

逢甲大學學生報告 ePaper

新北地區時價登錄交易

New Taipei City Real-Price Registration Transaction

作者：張瑋倫、陳郁妍、林禹彤、吳亭萱、末吉珠理

系級：統計三乙

學號：D1223101、D1242597、D1242608、D1242584、D1266866

開課老師：劉峰旗

課程名稱：迴歸分析

開課系所：統計系

開課學年：114 學年度 第 1 學期

摘要

隨著台灣房價持續上揚，特別是在新北市地區，住宅價格受到多項因素影響，使得民眾在購屋時面臨高度不確定性，尤其對於屋齡、住宅條件與區位等因素與單價之間的關聯性，常缺乏明確判斷依據。因此，本研究之目的在於透過實證分析，探討影響新北市住宅單價之主要因素，以協助購屋者進行更理性的決策，並加深對不動產市場特性的理解。

本研究以新北市於 2025 年 10 月 17 日至 11 月 17 日期間之實價登錄資料作為研究樣本，並以每平方公尺單價作為反應變數 (Y)。解釋變數 (X) 則包含屋齡、是否設有電梯、房間數、建物所在之鄉鎮市區以及建物總樓層數等住宅特性。研究過程中，先利用 Excel 進行資料整理與初步處理，再透過 SAS Enterprise Guide 進行統計分析與模型建構，以評估各變數對住宅單價之影響情形。

研究結果顯示，區位對新北市住宅單價具有穩定且顯著之影響，為影響房價的重要因素之一；屋齡與房間數則與單價呈現負向關係。相較之下，是否設有電梯及建物總樓層數對房價之影響較不明顯。整體而言，本研究結果可作為民眾購屋時評估住宅條件之參考，亦能提供後續不動產相關研究之基礎。

關鍵字：多元迴歸分析、殘差分析、離群值診斷、住宅單價

Abstract

As housing prices in Taiwan continue to rise, particularly in New Taipei City. As housing prices in Taiwan continue to rise, particularly in the New Taipei City area, residential property values are influenced by multiple factors, creating a high level of uncertainty for homebuyers. In particular, the relationships between housing age, residential characteristics, location, and unit price are often unclear to the general public. Therefore, the purpose of this study is to conduct an empirical analysis to identify the key factors affecting residential unit prices in New Taipei City, with the aim of assisting homebuyers in making more informed decisions and enhancing the understanding of local real estate market characteristics.

This study utilizes transaction price registration data from New Taipei City covering the period from October 17 to November 17, 2025. The unit price per square meter is selected as the dependent variable (Y), while the independent variables (X) include housing age, the presence of an elevator, number of rooms, the administrative district in which the property is located, and the total number of floors in the building. Data were first organized and preprocessed using Microsoft Excel, followed by statistical analysis and model construction conducted through SAS Enterprise Guide to evaluate the effects of various housing characteristics on residential unit prices.

The empirical results indicate that location has a stable and statistically significant impact on residential unit prices in New Taipei City and represents one of the most influential determinants of housing prices. Housing age and number of rooms exhibit negative relationships with unit prices. In contrast, the presence of an elevator and the total number of building floors do not have significant effects on housing prices. Overall, the findings of this study provide useful references for homebuyers when evaluating housing conditions and offer a foundation for future research in the field of real estate studies.

Keyword : Multiple regression analysis, Residuals analysis, Outliers detection, House unit price.

目次

| | |
|--------------------------|----|
| 第一章、前言 | 4 |
| 第一節、研究背景與動機 | 4 |
| 第二節、研究問題 | 4 |
| 第二章、文獻探討 | 6 |
| 第三章、資料分析 | 7 |
| 第一節、資料來源 | 7 |
| 第二節、變數說明 | 7 |
| 第三節、單一變數分析 | 8 |
| 一、有無電梯對住宅單價之影響 | 8 |
| 二、房間數對住宅單價之影響 | 10 |
| 三、總樓層對住宅單價之影響 | 11 |
| 四、鄉鎮市區對住宅單價之影響 | 13 |
| 五、屋齡對住宅單價之影響 | 16 |
| 第四節、多元迴歸分析(未刪除離群值) | 17 |
| 一、模型整體檢定與迴歸係數 | 17 |
| 二、殘差分析 | 19 |
| 三、影響值與離群值診斷 | 21 |
| 第五節、離群值檢測與刪除 | 23 |
| 第六節、刪除離群值後之多元迴歸分析 | 25 |
| 一、模型整體檢定與迴歸係數 | 25 |
| 二、殘差分析 | 26 |
| 三、影響值與離群值診斷 | 28 |
| 第七節、模型選擇與變數篩選 | 31 |
| 一、前進選擇法 | 31 |
| 二、向後剔除法 | 32 |
| 三、逐步選擇法 | 33 |
| 四、最大 R 平方改善法 | 34 |
| 五、最小 R 平方改善法 | 35 |
| 六、R 平方選取法 | 36 |
| 七、調整 R 平方選取法 | 37 |
| 八、Mallow's Cp 選取法 | 39 |
| 九、模型選擇與最終模型決定 | 40 |
| 第四章、結論 | 41 |
| 參考文獻 | 42 |

第一章 前言

第一節、研究背景與動機

近年來，台灣整體房價持續攀升，尤其新北市作為人口密集、交通便利的都會區，住宅價格呈現明顯上升趨勢。民眾在購屋決策過程中，除了考量房屋坪數與總價，亦會關注住宅特性如屋齡、房間數、是否設有電梯以及地理位置。然而，現有研究多集中於單一因素或特定區域，缺乏同時考量多項變數的系統性分析，因此對於影響住宅單價的主要因素仍缺乏清楚的理解。

鑑於此，本研究旨在透過實證分析探討新北市住宅單價的影響因素，並評估各住宅特性對價格之影響程度。研究結果可協助購屋者進行更理性決策，並增進對當地不動產市場結構與動態的理解。本研究以 2025 年 10 月 17 日至 11 月 17 日的新北市實價登錄資料為分析對象，將每平方公尺單價作為依變數，並將屋齡、房間數、電梯設置、建物總樓層及所在鄉鎮市區納入主要解釋變數，建立完整的實證模型，以系統性評估各因素對住宅價格之影響。

第二節、研究問題

本研究主要探討影響新北市住宅單價的關鍵因素，並針對五個核心研究問題進行分析。首先，探討屋齡是否會對住宅單價產生顯著影響，以及其影響方向與程度。其次，評估住宅是否設有電梯對房價的影響，檢視其是否為顯著因素。再者，研究房間數是否在坪數相同的情況下對單價造成變化。此外，考察建物的總樓層數是否與住宅價格呈現顯著關聯。最後，分析建物所在的鄉鎮市區對房價之影響，以了解地理位置在房價形成中的作用。透過這些研究問題，本研究將全面呈現各住宅特性對價格的影響情形。

在釐清研究問題之後，本文後續章節將依序說明研究相關文獻、資料來源與

分析方法，並透過實證分析探討各住宅特性對新北市住宅單價之影響。第二章為文獻探討，整理過去相關研究成果，說明住宅價格受建物特性與地理位置等因素影響的研究發現，並據此建立本研究之分析基礎。第三章為資料分析，首先介紹研究資料來源與變數定義，接著進行單一變數分析、多元迴歸分析以及離群值檢測，並比較刪除離群值前後模型之結果，最後透過多種變數選擇方法建立最適模型。第四章則彙整實證分析結果，提出研究結論，說明影響新北市住宅單價之主要因素及其市場意涵。



第二章 文獻探討

過去研究顯示，住宅價格受多種因素影響，其中房屋物理特性與地理位置為主要決定因素。Lin (2014) 以台北市與新北市的實價登錄資料分析住宅價格，發現建築完成年月、房間數及建物樓層為影響住宅單價的重要因素；此結果與本研究選擇的解釋變數高度一致，驗證了研究設計的合理性與有效性。另一方面，Huang (2016) 探討大眾捷運系統 (MRT) 對新北市住宅區房價的影響，指出行政區與地理位置對房價具有顯著作用，顯示交通可及性與區域差異對房價具有重要影響。

此外，Lee (2014) 利用台灣實價登錄資料進行空間分析與多層次模型研究，發現房屋本身特徵與區域因素同時影響住宅價格，且房價在不同地區之間存在空間差異。Lin (2013) 則從鄰里效應的角度分析台北市房價，指出房價不僅受到房屋本身條件影響，也會受到周邊環境與區域特性的影響，顯示地理位置與社區環境在房價形成過程中具有重要角色。另一方面，Shih (2024) 透過迴歸分析探討台灣整體房價的影響因素，指出總體經濟變數與住宅市場因素皆會對房價產生影響，進一步說明房價形成具有多面向特性。

綜合上述研究可知，房屋物理特性 (如房間數、屋齡與樓層) 與地理區位 (如行政區與周邊環境) 皆為影響住宅價格的重要因素。然而，現有研究多著重於特定因素或較大尺度的市場分析，較少同時針對多項住宅特性進行整合分析。因此，本研究以新北市 2025 年 10 月至 11 月的實價登錄資料為分析對象，納入屋齡、房間數、電梯設置、建物樓層及行政區等多項變數，建立多元迴歸模型，以全面探討各住宅特性對每平方公尺單價的影響，並補充既有研究在多變量分析上的不足。

第三章 資料分析

第一節、資料來源

本研究主要採用臺灣內政部提供的「不動產實價登錄資訊—買賣案件」作為研究資料來源（內政部，2025）。研究範圍限定於新北市住宅交易，資料期間自2025年10月17日至11月17日，總樣本數約200筆。資料中包含住宅單價、屋齡、房間數、建物總樓層、是否設有電梯及所在鄉鎮市區等變數，可作為本研究多元迴歸分析之依據。透過該資料，本研究得以系統性檢視住宅特性與單價之間的關聯，並提供可靠的實證分析基礎。

第二節、變數說明

本研究將住宅單價每平方公尺價格作為反應變數（Y），以量化住宅交易價格。其計算方式為總價除以建物面積所得之單價，屬於連續型變數。

主要解釋變數（X）如下：

1. 是否有電梯：以二元變數表示，1表示有電梯，0表示無電梯，用以檢驗便利性對價格之影響，為類別型變數。
2. 屋齡：建物完工至交易之年份差，衡量房屋老舊程度，為連續型變數。
3. 鄉鎮市區：住宅所在的行政區域，為類別型變數，用以觀察地理位置對單價的影響。
4. 房數：住宅內房間數量，反映住宅大小與空間配置，屬離散型變數。
5. 總樓層：房屋所在的樓層，為連續型變數，用於衡量建物規模及特性。

上述變數設計涵蓋住宅物理特性與區位因素，便於建構多元迴歸模型，以全面分析各因素對新北市住宅單價的影響。

第三節、單一變數分析

一、有無電梯對住宅單價之影響

為探討是否有電梯對住宅單價的影響，本研究先將「電梯」變數由字串類型轉換為 Dummy 變數 (1 = 有電梯，0 = 無電梯)，以便進行線性迴歸分析。簡單線性迴歸結果顯示，電梯變數的迴歸係數估計為 17,180 元，表示設有電梯的住宅單價平均每平方公尺比無電梯住宅高 17,180 元。對應的 p 值為 0.0260 (<0.05)，顯示該變數對單價具有統計顯著影響。然而，模型的決定係數 R^2 僅為 0.0248，即僅能解釋房價變異的 2.48%，顯示單一變數對房價的解釋力有限，暗示房價受其他因素影響甚大。

