

# 影響房價的因素

409120627 應數三 周家鴻

113 年 1 月

## 目錄

一、	引言.....	3
	研究背景與動機:.....	3
	研究目標:.....	3
	房地產交易實價登錄簡介:.....	3
	資料的獲取:.....	4
	資料的正確性:.....	4
二、	資料分析結果與討論.....	6
	各項變數概覽:.....	6
	所屬行政區域:.....	6
	土地類別.....	7
	建物型態.....	7
	屋齡 8	
	是否有管理組織.....	9
	ANOVA 簡介:.....	9
	多元迴歸簡介:.....	10
	分析結果:.....	11
	1. 單純考慮行政區變因的多元迴歸結果:.....	11
	2. 考慮行政區變因及土地類別變因的多元迴歸結果:.....	12
	3. 考慮行政區變因及房屋種類變因的多元迴歸結果:.....	14
	4. 考慮屋齡變因及管理狀況變因的多元迴歸結果:.....	16
三、	研究限制.....	17
四、	結論.....	17
	四組隨機變數組合的多元迴歸模型的比較:.....	17
	過大的 ANOVA 殘差值:.....	18

**表目錄：**

表格 1 台北行政區分佈.....	6
表格 2 土地類別分佈.....	7
表格 3 建物型態分佈.....	7
表格 4 屋齡分佈.....	8
表格 5 管理組織分佈.....	9
表格 6 迴歸統計結果 1.....	11
表格 7 ANOVA 結果 1.....	11
表格 8 係數結果 1.....	11
表格 9 迴歸統計結果 2.....	12
表格 10 ANOVA 結果 2.....	13
表格 11 係數結果 2.....	13
表格 12 迴歸統計結果 3.....	14
表格 13 ANOVA 結果 3.....	14
表格 14 係數結果 3.....	15
表格 15 迴歸統計結果 4.....	16
表格 16 ANOVA 結果 4.....	16
表格 17 係數結果 4.....	16

**圖目錄：**

圖表 1 甘特圖.....	錯誤! 尚未定義書籤。
圖表 2 台北行政區分佈.....	6
圖表 3 土地類別分佈.....	7
圖表 4 建物型態分佈.....	8
圖表 5 屋齡分佈.....	8
圖表 6 管理組織分佈.....	9

# 一、引言

## 研究背景與動機：

選擇以房地產作為期末報告的因素，主要可分為以下三點：

首先，房地產是一個社會經濟體系中至關重要的一環，直接關係到人們的生活品質和經濟發展。通過深入研究房地產市場的各種因素，我們可以更好地理解經濟運行的規律，從而提出更有針對性的政策建議，促進社會穩定和可持續發展。

其次，房地產市場涉及眾多變數，如房價、供需關係、貸款利率等，這些變數之間複雜的關聯性提供了豐富的研究材料。通過運用統計方法，我們可以深入挖掘這些變數之間的潛在模式，揭示其背後的規律和趨勢，從而對市場未來的發展趨勢有更為準確的預測。

此外，房地產市場的波動對個人和企業的財務狀況產生直接而深遠的影響。透過對房地產市場的統計分析，我們可以更好地評估風險，提高風險管理的效率，有助於個人和企業更好地應對市場變化。

期望透過本次報告，對房地產交易有更進一步的認識，而本報告將以台灣首都台北作為研究對象。

## 研究目標：

了解房地產的各項變因如何影響房地產價格，以及所使用統計方法在此題目上的優缺點，在實作的過程中對統計方法產生更進一步的認識。

## 房地產交易實價登錄簡介：

近年來，台灣政府推動房地產實價登錄制度，該政策推行的原因主要源自於房地產市場的不透明性、不公平性和投資風險。實價登錄是指將房地產交易實際成交價格公開登錄的一種制度。這一政策的推動旨在提高市場透明度，防範投機行為，促進房地產市場的健康發展。

背景上，過去房地產交易價格的不透明性成為社會關注的焦點。缺乏公開的價格資訊，容易讓市場出現不實報價、價格操縱等不當行為。此外，不同地區、不同房型的價格波動差異大，使得市場參與者難以正確評估房地產價格走勢，增加了投資風險。

