statistics Directorate.
- Felson, M., & Poulsen, E. (2003). Simple Indicators of Crime by Time of Day. *International Journal of Forecasting*, 19, 595-601.
doi:10.1016/S0169-2070(03)00093-1
- Gao, P., Guo, D., Liao, K., Webb, J. J., & Cutter, S. L. (2013). Early detection of terrorism outbreaks using prospective space-time scan statistics. *The Professional Geographer*, 65(4), 676-691.
- Gilbertson, D. L. (2007). Book Review: Lersch, K. M. (2004). Space, Time, and Crime. Durham, NC: Carolina Academic Press, pp. 282. *Criminal Justice*

- Review*, 32(4), 432-434. doi:10.1177/0734016807310632
- Goldstein, H. (1979). Improving Policing: A Problem-Oriented Approach. *Crime & Delinquency*, 25(2), 236-258. doi:10.1177/001112877902500207
- Grubestic, T. H., & Mack, E. A. (2008). Spatio-Temporal Interaction of Urban Crime. *Journal of Quantitative Criminology*, 24(3), 285-306. doi:10.1007/s10940-008-9047-5
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2013). *Multivariate Data Analysis*: Pearson Education Limited.
- Hambaba, M. L. (1996, 24-26 March 1996). *Intelligent hybrid system for data mining*. Paper presented at the IEEE/IAFE 1996 Conference on Computational Intelligence for Financial Engineering (CIFEr).
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques*: Morgan Kaufmann Publishers Inc.
- Han, J., Zhu, L., Kuldorff, M., Hostovich, S., Stinchcomb, D. G., Tatalovich, Z., . . . Feuer, E. J. (2016). Using Gini coefficient to determining optimal cluster reporting sizes for spatial scan statistics. *International Journal of Health Geographics*, 15(1), 27-27. doi:10.1186/s12942-016-0056-6
- Harcourt, B. E., & Ludwig, J. (2006). Broken Windows: New Evidence from New York City and a Five-City Social Experiment. *The University of Chicago Law Review*, 73(1), 271-320.
- Hino, K., & Amemiya, M. (2019). Spatiotemporal analysis of burglary in multifamily housing in Fukuoka City, Japan. *Cities*, 90, 15-23. doi:https://doi.org/10.1016/j.cities.2019.01.030
- Hirschi, T. (1979). Separate and Unequal Is Better. *Journal of Research in Crime and Delinquency*, 16(1), 34-38. doi:10.1177/002242787901600104
- Hoppe, L., & Gerell, M. (2019). Near-repeat burglary patterns in Malmö: Stability and change over time. *European Journal of Criminology*, 16(1), 3-17. doi:10.1177/1477370817751382
- Hosmer, D. W., Lemeshow, S., & Cook, E. D. (2001). *Applied Logistic Regression, Second Edition: Book and Solutions Manual Set*: Wiley.
- Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical statistics*, 15(3), 651-674.
- Jefferis, E. (1998). A multi-method exploration of crime hot spots: SaTScan results. *National Institute of Justice, Crime Mapping Research Center*.
- Jeffery, C. Ray. (1971). *Crime Prevention Through Environmental Design*. Beverly Hills, CA: Sage Publications.
- Johnson, S. D., Bernasco, W., Bowers, K. J., Elffers, H., Ratcliffe, J., Rengert, G., &

- Townsley, M. (2007). Space–Time Patterns of Risk: A Cross National Assessment of Residential Burglary Victimization. *Journal of Quantitative Criminology*, 23(3), 201-219. doi:10.1007/s10940-007-9025-3
- Johnson, S. D., Bowers, K., & Hirschfield, A. (1997). New Insights Into the Spatial and Temporal Distribution of Repeat Victimization. *The British Journal of Criminology*, 37(2), 224-241.