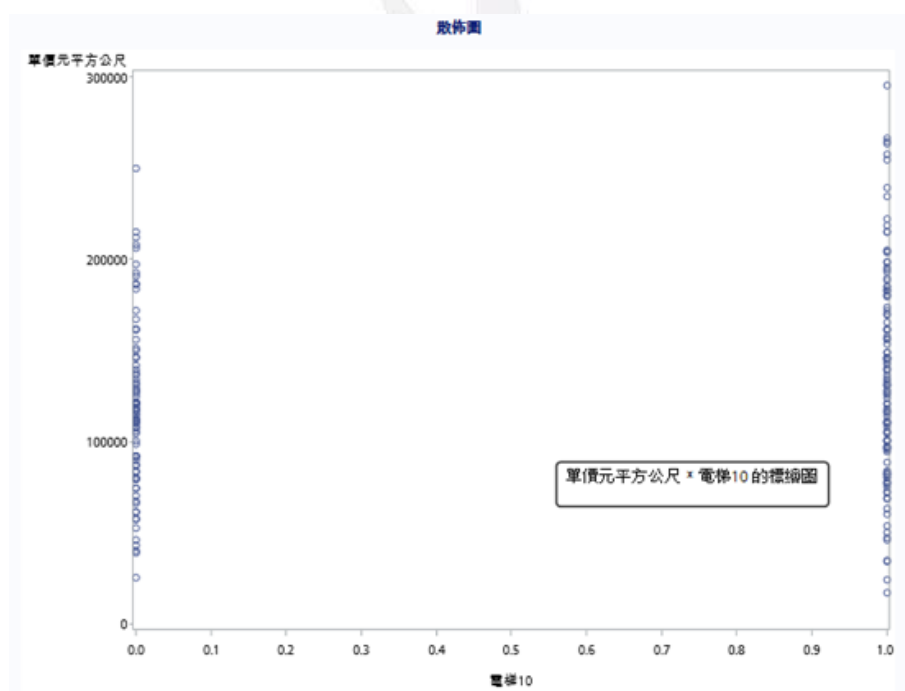


圖 1：電梯變數與每平方公尺單價的散布圖

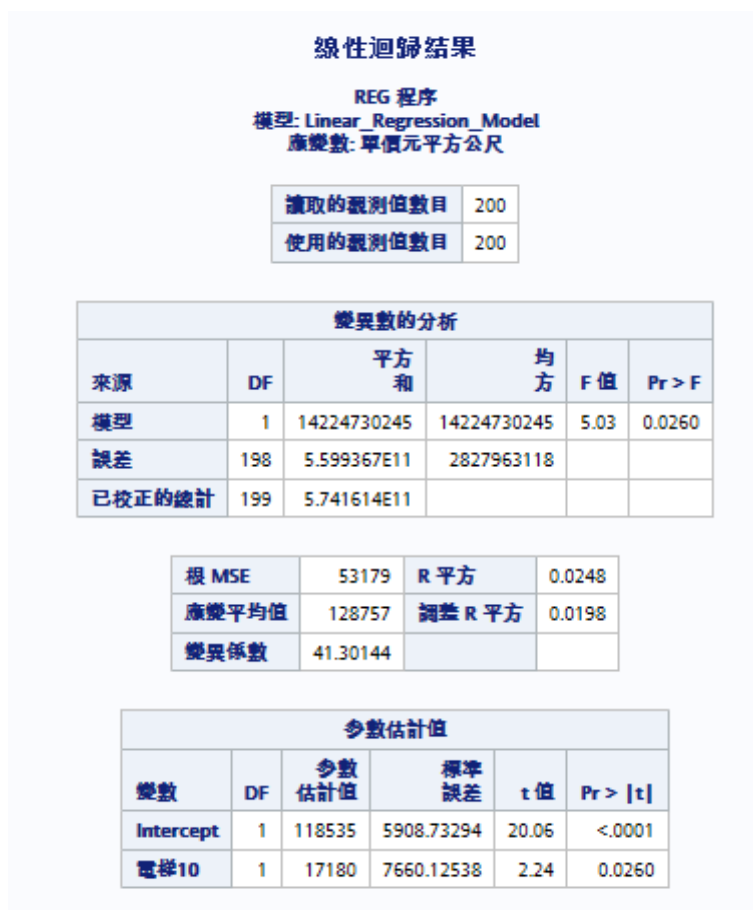


圖 2：ANOVA 與迴歸參數截圖



圖 3：Dummy 變數轉換後的數據截圖

二、 房間數對住宅單價之影響

本研究以房間數作為單一解釋變數，分析其對住宅單價的影響。簡單線性迴歸結果顯示，房間數變數的迴歸係數估計為 -9,730 元，表示房間數每增加一間，住宅單價平均每平方公尺下降 9,730 元。對應的 p 值為 0.0034 (<0.05)，顯示房間數對單價具有統計顯著影響。然而，模型的決定係數 R^2 僅為 0.0424，即僅能解釋房價變異的 4.24%，顯示單一變數對房價解釋力有限，暗示房價還受到坪數、地段、屋齡等其他因素的影響，後續需納入多變量模型以提升分析準確性。

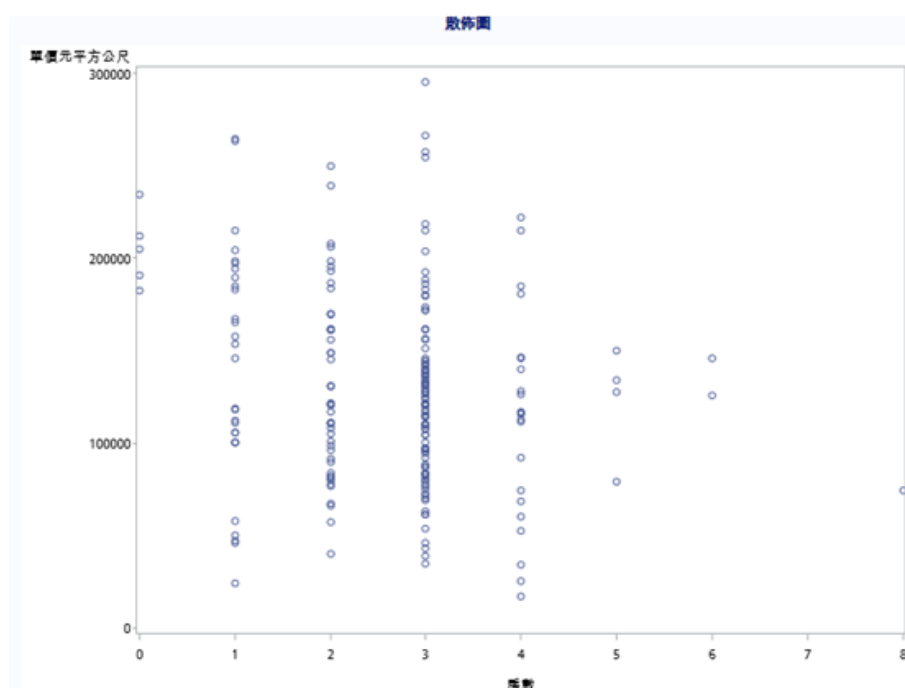


圖 4：房間數與每平方公尺單價的散佈圖

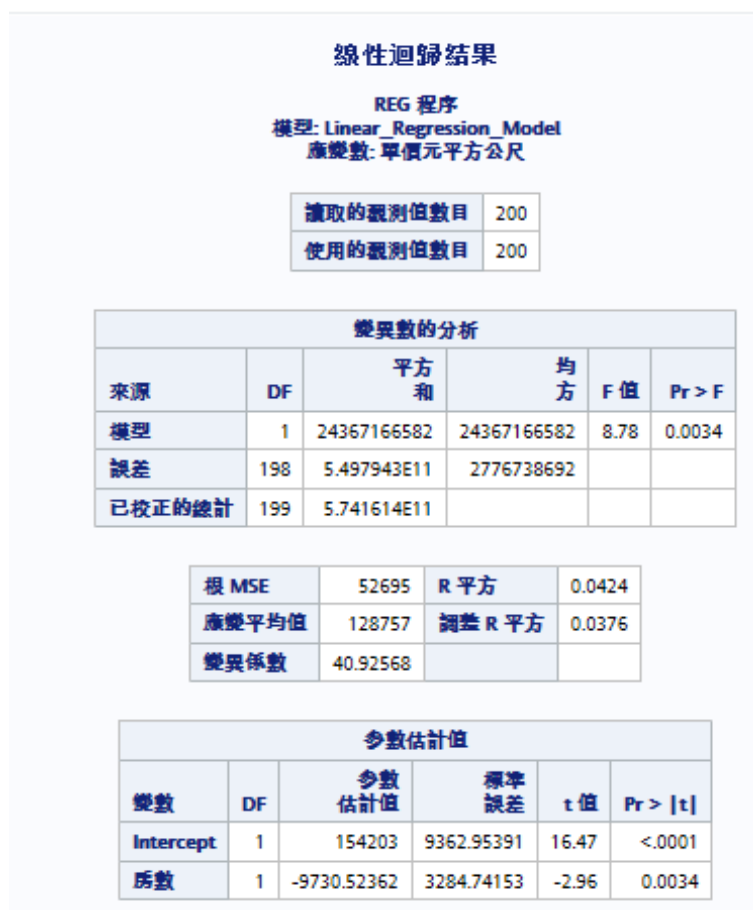


圖 5：ANOVA 與迴歸參數截圖

三、 總樓層數對住宅單價之影響

本研究以建物總樓層作為單一解釋變數，分析其對住宅單價的影響。簡單線性迴歸結果顯示，總樓層變數對住宅單價的影響並不顯著，對應的 p 值為 0.0725 (>0.05)，無法拒絕虛無假設，表示總樓層單獨對單價沒有統計顯著影響。模型的決定係數 R^2 為 0.0162，即僅能解釋房價變異的 1.62%，與前述單一變數分析結果一致，顯示單靠單一變數難以充分解釋房價變動，住宅價格仍受到其他因素如屋齡、房間數及地段等影響，後續需納入多變量迴歸模型進行分析。

