

## 西雅圖（金郡）房價分析

劉羿暄

# 目錄

目錄.....	2
表目錄.....	5
圖目錄.....	6
第壹章 緒論.....	7
第一節 前言.....	7
第二節 研究目的與動機.....	7
第三節 研究對象.....	8
第四節 研究方法.....	8
第五節 研究架構流程.....	9
第二章 基本資料分析.....	10
第一節 基本敘述統計量.....	10
一、房價(Y).....	10
二、臥室數量(X1).....	13
三、浴室數量(X2).....	14
四、居住面積(X3).....	15
五、土地面積(X4).....	16
六、地上居住面積(X5).....	17
七、建築年分(X6).....	18
八、最近 15 棟房屋的內部居住空間平均大小(X7).....	19
九、最近 15 棟房屋的土地平均大小(X8).....	20
第二節 CORRELATION.....	21
第三節 Variance Inflation Factor(VIF).....	23
第參章 原始模型檢定.....	24
第一節 建立迴歸模型.....	24
第二節 單一參數 t 檢定.....	25
一、 $\beta_1$ 之 t 檢定.....	26
二、 $\beta_2$ 之 t 檢定.....	26
三、 $\beta_3$ 之 t 檢定.....	26
四、 $\beta_4$ 之 t 檢定.....	27
五、 $\beta_5$ 之 t 檢定.....	27
六、 $\beta_6$ 之 t 檢定.....	28
七、 $\beta_7$ 之 t 檢定.....	28

八、 $\beta_8$ 之t檢定 .....	28
第三節 模型適合度檢定 .....	29
第四節 模型解釋能力 .....	30
第肆章 模型的選取方法 .....	31
第一節 向前選取法(Forward) .....	31
第二節 後退刪去法(Backward) .....	33
第三節 逐步迴歸法 .....	35
第四節 其他選取法 .....	40
一、複判定係數法( $R^2$ ) .....	53
二、 $C_p$ 準則 .....	53
三、SBC、AIC 法 .....	54
第五節 結論 .....	54
第伍章 離群值及影響點之檢定 .....	55
第一節 離群值 .....	55
一、標準化殘差值 .....	55
二、Y outlier .....	56
三、X outlier .....	56
第二節 影響點 .....	57
一、DFFITS .....	57
二、COOK'S D .....	59
三、DFBETAS .....	59
第三節 結論 .....	59
第陸章 殘差檢定 .....	61
第一節 均質性 .....	61
一、White test .....	61
二、殘差分析圖 .....	62
第二節 常態性 .....	65
一、常態性檢定 .....	65
二、常態圖 .....	66
第三節 獨立性 .....	67
一、一階自我相關檢定 .....	67
第柒章 模型確認 .....	68
第一節 最終模型解釋能力 .....	68
第二節 最終模型預測能力 .....	69

一、MAPE.....	69
二、最終模型的應用.....	70
第捌章 結論.....	70
第玖章 附錄.....	72
第一節 資料.....	72
一、原始資料(前二十筆).....	72
二、選取資料(前二十筆).....	72
三、訓練資料.....	73
四、確認資料.....	80
第二節 程式碼.....	85
一、SAS.....	85
二、R.....	90
第三節 資料來源.....	91

## 表目錄

表 2-1 簡單統計值 .....	10
表 2-2 log-likelihood 最大概似 $\lambda$ .....	12
表 2-3 各變數的相關性 .....	21
表 2-4 各變異數的 VIF .....	23
表 3-1 參數估計 .....	25
表 3-2 變異數分析 .....	29
表 3-3 模型解釋能力分析 .....	30
表 4-1 向前選取法 .....	31
表 4-2 參數估計 .....	33
表 4-3-1 後退刪去法 .....	33
表 4-3-2 後退刪去法 .....	34
表 4-3-3 後退刪去法 .....	34
表 4-4 參數估計 .....	34
表 4-5-1 變異數分析與參數估計 .....	35
表 4-5-2 變異數分析與參數估計 .....	35
表 4-5-3 變異數分析與參數估計 .....	36
表 4-5-4 變異數分析與參數估計 .....	37
表 4-5-5 變異數分析與參數估計 .....	37
表 4-5-6 變異數分析與參數估計 .....	38
表 4-6 逐步迴歸法摘要 .....	39
表 4-7 其他選取法 .....	40
表 4-8 模型選取方法與結論 .....	54
表 5-1 離群值分析 .....	55
表 5-2 帽子矩陣分析 .....	56
表 5-3 DFFITS 影響力分析 .....	57
表 5-4 離群值與影響點分析之結論 .....	59
表 5-5 離群值與影響點分析之結論 .....	60
表 6-1 均質性檢定 .....	61

表 6-2 分位數殘差值 .....	64
表 6-3 常態性檢定 .....	65
表 6-4 獨立性檢定 .....	67
表 7-1 配適模型解釋能力分析 .....	68
表 7-2 原始模型與配適模型解釋能力之比較.....	68
表 7-3 最終模型 ANOVA 表 .....	69
表 7-4 確認資料集 ANOVA 表 .....	69

## 圖目錄

圖 1-1 研究流程 .....	9
圖 2-1 房價相對直方圖 .....	10
圖 2-2 Box-Cox 分析 .....	11
圖 2-3 Box-Cox 轉換後房價相對直方圖 .....	12
圖 2-4 臥室數量相對直方圖 .....	13
圖 2-5 浴室數量相對直方圖 .....	14
圖 2-6 居住面積相對直方圖 .....	15
圖 2-7 土地面積相對直方圖 .....	16
圖 2-8 取對數後的土地面積相對直方圖 .....	16
圖 2-9 地上居住面積相對直方圖 .....	17
圖 2-10 建築年份相對直方圖 .....	18
圖 2-11 最近 15 棟房屋的內部居住空間平均大小相對直方圖.....	19
圖 2-12 最近 15 棟房屋的土地平均大小相對直方圖.....	20
圖 2-13 取對數後的最近 15 棟房屋的土地平均大小相對直方圖.....	20
圖 6-1 原始模型 Residuals vs. Fitted plot.....	62
圖 6-2 原始模型 Standardized Residuals vs. Fitted plot.....	62
圖 6-3 配適模型 Residuals vs. Fitted plot.....	63
圖 6-4 配適模型 Standardized Residuals vs. Fitted plot.....	63
圖 6-5 Q-Q plot-1 .....	66
圖 6-6 Q-Q plot-2.....	66

# 第壹章 緒論

## 第一節 前言

在社會發展的浪潮中，房價不僅僅是經濟活動中的一項重要指標，更關係到人們的居住狀態、生活品質，以及整個社會的穩定。近年來，台灣媒體對房價上升的報導常常佔據著重要的版面，其複雜性涉及多方面因素，包括政府政策、建築成本上升、房屋狀況、地理位置等，這些元素相互交織，共同構成了房地產市場的獨特格局。

在太平洋另一端的美國，作為全球經濟的重要角色，其房價以經濟活動旺盛的地區波動較為明顯。以西雅圖為例，這座城市位於美國西北部臨太平洋，以其優秀的地理位置和科技創新脫穎而出，在全球佔有一席之地。不僅擁有世界知名的科技巨頭總部，還孕育著眾多嶄露頭角的初創企業，使房價相對較高。對於購房者來說，擁有一套房產既是享受城市繁榮的機會。然而，這種高度的需求壓力也帶來房價上升的問題。

購房者在投資房屋時需考慮多方面因素，在同一城市，政策與經濟相狀況近的地區，若僅討論房屋條件對房價的影響，有哪些關鍵因素呢？本研究將以線性迴歸方法深入探究。

## 第二節 研究目的與動機

本研究的主要目的為分析影響西雅圖金郡房價的房屋因素，以臥室數量、浴室數量、居住面積、土地面積、地上居住面積、建築年份、以及最近 15 棟房屋的內部居住空間和最近 15 棟房屋的土地平均大小等多個變數進行分析。並進一步了解這些變數之間的關聯性，建立更全面的模型，進而準確地預測金郡房價。

而更深遠的意涵，則是希望能藉此了解構成高房價的主要原因，若能發現哪些選屋條件對房價有影響，我們不僅能夠把握當前的市場狀況，更能為未來的投資與居住規劃提供更具參考價值的資訊。

### 第三節 研究對象

隨著科技產業的蓬勃發展，金郡作為西雅圖的主要發展區域，不僅成為眾多企業的選址首選，更吸引了許多專業人才的湧入。其發展驅動了整個城市的經濟成長和就業機會的增多，帶動了人們對房屋的需求，使得金郡的房價一直處於市場的關注焦點。

因此，我們選取金郡的 500 筆房屋資料（房價、臥室數量、浴室數量、居住面積、土地面積、地上居住面積、建築年份、以及最近 15 棟房屋的內部居住空間和最近 15 棟房屋的土地平均大小），來進行研究。

### 第四節 研究方法

我們於 Kaggle 找尋金郡 2014 到 2015 年房價及影響要素的資料集，從中保留八個數值型變數作為研究變數。隨後，從 21613 筆資料以 R 程式隨機抽取 350 筆作為訓練集，再從剩餘資料中抽取 150 筆做為確認資料集，即隨機選取 500 筆資料以 7:3 分為兩份做分析與比較。（資料見第玖章 附錄）

首先，由訓練資料建模，分析其基本資料，檢測相關性以及是否符合各種前提假設，並對模型做適合度檢定，而得到初步的迴歸模型。接著，篩選掉不合適的變數，並將具離群值及影響點特性的資料刪除，得到最適迴歸模型。隨後，分析殘差值是否符合同質性、常態性與獨立性假設，並將此模型套入確認資料中，檢驗其平均絕對誤差(MAPE)，判斷預測能力的高低，以確認最終模型。