推動實價登錄的目的在於建立更透明、公正的房地產市場。透過公開實際成交價格，市場參與者可以更準確地評估市場價格，降低資訊不對稱造成的風險。同時，實價登錄可以有效防範價格操縱、逃漏稅等不當行為，提高市場的公平性和穩定性。

當然，推動房地產實價登錄也面臨著一些挑戰和爭議。首先，可能影響市場活動。一些人擔心，公開實際成交價格會讓市場參與者過於謹慎，導致交易減少，進而影響整體市場的流動性。其

次，可能引發價格上升。一旦實價登錄實施，市場上的實際成交價格一旦公開，賣方可能傾向於將價格抬高，增加了買方的購屋成本。

儘管此政策的推行有利有弊，然而無疑使得資料研究者可以取得相對公正的資料，用以進行研究。而本報告將以行政院內政部所提供的房地產實價登錄得資訊，作為資料來源。

## 資料的獲取：

由於房地產的資料需求龐大且變因眾多，第二手資料相較第一手資料來說實際得多，而以下有幾種常見的資料來源：

1. 不動產仲介公司：與不動產仲介公司合作，他們通常有完整的房地產資訊和市場動態。這些公司可能提供不同層面的資訊，包括房價、租金、區域資料等。
2. 政府機構：向相關政府機構（例如地政局、房地產稅務機關）查詢，這些機構擁有關於土地、建物登記、稅務等方面的資料。有些資訊可能需要支付相應的費用。
3. 不動產交易平台：使用線上不動產交易平台，這些平台通常提供即時的房價和租金資訊。你可以參考多個平台以獲取更全面的資訊。
4. 資訊服務提供商：有一些公司專門提供房地產市場的資料報告和分析。這些報告可能包含市場趨勢、投資建議等資訊。
5. 研究機構和學術機構：一些大學或研究機構可能進行相關領域的研究，並提供有價值的資料。你可以查詢相應的研究報告或聯絡相關的研究人員。
6. 地方性報紙和媒體：一些地方性報紙和媒體可能會提供當地房地產市場的報導和資訊。
7. 網路爬蟲和 API：一些網站可能提供 API，你可以使用它們來獲取實時的房地產資料。然而，使用網路爬蟲需要確保合法性和符合相關條款，以免觸犯法律。

本次報告的資料來源為政府機構。（[行政院內政部房地產公開資料](#)，使用 112 年登錄的資料，交易日期涵蓋 100 年至 112 年）

## 資料的正確性：

在進行下去前，我們須了解到，政府所提供的資料未必皆能反映市場真實的價格。實價登錄存在一些潛在的問題和漏洞，而這些問題主要來自於政策的執行和課稅的實際操作。實價登錄作為公開房地產成交價格的一種方式，旨在提高市場透明度，但曾有報章雜誌指出了一些政策漏洞，引起了人們的關注。

首先，內政部在公開資料前將交易價格最高和最低各 5% 去除，可能使資料失真。此外，同一筆交易中含有兩個或兩個以上的標的時，僅登錄第一個標的地址，對其他標的不予登錄，這也可能造成對整體交易情況的誤解。作業疏失方面，例如買方未熟悉房地契約內容而導致報告誤差，或是地政士或仲介未確實查證賣方提供的資料的正確性，都可能使得登錄資料不夠精確。最後，買賣雙方刻意提高或壓低價格，以達到逃漏稅或獲取更高貸款額的目的，這可能使交易缺乏市場代表性。

另一方面，由於建設公司自建自售不用登錄、非透過仲介的租賃案件不登錄、代銷公司登錄時間差異大等因素，導致資料失真。其中，建設公司自建自售不用登錄的情況可能使得相關交易在登錄資料中失去了可追蹤的依據，難以監控。非透過仲介的租賃案件不登錄，則可能造成租賃市場真實價格的不透明性。此外，代銷公司登錄時間差異大的情況，一些代銷公司可能為了逃避登錄而將契約終止期限延長，以拉長登錄時間。

最後是課稅的問題。政府並非以買賣價格為課稅基礎，而是以公告現值為依據，這可能引起一些疑慮。實價登錄被視為還原真實房價的手段，未來如何實施實價課稅成為一個重要的問題。曾有電視名嘴提出，低報交易價格可能是為了逃避稅收，買方可能與賣方串通，以低報價格來減少應繳的相關稅金。而高報交易價格可能是為了獲取更高的貸款，有人討論到在一些特定建案中，高報價格的現象可能與建商為了提高該建案的行情，或在實價登錄中提供更多的貸款擔保有關。