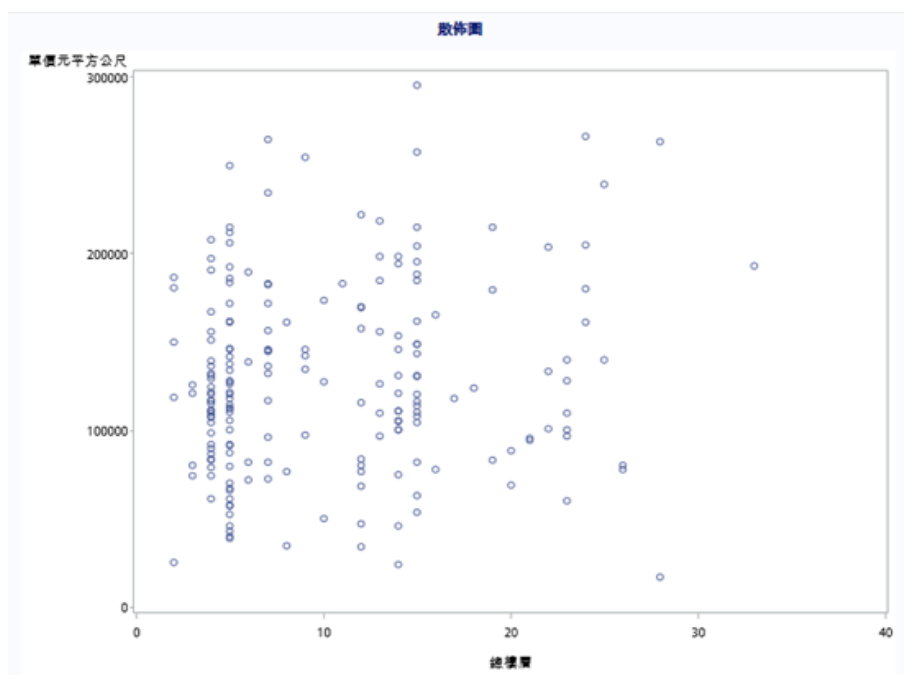


圖 6：總樓層數與每平方公尺單價的散布圖

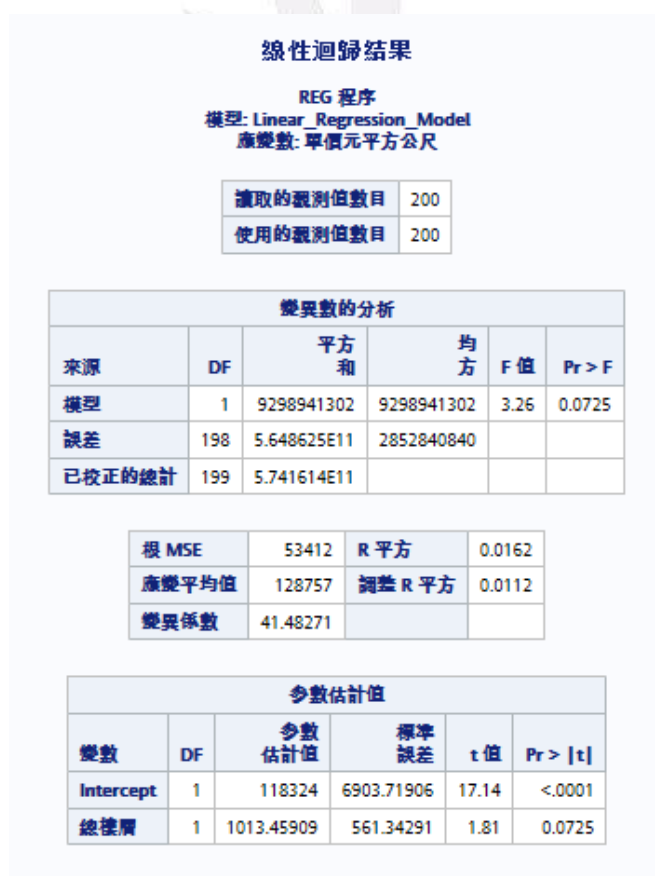


圖 7：ANOVA 與迴歸參數截圖

四、 鄉鎮市區對住宅單價之影響

本研究以鄉鎮市區作為單一解釋變數，分析其對住宅單價的影響。由於鄉鎮市區為字串類型，本研究先將新北市行政區分為三種類型：

- 蛋黃區：板橋區、永和區、中和區、新店區、三重區、新莊區
- 蛋白區：土城區、蘆洲區、汐止區、淡水區、林口區、樹林區、鶯歌區
- 蛋殼區：三峽區、五股區、泰山區、深坑區、石碇區、坪林區、三芝區、石門區、金山區、萬里區、瑞芳區、平溪區、雙溪區、貢寮區、烏來區

將三個區轉換為虛擬變數 (Dummy Variables) 後，定義：

- D1_蛋黃區 1：蛋黃區設為 1，其他兩區為 0
- D2_蛋白區 1：蛋白區設為 1，其他兩區為 0
- 蛋殼區作為基準組 (D1_蛋黃區 1=0, D2_蛋白區 1=0)

迴歸分析結果顯示，鄉鎮市區對住宅單價具有顯著影響，p 值為 0.0001 (<0.05)。模型的決定係數 R^2 為 0.1138，即約可解釋房價變異的 11.38%，相較於前述單一變數分析，解釋力較高，但仍不足以完全解釋房價的變動，顯示房價受其他因素如屋齡、房間數及電梯等影響。

新北地區時價登錄交易

開始頁面 處理流程 匯入資料 (2345...) 查詢產生器 查詢產生器 1 查詢產生器 2 x 線性迴歸

執行 修改任務 共用 屬性

程式碼 日誌 輸出資料 (1)

WORK_QUERY_FOR_0001 因為變置而鎖定資料，值可能已過期，或在重新整理之前無法使用。

| | 單價元平方公... | 單位移轉時面... | 單位時價元 | 主建物面積 | 附屬建物面積 | 陽台面積 | 電梯 | 電梯10 | D1_蛋黃區1 | D2_蛋白區1 |
|----|-----------|-----------|---------|--------|--------|-------|----|------|---------|---------|
| 1 | 160910 | 0 | 0 | 73.99 | 0 | 4.64 | 有 | 1 | 1 | 0 |
| 2 | 139313 | 0 | 0 | 83.84 | 0 | 0 | 無 | 0 | 1 | 0 |
| 3 | 264026 | 0 | 0 | 147 | 0 | 0 | 有 | 1 | 1 | 0 |
| 4 | 234328 | 0 | 0 | 30 | 0 | 0 | 有 | 1 | 1 | 0 |
| 5 | 156452 | 0 | 0 | 75.09 | 0 | 8.38 | 有 | 1 | 0 | 1 |
| 6 | 131279 | 33.27 | 1800000 | 41.79 | 0 | 3.49 | 有 | 1 | 1 | 0 |
| 7 | 211889 | 0 | 0 | 47.79 | 4.36 | 0 | 無 | 0 | 1 | 0 |
| 8 | 165510 | 0 | 0 | 36.89 | 0 | 5.22 | 有 | 1 | 0 | 1 |
| 9 | 167277 | 0 | 0 | 76.52 | 0 | 0 | 無 | 0 | 1 | 0 |
| 10 | 148880 | 35.18 | 1800000 | 42.65 | 0 | 4.98 | 有 | 1 | 0 | 1 |
| 11 | 110564 | 0 | 0 | 104.79 | 0 | 10.98 | 無 | 0 | 0 | 1 |
| 12 | 92124 | 0 | 0 | 57.64 | 0 | 0 | 無 | 0 | 0 | 1 |
| 13 | 132223 | 0 | 0 | 19.21 | 0 | 0 | 無 | 0 | 0 | 1 |
| 14 | 110000 | 0 | 0 | 60 | 0 | 0 | 無 | 0 | 1 | 0 |
| 15 | 57637 | 0 | 0 | 16.08 | 0 | 0 | 無 | 0 | 0 | 1 |
| 16 | 136426 | 0 | 0 | 92.97 | 0 | 15.98 | 無 | 0 | 1 | 0 |
| 17 | 157895 | 0 | 0 | 21.73 | 0 | 8.67 | 有 | 1 | 1 | 0 |
| 18 | 198098 | 0 | 0 | 32.94 | 0 | 5.72 | 有 | 1 | 1 | 0 |
| 19 | 113592 | 27.68 | 2000000 | 75.22 | 2.05 | 12.11 | 有 | 1 | 1 | 0 |
| 20 | 128205 | 0 | 0 | 70 | 0 | 8 | 無 | 0 | 1 | 0 |
| 21 | 16804 | 41.36 | 0 | 92.02 | 1.22 | 12.83 | 有 | 1 | 1 | 0 |
| 22 | 114647 | 0 | 0 | 90.78 | 0 | 9.3 | 無 | 0 | 1 | 0 |
| 23 | 72333 | 0 | 0 | 82.73 | 0 | 5.8 | 有 | 1 | 0 | 1 |

圖 8：Dummy 變數轉換後的數據截圖

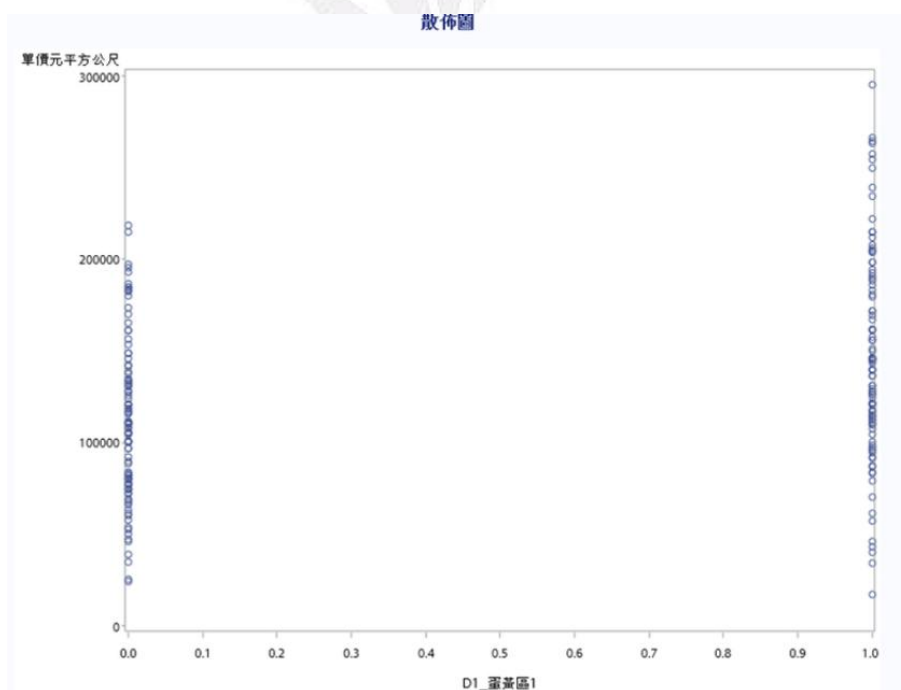


圖 9：D1_蛋黃區 1 散佈圖

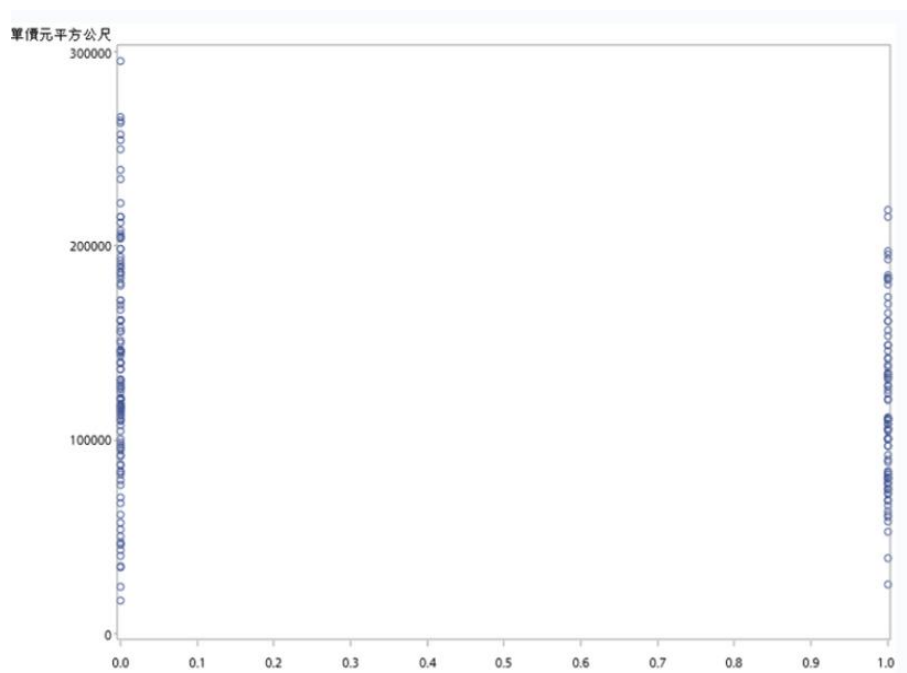


圖 10：D2_蛋白區 1 散布圖



圖 11：ANOVA 與迴歸參數截圖

五、 屋齡對住宅單價之影響

本研究以建物屋齡作為單一解釋變數，探討其對住宅單價之影響。簡單線性迴歸分析結果顯示，屋齡對住宅單價具有顯著影響，其對應之 p 值為 0.0009 (< 0.05)，顯示屋齡與住宅單價之間存在統計上顯著的關聯性。迴歸係數估計結果顯示，屋齡每增加一年，住宅單價平均每平方公尺約下降 780 元，呈現明顯之負向關係，反映房屋隨著使用年限增加，其市場評價及價格水準逐漸降低之現象。