## 第五節 研究架構流程

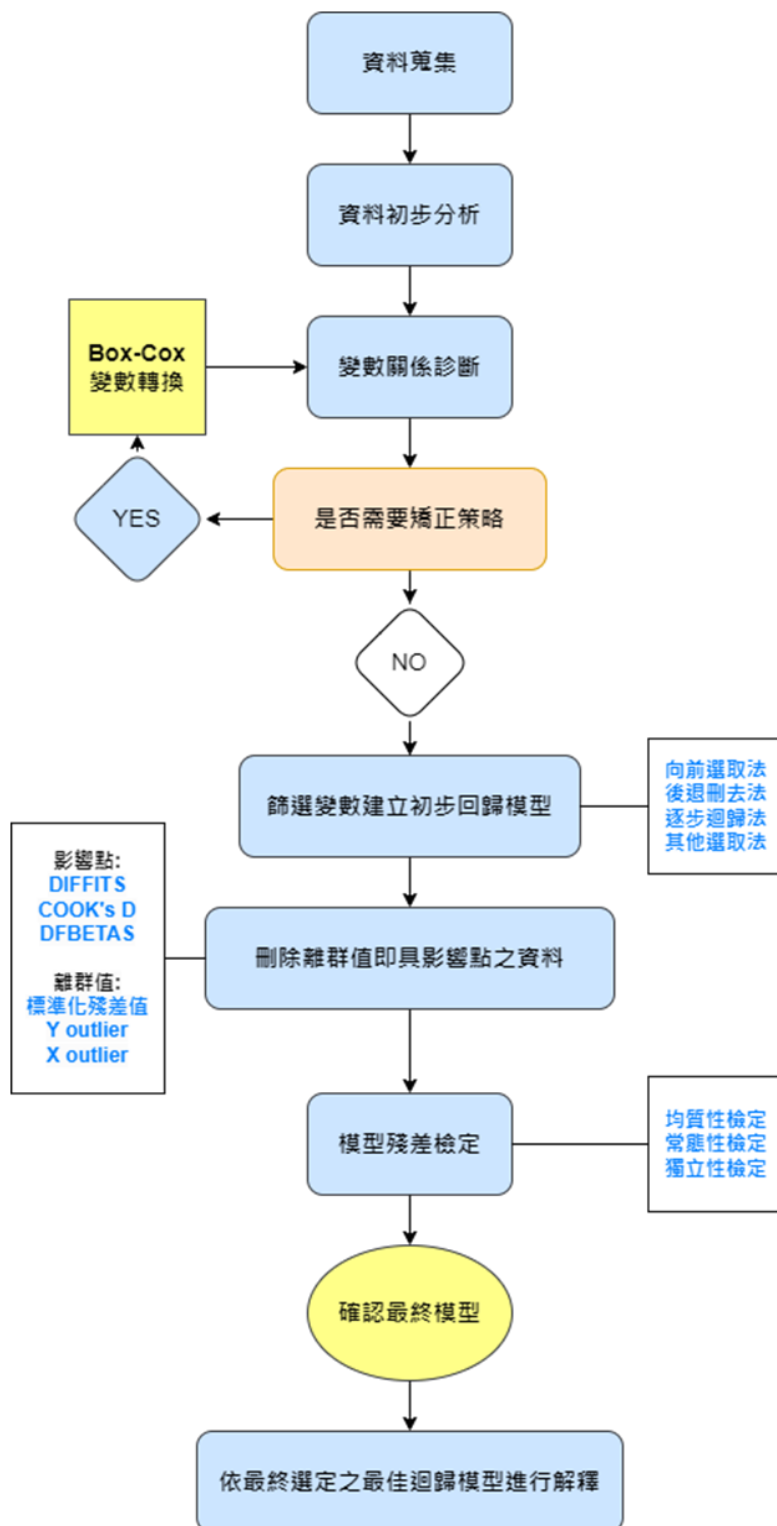


圖 1-1 研究流程

## 第二章 基本資料分析

### 第一節 基本敘述統計量

表 2-1 簡單統計值

變數	標籤	N	平均值	標準差	最小值	最大值
Ynew	房價	350	8.5735603	0.2012002	8.0713421	9.27025
x1	臥室數量	350	3.4314286	0.954336	1	8
x2	浴室數量	350	2.1628571	0.760254	0.75	5
x3	居住面積	350	2167.66	942.7399989	610	5774
x4	土地面積	350	15520.31	33766.26	867	322188
x5	地上居住面積	350	1875.79	858.8765121	610	5010
x6	建築年份	350	1971.41	29.8699631	1901	2015
x7	最近 15 棟房屋的內部居住空間平均大小	350	2018.01	675.847584	840	4760
x8	最近 15 棟房屋的土地平均大小	350	12127.57	21357.06	1021	207781

(資料見第玖章 附錄)

#### 一、房價(Y)

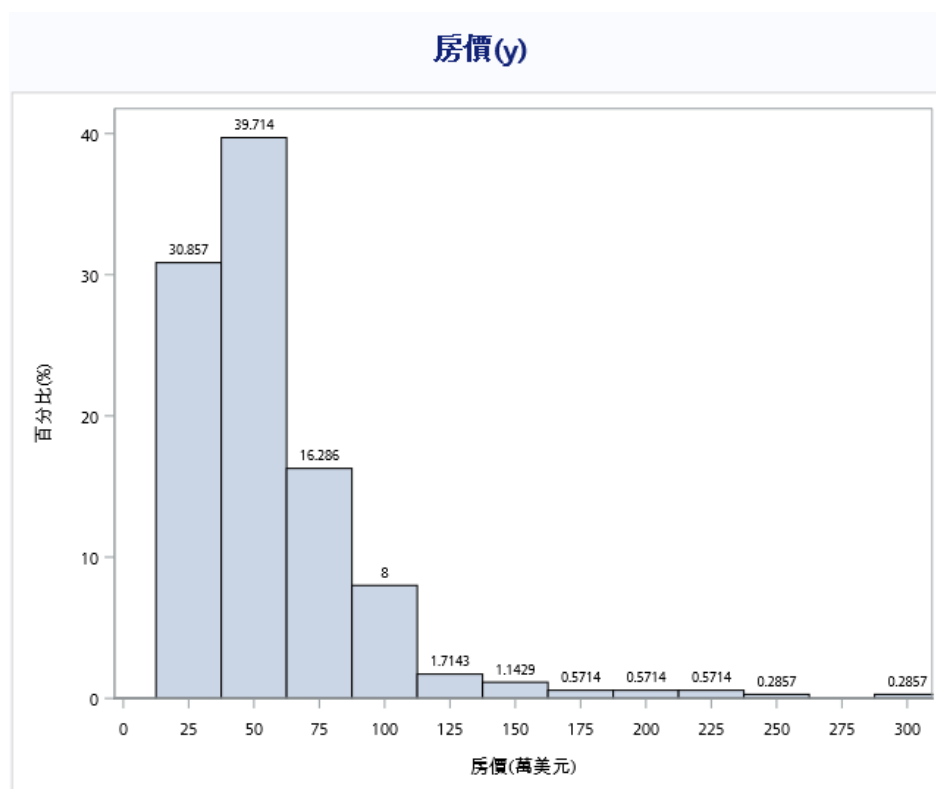


圖 2-1 房價相對直方圖

透過圖 2-1 可以了解到：

房價主要落在 50 萬美元左右(37.5~62.5 萬)，佔所有房屋 40%。由直方圖可見整體呈右偏分布，故我們以 Box-Cox 轉換選取 log-likelihood 之最大  $\lambda$ ，使  $Y \rightarrow Y_{\text{new}}$ 。

轉換公式：

$$Y_{\text{new}}(\lambda) = \begin{cases} \frac{(Y^\lambda - 1)}{\lambda} & , \text{ if } \lambda \neq 0 \\ \log(Y) & , \text{ if } \lambda = 0 \end{cases}$$

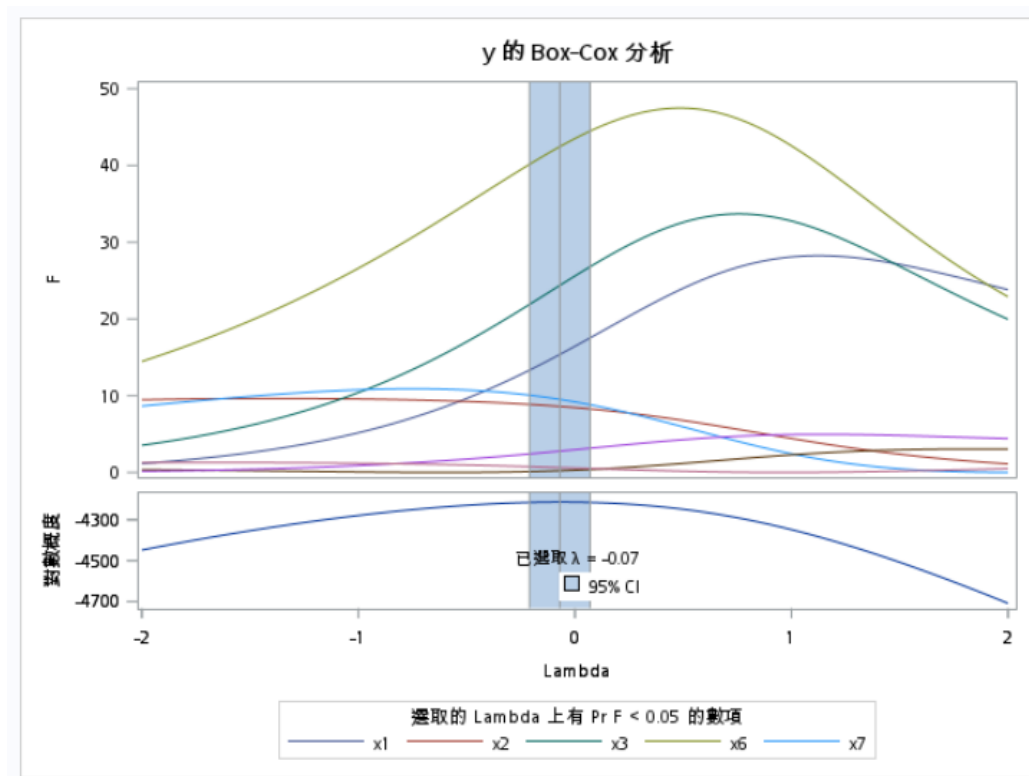


圖 2-2 Box-Cox 分析

表 2-2 log-likelihood 最大概似  $\lambda$

模型陳述式規格詳細資料				
類型	DF	變數	描述	值
Dep	1	BoxCox(y)	使用的 Lambda	-0.07
			Lambda	-0.07
			對數概度	-4213.3
			Conv. Lambda	0
			Conv. Lambda LL	-4213.7
			CI 界限	-4215.2
			Alpha	0.05
Ind	1	Identity(x1)	DF	1
Ind	1	Identity(x2)	DF	1
Ind	1	Identity(x3)	DF	1
Ind	1	Identity(x4)	DF	1
Ind	1	Identity(x5)	DF	1
Ind	1	Identity(x6)	DF	1
Ind	1	Identity(x7)	DF	1
Ind	1	Identity(x8)	DF	1

根據上表 2-2，求得最適  $\lambda$  值為 -0.07，即  $Y_{\text{new}} = \frac{(Y^{-0.17}-1)}{-0.17}$