綜上所述，實價登錄政策在執行中遇到了種種問題。資料不完整、政策執行上的漏洞、課稅基礎的選擇等都需要被更仔細地考慮和解決。

## 二、 資料分析結果與討論

### 各項變數概覽：

在房地產的公開資料中，由於變數眾多，而筆者進行資料分析的意圖僅須了解到未來一般民眾（係指非政府人員）購屋時，能夠作為參考的變數之於房價的影響即可，同時也考慮到資料分析執行上的困難，不考慮將類別過多的資料納入分析。基於以上理由，筆者選定了五項變因，其分別為：所屬行政區域、土地類別（非都市土地）、建物型態、屋齡、是否有管理組織。

### 所屬行政區域：

表格 1 台北行政區分佈

地區	士林區	大同區	大安區	中山區	中正區	內湖區	文山區	北投區	松山區
交易筆數	1370	716	1641	2495	850	1788	1274	1291	993

地區	信義區	南港區	萬華區
交易筆數	1266	689	1166



圖表 1 台北行政區分佈

## 土地類別

表格 2 土地類別分佈

土地類別	工	住	商	農
交易筆數	328	9750	5459	2



圖表 2 土地類別分佈

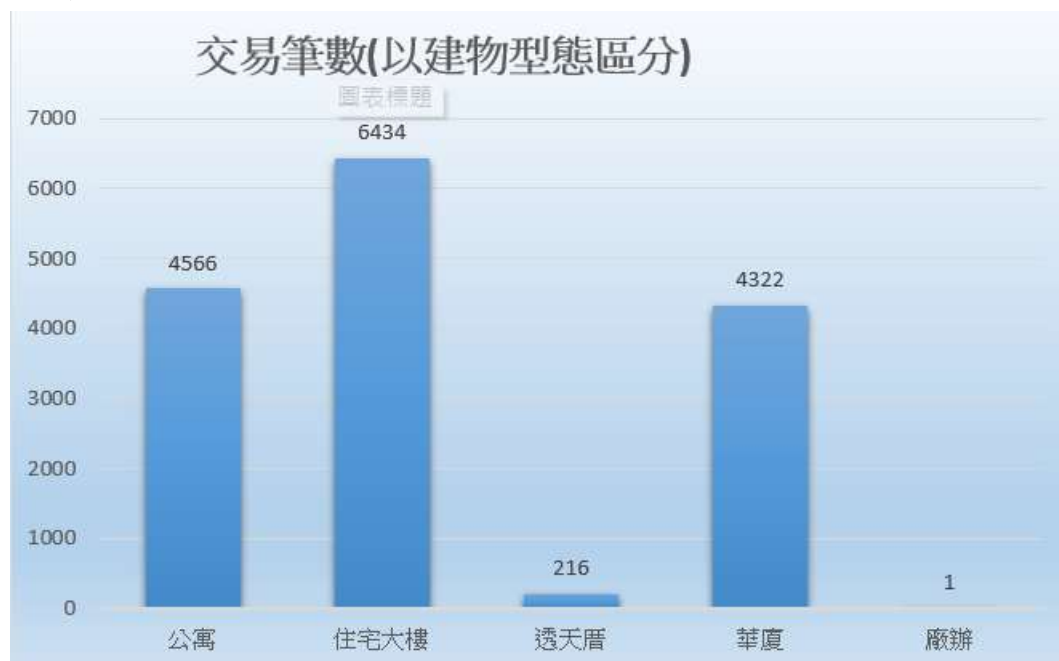
註:都市用土地受政策影響，故選擇剔除。

## 建物型態

表格 3 建物型態分佈

建物型態	公寓	住宅大樓	透天厝	華廈	廠辦
交易筆數	4566	6434	216	4322	1





圖表 3 建物型態分佈

註:其他類別由於組成複雜難以分析，故選擇剔除。

## 屋齡

表格 4 屋齡分佈

屋齡	10 年以下	10~20 年	20~30 年	30~40 年	40~50 年	50~60 年	60 年以上
交易筆數	2109	2566	1981	2881	4645	1302	55