然而，模型之決定係數 R^2 僅為 0.0544，表示屋齡單一變數僅能解釋約 5.44% 的住宅單價變異顯示雖然屋齡對房價具有顯著影響，但其解釋力仍屬有限。此結果亦說明住宅價格之形成並非僅由單一因素所決定，仍須納入其他關鍵變數，如區位、房間數及住宅設備條件等，方能提升整體模型之解釋能力，後續將透過多元迴歸分析進一步探討。

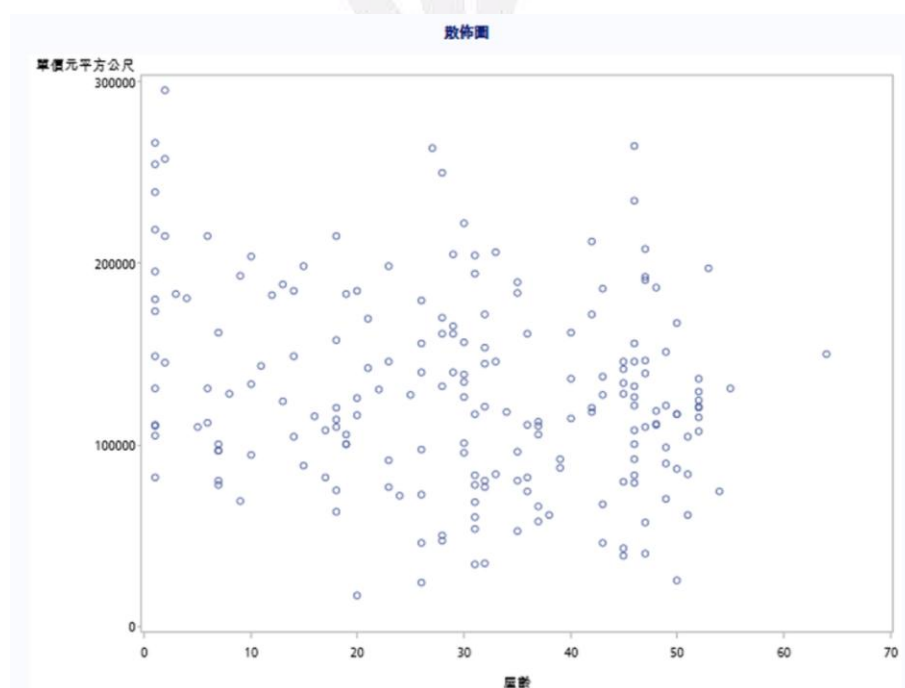


圖 12：屋齡變數與每平方公尺單價的散布圖



圖 13：ANOVA 與迴歸參數截圖

第四節、多元迴歸分析(未刪除離群值)

本研究首先將所有解釋變數納入多元迴歸模型，以評估其對新北市住宅單價之影響。反應變數 (Y) 為每平方公尺單價 (元/平方公尺) 解釋變數 (X) 包括屋齡、房間數、總樓層、是否有電梯及鄉鎮市區等五項變數 (未刪除離群值)。

一、模型整體檢定與迴歸係數

多元迴歸分析結果顯示，模型整體具有統計顯著性 ($F = xx, P < 0.0001$)， $R^2 = 0.2546$ ，表示所有解釋變數僅能解釋約 25% 的住宅單價波動。

個別變數方面：

- 屋齡：迴歸係數為 -1,087.39， $P = 0.0012$ ，呈現顯著負向影響，表示屋齡每增加一年，平均單價下降約 1,087 元／平方公尺。
- D1_蛋黃區：迴歸係數為 61,600， $P < 0.0001$ ，為模型中最顯著之加價因素，位於蛋黃區的物件單價平均高於蛋殼區約 6.16 萬元／平方公尺。
- 房間數：迴歸係數為 -9,730， $P = 0.0024$ ，呈負相關，反映在相同坪數下房數增加會降低住宅單價。
- 電梯、D2_蛋白區及總樓層：P 值均大於 0.05，未達統計顯著性，對單價影響有限。

模型之變異係數 (CV) 為 36.57，均方誤差 (MSE) 為 47,090，顯示存在高影響離群值，導致模型穩定性與預測精準度受限。變數間的多重共線性檢測結果 ($VIF = 1-5$) 顯示共線性問題不明顯。



圖 14：多元迴歸模型 ANOVA、參數估計及 R²結果截圖

二、殘差分析

為檢驗模型假設，進行殘差分析，得到以下的發現：

- 殘差直方圖與 Q-Q 圖顯示殘差呈鐘形分佈但右偏，部分觀測值實際單價高於預測值，反映市場中存在單價極高的案例。
- 殘差預測圖顯示數據點分布大致對稱，但隨預測單價增加，散布幅度擴大，部分預測誤差超過 10 萬元／平方公尺。

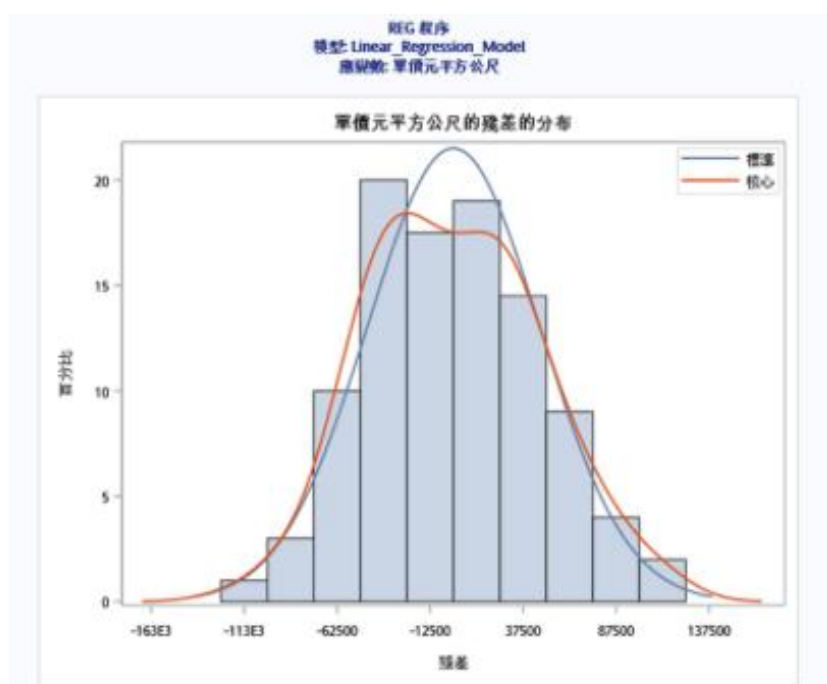


圖 15：殘差直方圖

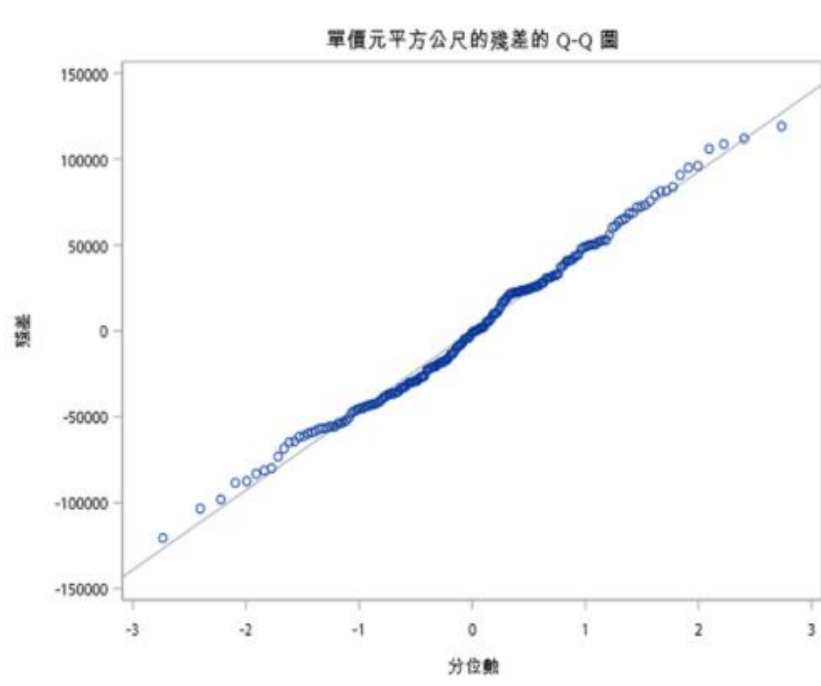


圖 16：殘差 Q-Q 圖

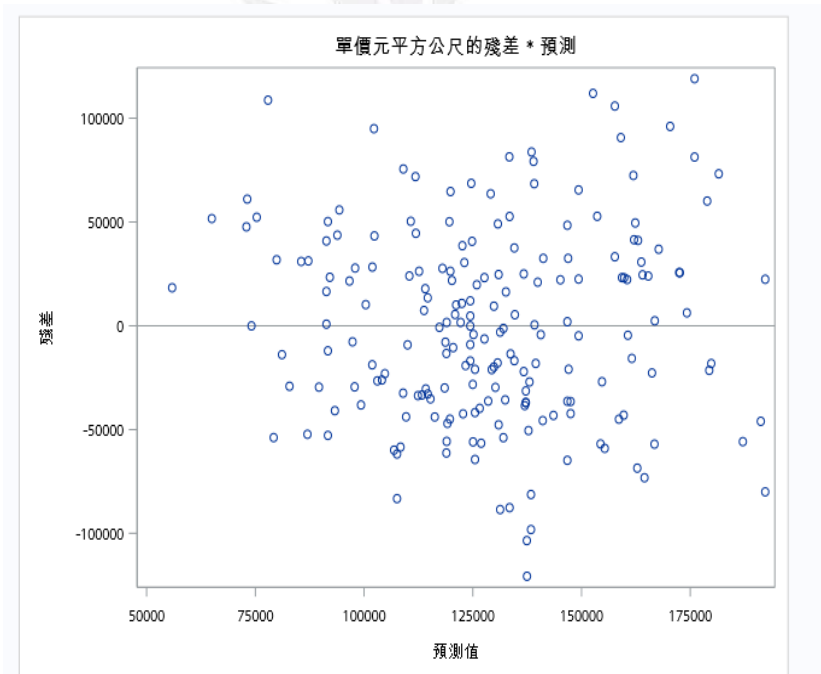


圖 17：殘差預測圖

三、影響值與離群值診斷

- Rstudent 預測圖顯示大部分觀測值落在 ± 2 區間，但少數點接近 3，為高影響值。
- Cook's D 圖指出若刪除突出點，模型迴歸係數可能顯著改變。
- 離群值與槓桿作用診斷圖顯示咖啡色點（高離群值及高槓桿）為模型穩定性受損的主要原因。
- DFFITS 影響診斷圖確認編號 21 與 42 等觀測值對模型產生顯著干擾。