轉換後  $Y_{\text{new}}$  如圖 2-3 所示，呈現常態分配。

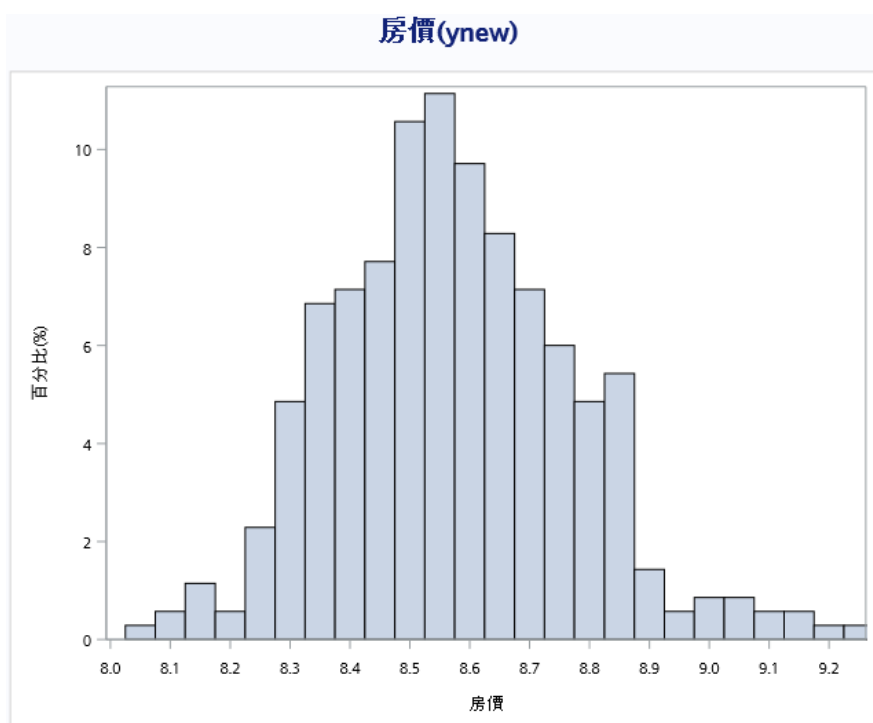


圖 2-3 Box-Cox 轉換後房價相對直方圖

## 二、臥室數量( $X_1$ )

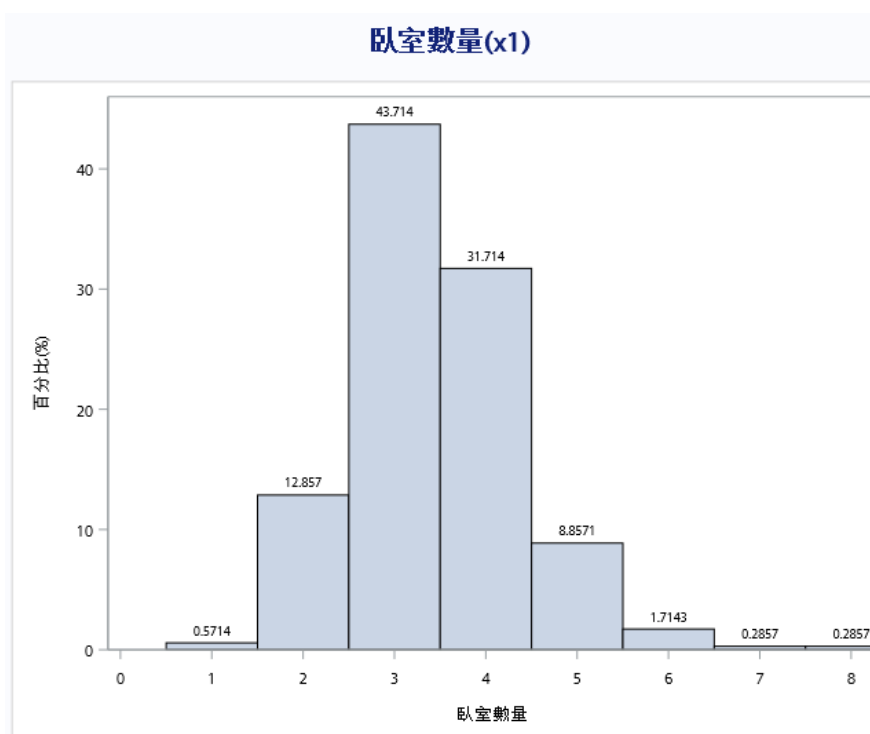


圖 2-4 臥室數量相對直方圖

透過圖 2-4 可以了解到：

臥室數量分布在 1 到 8 間，大部分的房屋擁有 3 間臥室(約占 44%)，其次是 4 間臥室(約占 31%)。

### 三、浴室數量( $X_2$ )

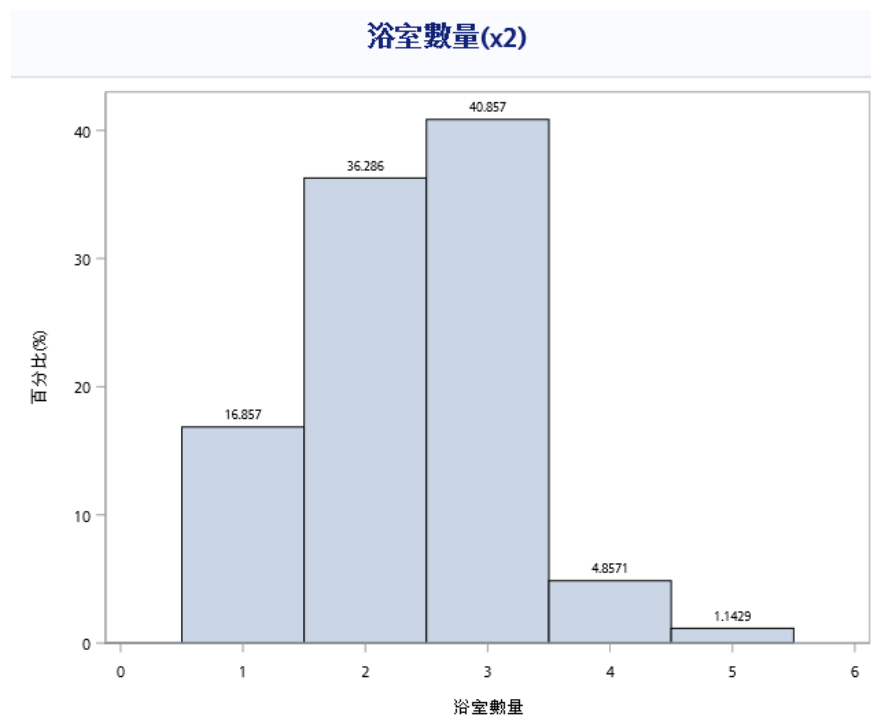


圖 2-5 浴室數量相對直方圖

透過圖 2-5 可以了解到：

大多數房屋擁有 2 到 3 間浴室，約占整體的 77%，也有少部分房屋具有 4 到 5 間浴室。

#### 四、居住面積(X<sub>3</sub>)

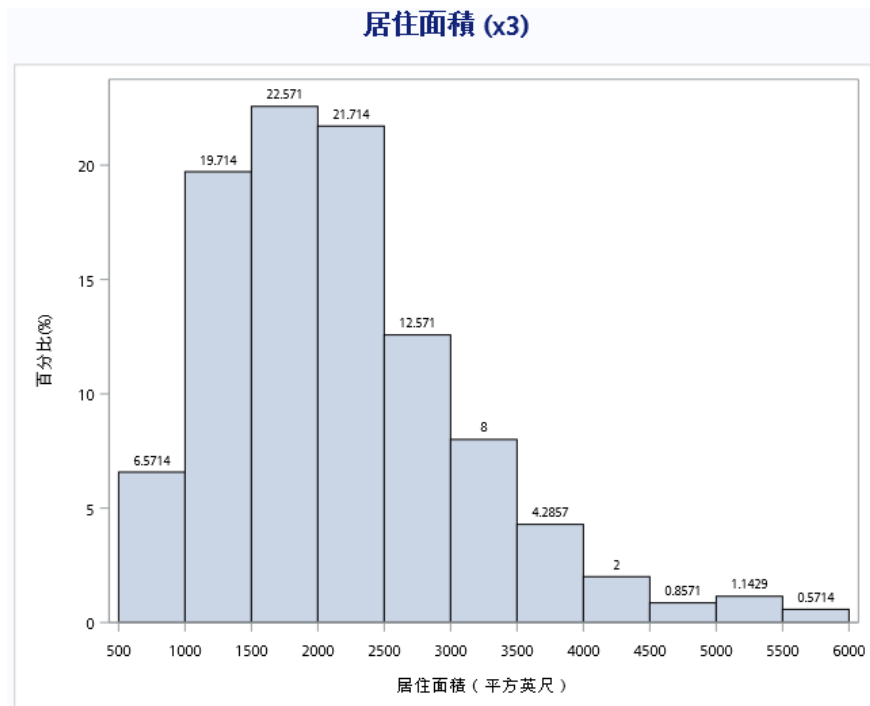


圖 2-6 居住面積相對直方圖

居住面積最小值為 610 平方英尺，最大值為 5774 平方英尺。由圖 2-6 可見，居住面積呈右偏分布，大部分落在 1000 至 2500 平方英尺，其中以 1500 到 2000 區間最多，約占 23%。

## 五、土地面積(X<sub>4</sub>)

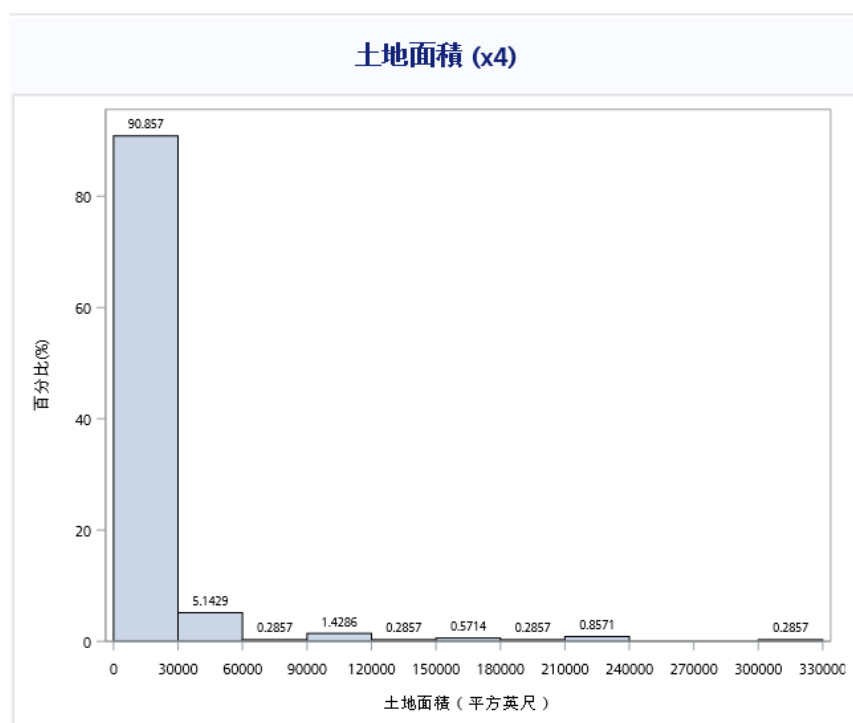


圖 2-7 土地面積相對直方圖

由圖 2-7 發現超過 90% 的土地面積小於 3 萬平方公尺，因為有離群值且全距很大(最大值為 322188，最小值為 867)，故難以由上圖看出確切分布。

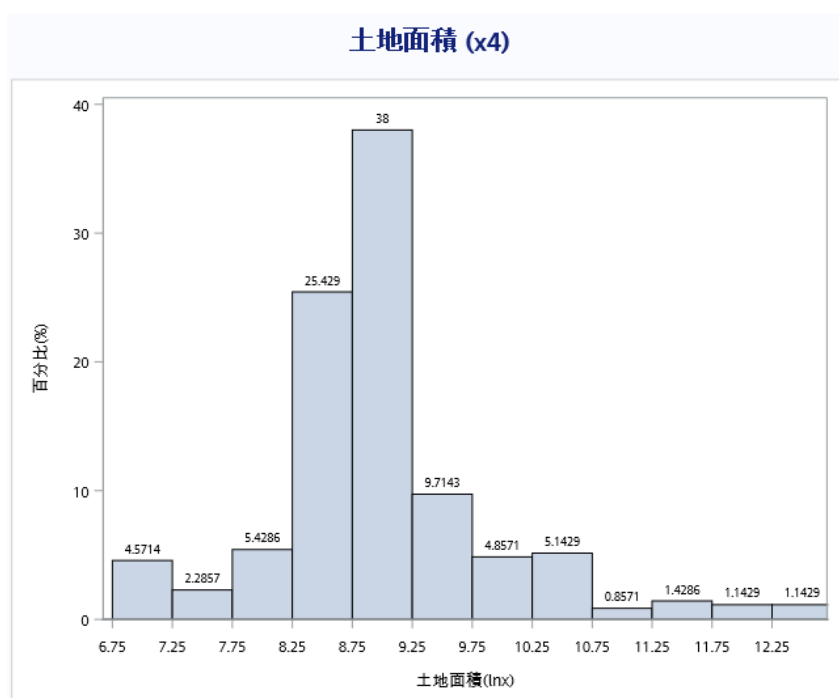


圖 2-8 取對數後的土地面積相對直方圖



由圖 2-8 可看出取對數後的分布，以 8.75 至 9.25 占 38% 最多，即土地面積大多介於 6310 到 10405 平方英尺。然而，中位數為 7500 也落在此區間。次為 8.25 至 8.75 相當於 3827 到 6310 平方英尺，約占 25%。

$$(e^{8.25} \approx 3827.63, e^{8.75} \approx 6310.69, e^{9.25} \approx 10404.57)$$

## 六、地上居住面積( $X_5$ )

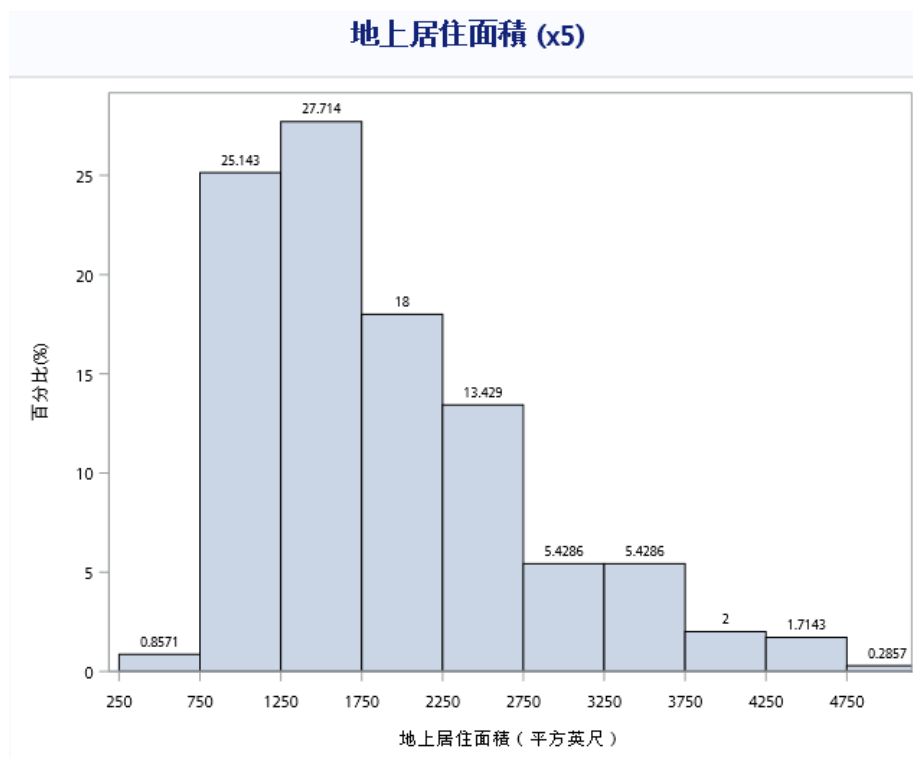


圖 2-9 地上居住面積相對直方圖

地上居住面積最小值為 610，最大值為 5010，其值與居住面積相近且都呈現右偏分配，但由於兩者的集中趨勢和變異皆有所差異，故仍呈現不同分布情形。透過圖 2-9 可見，大部分面積落在 750 至 1750 平方英尺，其中以 1250 到 1750 區間最多，約占 28%。