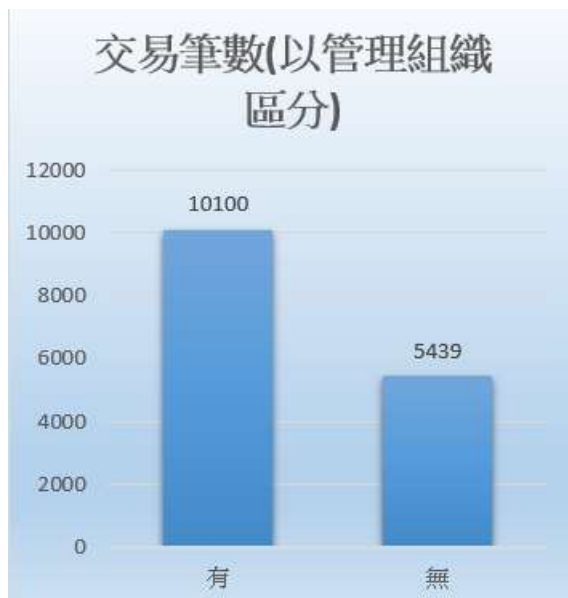


圖表 4 屋齡分佈

## 是否有管理組織

表格 5 管理組織分佈

管理組織	有	無
交易筆數	10100	5439



圖表 5 管理組織分佈

## ANOVA 簡介：

ANOVA，全名為變異數分析 (Analysis of Variance)，是一種統計方法，用來比較三個或三個以上分布的平均數是否存在顯著差異。ANOVA 的核心思想是將總體變異分解成各個分布內部的變異和分布之間的變異，然後比較這兩種變異的大小。

首先，非常重要的，ANOVA 的基本假設包括：

1. 隨機變數(每項變因)相互獨立：ANOVA 假設來自不同分布的樣本是相互獨立的，即來自不同分布的觀察值不受來自其他不同分布的觀察值的影響。
2. 樣本內的觀察值是獨立且隨機抽樣的：ANOVA 假設每個分布的觀察值是獨立且隨機抽樣的，確保分部內差異是真實的而非由抽樣方法引起的。
3. 所有分布的變異數相等 (同方差性)：ANOVA 假設不同分布的總體變異數應該是相等的。同方差性的存在確保 ANOVA 統計的準確性。
4. 觀察值呈常態分佈：ANOVA 假設所有隨機樣本應該都是來自常態分佈。這個假設在樣本容量足夠大時可以忽略，因為中央極限定理可以使二階動差存在的分布都趨近於常態分布。

而 ANOVA 的基本思想是變異數的分解。ANOVA 將總體變異分為組內變異(SSW)和組間變異(SSB)。組內變異度量來自同一分布的隨機樣本(觀察值)之間的差異，而組間變異度量來自不同分布之隨機樣本之間的差異。

在 ANOVA 中，使用 F 分佈進行假設檢定， $H_0$  為所有隨機變數的平均值相等。F 值表示組間變異和組內變異的比例，當組間變異大於組內變異時，F 值將較大，顯示組別之間可能存在顯著差異。

## 多元迴歸簡介：

多元迴歸分析是一種統計方法，用來研究多個變數之間的關係。在統計學中，我們經常想知道一個變數如何受到其他變數的影響。而多元迴歸正是幫助我們解答這個問題的工具。多元迴歸的目的是建立一個數學模型，該模型描述了多個自變數和一個因變數之間的關係。這使我們能夠理解每個自變數對因變數的獨立和共同影響。

舉例來說，如果想知道學生的考試成績（因變數）是如何受到學習時間、睡眠時間和營養攝取（自變數）的影響。透過多元迴歸，我們可以建立一個模型，量化這些變數之間的關係，並預測一個學生的可能考試成績。

另外，在迴歸分析中，解釋力是一個極其重要的概念，它衡量了模型對因變數的變異能力以及模型中自變數的解釋效果。一個具有高解釋力的模型能夠有效地捕捉到樣本數據中的變異，提供對因變數的良好解釋。以下介紹在衡量模型的有效性時，所牽涉到的名詞：