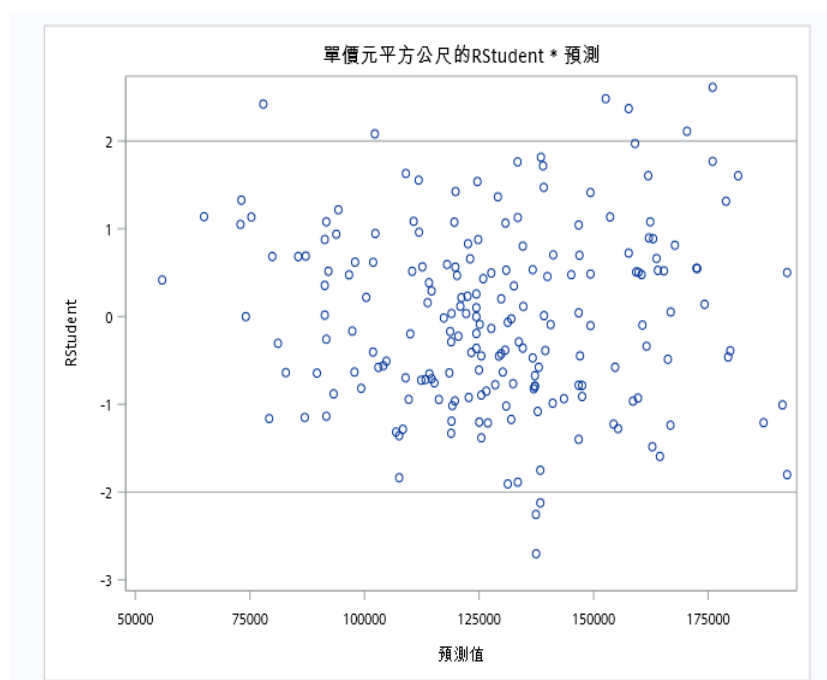


圖 18：Rstudent 預測圖

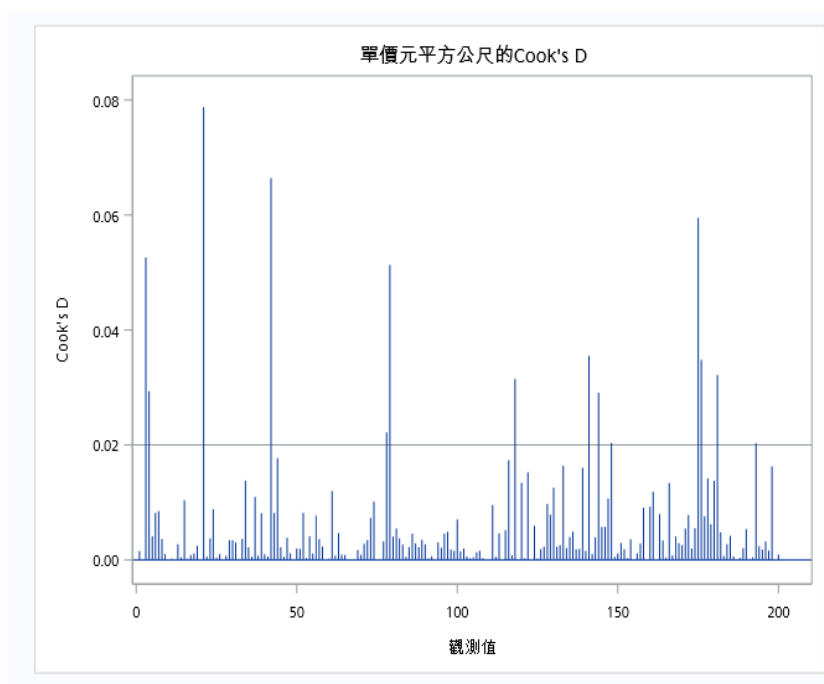


圖 19：Cook's D 圖

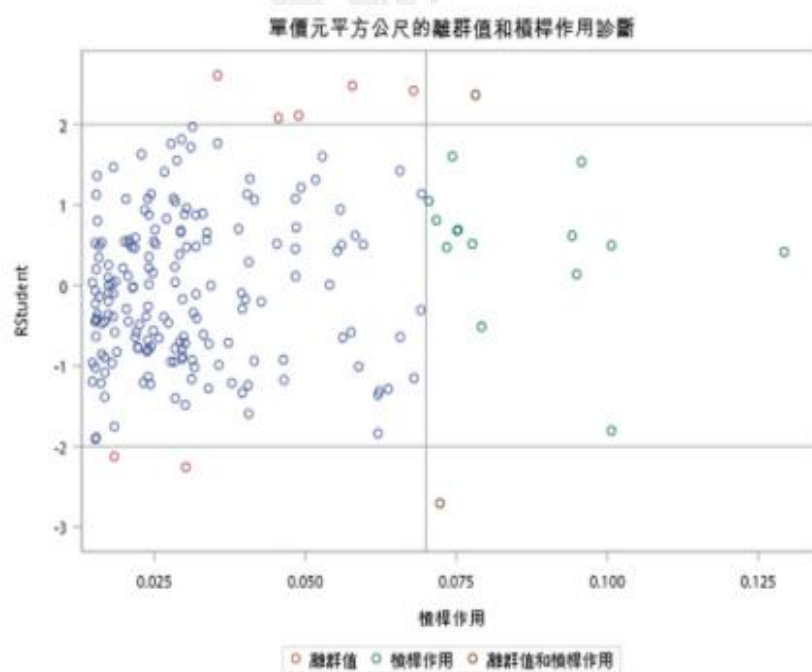


圖 20：離群值與槓桿作用診斷圖

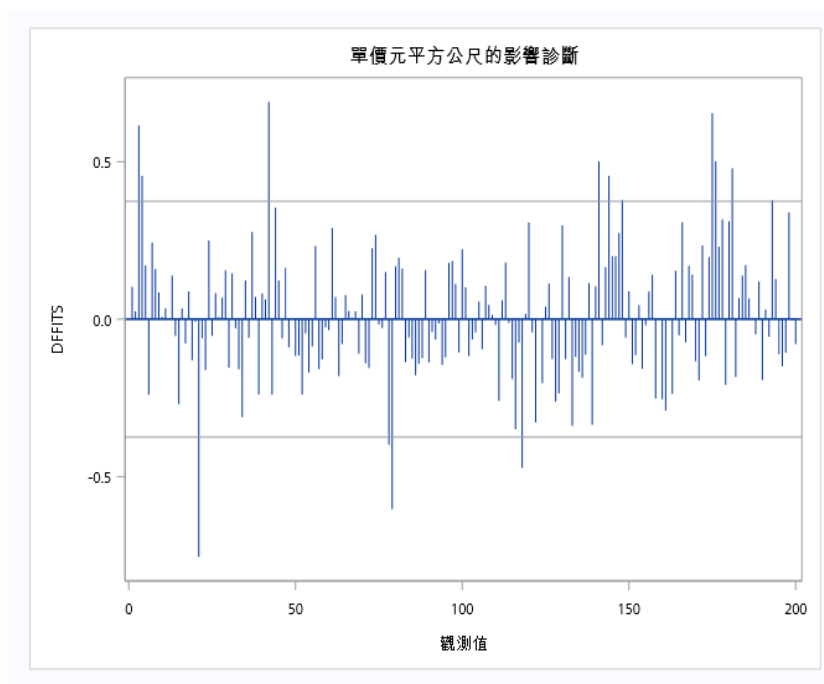


圖 21：DFFITS 影響診斷圖

第五節、離群值檢測與刪除

為確保多元迴歸模型的穩定性及預測精準度，本研究針對所有觀測值進行離群值診斷。離群值判定依據如下三個統計指標：

1. Cook's $D > 0.02$ ：表示該觀測值對迴歸係數估計具有不成比例的影響力。
2. RStudent 絕對值 > 2 ：表示該觀測值的標準化殘差過大，對模型擬合影響顯著。
3. DFFITS 絕對值 > 0.37 ：表示該觀測值對模型預測值的影響超過統計門檻。

| | stdi_單價元平方公尺 | student_單價元平方公尺 | cook_單價元平方公尺 | h_單價元平方公尺 | rstudent_單價元平方公尺 | dffits_單價元平方公尺 | stdi_單價元... |
|----|--------------|-----------------|--------------|--------------|------------------|----------------|--------------|
| 1 | 45937.310738 | 0.456982455 | 0.0015157103 | 0.0483495715 | 0.4560438198 | 0.1027931479 | 48214.748355 |
| 2 | 46726.972325 | 0.2027026612 | 0.0000915095 | 0.0153506538 | 0.2021983673 | 0.025246448 | 47449.852835 |
| 3 | 45708.491453 | 2.4501265522 | 0.0526157554 | 0.0578065307 | 2.4826868908 | 0.6149507833 | 48431.72809 |
| 4 | 45305.266606 | 1.5985545433 | 0.0293246537 | 0.0743566316 | 1.6050690313 | 0.4549165347 | 48809.131261 |
| 5 | 46370.248702 | 0.9609705821 | 0.0041260159 | 0.0303273115 | 0.960779107 | 0.169913515 | 47798.519978 |
| 6 | 46190.960856 | -1.20811098 | 0.008193615 | 0.0378111806 | -1.209559325 | -0.239776789 | 47971.800174 |
| 7 | 45937.361394 | 1.0785027476 | 0.008441893 | 0.0483474727 | 1.0789613141 | 0.2431943939 | 48214.700091 |
| 8 | 46332.756798 | 0.8782724849 | 0.0036304168 | 0.0318947034 | 0.8777500193 | 0.1593194629 | 47834.863062 |
| 9 | 46369.90598 | 0.4778217876 | 0.0010205969 | 0.0303416451 | 0.4768644419 | 0.0843538929 | 47798.852456 |
| 10 | 46415.078804 | 0.041834645 | 7.3217501E-6 | 0.0284515054 | 0.0417263136 | 0.0071405292 | 47754.989344 |
| 11 | 46518.671883 | 0.2191239649 | 0.0001694633 | 0.0241098663 | 0.2183827407 | 0.0343568077 | 47654.083172 |
| 12 | 46518.765088 | 0.0170915333 | 1.0308268E-6 | 0.0241059557 | 0.0170472101 | 0.0026792557 | 47653.992188 |
| 13 | 46518.765088 | 0.8790877605 | 0.0027270187 | 0.0241059557 | 0.878568084 | 0.1380817446 | 47653.992188 |
| 14 | 46726.972325 | -0.424622383 | 0.0004015624 | 0.0153506538 | -0.423718866 | -0.052805454 | 47449.852835 |
| 15 | 46149.143707 | -1.327759019 | 0.0103714932 | 0.0395525207 | -1.330404943 | -0.259981679 | 48012.029866 |
| 16 | 46678.376154 | 0.2575416754 | 0.0001677684 | 0.0173976645 | 0.2569177533 | 0.0341861895 | 47497.659693 |
| 17 | 46442.52532 | -0.462128241 | 0.0008563385 | 0.0273021268 | -0.461184693 | -0.077265237 | 47728.296833 |
| 18 | 46499.153318 | 0.5519907962 | 0.0011128271 | 0.0249286333 | 0.5509940168 | 0.0881004073 | 47673.128882 |
| 19 | 46664.225581 | -0.964116771 | 0.0024330905 | 0.017993327 | -0.963939867 | -0.130481355 | 47511.562042 |
| 20 | 46729.510679 | -0.065618401 | 9.5216909E-6 | 0.0152436726 | -0.065448915 | -0.008142971 | 47447.353018 |
| 21 | 45355.629446 | -2.660092835 | 0.0787789762 | 0.0722975359 | -2.703210031 | -0.754635416 | 48762.335406 |
| 22 | 46704.22706 | -0.472361127 | 0.0005284687 | 0.0163090151 | -0.471408378 | -0.060699039 | 47472.240851 |
| 23 | 46417.881249 | -0.945930973 | 0.0037274695 | 0.0283341497 | -0.945671898 | -0.161486829 | 47752.264628 |
| 24 | 46548.11368 | 1.6231568039 | 0.0088108563 | 0.0228741882 | 1.6301108061 | 0.2494104973 | 47625.325086 |
| 25 | 46660.089575 | -0.388374211 | 0.0003987108 | 0.0181673963 | -0.387518211 | -0.052713243 | 47515.623931 |
| 26 | 46492.02688 | 0.5166263576 | 0.0008667925 | 0.0252274884 | 0.5156428779 | 0.0829534388 | 47680.078788 |