## 七、建築年分(X<sub>6</sub>)

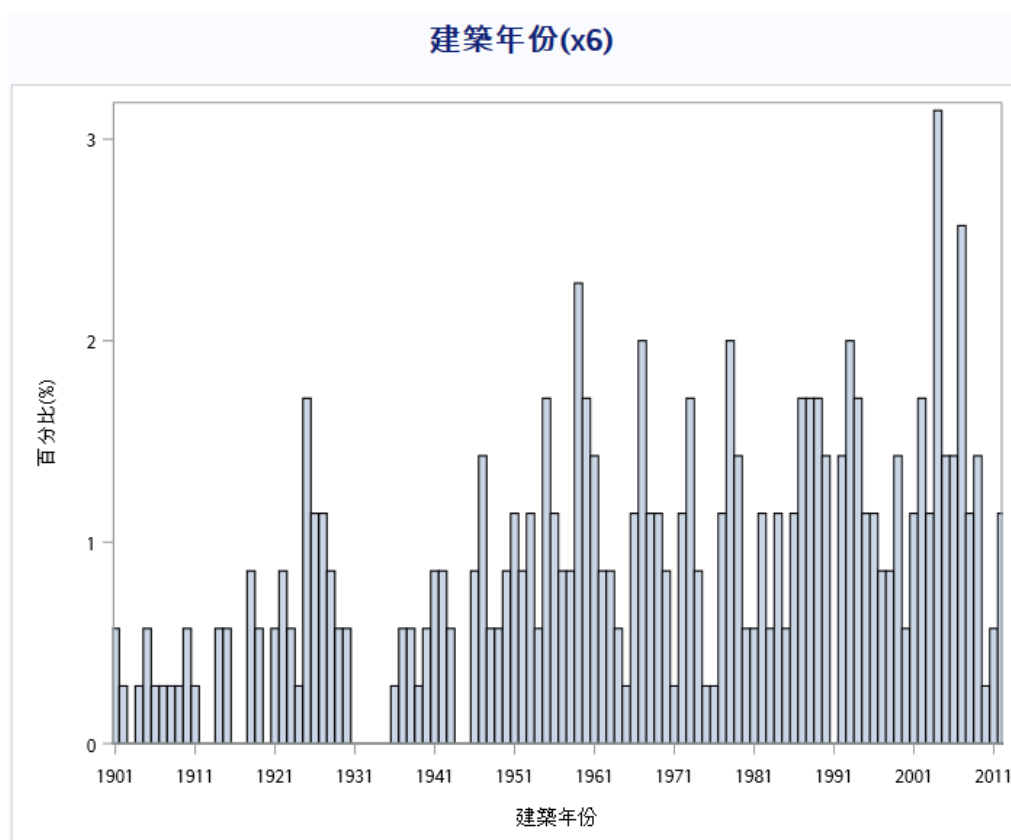


圖 2-10 建築年份相對直方圖

x6	次數	百分比
1959	8	2.29
2004	11	3.14
2007	9	2.57
2014	10	2.86

建築年份為房屋建造的年份，最大屋齡為 1901 年，最小則為 2015 年。透過圖 2-10 可大致看出分布。由上表可迅速看出哪些年份的房屋數量較多，占比較高的為 1959、2004、2007 及 2014 年。

## 八、最近 15 棟房屋的內部居住空間平均大小( $X_7$ )

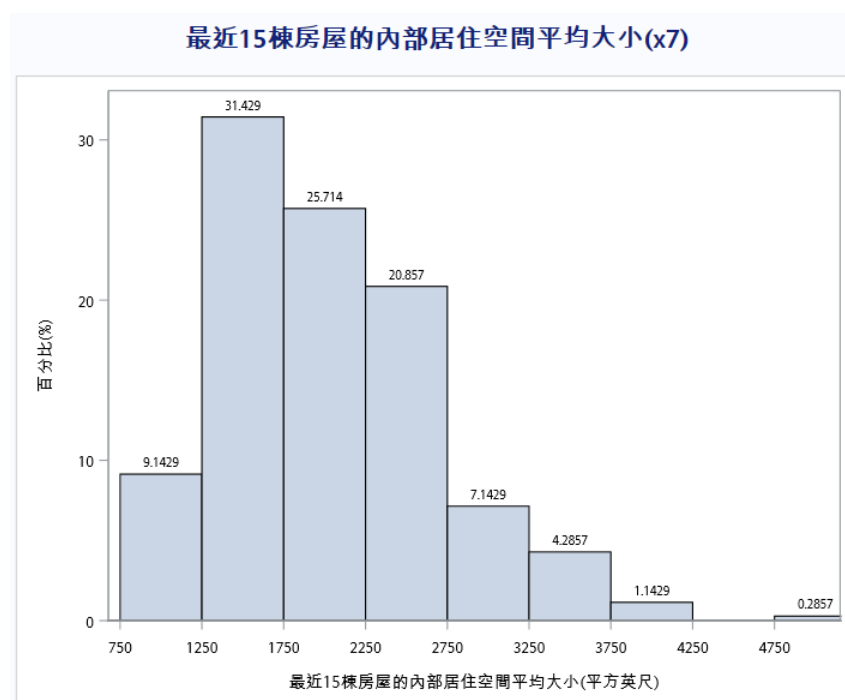


圖 2-11 最近 15 棟房屋的內部居住空間平均大小相對直方圖

透過圖 2-11 可以了解到離目標房子最近的 15 棟房屋之內部居住空間平均大小的分布情形，以 1250 至 1750 平方英尺最多，約占 31%；1750 至 2250 平方英尺次之，約占 26%。

## 九、最近 15 棟房屋的土地平均大小( $X_8$ )

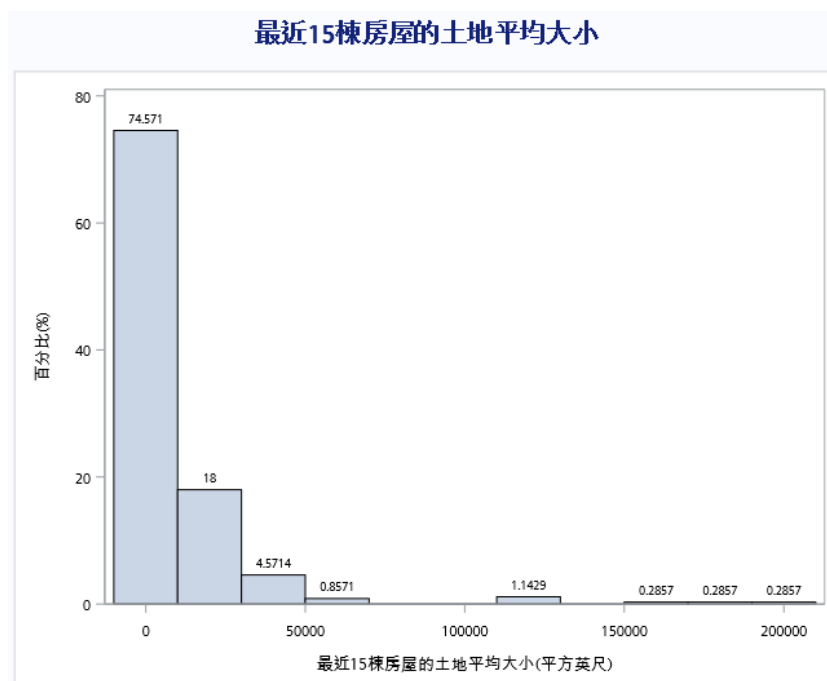


圖 2-12 最近 15 棟房屋的土地平均大小相對直方圖

由圖 2-12 發現大部分數據都小於 5 萬平方英尺，因為有離群值且全距很大（最大值為 207781，最小值為 1021），故難以由上圖看出確切分布。

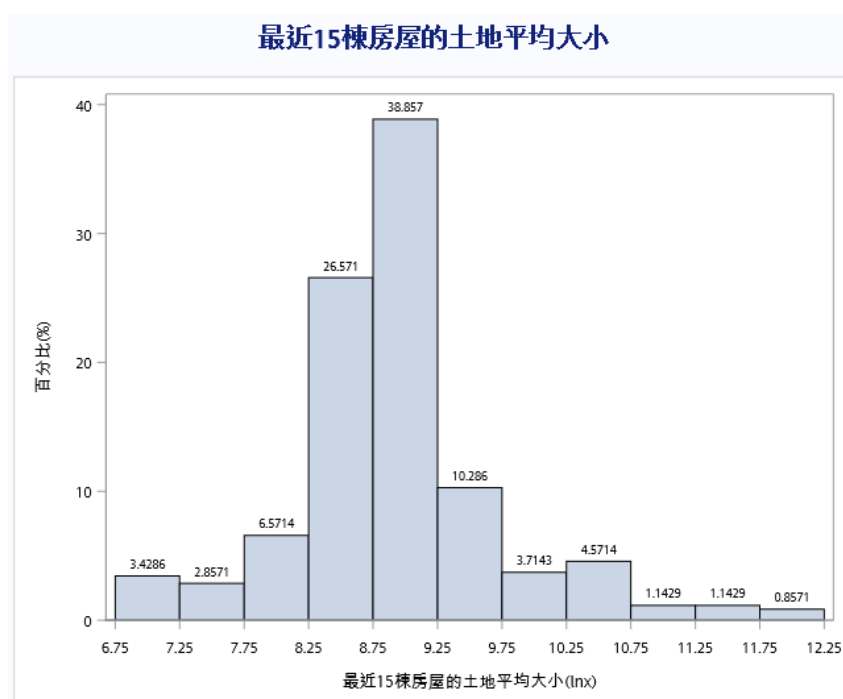


圖 2-13 取對數後的最近 15 棟房屋的土地平均大小相對直方圖

由圖 2-13 可看出取對數後的分布，以 8.75 至 9.25 占 39% 最多，即最近的 15 棟房屋的土地平均大小大多介於 6310 到 10405 平方英尺。然而，中位數為 7500 也落在此區間。次為 8.25 至 8.75 相當於 3827 到 6310 平方英尺，約占 27%。

$$(e^{8.25} \approx 3827.63, e^{8.75} \approx 6310.69, e^{9.25} \approx 10404.57)$$

## 第二節 CORRELATION

相關係數是一種衡量兩個變數之間線性關係強度和方向的統計指標，介於 -1 與 1 之間。相關係數的正負符號代表變數間關係的方向。正為「正相關」，表示兩變數為正向關係；負則為「負相關」，意指兩變數為反向關係。判定原則如下：

皮爾森相關係數	
$ r  = 1$	完全相關
$0.7 <  r  < 1$	高度相關
$0.3 <  r  \leq 0.7$	中度相關
$ r  \leq 0.3$	低度相關
$r = 0$	無線性關係

表 2-3 各變數的相關性

Pearson 相關係數, N = 350 Prob >  r  (低於 H0): Rho=0									
	Ynew	x1	x2	x3	x4	x5	x6	x7	x8
Ynew	1.00000	0.30862 <.0001	0.49720 <.0001	0.67424 <.0001	0.10958 0.0405	0.60982 <.0001	0.03055 0.5690	0.55324 <.0001	0.10434 0.0511
x1 臥室數量	0.30862 <.0001	1.00000	0.52094 <.0001	0.60666 <.0001	0.03866 0.4709	0.52600 <.0001	0.18198 0.0006	0.40490 <.0001	0.04805 0.3702
x2 浴室數量	0.49720 <.0001	0.52094 <.0001	1.00000	0.73839 <.0001	0.05880 0.2726	0.67374 <.0001	0.49164 <.0001	0.51494 <.0001	0.10354 0.0530
x3 居住面積	0.67424 <.0001	0.60666 <.0001	0.73839 <.0001	1.00000	0.25350 <.0001	0.88189 <.0001	0.30544 <.0001	0.72119 <.0001	0.23561 <.0001
x4 土地面積	0.10958 0.0405	0.03866 0.4709	0.05880 0.2726	0.25350 <.0001	1.00000	0.13445 0.0118	-0.00348 0.9483	0.18947 0.0004	0.82368 <.0001
x5 地上居住面積	0.60982 <.0001	0.52600 <.0001	0.67374 <.0001	0.88189 <.0001	0.13445 0.0118	1.00000	0.41476 <.0001	0.72209 <.0001	0.16542 0.0019
x6 建築年份	0.03055 0.5690	0.18198 0.0006	0.49164 <.0001	0.30544 <.0001	-0.00348 0.9483	0.41476 <.0001	1.00000	0.31260 <.0001	0.03655 0.4955
x7 最近15棟房屋的內部居住空間平均大小	0.55324 <.0001	0.40490 <.0001	0.51494 <.0001	0.72119 <.0001	0.18947 0.0004	0.72209 <.0001	0.31260 <.0001	1.00000	0.23785 <.0001
x8 最近15棟房屋的土地平均大小	0.10434 0.0511	0.04805 0.3702	0.10354 0.0530	0.23561 <.0001	0.82368 <.0001	0.16542 0.0019	0.03655 0.4955	0.23785 <.0001	1.00000