1. R 平方 ( $R^2$ )：R 平方是一個衡量模型解釋力的統計指標。它表示模型所解釋的變異在總變異中的比例。高 R 平方值意味著模型能夠解釋較多的變異，提供較好的解釋力。然而，R 平方也應謹慎使用，因為它可能受到過度擬合的影響。
2. 係數解釋：模型中的迴歸係數提供了每個自變數對因變數的解釋效果。當迴歸係數的絕對值越大，表示相應的自變數對因變數的影響越大，這增加了模型的解釋力。
3. 統計顯著性：自變數的統計顯著性與解釋力直接相關。統計上顯著的自變數對因變數的解釋力較大，而非顯著的自變數則對解釋力的提升有限。
4. 殘差分析：通過對殘差的分析，我們可以評估模型對於未解釋的變異是否存在模式。若殘差呈現隨機分佈，表示模型的解釋力較好。

解釋力在迴歸分析中扮演了至關重要的角色，幫助我們了解模型的可信程度，確保模型的解釋結果是合理而有價值的。良好的解釋力使我們能夠更準確地解釋變數間的關係，並在需要的情況下進行預測。以下將以選定的五個變數，建構房地產價格的多元迴歸模型，並探討其效果。

## 分析結果：

由於 EXCEL 軟體的限制，至多只能選擇 16 筆資料進行迴歸分析，於是將類別資料數值化後，筆者進行了以下幾組分析：

### 1. 單純考慮行政區變因的多元迴歸結果：

表格 6 迴歸統計結果 1

迴歸統計	
R 的倍數	0.459804725
R 平方	0.211420385
調整的 R 平方	0.210796353
標準誤	81465.68406
觀察值個數	15515

表格 7 ANOVA 結果 1

ANOVA					
	自由度	SS	MS	F	顯著值
迴歸	12	2.76E+13	2.3E+12	377.8543	0
殘差	15503	1.03E+14	6.64E+09		
總和	15515	1.3E+14			

表格 8 係數結果 1

係數								
	係數	標準誤	t 統計	P-值	下限 95%	上限 95%	下限 95.0%	上限 95.0%
截距	204202.4	3103.595	65.79543	0	198119	210285.8	198119	210285.8
是否為士林區	-3617.46	3812.325	-0.94888	0.342694	-11090.1	3855.147	-11090.1	3855.147
是否為大同區	-992.879	4347.572	-0.22838	0.819357	-9514.63	7528.87	-9514.63	7528.87
是否為大安區	98780.06	3698.52	26.708	1.1E-153	91530.52	106029.6	91530.52	106029.6

是否為中山區	21625.41	3506.036	6.168051	7.09E-10	14753.17	28497.65	14753.17	28497.65
是否為中正區	61784.78	4177.24	14.79082	3.63E-49	53596.91	69972.66	53596.91	69972.66
是否為內湖區	-12726.7	3653.524	-3.48341	0.000496	-19888	-5565.38	-19888	-5565.38
是否為文山區	-36581	3854.077	-9.4915	2.6E-21	-44135.4	-29026.5	-44135.4	-29026.5
是否為北投區	-34050.7	3844.086	-8.85793	9.02E-19	-41585.5	-26515.8	-41585.5	-26515.8
是否為松山區	47239.08	4039.275	11.69494	1.83E-31	39321.63	55156.54	39321.63	55156.54
是否為信義區	44348.03	3856.751	11.49881	1.78E-30	36788.35	51907.71	36788.35	51907.71
是否為南港區	0	0	65535	#NUM!	0	0	0	0
是否為萬華區	-38663	3914.603	-9.8766	#NUM!	-46336	-30989.9	-46336	-30989.9

### 結果解釋：

- i. **迴歸統計**:R 的倍數為 0.4598。這表示模型中的自變數能夠解釋因變數約 46%的變異。R 平方為 0.2114，表示模型能夠解釋因變數變異的 21%。
- ii. **迴歸**:顯示模型中的變數是否統計顯著。F 統計量為 377.85，p 值為 0，顯示至少有一個自變數對因變數有統計上的顯著影響。殘差值為 1.03E+14 及 6.64E+09，顯示有許多未解釋的變異。
- iii. **係數**:是否為南港區及是否為萬華區係數的 P 值顯示錯誤，已知並非由資料缺失所導致，其原因不明。是否為士林區及是否為大同區的 P 值大於選定的信賴水準(0.05)，代表其係數在統計上並不顯著。