圖 22：離群值判斷標準示意圖（Cook's D、RStudent、DFFITS 閾值）

依上述判斷，模型識別出以下觀測值為離群值：3、21、42、78、79、118、144、175、176、181 及後續刪除的觀測值 3。這些觀測值均顯示其 Cook's D、RStudent 或 DFFITS 指標超出閾值，對模型穩定性造成顯著影響。刪除上述離群值後，將進行多元迴歸分析，以提升模型擬合度及結果解釋力。

| | residual_單價元平方公尺 | stdi_單價元平方公尺 | student_單價元平方公尺 | cook_單價元平方公尺 | h_單價元平方公尺 | rstudent_單價元平方公尺 | dffits_單價元平方公尺 |
|-----|------------------|--------------|-----------------|--------------|--------------|------------------|----------------|
| 3 | 111991.58857 | 45708.491453 | 2.4501265522 | 0.0526157554 | 0.0578065307 | 2.4826868908 | 0.6149507833 |
| 21 | -120650.1849 | 45355.629446 | -2.660092835 | 0.0787789762 | 0.0722975359 | -2.703210031 | -0.754635416 |
| 42 | 105863.59244 | 45211.07339 | 2.341541231 | 0.0664485144 | 0.0782016013 | 2.3693945526 | 0.690115426 |
| 78 | -103499.3976 | 46372.942679 | -2.231891954 | 0.022171244 | 0.0302146378 | -2.255398599 | -0.398101845 |
| 79 | -80001.78874 | 44655.982259 | -1.791513358 | 0.0513401199 | 0.1006979023 | -1.801911484 | -0.602963276 |
| 118 | -83254.60962 | 45606.004154 | -1.825518617 | 0.0314821652 | 0.0630269547 | -1.83670937 | -0.472319083 |
| 144 | 94966.503198 | 46004.761146 | 2.0642755409 | 0.0290536849 | 0.045552876 | 2.0820335569 | 0.4548515606 |
| 175 | 108708.97871 | 45462.313232 | 2.3911888988 | 0.059529249 | 0.0679281903 | 2.4211195292 | 0.6536068073 |
| 176 | 119053.01174 | 46246.831286 | 2.5742955447 | 0.0348271747 | 0.0354821355 | 2.6128682917 | 0.5011494473 |
| 181 | 96075.649905 | 45924.019393 | 2.0920566443 | 0.0321464675 | 0.0489001864 | 2.110699394 | 0.4785954821 |

圖 23：被判定為離群值之觀測數據各指標截圖（包含 Cook's D、RStudent 及 DFFITS）

第六節、刪除離群值後之多元迴歸分析

一、模型整體檢定與迴歸係數

為提升模型的穩定性與解釋力，本研究在前述離群值診斷後，刪除影響顯著的觀測值，並針對剩餘樣本進行多元迴歸分析。分析之反應變數（Y）為住宅單價（元／平方公尺），解釋變數（X）為屋齡、房數、總樓層、是否有電梯及鄉鎮市區等所有解釋變數（已刪除離群值）。

多元迴歸分析結果顯示，模型整體具有高度統計顯著性（ $P < 0.0001$ ）， $R^2 = 0.2894$ ，表示刪除離群值後，所有解釋變數能解釋約 28.94% 的住宅單價波動，相較於刪除前的 25% 略有提升。

個別變數分析如下：

- 屋齡：迴歸係數為 -1,329.72， $P < 0.0001$ ，呈顯著負向影響，表示屋齡每增加一年，平均單價下降約 1,330 元／平方公尺。
- 房間數：迴歸係數為 -9,730， $P = 0.039 < 0.05$ ，呈負相關，反映相同坪數下房數增加會降低住宅單價。
- D1_蛋黃區：迴歸係數為 58,800， $P < 0.0001$ ，為模型中最顯著的加價因子，位於蛋黃區的物件單價平均高於蛋殼區約 5.88 萬元／平方公尺。
- 電梯、D2_蛋白區及總樓層：P 值均大於 0.05，未達統計顯著性，對單價影響有限。

模型之變異係數（CV）為 32.80，均方誤差（MSE）為 41,402，顯示刪除離群值後，模型穩定性與預測精準度提升。變數間多重共線性檢測（ $VIF = 1-5$ ）結果顯示共線性問題仍不明顯。



圖 24：多元迴歸模型 ANOVA、參數估計及 R² 結果截圖

二、殘差分析

為檢驗刪除離群值後多元迴歸模型之假設，本研究進行殘差分析，以評估模型對住宅單價的擬合情形及預測精準度。刪除離群值後的殘差直方圖顯示，殘差分布較為集中並呈鐘形，誤差範圍縮小，部分極端高價房屋的影響已排除，模型對大多數觀測值的預測較為準確。殘差 Q-Q 圖進一步顯示，大部分數據點接近理想的 45 度斜線，尾端偏離程度降低，代表刪除離群值後，模型殘差更符合常態分布假設。

從殘差預測圖觀察，數據點大致隨機分布於 0 上下，雖隨預測單價增加散布幅度略有擴大，但整體誤差幅度縮小，模型對高價物件的預測誤差明顯降低。

整體而言，刪除離群值後的模型在殘差分布、偏態及預測穩定性上均有所改善，顯示模型更適合用於解釋新北市住宅單價波動。

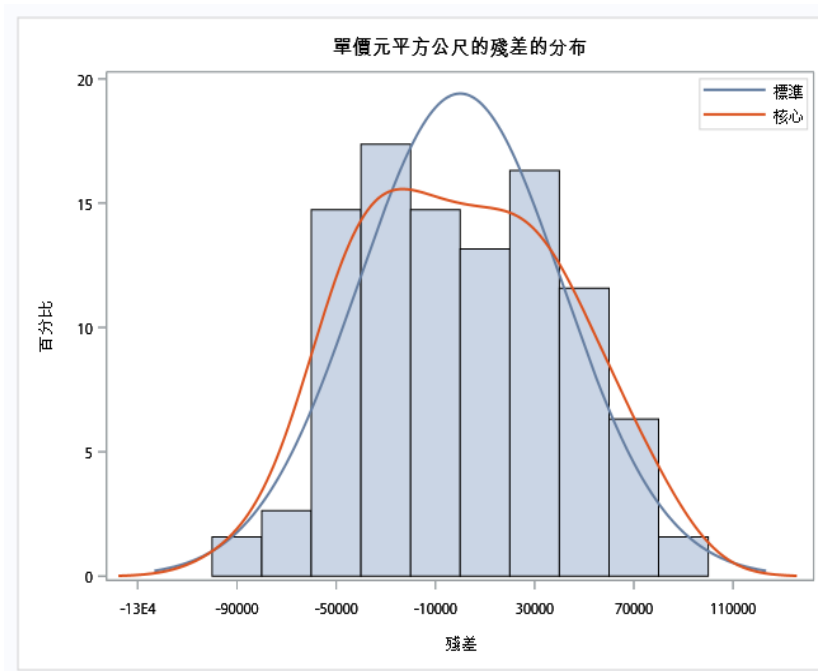


圖 25：刪除離群值後殘差直方圖

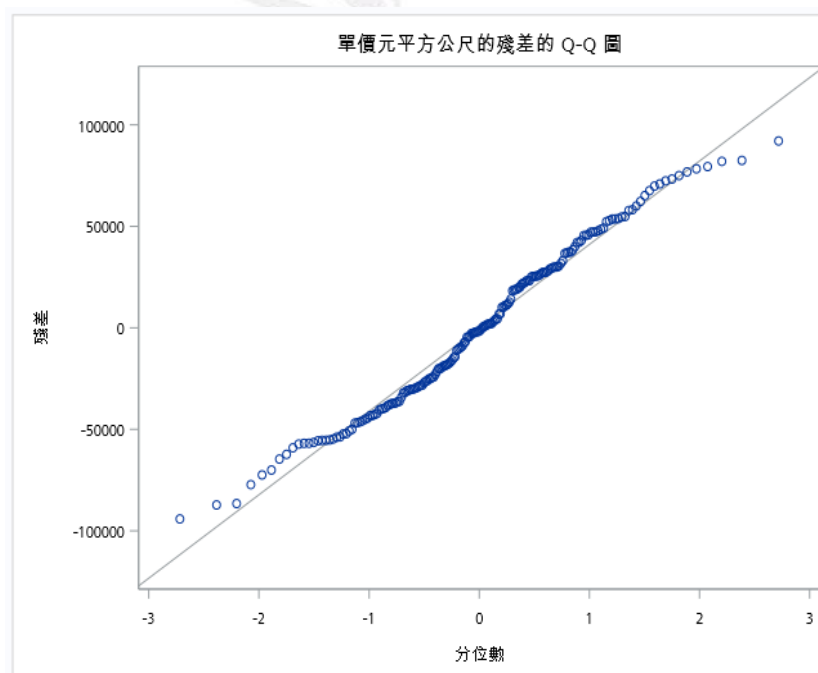


圖 26：刪除離群值後殘差 Q-Q 圖

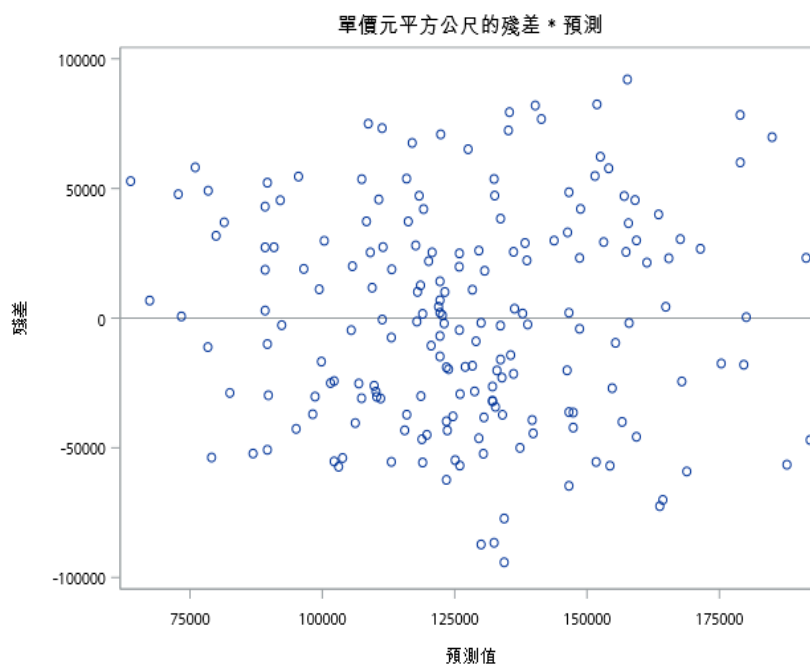


圖 27：刪除離群值後殘差預測圖

三、影響值與離群值診斷

為評估刪除離群值後模型的穩定性與各觀測值對迴歸結果的影響，本研究進行影響值與離群值診斷。刪除離群值後的 Rstudent 預測圖顯示，大多數觀測值仍落在 ± 2 的合理區間內，顯示模型對整體數據擬合良好，僅有少數點接近 3，為高影響值。Cook's D 圖則指出，若刪除極端影響值，模型迴歸係數可能發生顯著變化，但整體影響已明顯降低。