- Johnson, S. D., & Bowers, K. J. (2010). Permeability and Burglary Risk: Are Cul-de-Sacs Safer? *Journal of Quantitative Criminology*, 26(1), 89-111. doi:10.1007/s10940-009-9084-8
- Johnson, S. D., Bowers, K. J., Birks, D. J., & Pease, K. (2009). Predictive Mapping of Crime by ProMap: Accuracy, Units of Analysis, and the Environmental Backcloth. In D. Weisburd, W. Bernasco, & G. J. N. Bruinsma (Eds.), *Putting Crime in its Place: Units of Analysis in Geographic Criminology* (pp. 171-198). New York, NY: Springer New York.
- Kennedy, L. W., Caplan, J. M., & Piza, E. (2011). Risk Clusters, Hotspots, and Spatial Intelligence: Risk Terrain Modeling as an Algorithm for Police Resource Allocation Strategies. *Journal of Quantitative Criminology*, 27(3), 339-362. doi:10.1007/s10940-010-9126-2
- Klauber, M. R. (1971). Two-sample randomization tests for space-time clustering. *Biometrics*, 129-142.
- Knox, G. (1964). Epidemiology of Childhood Leukaemia in Northumberland and Durham. *British journal of preventive & social medicine*, 18(1), 17-24.
- Kornhauser, R. R. (1984). *Social Sources of Delinquency: An Appraisal of Analytic Models*: University of Chicago Press.
- Kubrin, C. (2009). Social Disorganization Theory: Then, Now, and in the Future. In (pp. 225-236).
- Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics - Theory and Methods*, 26(6), 1481-1496. doi:10.1080/03610929708831995
- Kulldorff, M. (1998). *Statistical Methods for Spatial Epidemiology: Tests for Randomness*.
- Kulldorff, M. (1999). Spatial Scan Statistics: Models, Calculations, and Applications. In J. Glaz & N. Balakrishnan (Eds.), *Scan Statistics and Applications* (pp. 303-322). Boston, MA: Birkhäuser Boston.
- Kulldorff, M. (2001). Prospective Time Periodic Geographical Disease Surveillance Using A Scan Statistic. *Journal of the Royal Statistical Society Series A*, 164, 61-72. doi:10.1111/1467-985X.00186
- Kulldorff, M. (2006). Tests of Spatial Randomness Adjusted for an Inhomogeneity: A General Framework. *Journal of the American Statistical Association*, 101,

- 1289-1305. doi:10.2307/27590802
- Kulldorff, M. (2018). SaTScan User Guide TM for version 9.6. Retrieved from <http://www.satscan.org/>
- Kulldorff, M., Athas, W. F., Feurer, E. J., Miller, B. A., & Key, C. R. (1998). Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico. *Am J Public Health, 88*(9), 1377-1380.
- Kulldorff, M., Feuer, E. J., Miller, B. A., & Freedman, L. S. (1997). Breast Cancer Clusters in the Northeast United States: A Geographic Analysis. *American Journal of Epidemiology, 146*(2), 161-169.
- Kulldorff, M., Heffernan, R., Hartman, J., Assuncao, R., & Mostashari, F. (2005). A space-time permutation scan statistic for disease outbreak detection. *PLoS Med, 2*(3), e59. doi:10.1371/journal.pmed.0020059
- Kulldorff, M., Huang, L., Pickle, L., & Duczmal, L. (2006). An elliptic spatial scan statistic. *Stat Med, 25*(22), 3929-3943. doi:10.1002/sim.2490
- Kuo, S.-Y. (2014). Opportunity, Choice, and Burglary Victimization in Taiwan. *International journal of offender therapy and comparative criminology, 59*. doi:10.1177/0306624X13520439
- Leitner, M., & Helbich, M. (2011). The Impact of Hurricanes on Crime: A Spatio-Temporal Analysis in the City of Houston, Texas. *Cartography and Geographic Information Science, 38*(2), 213-221. doi:10.1559/15230406382213
- Liggett, R., Loukaitou-Sideris, A., & Iseki, H. (2003). Journeys to crime: Assessing the effects of a light rail line on crime in the neighborhoods.
- Lloyd, S., Farrell, G., & Pease, K. (1994). *Preventing repeated domestic violence: A demonstration project on Merseyside*.