## 2. 考慮行政區變因及土地類別變因的多元迴歸結果：

表格 9 迴歸統計結果 2

迴歸統計	
R 的倍數	0.471269
R 平方	0.222095

調整的 R 平方	0.221327
標準誤	80920.28
觀察值個數	15515

表格 10 ANOVA 結果 2

ANOVA					
	自由度	SS	MS	F	顯著值
迴歸	15	2.89773E+13	1.93182E+12	316.0928858	0
殘差	15500	1.01495E+14	6548091514		
總和	15515	1.30473E+14			

表格 11 係數結果 2

係數								
	係數	標準誤	t 統計	P-值	下限 95%	上限 95%	下限 95.0%	上限 95.0%
截距	158207	57347.93	2.758721	0.00581	45798.3	270615.6	45798.3	270615.6
是否為士林區	6054.304	3798.141	1.594018	0.110952	-1390.5	13499.11	-1390.5	13499.11
是否為大同區	0	0	65535	#NUM!	0	0	0	0
是否為大安區	103378.2	3638.676	28.41094	#NUM!	96245.95	110510.4	96245.95	110510.4
是否為中山區	21366.82	3432.512	6.224836	4.94E-10	14638.7	28094.95	14638.7	28094.95
是否為中正區	64684.06	4109.15	15.74147	2.11E-55	56629.65	72738.47	56629.65	72738.47
是否為內湖區	-967.6	3656.337	-0.26464	0.791293	-8134.45	6199.249	-8134.45	6199.249
是否為文山區	-29314	3821.744	-7.67033	1.82E-14	-36805.1	-21823	-36805.1	-21823
是否為北投區	-23321.5	3839.208	-6.07455	1.27E-09	-30846.8	-15796.2	-30846.8	-15796.2
是否為松山區	50652.27	3973.101	12.7488	4.86E-37	42864.53	58440.01	42864.53	58440.01
是否為信義區	49850.6	3804.25	13.10392	5.04E-39	42393.82	57307.37	42393.82	57307.37

是否為南港區	13453.7	4402.554	3.055886	0.002248	4824.181	22083.22	4824.181	22083.22
是否為萬華區	-36713.1	3842.969	-9.55331	1.44E-21	-44245.7	-29180.4	-44245.7	-29180.4
是否為工用地	-5168.63	57434.69	-0.08999	0.928295	-117747	107410.1	-117747	107410.1
是否為住用地	35167.87	57264.41	0.614131	0.539138	-77077.1	147412.8	-77077.1	147412.8
是否為商用地	52054.26	57278.96	0.908785	0.363478	-60219.2	164327.7	-60219.2	164327.7

### 結果解釋：

- i. **迴歸統計：**R 的倍數為 0.4713。這表示模型中的自變數能夠解釋因變數約 47.13% 的變異。R 平方為 0.2221，表示模型能夠解釋因變數變異的 22.21%。
- ii. **ANOVA：**顯示模型中的變數是否統計顯著。F 統計量為 316.09，p 值為 0，顯示至少有一個自變數對因變數有統計上的顯著影響。殘差值為 1.01495E+14 及 6548091514，過大的殘差值顯示有許多未解釋的變異。
- iii. **係數：**是否為士林區及是否為內湖區的係數，P 值都大於 0.05，因此係數在此信賴水準下不顯著。是否為大同區的係數為 0，代表其對因變數的影響並不顯著。另外在工用地、住用地、商用地中，P 值都較大，代表其在統計上不顯著

### 3. 考慮行政區變因及房屋種類變因的多元迴歸結果：

表格 12 迴歸統計結果 3

迴歸統計	
R 的倍數	0.540135
R 平方	0.291746
調整的 R 平方	0.290996
標準誤	77215.17
觀察值個數	15515

表格 13 ANOVA 結果 3

ANOVA					
	自由度	SS	MS	F	顯著值
迴歸	16	3.80648E+13	2.37905E+12	425.6252574	0
殘差	15499	9.24079E+13	5962182133		
總和	15515	1.30473E+14			