離群值與槓桿作用診斷圖中，咖啡色點（同時具有高離群值與高槓桿作用）對模型穩定性具有最大影響，但數量已減少，顯示刪除離群值後模型的預測更穩定。DFITS 影響診斷圖進一步確認，原先高影響力的編號 21 與 42 等觀測值對模型變動的影響已降低，沒有單一樣本能顯著扭曲整體預測結果。

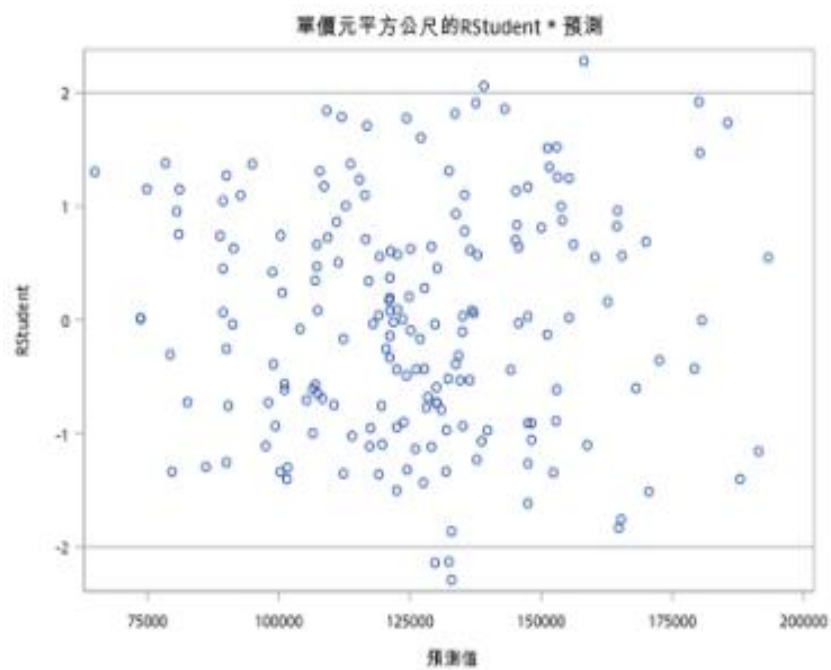


圖 28：刪除離群值後 Rstudent 預測圖

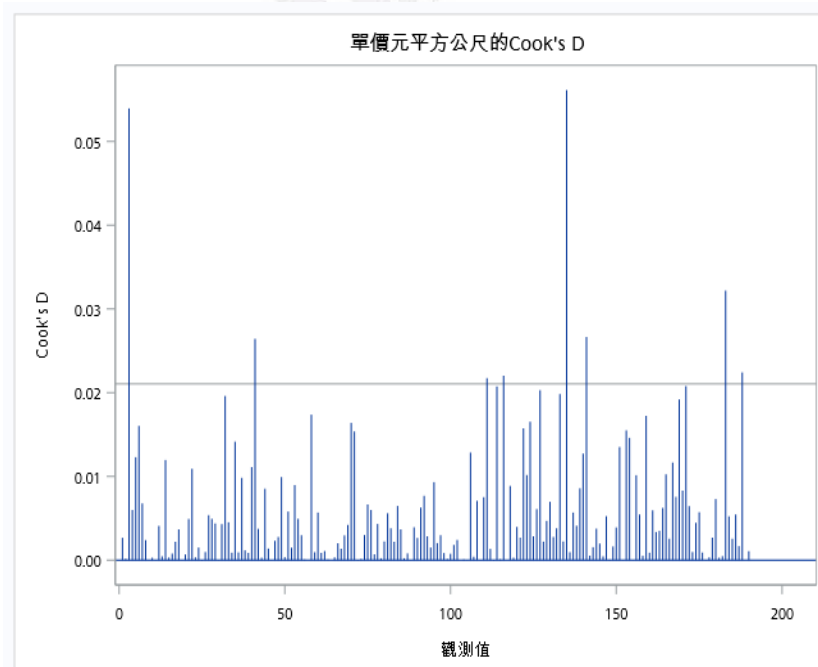


圖 29：刪除離群值後 Cook's D 圖

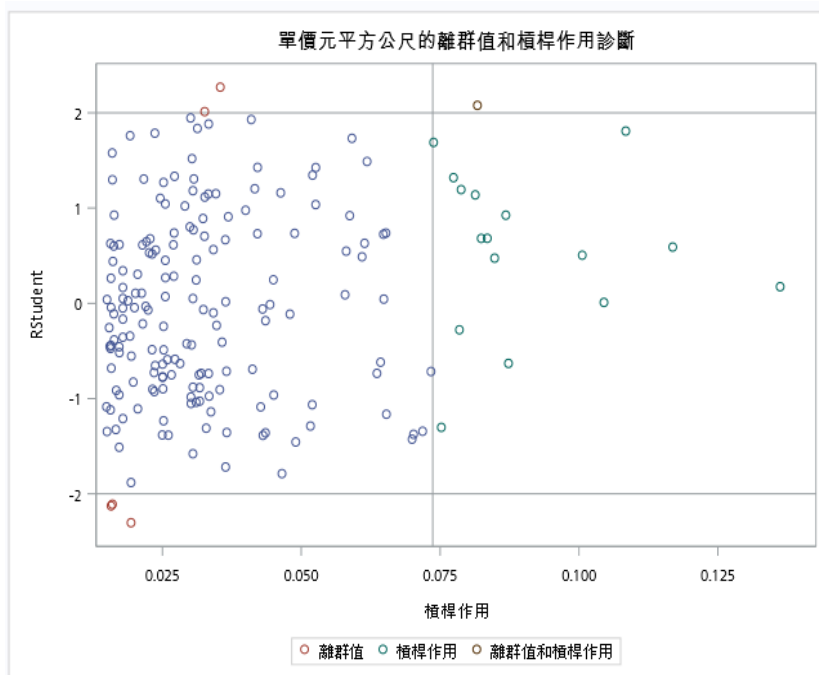


圖 30：刪除離群值後離群值與槓桿作用診斷圖

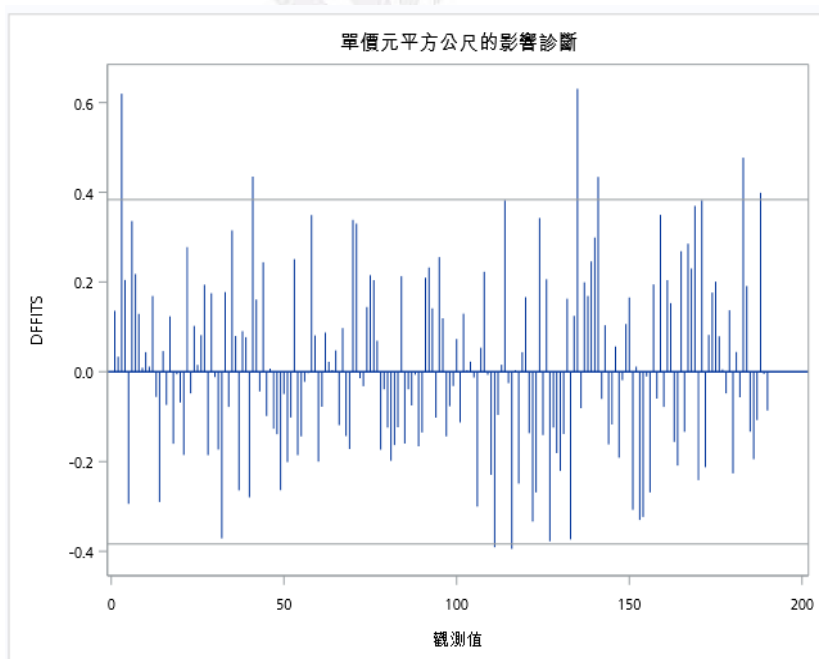


圖 31：刪除離群值後 DFFITS 影響診斷圖

第七節、模型選擇與變數篩選

為了建立最穩健且解釋力最佳的多元迴歸模型，本研究採用多種變數選擇方法，包括前進選擇法 (Forward Selection)、向後剔除法 (Backward Elimination)、逐步選擇法 (Stepwise Selection)、最大 R 平方改善法、最小 R 平方改善法、R 平方選取法、調整 R 平方選取法以及 Mallows's Cp 選取法，以比較不同策略下的模型表現與變數重要性。

一、前進選擇法(Forward Selection)

模型最先選入「D1_蛋黃區 1」，單一變數即可提供 $R^2 = 0.1074$ ，顯示地段為影響單價的核心因素。加入屋齡後，模型 R^2 大幅提升至 0.2592，顯示屋齡是單價的第二大關鍵變數。隨後依序加入房數、D2_蛋白區 1 及總樓層，最終模型 R^2 提升至 0.2875。步驟 5 的 Mallows's Cp = 5.0000，接近最終變數數量 (5 個)，顯示模型達到平衡，既不過度擬合，也未遺漏重要變數。

核心變數統計結果：

- 屋齡：F = 24.44， $p < 0.0001$
- D1_蛋黃區 1：F = 27.09， $p < 0.0001$



圖 32：前進選擇法結果截圖

二、向後剔除法 (Backward Elimination)

初始模型納入所有變數後，首先剔除「電梯 10」，其 p 值高達 0.9966，對單價影響不顯著，R² 維持 0.2875。接著剔除「總樓層」，p = 0.2781，剔除後 R² 輕微下降至 0.2829。最終模型步驟 2 的 Mallows' Cp = 4.1770，接近變數數量 (4 個)，顯示向後剔除法成功簡化模型，降低過度擬合風險。