由表 2-3 可知  $Y_{new}$ （房價）與其他變數的相關性：

$Y_{new}$  與  $X_1$ （臥室數量）的相關係數為 0.30862 且， $p$  值極低（ $<.0001$ ），表示存在中度正相關。

$Y_{new}$  與  $X_2$ （浴室數量）的相關係數為 0.49720，且  $p$  值極低（ $<.0001$ ），表示存在中度正相關。

$Y_{new}$  與  $X_3$ （居住面積）的相關係數為 0.67424，且  $p$  值極低（ $<.0001$ ），表示存在中度正相關。

$Y_{new}$  與  $X_4$ （土地面積）的相關係數為 0.10958，且  $p$  值為 0.0405，表示存在低度正相關。

$Y_{new}$  與  $X_5$ （地上居住面積）的相關係數為 0.60982，且  $p$  值極低（ $<.0001$ ），表示存在中度正相關。

$Y_{new}$  與  $X_6$ （建築年份）的相關係數為 0.03055， $p$  值為 0.5690，表示相關性較弱。

$Y_{new}$  與  $X_7$ （最近 15 棟房屋的內部居住空間平均大小）的相關係數為 0.55324，且  $p$  值極低（ $<.0001$ ），表示存在中度正相關。

$Y_{new}$  與  $X_8$ （最近 15 棟房屋的土地平均大小）的相關係數為 0.10434， $p$  值為 0.0511，表示存在較弱的正相關。

### 第三節 Variance Inflation Factor(VIF)

變異數膨脹因子(VIF)為診斷多元共線性嚴重程度的指標。

表 2-4 各變異數的 VIF

變數	標籤	變異數膨脹
Intercept	Intercept	0
X1	臥室數量	1.65087
X2	浴室數量	2.86015
X3	居住面積	7.99329
X4	土地面積	3.45053
X5	地上居住面積	5.58845
X6	建築年份	1.51007
X7	最近 15 棟房屋的 內部居住空間 平均大小	2.32102
X8	最近 15 棟房屋的 土地平均大小	3.25018

$$VIF_i = \frac{1}{1 - R_i^2}, i=1,2, \dots, 8$$

觀察自變數  $X_j$  的迴歸模型時，複判定係數  $R_j^2$  反映了  $X_j$  與其他自變數的相依程度。 $R_j^2$  愈大，VIF 值愈大，可能導致迴歸參數估計不穩定，增加係數估計的變異性。通常，VIF 值大於 10 被視為存在共線性的警告信號。

## 第參章 原始模型檢定

### 第一節 建立迴歸模型

為了解影響西雅圖金郡房價的因素，我們將探討臥室數量、浴室數量、居住面積、土地面積、地上居住面積、建築年份、以及最近 15 棟房屋的內部居住空間和土地平均大小之相關性。

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + \beta_7 X_{7i} + \beta_8 X_{8i} + \varepsilon_i \\ i = 1, 2, \dots, 350$$

$$E(Y_i) = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{\beta}_4 X_{4i} + \hat{\beta}_5 X_{5i} + \hat{\beta}_6 X_{6i} + \hat{\beta}_7 X_{7i} + \hat{\beta}_8 X_{8i} \\ i = 1, 2, \dots, 350$$

為了印證以上的基本假設，我們將所有的可能變數皆列入解釋變數。

首先令

$X_1$ : 臥室數量

$X_2$ : 浴室數量

$X_3$ : 居住面積

$X_4$ : 土地面積

$X_5$ : 地上居住面積

$X_6$ : 建築年份

$X_7$ : 最近 15 棟房屋的內部居住空間平均大小

$X_8$ : 最近 15 棟房屋的土地平均大小

以  $Y$  為反應變數， $Y$ : 房價



以下用 SAS 建立迴歸模型

表 3-1 參數估計

變數	標籤	DF	參數估計值	標準誤差	t 值	Pr >  t
<b>Intercept</b>	<b>Intercept</b>	<b>1</b>	12.10100	0.58484	20.69	<.0001
X1	臥室數量	<b>1</b>	-0.03888	0.00992	-3.92	0.0001
X2	浴室數量	<b>1</b>	0.04808	0.01638	2.93	0.0036
X3	居住面積	<b>1</b>	0.00010904	0.00002209	4.94	<.0001
X4	土地面積	<b>1</b>	-1.95684E-7	4.051799E-7	-0.48	0.6294
X5	地上居住面積	<b>1</b>	0.00003380	0.00002027	1.67	0.0963
X6	建築年份	<b>1</b>	-0.00197	0.00030301	-6.52	<.0001
X7	最近 15 棟房屋的內部居住空間平均大小	<b>1</b>	0.00005125	0.00001660	3.09	0.0022
X8	最近 15 棟房屋的土地平均大小	<b>1</b>	-4.99694E-7	6.217282E-7	-0.80	0.4221

經表 3-1，我們可得：

$$\hat{\beta}_0=12.10100 \quad \hat{\beta}_1=-0.03888 \quad \hat{\beta}_2=0.04808 \quad \hat{\beta}_3=0.00010904 \quad \hat{\beta}_4=-1.95684E-7$$

$$\hat{\beta}_5=0.00003380 \quad \hat{\beta}_6=-0.00197 \quad \hat{\beta}_7=0.00005125 \quad \hat{\beta}_8=-4.99694E-7$$

建立原始模型：

$$\hat{Y}=12.101 - 0.03888 X_1 + 0.04808 X_2 + 0.00010904 X_3 - 1.95684E-7 X_4$$

$$+ 0.0000338 X_5 - 0.00197 X_6 + 0.00005125 X_7 - 4.99694E-7 X_8$$

## 第二節 單一參數 t 檢定

得知  $\beta$  值之後，接著判斷各解釋變數  $X_1$  (臥室數量)、 $X_2$  (浴室數量)、 $X_3$  (居住面積)、 $X_4$  (土地面積)、 $X_5$  (地上居住面積)、 $X_6$  (建築年份)、 $X_7$  (最近 15 棟房屋的內部居住空間平均大小)、 $X_8$  (最近 15 棟房屋的土地平均大小) 與  $Y$  (房價) 是否存在線性關係。

### 一、 $\beta_1$ 之 t 檢定

欲了解  $X_1$ (臥室數量)與  $Y$ (房價)是否存在線性相關，首先我們先假定其他變數為固定的情況下統計基本假設：

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

在顯著水準  $\alpha=0.05$  下虛無假設拒絕域為： $p\text{-value} < \alpha=0.05$  檢定如下：

因為  $p\text{-value}=0.0001 < \alpha=0.05$ ，因此拒絕  $H_0$  的假設，表示有充分證據顯示  $\beta_1 \neq 0$ 。即表示  $X_1$ (臥室數量)與  $Y$ (房價)有存在線性相關。

### 二、 $\beta_2$ 之 t 檢定

我們想要判斷  $X_2$ (浴室數量)與  $Y$ (房價)是否存在線性相關，我們假定其他變數為固定的情況下統計基本假設：

$$H_0 : \beta_2 = 0$$

$$H_1 : \beta_2 \neq 0$$

虛無假設拒絕域為： $p\text{-value} < \alpha=0.05$  檢定如下：

因為  $p\text{-value}=0.0036 < \alpha=0.05$ ，因此拒絕  $H_0$  的假設，表示有充分證據顯示  $\beta_2 \neq 0$ 。即表示  $X_2$ (浴室數量)與  $Y$ (房價)有存在線性相關。

### 三、 $\beta_3$ 之 t 檢定

我們想要判斷  $X_3$ (居住面積)與  $Y$ (房價)是否存在線性相關，我們假定其他變數為固定的情況下統計基本假設：

$$H_0 : \beta_3 = 0$$

$$H_1 : \beta_3 \neq 0$$

虛無假設拒絕域為：p-value <  $\alpha=0.05$  檢定如下：

因為 p-value < 0.0001，所以 p-value <  $\alpha=0.05$ ，因此拒絕  $H_0$  的假設，表示有充分證據顯示  $\beta_3 \neq 0$ 。即表示  $X_3$ (居住面積)與  $Y$ (房價)有存在線性相關。

#### 四、 $\beta_4$ 之 t 檢定

我們想要判斷  $X_4$ (土地面積)與  $Y$ (房價)是否存在線性相關，我們假定其他變數為固定的情況下統計基本假設：

$$H_0 : \beta_4 = 0$$

$$H_1 : \beta_4 \neq 0$$

虛無假設拒絕域為：p-value <  $\alpha=0.05$  檢定如下：

因為 p-value = 0.6294 >  $\alpha=0.05$ ，因此不拒絕  $H_0$  的假設，表示沒有充分證據顯示  $\beta_4 \neq 0$ 。即表示  $X_4$ (土地面積)與  $Y$ (房價)沒有存在線性相關。

#### 五、 $\beta_5$ 之 t 檢定

我們想要判斷  $X_5$ (地上居住面積)與  $Y$ (房價)是否存在線性相關，我們假定其他變數為固定的情況下統計基本假設：

$$H_0 : \beta_5 = 0$$

$$H_1 : \beta_5 \neq 0$$

虛無假設拒絕域為：p-value <  $\alpha=0.05$  檢定如下：

因為 p-value = 0.0963 >  $\alpha=0.05$ ，因此不拒絕  $H_0$  的假設，表示沒有充分證據顯示  $\beta_5 \neq 0$ 。即表示  $X_5$ (地上居住面積)與  $Y$ (房價)沒有存在線性相關。

## 六、 $\beta_6$ 之 t 檢定

我們想要判斷  $X_6$ (建築年份)與  $Y$ (房價)是否存在線性相關，我們假定其他變數為固定的情況下統計基本假設：

$$H_0 : \beta_6 = 0$$

$$H_1 : \beta_6 \neq 0$$

虛無假設拒絕域為： $p\text{-value} < \alpha=0.05$  檢定如下：

因為  $p\text{-value} < 0.0001$ ，所以  $p\text{-value} < \alpha=0.05$ ，因此拒絕  $H_0$  的假設，表示有充分證據顯示  $\beta_6 \neq 0$ 。即表示  $X_6$ (建築年份)與  $Y$ (房價)有存在線性相關。

## 七、 $\beta_7$ 之 t 檢定

我們想要判斷  $X_7$ (最近 15 棟房屋的內部居住空間平均大小)與  $Y$ (房價)是否存在線性相關，我們假定其他變數為固定的情況下統計基本假設：

$$H_0 : \beta_7 = 0$$

$$H_1 : \beta_7 \neq 0$$

虛無假設拒絕域為： $p\text{-value} < \alpha=0.05$  檢定如下：

因為  $p\text{-value} = 0.0022 < \alpha=0.05$ ，因此拒絕  $H_0$  的假設，表示有充分證據顯示  $\beta_7 \neq 0$ 。即表示  $X_7$ (最近 15 棟房屋的內部居住空間平均大小)與  $Y$ (房價)有存在線性相關。

## 八、 $\beta_8$ 之 t 檢定

我們想要判斷  $X_8$ (最近 15 棟房屋的土地平均大小)與  $Y$ (房價)是否存在線性相關，我們假定其他變數為固定的情況下統計基本假設：

$$H_0 : \beta_8 = 0$$

$$H_1 : \beta_8 \neq 0$$

虛無假設拒絕域為：p-value <  $\alpha=0.05$  檢定如下：

因為 p-value = 0.4221 >  $\alpha=0.05$ ，因此不拒絕  $H_0$  的假設，表示沒有充分證據顯示  $\beta_8 \neq 0$ 。即表示  $X_8$  (最近 15 棟房屋的土地平均大小) 與  $Y$  (房價) 沒有存在線性相關。