- Lloyd, S., & Roberts, C. (1973). A test for space clustering and its application to congenital limb defects in Cardiff. *British journal of preventive & social medicine, 27*(3), 188.
- Loader, C. R. (1991). Large-Deviation Approximations to the Distribution of Scan Statistics. *Advances in Applied Probability, 23*(4), 751-771. doi:10.2307/1427674
- Lober, W., Karras, B., Wagner, M., Overhage, J. M., Davidson, A., Fraser, H., . . . Tsui, F. (2002). Roundtable on Bioterrorism Detection: Information System-based Surveillance. *Journal of the American Medical Informatics Association : JAMIA, 9*, 105-115. doi:10.1197/jamia.M1052
- Lohr, S. (2012). The Age of Big Data. *The New York Times*. Retrieved from <https://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html>

- Mahfoud, M., Bhulai, S., & Mei, R. D. v. d. (2017). *Spatio-temporal modeling for residential burglary*. Paper presented at the 6th International Conference on Data Analytics, Barcelona (Spain).
- Malczewski, J., & Poetz, A. (2005). *Residential Burglaries and Neighborhood Socioeconomic Context in London, Ontario: Global and Local Regression Analysis** (Vol. 57).
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer research*, 27(2 Part 1), 209-220.
- Martin, D. (2002). Spatial Patterns in Residential Burglary: Assessing the Effect of Neighborhood Social Capital. *Journal of Contemporary Criminal Justice*, 18(2), 132-146. doi:10.1177/1043986202018002002
- Matsueda, R. L. (2017). Social Structure, Culture, and Crime: Assessing Kornhauser's Challenge to Criminology. Retrieved from <https://www.semanticscholar.org/paper/Social-Structure%2C-Culture%2C-and-Crime%3A-Assessing-to-Matsueda/d9505d30dbc23847b4b1da588a010f796c58dc42>
- Mayhew, P. (Ed.) (1990). *Opportunity and Vehicle Crime*. Aldershot UK: Gower.
- Meier, R. F., & Miethe, T. D. (1993). Understanding Theories of Criminal Victimization. *Crime and Justice*, 17, 459-499.
- Messner, S. F., Krohn, M. D., & Liska, A. E. (1989). *Theoretical integration in the study of deviance and crime : problems and prospects*. Albany, N.Y.: State University of New York Press.
- Midi, H., Sarkar, S., & Rana, S. (2013). Collinearity diagnostics of binary logistic regression model. *Journal of Interdisciplinary Mathematics*, 13, 253-267. doi:10.1080/09720502.2010.10700699
- Miethe, T. D., & McDowall, D. (1993). Contextual Effects in Models of Criminal Victimization*. *Social Forces*, 71(3), 741-759. doi:10.1093/sf/71.3.741
- Miethe, T. D., & Meier, R. F. (1994). *Crime and its Social Context: Toward an Integrated Theory of Offenders, Victims, and Situations*. Albany, NY: State University of New York.
- Miller, H. J. (2004). Tobler's First Law and Spatial Analysis. *Annals of the Association of American Geographers*, 94(2), 284-289.
- Moraga, P., & Kulldorff, M. (2013). Detection of spatial variations in temporal trends with a quadratic function. *Statistical Methods in Medical Research*, 25(4), 1422-1437. doi:10.1177/0962280213485312
- Nakaya, T., & Yano, K. (2010). Visualising crime clusters in a space-time cube: An exploratory data-analysis approach using space-time kernel density estimation and scan statistics. *Transactions in GIS*, 14(3), 223-239.