核心變數統計結果：

- 屋齡：p < 0.0001
- D1_蛋黃區 1：p < 0.0001
- 房數：p = 0.0448

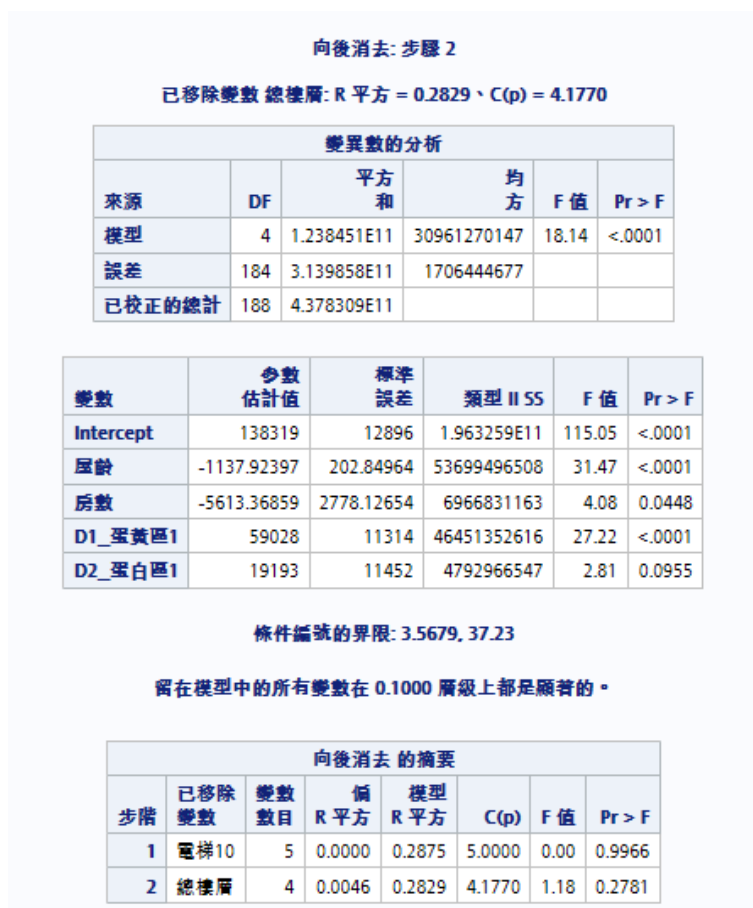


圖 33：向後剔除法結果截圖

三、逐步選擇法 (Stepwise Selection)

模型先選入「D1_蛋黃區 1」與「屋齡」， R^2 提升至 0.2592，再加入「房數」與「D2_蛋白區 1」，最終停在步驟 4，最終 $R^2 = 0.2829$ ，Mallow's $C_p = 4.1770$ ，顯示在解釋力與簡潔性間取得平衡。

變數係數：

- 屋齡：約 -1,137 元／平方公尺
- D1_蛋黃區 1：約 59,028 元／平方公尺
- 房數：係數 -5,613 元／平方公尺， $p = 0.0448$

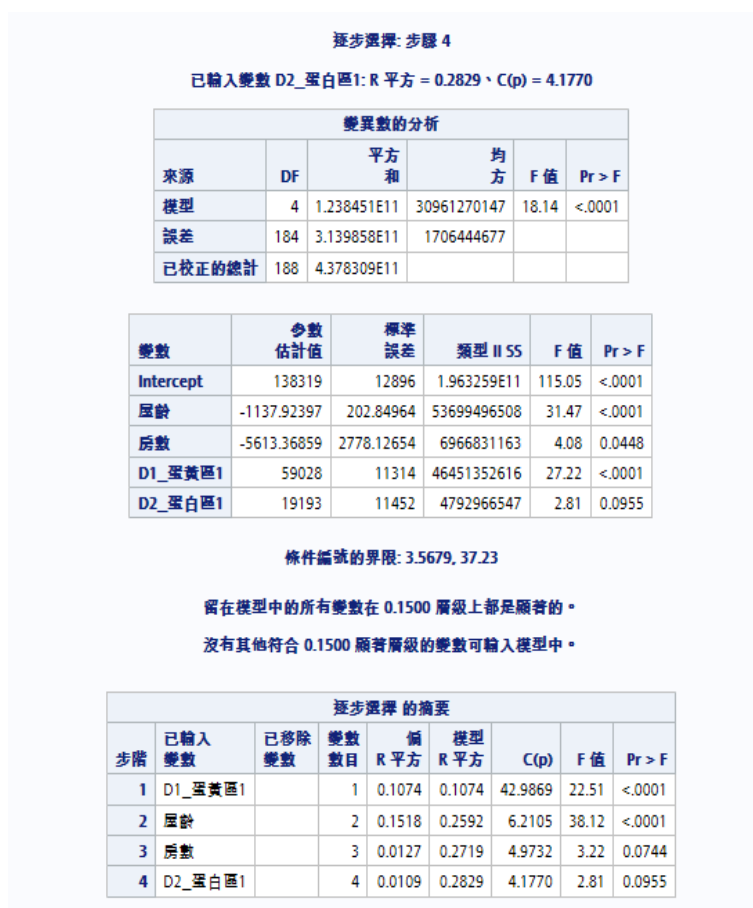


圖 34：逐步選擇法結果截圖

四、最大 R 平方改善法 (Max R² Improvement)

最終 R² = 0.2875，Pr < 0.0001，顯示模型顯著。核心加價與減價因子如下：

- D1_蛋黃區 1：係數 58,875，p < 0.0001
- 屋齡：係數 -1,329，p < 0.0001
- 房數：係數負向影響，p = 0.0390

模型同時保留 D2_蛋白區 1 (p = 0.0995)、總樓層 (p = 0.3300) 及電梯 10 (p = 0.9966)，因其微小貢獻可增加 R²。模型 C(p) = 7.0000，顯示統計上無偏誤，為最佳預測組合。



圖 35：最大 R² 改善法結果截圖

五、最小 R 平方改善法 (Min R² Improvement)

最終模型 $R^2 = 0.2875$, $Pr < 0.0001$ 。由 MINR 與 MAXR 選出的變數一致，顯示變數地位穩固：

- D1_蛋黃區 1：係數 58,875， $p < 0.0001$
- 屋齡：係數 -1,329， $p < 0.0001$
- 房數：係數負向影響， $p = 0.0390$

D2_蛋白區 1、總樓層、電梯 10 由於對 R^2 幾乎無改善，未被剔除。模型 C(p) = 7.0000，與變數總數相等，顯示模型無偏誤。



圖 36：最小 R² 改善法結果截圖

六、R 平方選取法

- 1 個變數最佳模型：D1_蛋黃區，R² = 0.1074
- 2 個變數最佳模型：D1_蛋黃區 + 屋齡，R² = 0.2592
- 3 個變數最佳模型：D1_蛋黃區 + 屋齡 + 房數，R² = 0.2719
- 6 個變數最佳模型：納入所有變數，R² = 0.2875

觀察解釋力提升的邊際遞減：從 1 → 2 個變數，R² 增加約 0.15；從 4 → 6 個變數，R² 只增加 <0.005。



圖 37：R 平方選取法結果截圖

七、調整 R 平方選取法

1. 最佳變數組合：屋齡、房數、D1_蛋黃區 1、D2_蛋白區 1、總樓層，調整 $R^2 = 0.2680$
2. 精簡版最佳組合：屋齡、房數、D1_蛋黃區 1、D2_蛋白區 1(4 個變數)，

調整 $R^2 = 0.2673$

觀察結果顯示，增加到 6 個變數時，雖 R^2 維持 0.2875，但調整 R^2 下降至 0.2640，顯示部分額外變數貢獻有限。



圖 38：調整 R^2 選取法結果截圖

八、Mallow's Cp 選取法

- 最佳組合：屋齡、房數、D1_蛋黃區 1、D2_蛋白區 1(4 個變數)， $C(p) = 4.1770$ ，接近變數數量，模型無偏誤，為首選預測組合。
- 精簡組合：屋齡、房數、D1_蛋黃區 1 (3 個變數)， $C(p) = 4.9732$ ，略偏離變數數量，省略 D2_蛋白區可能造成小幅預測偏差。
- 全部變數組合： $C(p) = 7.0000$ ，雖準確，但電梯變數對模型加分有限，造成模型複雜化。



圖 39：Mallow's Cp 選取法結果截圖

九、模型選擇與最終模型決定

綜合八種模式的變數選取結果，本研究最終決定採用 向後消去法 (Backward Elimination) 與逐步選擇法 (Stepwise Selection) 所得到的模型。兩種方法所得的變數組合完全一致，顯示模型選取結果具有高度一致性與穩定性。經過刪除不顯著變數後，模型仍能維持接近最大 R^2 的解釋力，顯示精簡變數並未造成模型效能的明顯犧牲。這一結果說明，留下的變數為影響新北市住宅單價的核心因素，而被剔除的變數對模型貢獻有限。

此外，Mallow's C_p 選取法所選模型的 C_p 值與變數數量非常接近，顯示模型偏誤程度低，結構合理。此方法在控制模型複雜度的同時，仍能有效保留對解釋力有實質貢獻的變數，使得即便在使用較少變數的情況下，模型仍能掌握住宅單價變動的主要結構特徵。

綜上所述，考量模型解釋力、穩定性以及簡潔性，本研究最終選定由 屋齡、房數、D1_蛋黃區 1 與 D2_蛋白區 1 所組成的模型作為分析基礎，作為後續單價預測與解釋的核心模型。

第四章 結論

本研究針對新北市住宅單價之影響因素進行多元迴歸分析，主要結論如下：

1. 區位為主要影響因素：

研究結果顯示，蛋黃區物件之單價顯著高於蛋殼區，且在各種模型中均維持高度顯著性。此結果說明，區位對新北市住宅單價具有穩定且強大的解釋力，是決定房價的核心因素。

2. 屋齡與房數對單價呈負向影響：

屋齡每增加一年，住宅單價顯著下降；在坪數固定情況下，房數增加亦會降低市場評價。上述影響在完整模型與刪除離群值後的模型中均保持一致，顯示其對房價具有穩定且可靠的影響。

3. 電梯與總樓層對單價影響有限：

雖然電梯在單一變數分析中曾達統計顯著，但在多元迴歸控制其他變數後，電梯與總樓層對單價的影響均未達統計顯著水準，顯示其非主要價格決定因素。

4. 離群值影響模型穩定性，但不改變核心結論：

高影響離群值會放大模型誤差並降低預測精準度。刪除離群值後，模型整體表現明顯改善，且核心變數的影響方向及統計顯著性保持不變，證實模型結果的穩健性。

綜上所述，本研究確認區位、屋齡與房數為新北市住宅單價之關鍵影響因素，而電梯與總樓層影響有限；經離群值調整後，模型仍能有效反映住宅市場價格結構，具實務應用價值。

參考文獻

- Lin, Y. Y. (2014). *The determinants of house pricing in the Greater Taipei Area* [Master's thesis, National Taiwan University]. National Digital Library of Theses and Dissertations in Taiwan. <https://ndltd.ncl.edu.tw/handle/77039297194385721490>
- Huang, Y. (2016). *The effect of mass rapid transit on residential area property value: Evidence from New Taipei City* [Master's thesis, National Taiwan University]. Airiti Library. https://www.airitilibrary.com/Article/Detail/U0001-3006201712503000?utm_source
- Lee, L. S. (2014). *Determinants of the housing prices: Application of spatial and multilevel analysis* (Doctoral dissertation, National Taiwan University). National Taiwan University Institutional Repository.
- Lin, Y. C. (2013). *Neighborhood effect and housing prices: A case study in Taipei* (Master's thesis, National Taiwan University). National Taiwan University Institutional Repository.
- Shih, C. L. (2024). *A study on the determinants of housing prices in Taiwan* (Master's thesis, Chaoyang University of Technology). CYUT Institutional Repository.
- 內政部不動產資訊平台 . (2025). 新北市住宅實價登錄資料 <https://plvr.land.moi.gov.tw/>