### 第三節 模型適合度檢定

以下為我們利用 SAS 做適合度檢定，建立迴歸模型

表 3-2 變異數分析

變異數的分析					
來源	DF	平方和	均方	F 值	Pr > F
模型	8	7.67217	0.95902	50.66	<.0001
誤差	341	6.45589	0.01893		
已校正的總計	349	14.12806			

統計假設如下：

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$$

$$H_1 : \beta_i \text{ 不全為 } 0, i = 1, 2, \dots, 8$$

虛無假設之拒絕域為：p-value <  $\alpha=0.05$

因為 p-value < 0.0001，所以 p-value <  $\alpha=0.05$ ，拒絕  $H_0$  的假設，表示有充分證據顯示  $\beta_i$  不全為 0，即表示根據資料所配適之迴歸模型是合適的。

#### 第四節 模型解釋能力

表 3-3 模型解釋能力分析

根 MSE	0.13759	$R^2$	0.5430
應變平均值	8.57356	調整 $R^2$	0.5323
變異係數	1.60487		

$$R^2 = SSR/SSTO = 0.5430 > 0.5$$

$$\text{調整 } R^2 = 1 - MSE/MSTO = 0.5323$$

全模型下的  $R^2$ ，表示此迴歸能解釋 54.3% 的 Y 變異，而經過校正的調整  $R^2$  只低 0.0107，所以代表  $X_i, i=1,2,\dots,8$  對 Y (房價) 具有相當程度的解釋能力，因此下一步我們將進行模型的篩選，去除掉影響力較小的變數，以精簡模型。

## 第肆章 模型的選取方法

令

$X_1$ : 臥室數量

$X_2$ : 浴室數量

$X_3$ : 居住面積

$X_4$ : 土地面積

$X_5$ : 地上居住面積

$X_6$ : 建築年份

$X_7$ : 最近 15 棟房屋的內部居住空間平均大小

$X_8$ : 最近 15 棟房屋的土地平均大小

### 第一節 向前選取法(Forward)

將每個特徵單獨添加到模型中，分別估計回歸係數。選擇能夠最大程度提升模型性能的特徵，將其添加到特徵集合中，即 p-value 最小者優先進入此模型內，直到模型外的每一個變數加進模型內其 F 統計量皆不顯著時，則選取解釋變數進入模型的工作停止，以此過程提高了模型的簡潔性。（在此使用 PROC REG 程序）

表 4-1 向前選取法

步驟	變數	標籤	偏 $R^2$	模型 $R^2$	$C_{(p)}$	F 值	PR>F
1	x3	居住面積	0.4546	0.4546	60.9969	290.07	<.0001
2	x6	建築年份	0.0339	0.4885	37.6781	23.02	<.0001
3	x1	臥室數量	0.0162	0.5047	27.6181	11.29	0.0009
4	x7	最近 15 棟房屋的內部居住空間平均大小	0.0135	0.5182	19.5139	9.70	0.0020
5	x2	浴室數量	0.0133	0.5315	11.5973	9.76	0.0019
6	x4	土地面積	0.0072	0.5387	8.2368	5.34	0.0214

**Step1:**

系統所選擇的解釋變數是  $X_3$ （居住面積），其 F 統計量為 290.07 ( $p\text{-value} < 0.0001 < \alpha = 0.05$ ) 是八個變數中最顯著的，所以優先進入此模型之內。

**Step2:**

系統在其他尚未進入模型中的七個變數中選取最顯著的變數為  $X_6$ （建築年份），其 F 統計量為 23.02 ( $p\text{-value} < 0.0001 < \alpha = 0.05$ ) 進入模型之內。

**Step 3:**

系統在其他尚未進入模型中的六個變數中選取最顯著的變數為  $X_1$ （臥室數量），其 F 統計量為 11.29 ( $p\text{-value} = 0.0009 < \alpha = 0.05$ ) 進入模型之內。

**Step 4:**

系統在其他尚未進入模型中的五個變數中選取最顯著的變數為  $X_7$ （最近 15 棟房屋的內部居住空間平均大小），其 F 統計量為 9.70 ( $p\text{-value} = 0.002 < \alpha = 0.05$ ) 進入模型之內。

**Step 5:**

系統在其他尚未進入模型中的四個變數中選取最顯著的變數為  $X_2$ （浴室數量），其 F 統計量為 9.76 ( $p\text{-value} = 0.0019 < \alpha = 0.05$ ) 進入模型之內。

**Step 6:**

系統在其他尚未進入模型中的三個變數中選取最顯著的變數為  $X_4$ （土地面積），其 F 統計量為 5.34 ( $p\text{-value} = 0.0214 < \alpha = 0.05$ ) 進入模型之內。

最後由於變數  $X_5$ （地上居住面積）、 $X_8$ （最近 15 棟房屋的土地平均大小），在給定  $\alpha = 0.05$  之下不顯著（也就是其  $p\text{-value} > 0.05$ ），因此系統並不考慮將此這兩個變數放入模型中。



表 4-2 參數估計

變數	參數估計值	標準誤差	類型 II SS	F 值	PR > F
<b>INTERCEPT</b>	11.80876	0.55920	8.47290	445.93	<.0001
<b>X1</b>	- 0.03934	0.00993	0.29820	15.69	<.0001
<b>X2</b>	0.04462	0.01628	0.14278	7.51	0.0064
<b>X3</b>	0.00013531	0.00001613	1.33699	70.37	<.0001
<b>X4</b>	-5.3687E-7	2.322988E-7	0.10149	5.34	0.0214
<b>X6</b>	-0.00182	0.00028985	0.75154	39.55	<.0001
<b>X7</b>	0.00005548	0.00001602	0.22785	11.99	0.0006

所以由向前選取法可得其線性迴歸模型為：

$$\hat{Y} = 11.80876 - 0.03934 X_1 + 0.04462 X_2 + 0.00013531 X_3 - 5.3687E-7 X_4 - 0.00182 X_6 + 0.00005548 X_7$$

## 第二節 後退刪去法(Backward)

與向前選取法相反，後退刪去法是從包含所有特徵的模型開始，然後逐步刪除對模型性能影響最小的特徵。能夠快速減少特徵數量，提高模型的簡潔性和解釋性。（在此使用 PROC REG 程序）。

表 4-3-1 後退刪去法

步驟	已移除變數	標籤	偏 $R^2$	模型 $R^2$	$C_{(p)}$	F 值	PR>F
<b>1</b>	x4	土地面積	0.0003	0.5427	7.2332	0.23	0.6294
<b>2</b>	x5	地上居住面積	0.0043	0.5385	8.4168	3.19	0.0749

### Step 1:

系統先將所有的解釋變數放入模型內，而由模型所求得：

$$R^2 = 0.5430, C(p) = 9.0000$$

**Step 2:**

表 4-3-2 後退刪去法

已移除變數	標籤	偏 $R^2$	模型 $R^2$	$C_{(P)}$	F 值	PR>F
X4	土地面積	0.0003	0.5427	7.2332	0.23	0.6294

由模型所求得的 F 統計量可知，變數 X<sub>4</sub> 的統計量最不顯著，其 F=0.23 (p - value = 0.6294 > 0.05)，所以變數 X<sub>4</sub> 最先被剔除出此模型之外。而將其剔除後， $R^2 = 0.5427$ ， $C(p) = 7.2332$ 。

**Step 3:**

表 4-3-3 後退刪去法

已移除變數	標籤	偏 $R^2$	模型 $R^2$	$C_{(P)}$	F 值	PR>F
X5	地上居住面積	0.0043	0.5385	8.4168	3.19	0.0749

由模型所求得的 F 統計量可知，變數 X<sub>5</sub> 的統計量其次不顯著，其 F=3.19 (p - value = 0.0749 > 0.05)，所以變數 X<sub>5</sub> 也被剔除出此模型之外。而將其剔除後， $R^2 = 0.5385$ ， $C(p) = 8.4168$ 。

**Step 4:**

最後系統由剩餘變數 X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub>, X<sub>6</sub>, X<sub>7</sub>, X<sub>8</sub> 重新建立模型，而由模型所求得的 F 統計量值可知，在給定的顯著水準  $\alpha = 0.05$  之下，皆具顯著性。

表 4-4 參數估計

變數	參數估計值	標準誤差	類型 II SS	F 值	PR > F
INTERCEPT	11.80544	0.55932	8.46894	445.49	<.0001
X1	-0.03880	0.00991	0.29159	15.34	0.0001
X2	0.04781	0.01613	0.16709	8.79	0.0032
X3	0.00013080	0.00001571	1.31713	69.28	<.0001
X6	-0.00182	0.00028993	0.75194	39.55	<.0001
X7	0.00005897	0.00001610	0.25515	13.42	0.0003
X8	-8.20927E-7	3.614201E-7	0.09808	5.16	0.0237

因此由後退刪去法選取的解釋變數建立的線性迴歸模型如下：

$$\hat{Y} = 11.80544 - 0.03880 X_1 + 0.04781 X_2 + 0.00013080 X_3 - 0.00182 X_6 + 0.00005897 X_7 - 8.20927E-7 X_8$$

### 第三節 逐步迴歸法

結合向前選取法和後退刪去法的特徵選取方法，逐步地選擇和排除特徵，以優化模型的性能。

#### Step 1:

已輸入變數 x3:  $R^2 = 0.4546$ 、 $C(p) = 60.9969$

表 4-5-1 變異數分析與參數估計

來源	DF	平方和	均方	F 值	PR > F
模型	1	6.42270	6.42270	290.07	<.0001
誤差	348	7.70536	0.02214		
已校正的總計	349	14.12806			
變數	參數估計值	標準誤差	類型 II SS	F 值	Pr > F
INTERCEPT	8.26164	0.01997	3790.70430	171201	<.0001
X3	0.00014390	0.00000845	6.42270	290.07	<.0001

所選取的變數為  $X_3$ ，其統計量  $F = 290.07$  ( $p\text{-value} < 0.0001 < \alpha = 0.05$ ) 是八個解釋變數中最顯著的，接著系統檢定模式內的唯一解釋變數  $X_3$ ，結果發現是顯著的，因此不再將其從模式中剔除。

#### Step 2:

已輸入變數 x6:  $R^2 = 0.4885$ 、 $C(p) = 37.6781$

表 4-5-2 變異數分析與參數估計

來源	DF	平方和	均方	F 值	PR > F
模型	2	6.90204	3.45102	165.72	<.0001
誤差	347	7.22602	0.02082		

已校正的總計	349	14.12806			
變數	參數估計值	標準誤差	類型 II SS	F 值	Pr > F
INTERCEPT	10.80304	0.53006	8.64989	415.38	<.0001
X3	0.00015651	0.00000860	6.88886	330.81	<.0001
X6	-0.00130	0.00027158	0.47934	23.02	<.0001

再由其餘的七個解釋變數  $X_1$ 、 $X_2$ 、 $X_4$ 、 $X_5$ 、 $X_6$ 、 $X_7$ 、 $X_8$  中選取最顯著的  $X_6$ ，其統計檢定量  $F = 23.02$  ( $p\text{-value} < 0.0001 < \alpha = 0.05$ )，為七個解釋變數中最顯著的，接著系統檢定模式內的解釋變數  $X_3$ 、 $X_6$ ，結果發現是顯著的，因此不再將其從模式中剔除。

### Step 3:

已輸入變數 x1:  $R^2 = 0.5047$ 、 $C(p) = 27.6181$

表 4-5-3 變異數分析與參數估計

來源	DF	平方和	均方	F 值	PR > F
模型	3	7.13036	2.37679	117.52	<.0001
誤差	346	6.99769	0.02022		
已校正的總計	349	14.12806			
變數	參數估計值	標準誤差	類型 II SS	F 值	Pr > F
INTERCEPT	10.88153	0.52289	8.75852	433.06	<.0001
X1	-0.03371	0.01003	0.22832	11.29	0.0009
X3	0.00017725	0.00001049	5.77517	285.55	<.0001
X6	-0.00131	0.00026765	0.48223	23.84	<.0001

由其餘的六個解釋變數  $X_1$ 、 $X_2$ 、 $X_4$ 、 $X_5$ 、 $X_7$ 、 $X_8$  中選取最顯著的  $X_8$ ，其統計檢定量  $F = 11.29$  ( $p\text{-value} = 0.0009 < \alpha = 0.05$ )，為六個解釋變數中最顯著的，接著系統檢定模式內的解釋變數  $X_1$ 、 $X_3$ 、 $X_6$ ，結果發現是顯著的，因此不再將其從模式中剔除。