- Naus, J. I. (1965). Clustering of Random Points in Two Dimensions. *Biometrika*, 52(1/2), 263-267. doi:10.2307/2333829
- Newman, O. (1995). Defensible Space: A New Physical Planning Tool for Urban Revitalization. *Journal of the American Planning Association*, 61(2), 149-155. doi:10.1080/01944369508975629
- Nissan, E. (2012). An Overview of Data Mining for Combating Crime. *Applied Artificial Intelligence*, 26(8), 760-786. doi:10.1080/08839514.2012.713309
- Odoi, A., Martin, S. W., Michel, P., Middleton, D., Holt, J., & Wilson, J. (2004). Investigation of clusters of giardiasis using GIS and a spatial scan statistic. *International Journal of Health Geographics*, 3(1), 11. doi:10.1186/1476-072x-3-11
- Openshaw, S., Charlton, M., Wymer, C., & Craft, A. (1987). A Mark 1 Geographical Analysis Machine for the automated analysis of point data sets. *International Journal of Geographical Information Systems*, 1(4), 335-358. doi:10.1080/02693798708927821
- Popper, K. (1959). *The Logic of Scientific Discovery*. In. Retrieved from <http://strangebeautiful.com/other-texts/popper-logic-scientific-discovery.pdf>
- Poulsen, E., & Kennedy, L. W. (2004). Using dasymetric mapping for spatially aggregated crime data. *Journal of Quantitative Criminology*, 20(3), 243-262.
- Prates, M. O., Kulldorff, M., & Assunção, R. M. (2014). Relative risk estimates from spatial and space-time scan statistics: are they biased? *Statistics in medicine*, 33(15), 2634-2644. doi:10.1002/sim.6143
- Ramasamy, J., & Janarthanan, K. (2019). A Study On Spatial Variations In Temporal Trends Of Dengue Incidences In Tamil Nadu, India. *International Journal of Scientific & Technology Research*, 8, 788-792.
- Ratcliffe, J. H. (2004). Geocoding crime and a first estimate of a minimum acceptable hit rate. *International Journal of Geographical Information Science*, 18(1), 61-72. doi:10.1080/13658810310001596076
- Ratcliffe, J. H. (2005). Detecting Spatial Movement of Intra-Region Crime Patterns Over Time. *Journal of Quantitative Criminology*, 21(1), 103-123.
- Rice, K. J., & Smith, W. R. (2002). Socioecological Models of Automotive Theft: Integrating Routine Activity and Social Disorganization Approaches. *Journal of Research in Crime and Delinquency*, 39(3), 304-336. doi:10.1177/002242780203900303
- Rountree, P., Land, K., & Miethe, T. (1994). Macro-micro integration in the study of victimization: A hierarchical logistic model analysis across Seattle neighborhoods. *Criminology*, 32(3), 387. doi:10.1111/j.1745-9125.1994.tb01159.x

- Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *SCIENCE*, 277(5328), 918-924.
- Schapire, R. E. (1990). The strength of weak learnability. *Machine learning*, 5(2), 197-227. doi:10.1007/bf00116037
- Schapire, R. E., & Singer, Y. (1999). Improved Boosting Algorithms Using Confidence-rated Predictions. *Machine learning*, 37(3), 297-336. doi:10.1023/a:1007614523901
- Shaw, C. R., & McKay, H. D. (1942). *Juvenile delinquency and urban areas*. Chicago, IL, US: University of Chicago Press.
- Siemiatycki, J. (1978). Mantel's space-time clustering statistic: computing higher moments and a comparison of various data transforms. *Journal of statistical Computation and Simulation*, 7(1), 13-31.
- Smith, W., Frazee, S., & Davison, E. (2006). Furthering the Integration of Routine Activity and Social Disorganization Theories: Small Units of Analysis and the Study of Street Robbery as a Diffusion Process. *Criminology*, 38, 489-524. doi:10.1111/j.1745-9125.2000.tb00897.x
- Snook, B., Dhimi, M. K., & Kavanagh, J. M. (2011). Simply Criminal: Predicting Burglars' Occupancy Decisions with a Simple Heuristic. *Law and Human Behavior*, 35(4), 316-326. doi:10.1007/s10979-010-9238-0
- Stibbs, A. (1994). Deconstructing Burglars: Formal analysis as a Pedagogic Jemmy. *Children's Literature in Education*, 25(4), 213-224. doi:10.1007/bf02355302
- Sugumaran, R., Larson, S. R., & Degroote, J. P. (2009). Spatio-temporal cluster analysis of county-based human West Nile virus incidence in the continental United States. *Int J Health Geogr*, 8, 43. doi:10.1186/1476-072X-8-43
- Swartz, C. (2000). *Analyzing Crime Patterns: Frontiers of Practice*. In V. Goldsmith, P. G. McGuire, J. H. Mollenkopf, & T. A. Ross (Eds.). Retrieved from <http://sk.sagepub.com/books/analyzing-crime-patterns> doi:10.4135/9781452220369
- Tango, T. (2010). *Statistical Methods for Disease Clustering*. New York Springer.