表格 14 係數結果 3

係數								
	係數	標準 誤	t 統 計	P-值	下限 95%	上限 95%	下限 95.0%	上限 95.0%
截距	120313.8	77300	1.556453	0.119621	-31203.2	271830.9	-31203.2	271830.9
是否為士林區	-658.83	3620.564	-0.18197	0.85561	-7755.56	6437.9	-7755.56	6437.9
是否為大同區	-9329.6	4127.507	-2.26035	0.023814	-17420	-1239.2	-17420	-1239.2
是否為大安區	90517.34	3515.549	25.74771	3.4E-143	83626.46	97408.23	83626.46	97408.23
是否為中山區	12405.86	3334.928	3.719979	0.0002	5869.013	18942.71	5869.013	18942.71
是否為中正區	55212.26	3964.072	13.92817	7.85E-44	47442.21	62982.3	47442.21	62982.3
是否為內湖區	-13473.4	3465.073	-3.88836	0.000101	-20265.4	-6681.5	-20265.4	-6681.5
是否為文山區	-36271	3653.681	-9.92724	3.72E-23	-43432.6	-29109.3	-43432.6	-29109.3
是否為北投區	-31643.2	3649.201	-8.67128	4.69E-18	-38796.1	-24490.4	-38796.1	-24490.4
是否為松山區	36664.41	3840.105	9.547763	1.52E-21	29137.36	44191.47	29137.36	44191.47
是否為信義區	44477.63	3656.001	12.16565	6.76E-34	37311.44	51643.82	37311.44	51643.82
是否為南港區	0	0	65535	#NUM!	0	0	0	0
是否為萬華區	-41050.3	3712.428	-11.0575	#NUM!	-48327.1	-33773.5	-48327.1	-33773.5
是否為公寓	51645.75	77247.6	0.668574	0.503777	-99768.6	203060.1	-99768.6	203060.1
是否為住宅大樓	109012.9	77249.65	1.411176	0.158213	-42405.5	260431.2	-42405.5	260431.2
是否為透天厝	183705	77412.92	2.373053	0.017654	31966.58	335443.3	31966.58	335443.3
是否為華夏	88208.09	77251.52	1.14183	0.253542	-63213.9	239630.1	-63213.9	239630.1

結果解釋：



- i. **迴歸統計**: R 的倍數為 0.540135。這表示模型解釋了因變數變異性的 54.01%。調整的 R 平方為 0.290996，表示模型能夠解釋因變數變異的 29.09%。
- ii. **ANOVA**: 顯示模型中的變數是否統計顯著。F 統計量為 425.6252574，p 值為 0，顯示至少有一個自變數對因變數有統計上的顯著影響。殘差值為 9.24079E+13 及 5962182133，過大的殘差值顯示有許多未解釋的變異。
- iii. **係數**: 是否為士林區的係數，P 值大於 0.05，因此係數在此信賴水準下不顯著。是否為南港區的係數為 0，代表其對因變數的影響並不顯著。另外在房屋種類的變因中，僅有是否為透天厝係數的 P 值達到顯著，其餘三者皆不顯著。

#### 4. 考慮屋齡變因及管理狀況變因的多元迴歸結果:

表格 15 迴歸統計結果 4

迴歸統計	
R 的倍數	0.364494
R 平方	0.132856
調整的 R 平方	0.1324
標準誤	85416.44
觀察值個數	15515

表格 16 ANOVA 結果 4

ANOVA					
	自由度	SS	MS	F	顯著值
迴歸	8	1.73E+13	2.17E+12	339.4067	0
殘差	15507	1.13E+14	7.3E+09		
總和	15515	1.3E+14			

表格 17 係數結果 4

係數								
	係數	標準誤	t 統計	P-值	下限 95%	上限 95%	下限 95.0%	上限 95.0%
截距	173687.9	2551.064	68.08448	0	168687.5	178688.3	168687.5	178688.3
屋齡 10 年以下	71264.54	2674.496	26.64597	5.4E-153	66022.21	76506.86	66022.21	76506.86
屋齡 11~20 年	43551.31	2557.51	17.02879	1.93E-64	38538.29	48564.33	38538.29	48564.33
屋齡 21~30 年	0	0	65535	#NUM!	0	0	0	0