### Step 4:

已輸入變數 x7:  $R^2 = 0.5182$ 、 $C(p) = 19.5139$

表 4-5-4 變異數分析與參數估計

來源	DF	平方和	均方	F 值	PR > F
模型	4	7.32166	1.83041	92.78	<.0001
誤差	345	6.80640	0.01973		
已校正的總計	349	14.12806			
變數	參數估計值	標準誤差	類型 II SS	F 值	Pr > F
INTERCEPT	11.05921	0.51959	8.93778	453.03	<.0001
X1	-0.03188	0.00993	0.20349	10.31	0.0014
X3	0.00015110	0.00001334	2.53211	128.35	<.0001
X6	-0.00142	0.00026697	0.56071	28.42	<.0001
X7	0.00005059	0.00001625	0.19129	9.70	0.0020

由其餘的五個解釋變數  $X_2$ 、 $X_4$ 、 $X_5$ 、 $X_7$ 、 $X_8$  選取最顯著的  $X_7$ ，其統計檢定量  $F = 9.7$  ( $p\text{-value} = 0.0020 < \alpha = 0.05$ )，為五個解釋變數中最顯著的，接著系統檢定模式內的解釋變數  $X_1$ 、 $X_3$ 、 $X_6$ 、 $X_7$ ，結果發現是顯著的，因此不再將其從模式中剔除。

#### Step 5:

已輸入變數 x2:  $R^2 = 0.5315$ 、 $C(p) = 11.5973$

表 4-5-5 變異數分析與參數估計

來源	DF	平方和	均方	F 值	PR > F
模型	5	7.50940	1.50188	78.06	<.0001
誤差	344	6.61866	0.01924		
已校正的總計	349	14.12806			
變數	參數估計值	標準誤差	類型 II SS	F 值	Pr > F
INTERCEPT	11.77977	0.56258	8.43559	438.43	<.0001

<b>X1</b>	-0.03642	0.00991	0.25977	13.50	0.0003
<b>X2</b>	0.05053	0.01618	0.18774	9.76	0.0019
<b>X3</b>	0.00012499	0.00001560	1.23550	64.21	<.0001
<b>X6</b>	-0.00181	0.00029164	0.74333	38.63	<.0001
<b>X7</b>	0.00005555	0.00001612	0.22839	11.87	0.0006

由其餘的四個解釋變數  $X_2$ 、 $X_4$ 、 $X_5$ 、 $X_8$  選取最顯著的  $X_2$ ，其統計檢定量  $F = 9.76$  ( $p - \text{value} = 0.0019 < \alpha = 0.05$ )，為四個解釋變數中最顯著的，接著系統檢定模式內的解釋變數  $X_1$ 、 $X_2$ 、 $X_3$ 、 $X_6$ 、 $X_7$ ，結果發現是顯著的，因此不再將其從模式中剔除。

#### Step 6:

已輸入變數 x4:  $R^2 = 0.5387$ 、 $C(p) = 8.2368$

表 4-5-6 變異數分析與參數估計

來源	DF	平方和	均方	F 值	PR > F
模型	6	7.61089	1.26848	66.76	<.0001
誤差	343	6.51717	0.01900		
已校正的總計	349	14.12806			
變數	參數估計值	標準誤差	類型 II SS	F 值	Pr > F
<b>INTERCEPT</b>	11.80876	0.55920	8.47290	445.93	<.0001
<b>X1</b>	-0.03934	0.00993	0.29820	15.69	<.0001
<b>X2</b>	0.04462	0.01628	0.14278	7.51	0.0064
<b>X3</b>	0.00013531	0.00001613	1.33699	70.37	<.0001
<b>X4</b>	-5.3687E-7	2.322988E-7	0.10149	5.34	0.0214
<b>X6</b>	-0.00182	0.00028985	0.75154	39.55	<.0001
<b>X7</b>	0.00005548	0.00001602	0.22785	11.99	0.0006

由其餘的三個解釋變數  $X_4$ 、 $X_5$ 、 $X_8$  選取最顯著的  $X_4$ ，其統計檢定量  $F = 5.34$  ( $p\text{-value} = 0.0214 < \alpha = 0.05$ )，為三個解釋變數中最顯著的，接著系統檢定模式內的解釋變數  $X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_6$ 、 $X_7$ ，結果發現是顯著的，因此不再將其從模式中剔除。

#### Step 7:

因為剩餘的變數  $X_5$ 、 $X_8$ ，其  $F$  統計檢定量不顯著，所以不再選入模型之中，停止變數的選取。

表 4-6 逐步迴歸法摘要

步驟	已輸入 變數	標籤	偏 $R^2$	模型 $R^2$	$C_{(p)}$	F 值	Pr>F
1	x3	居住面積	0.4546	0.4546	60.9969	290.07	<.0001
2	x6	建築年份	0.0339	0.4885	37.6781	23.02	<.0001
3	x1	臥室數量	0.0162	0.5047	27.6181	11.29	0.0009
4	x7	最近 15 棟 房屋的內部 居住空間平 均大小	0.0135	0.5182	19.5139	9.70	0.0020
5	x2	浴室數量	0.0133	0.5315	11.5973	9.76	0.0019
6	x4	土地面積	0.0072	0.5387	8.2368	5.34	0.0214

因此，在運用逐步迴歸法所建立的線性迴歸模式終告確立，其模型如下所示：

$$\hat{Y} = 11.80876 - 0.03934 X_1 + 0.04462 X_2 + 0.00013531 X_3 - 5.3687E-7 X_4 \\ - 0.00182 X_6 + 0.00005548 X_7$$

#### 第四節 其他選取法

表 4-7 其他選取法

數目	$R^2$	調整 $R^2$	C(p)	AIC	MSE	SBC	模型中的變數
1	0.4546	0.4530	60.9969	-1331.6060	0.02214	-1323.89015	x3
1	0.3719	0.3701	122.7331	-1282.1762	0.02550	-1274.46030	x5
1	0.3061	0.3041	171.8409	-1247.3041	0.02817	-1239.58828	x7
1	0.2472	0.2450	215.7700	-1218.8055	0.03056	-1211.08968	x2
1	0.0952	0.0926	329.1687	-1154.4510	0.03673	-1146.73517	x1
1	0.0120	0.0092	391.2827	-1123.6480	0.04011	-1115.93209	x4
1	0.0109	0.0080	392.1198	-1123.2508	0.04016	-1115.53492	x8
1	0.0009	-.0019	399.5474	-1119.7464	0.04056	-1112.03050	x6
2	0.4885	0.4856	37.6781	-1352.0858	0.02082	-1340.51199	x3 x6
2	0.4706	0.4675	51.0897	-1339.9985	0.02156	-1328.42468	x1 x3
2	0.4640	0.4609	56.0210	-1335.6570	0.02183	-1324.08320	x3 x7
2	0.4586	0.4555	59.9962	-1332.1961	0.02204	-1320.62226	x3 x4
2	0.4578	0.4546	60.6484	-1331.6314	0.02208	-1320.05765	x3 x8
2	0.4556	0.4525	62.2203	-1330.2745	0.02216	-1318.70071	x3 x5
2	0.4546	0.4515	62.9962	-1329.6066	0.02221	-1318.03283	x2 x3
2	0.4316	0.4283	80.1620	-1315.1476	0.02314	-1303.57376	x5 x6
2	0.3985	0.3950	104.8604	-1295.3387	0.02449	-1283.76492	x5 x7
2	0.3855	0.3820	114.5466	-1287.8663	0.02502	-1276.29249	x2 x5
2	0.3727	0.3690	124.1545	-1280.6085	0.02554	-1269.03469	x4 x5
2	0.3721	0.3685	124.5808	-1280.2899	0.02557	-1268.71612	x1 x5



數目	$R^2$	調整 $R^2$	C(p)	AIC	MSE	SBC	模型中的變數
2	<b>0.3719</b>	0.3683	124.7239	-1280.1830	0.02557	-1268.60924	x5 x8
2	<b>0.3674</b>	0.3638	128.0651	-1277.6970	0.02576	-1266.12323	x2 x7
2	<b>0.3285</b>	0.3247	157.0707	-1256.8264	0.02734	-1245.25260	x6 x7
2	<b>0.3146</b>	0.3107	167.4509	-1249.6499	0.02790	-1238.07608	x1 x7
2	<b>0.3075</b>	0.3035	172.7467	-1246.0445	0.02819	-1234.47066	x2 x6
2	<b>0.3069</b>	0.3029	173.2535	-1245.7014	0.02822	-1234.12756	x7 x8
2	<b>0.3061</b>	0.3021	173.8233	-1245.3160	0.02825	-1233.74220	x4 x7
2	<b>0.2537</b>	0.2494	212.9361	-1219.8303	0.03039	-1208.25645	x2 x4
2	<b>0.2506</b>	0.2463	215.2496	-1218.3793	0.03051	-1206.80555	x1 x2
2	<b>0.2500</b>	0.2457	215.6624	-1218.1212	0.03054	-1206.54735	x2 x8
2	<b>0.1048</b>	0.0996	324.0422	-1156.1650	0.03645	-1144.59118	x1 x4
2	<b>0.1033</b>	0.0981	325.1759	-1155.5715	0.03651	-1143.99769	x1 x8
2	<b>0.0959</b>	0.0907	330.6622	-1152.7137	0.03681	-1141.13988	x1 x6
2	<b>0.0130</b>	0.0073	392.5689	-1121.9870	0.04019	-1110.41315	x4 x6
2	<b>0.0126</b>	0.0069	392.8228	-1121.8664	0.04020	-1110.29256	x4 x8
2	<b>0.0116</b>	0.0059	393.5858	-1121.5041	0.04024	-1109.93028	x6 x8
3	<b>0.5047</b>	0.5004	27.6181	-1361.3233	0.02022	-1345.89161	x1 x3 x6
3	<b>0.5038</b>	0.4995	28.2625	-1360.7137	0.02026	-1345.28192	x3 x6 x7
3	<b>0.4980</b>	0.4936	32.6347	-1356.6050	0.02050	-1341.17324	x3 x5 x6
3	<b>0.4954</b>	0.4910	34.5626	-1354.8085	0.02060	-1339.37674	x2 x3 x6
3	<b>0.4949</b>	0.4905	34.9140	-1354.4819	0.02062	-1339.05020	x3 x4 x6

數目	$R^2$	調整 $R^2$	C(p)	AIC	MSE	SBC	模型中的變數
3	<b>0.4925</b>	0.4881	36.6980	-1352.8293	0.02072	-1337.39752	x3 x6 x8
3	<b>0.4785</b>	0.4740	47.1306	-1343.3177	0.02129	-1327.88594	x1 x3 x7
3	<b>0.4775</b>	0.4730	47.9163	-1342.6117	0.02134	-1327.17993	x1 x3 x4
3	<b>0.4758</b>	0.4712	49.2041	-1341.4577	0.02141	-1326.02595	x1 x3 x8
3	<b>0.4734</b>	0.4688	50.9829	-1339.8698	0.02150	-1324.43806	x2 x5 x6
3	<b>0.4714</b>	0.4668	52.4518	-1338.5640	0.02158	-1323.13227	x1 x3 x5
3	<b>0.4708</b>	0.4662	52.8896	-1338.1757	0.02161	-1322.74399	x1 x2 x3
3	<b>0.4683</b>	0.4637	54.7513	-1336.5295	0.02171	-1321.09781	x3 x7 x8
3	<b>0.4681</b>	0.4635	54.9285	-1336.3733	0.02172	-1320.94154	x3 x4 x7
3	<b>0.4640</b>	0.4594	57.9841	-1333.6892	0.02189	-1318.25749	x3 x5 x7
3	<b>0.4640</b>	0.4593	58.0157	-1333.6616	0.02189	-1318.22989	x2 x3 x7
3	<b>0.4599</b>	0.4553	61.0212	-1331.0417	0.02205	-1315.60997	x5 x6 x7
3	<b>0.4590</b>	0.4543	61.6898	-1330.4615	0.02209	-1315.02980	x3 x4 x5
3	<b>0.4588</b>	0.4541	61.8555	-1330.3179	0.02210	-1314.88616	x2 x3 x4
3	<b>0.4587</b>	0.4540	61.9503	-1330.2358	0.02210	-1314.80405	x3 x4 x8
3	<b>0.4585</b>	0.4538	62.0967	-1330.1089	0.02211	-1314.67721	x3 x5 x8
3	<b>0.4578</b>	0.4531	62.6113	-1329.6635	0.02214	-1314.23178	x2 x3 x8
3	<b>0.4557</b>	0.4509	64.2122	-1328.2814	0.02223	-1312.84968	x2 x3 x5
3	<b>0.4425</b>	0.4377	73.9972	-1319.9505	0.02276	-1304.51875	x2 x6 x7
3	<b>0.4323</b>	0.4273	81.6675	-1313.5558	0.02318	-1298.12410	x1 x5 x6
3	<b>0.4317</b>	0.4268	82.0578	-1313.2335	0.02320	-1297.80178	x4 x5 x6