- Tango, T., & Takahashi, K. (2012). A flexible spatial scan statistic with a restricted likelihood ratio for detecting disease clusters. *Stat Med*, 31(30), 4207-4218. doi:10.1002/sim.5478
- Taylor, R. B., & Harrell, A. (1996). *Physical environment and crime*: US Department of Justice, Office of Justice Programs, National Institute of ...
- Tonini, M., Tuia, D., & Ratle, F. (2009). Detection of clusters using space - time scan statistics. *International Journal of Wildland Fire*, 18(7). doi:10.1071/wf07167
- Townsley, M. (2008). Visualising space time patterns in crime: the hotspot plot. *Crime*

- patterns and analysis*, 1(1), 61-74.
- Townsley, M., Birks, D., Ruiter, S., Bernasco, W., & White, G. (2016). Target Selection Models with Preference Variation Between Offenders. *Journal of Quantitative Criminology*, 32(2), 283-304. doi:10.1007/s10940-015-9264-7
- Townsley, M., Homel, R., & Chaseling, J. (2000). Repeat Burglary Victimization: Spatial and Temporal Patterns. *Australian & New Zealand Journal of Criminology*, 33(1), 37-63. doi:10.1177/000486580003300104
- Triplett, R. A., Sun, I. Y., & Gainey, R. R. (2005). Social Disorganization and the Ability and Willingness to Enact Control: A Preliminary Test *Western Criminology Review*, 6(1), 89-103
- Turnbull, B., J. Iwano, E., S. Burnett, W., L. Howe, H., & C. Clark, L. (1990). *Monitoring for Clusters of Disease: Application to Leukemia Incidence in Upstate New York* (Vol. 132).
- Vesey, B. M., & Messner, S. F. (1999). Further Testing of Social Disorganization Theory: An Elaboration of Sampson and Groves's "Community Structure and Crime". *Journal of Research in Crime and Delinquency*, 36(2), 156-174. doi:10.1177/0022427899036002002
- Weisburd, D., Bushway, S., Lum, C., & Yang, S.-M. (2004). Trajectories of Crime at Places: A Longitudinal Study of Street Segments in the City of Seattle. *Criminology*, 42, 283-322. doi:10.1111/j.1745-9125.2004.tb00521.x
- Weisburd, D., Groff, E., & Yang, S.-M. (2013). The Criminology of Place: Street Segments and Our Understanding of the Crime Problem. *The Criminology of Place: Street Segments and Our Understanding of the Crime Problem*, 1-288. doi:10.1093/acprof:oso/9780195369083.001.0001
- Williams, G. J. (2011). *Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery*. New York Dordrecht Heidelberg London: Springer.
- Wilson, J. Q., & Kelling, G. L. (1982). Broken Windows: The Police and Neighbourhood Safety. *The Atlantic Monthly*(March).
- Wortley, R., & Ann Mazerolle, L. (2009). *Environmental Criminology and Crime Analysis: Situating the Theory, Analytic Approach and Application* (Vol. 11).
- Xu, Y., Fu, C., Kennedy, E., Jiang, S., & Owusu-Agyemang, S. (2018). The impact of street lights on spatial-temporal patterns of crime in Detroit, Michigan. *Cities*, 79, 45-52. doi:10.1016/j.cities.2018.02.021
- Zhang, L., Messner, S. F., & Liu, J. (2007). A multilevel analysis of the risk of household burglary in the city of Tianjin, China. *The British Journal of Criminology*, 47(6), 918-937.