屋齡 31~40 年	4547.913	2554.859	1.780103	#NUM!	-459.91	9555.737	-459.91	9555.737
屋齡 41~50 年	3878.656	2480.974	1.56336	0.117988	-984.343	8741.656	-984.343	8741.656
屋齡 51~60 年	40415.07	3406.067	11.86561	2.46E-32	33738.78	47091.36	33738.78	47091.36
屋齡 60 年以上	145680.6	11796.69	12.34928	7.17E-35	122557.7	168803.5	122557.7	168803.5
有無管理組織	32455.75	1821.388	17.81924	2.52E-70	28885.61	36025.88	28885.61	36025.88

### 結果解釋：

- i. **迴歸統計:**R 的倍數為 0.364494，表示模型解釋了因變數變異性的 36.45%。調整的 R 平方為 0.1324，表示模型能夠解釋因變數變異的 13.24%。
- ii. **ANOVA:**F 統計量為 339.4067，P 值為 0，顯示至少有一個自變數對因變數有統計上的顯著影響。殘差的平方和為 1.13E+14，殘差的平均平方為 7.3E+09，顯示有相當的未解釋變異存在。
- iii. **係數:**屋齡 41~50 年的係數，P 值大於 0.05，因此係數在此信賴水準下不顯著。屋齡 21~30 年的係數為 0，代表其對因變數的影響並不顯著。其餘變數，除屋齡 31~40 年出現錯誤外，在此信賴水準下皆為顯著。

## 三、研究限制

本次研究受限於 EXCEL 軟體本身的功能，並不能夠探討太過複雜的模型，以及篇幅限制，無法進行太過複雜的研究。亦受限於資料的真實性(如引言所述)。

## 四、結論

### 四組隨機變數組合的多元迴歸模型的比較：

在這四組分析之中，無一例外，ANOVA 的殘差值皆非常大。一般來說，在多元迴歸的模型中，ANOVA 殘差值很大時，會影響迴歸模型的結果，並且被視為一種不良的情況。殘差值大表示模型無法解釋的變異性很高，這可能來自未考慮的重要變數、模型規範不當、或者樣本中存在異常值等因素。具體的影響可能包括：模型解釋力下降、係數估計不準確、模型預測不可靠等因素。

首先，前三組皆牽涉到區域變數的變因，而其迴歸模型的解釋力(調整的 R 平方)依序為：0.210796353、0.221327398、及 0.290995599。而最後一組迴歸模型並未牽涉到區域變數，其迴歸模型的解釋力為：0.13240028。可見存在著解釋力普遍不足的情形，在這其中，有考慮區域變數的

模型其解釋力要明顯高於未考慮區域變數的模型，在這其中，又以區域變數與房屋種類變數的模型來的最高，具有大約 30% 的解釋力。

## 過大的 ANOVA 殘差值：

在先前曾提過，ANOVA 的基本假設為隨機變數(每項變因)相互獨立、樣本內的觀察值是獨立且隨機抽樣的、所有分布的變異數相等（同方差性）、觀察值呈常態分佈。

首先，考慮隨機變數(每項變因)相互獨立的假設，在未數值化類別資料前，應當先對各個變因進行卡方獨立性檢驗，確保看似無關聯的變因是真的並無關連。另外，在數值化類別資料後，顯然區域變數(是否為士林區、是否為大同區…)之間是彼此不獨立的，因這些區域變數服從多項分布。應當選擇更加正確的類別資料數值化方法，而非單純考慮二項分布。

再者，考慮所有分布的變異數相等的假設。顯然對於數值化後的類別資料，應當檢查其變異數是否存在顯著區別，因為分部相對單純(二項分布)，可以使用點估計或者區間估計。若變異數存在顯著差異，則應當使用其他方法調整變異數。

最後，考慮觀察值呈常態分佈的假設。根據中央極限定理，當樣本數夠大時，來自二項分布的樣本和會服從常態分布。並非來自二項分布的隨機樣本會服從常態分布。所以很顯然地，該假設並未成立。

由於選擇了不正確的類別資料數值化方法，以及以上的原因，導致了 ANOVA 有過大的殘差值，以及後續迴歸模型的不完善。除了以上原因外，由於房地產本身就是一項極為複雜的研究項目，提出完善迴歸模型的難度也高，因為本次報告篇幅有限，暫時不再繼續深究。