數目	$R^2$	調整 $R^2$	C(p)	AIC	MSE	SBC	模型中的變數
3	<b>0.4316</b>	0.4267	82.1416	-1313.1643	0.02321	-1297.73261	x5 x6 x8
3	<b>0.4102</b>	0.4050	98.1686	-1300.1827	0.02408	-1284.75098	x2 x5 x7
3	<b>0.3991</b>	0.3939	106.3869	-1293.7082	0.02453	-1278.27642	x5 x7 x8
3	<b>0.3990</b>	0.3937	106.5231	-1293.6019	0.02454	-1278.17013	x1 x5 x7
3	<b>0.3985</b>	0.3933	106.8344	-1293.3590	0.02456	-1277.92730	x4 x5 x7
3	<b>0.3877</b>	0.3824	114.9033	-1287.1228	0.02500	-1271.69112	x1 x2 x5
3	<b>0.3866</b>	0.3813	115.7365	-1286.4851	0.02505	-1271.05341	x2 x4 x5
3	<b>0.3855</b>	0.3802	116.5295	-1285.8793	0.02509	-1270.44758	x2 x5 x8
3	<b>0.3738</b>	0.3684	125.3036	-1279.2452	0.02557	-1263.81346	x4 x5 x8
3	<b>0.3728</b>	0.3674	126.0238	-1278.7063	0.02561	-1263.27453	x1 x4 x5
3	<b>0.3721</b>	0.3666	126.5748	-1278.2945	0.02564	-1262.86273	x1 x5 x8
3	<b>0.3679</b>	0.3624	129.6898	-1275.9754	0.02581	-1260.54369	x2 x7 x8
3	<b>0.3677</b>	0.3622	129.8675	-1275.8435	0.02582	-1260.41181	x2 x4 x7
3	<b>0.3675</b>	0.3620	130.0311	-1275.7222	0.02583	-1260.29050	x1 x2 x7
3	<b>0.3390</b>	0.3333	151.2599	-1260.3252	0.02699	-1244.89350	x1 x6 x7
3	<b>0.3297</b>	0.3239	158.1965	-1255.4375	0.02737	-1240.00582	x6 x7 x8
3	<b>0.3286</b>	0.3227	159.0502	-1254.8407	0.02742	-1239.40900	x4 x6 x7
3	<b>0.3152</b>	0.3092	169.0540	-1247.9216	0.02796	-1232.48984	x1 x7 x8
3	<b>0.3147</b>	0.3088	169.3934	-1247.6892	0.02798	-1232.25750	x1 x4 x7
3	<b>0.3126</b>	0.3067	170.9443	-1246.6294	0.02807	-1231.19765	x2 x4 x6
3	<b>0.3099</b>	0.3040	172.9490	-1245.2642	0.02818	-1229.83244	x2 x6 x8

數目	$R^2$	調整 $R^2$	C(p)	AIC	MSE	SBC	模型中的變數
3	<b>0.3092</b>	0.3032	173.5299	-1244.8696	0.02821	-1229.43784	x4 x7 x8
3	<b>0.3087</b>	0.3027	173.8948	-1244.6219	0.02823	-1229.19018	x1 x2 x6
3	<b>0.2570</b>	0.2505	212.4808	-1219.3766	0.03034	-1203.94491	x1 x2 x4
3	<b>0.2542</b>	0.2477	214.5366	-1218.0814	0.03045	-1202.64968	x2 x4 x8
3	<b>0.2534</b>	0.2470	215.1098	-1217.7211	0.03048	-1202.28940	x1 x2 x8
3	<b>0.1054</b>	0.0977	325.5755	-1154.4096	0.03653	-1138.97783	x1 x4 x6
3	<b>0.1051</b>	0.0973	325.8482	-1154.2666	0.03654	-1138.83489	x1 x4 x8
3	<b>0.1041</b>	0.0963	326.5655	-1153.8909	0.03658	-1138.45916	x1 x6 x8
3	<b>0.0135</b>	0.0049	394.1833	-1120.1702	0.04028	-1104.73849	x4 x6 x8
4	<b>0.5182</b>	0.5126	19.5139	-1369.0244	0.01973	-1349.73476	x1 x3 x6 x7
4	<b>0.5154</b>	0.5097	21.6611	-1366.9403	0.01985	-1347.65065	x1 x2 x3 x6
4	<b>0.5147</b>	0.5091	22.1259	-1366.4908	0.01987	-1347.20112	x1 x3 x4 x6
4	<b>0.5136</b>	0.5079	23.0040	-1365.6430	0.01992	-1346.35337	x1 x3 x5 x6
4	<b>0.5131</b>	0.5075	23.3184	-1365.3401	0.01994	-1346.05041	x2 x3 x6 x7
4	<b>0.5110</b>	0.5053	24.9083	-1363.8118	0.02002	-1344.52211	x1 x3 x6 x8
4	<b>0.5107</b>	0.5050	25.1579	-1363.5725	0.02004	-1344.28279	x3 x4 x6 x7
4	<b>0.5097</b>	0.5041	25.8470	-1362.9126	0.02008	-1343.62292	x3 x6 x7 x8
4	<b>0.5088</b>	0.5031	26.5775	-1362.2144	0.02012	-1342.92478	x3 x5 x6 x7
4	<b>0.5061</b>	0.5004	28.5666	-1360.3204	0.02023	-1341.03076	x2 x3 x5 x6
4	<b>0.5020</b>	0.4963	31.6068	-1357.4452	0.02039	-1338.15552	x3 x4 x5 x6
4	<b>0.5010</b>	0.4952	32.3638	-1356.7330	0.02043	-1337.44332	x3 x5 x6 x8

數目	$R^2$	調整 $R^2$	C(p)	AIC	MSE	SBC	模型中的變數
4	<b>0.4998</b>	0.4940	33.2825	-1355.8705	0.02048	-1336.58080	x2 x3 x4 x6
4	<b>0.4984</b>	0.4926	34.2942	-1354.9232	0.02054	-1335.63357	x2 x3 x6 x8
4	<b>0.4983</b>	0.4925	34.3645	-1354.8575	0.02054	-1335.56782	x2 x5 x6 x7
4	<b>0.4949</b>	0.4891	36.9072	-1352.4883	0.02068	-1333.19864	x3 x4 x6 x8
4	<b>0.4855</b>	0.4795	43.9454	-1346.0128	0.02107	-1326.72312	x1 x3 x4 x7
4	<b>0.4851</b>	0.4791	44.2385	-1345.7457	0.02109	-1326.45607	x1 x3 x7 x8
4	<b>0.4815</b>	0.4755	46.9269	-1343.3054	0.02123	-1324.01571	x1 x2 x5 x6
4	<b>0.4789</b>	0.4729	48.8597	-1341.5614	0.02134	-1322.27173	x1 x2 x3 x7
4	<b>0.4786</b>	0.4725	49.1040	-1341.3416	0.02135	-1322.05195	x1 x3 x5 x7
4	<b>0.4777</b>	0.4716	49.7953	-1340.7203	0.02139	-1321.43066	x1 x3 x4 x5
4	<b>0.4776</b>	0.4715	49.8672	-1340.6558	0.02139	-1321.36611	x1 x3 x4 x8
4	<b>0.4775</b>	0.4714	49.9149	-1340.6130	0.02140	-1321.32331	x1 x2 x3 x4
4	<b>0.4763</b>	0.4702	50.8296	-1339.7929	0.02145	-1320.50321	x1 x3 x5 x8
4	<b>0.4759</b>	0.4698	51.1331	-1339.5212	0.02146	-1320.23154	x1 x2 x3 x8
4	<b>0.4737</b>	0.4676	52.7745	-1338.0555	0.02155	-1318.76579	x2 x4 x5 x6
4	<b>0.4734</b>	0.4673	52.9619	-1337.8885	0.02156	-1318.59887	x2 x5 x6 x8
4	<b>0.4716</b>	0.4655	54.3009	-1336.6979	0.02164	-1317.40827	x1 x2 x3 x5
4	<b>0.4687</b>	0.4625	56.5113	-1334.7414	0.02176	-1315.45172	x3 x4 x7 x8
4	<b>0.4684</b>	0.4622	56.7378	-1334.5415	0.02177	-1315.25185	x2 x3 x7 x8
4	<b>0.4683</b>	0.4622	56.7503	-1334.5305	0.02177	-1315.24082	x3 x5 x7 x8
4	<b>0.4682</b>	0.4620	56.8510	-1334.4416	0.02178	-1315.15197	x2 x3 x4 x7

數目	$R^2$	調整 $R^2$	C(p)	AIC	MSE	SBC	模型中的變數
4	<b>0.4681</b>	0.4620	56.8987	-1334.3996	0.02178	-1315.10993	x3 x4 x5 x7
4	<b>0.4640</b>	0.4578	59.9809	-1331.6920	0.02195	-1312.40238	x2 x3 x5 x7
4	<b>0.4612</b>	0.4549	62.1008	-1329.8419	0.02207	-1310.55228	x5 x6 x7 x8
4	<b>0.4610</b>	0.4548	62.2026	-1329.7533	0.02207	-1310.46368	x1 x5 x6 x7
4	<b>0.4601</b>	0.4538	62.9266	-1329.1239	0.02211	-1309.83420	x4 x5 x6 x7
4	<b>0.4592</b>	0.4530	63.5347	-1328.5961	0.02214	-1309.30642	x2 x3 x4 x5
4	<b>0.4591</b>	0.4529	63.6113	-1328.5296	0.02215	-1309.23997	x3 x4 x5 x8
4	<b>0.4589</b>	0.4526	63.8230	-1328.3461	0.02216	-1309.05645	x2 x3 x4 x8
4	<b>0.4586</b>	0.4523	64.0397	-1328.1583	0.02217	-1308.86867	x2 x3 x5 x8
4	<b>0.4441</b>	0.4376	74.8565	-1318.9115	0.02277	-1299.62181	x1 x2 x6 x7
4	<b>0.4436</b>	0.4371	75.2346	-1318.5927	0.02279	-1299.30301	x2 x6 x7 x8
4	<b>0.4425</b>	0.4361	75.9950	-1317.9523	0.02283	-1298.66263	x2 x4 x6 x7
4	<b>0.4324</b>	0.4258	83.5514	-1311.6517	0.02324	-1292.36204	x4 x5 x6 x8
4	<b>0.4324</b>	0.4258	83.5810	-1311.6272	0.02324	-1292.33757	x1 x4 x5 x6
4	<b>0.4323</b>	0.4257	83.6362	-1311.5816	0.02325	-1292.29197	x1 x5 x6 x8
4	<b>0.4128</b>	0.4060	98.1899	-1299.7596	0.02405	-1280.46991	x1 x2 x5 x7
4	<b>0.4107</b>	0.4038	99.7757	-1298.4952	0.02413	-1279.20557	x2 x5 x7 x8
4	<b>0.4103</b>	0.4034	100.0703	-1298.2609	0.02415	-1278.97119	x2 x4 x5 x7
4	<b>0.4012</b>	0.3943	106.8186	-1292.9345	0.02452	-1273.64484	x4 x5 x7 x8
4	<b>0.3997</b>	0.3927	108.0033	-1292.0078	0.02458	-1272.71809	x1 x5 x7 x8
4	<b>0.3990</b>	0.3920	108.5047	-1291.6163	0.02461	-1272.32659	x1 x4 x5 x7

數目	$R^2$	調整 $R^2$	C(p)	AIC	MSE	SBC	模型中的變數
4	<b>0.3887</b>	0.3816	116.1549	-1285.6965	0.02503	-1266.40688	x1 x2 x4 x5
4	<b>0.3881</b>	0.3810	116.6063	-1285.3504	0.02506	-1266.06070	x2 x4 x5 x8
4	<b>0.3877</b>	0.3806	116.8980	-1285.1269	0.02507	-1265.83724	x1 x2 x5 x8
4	<b>0.3740</b>	0.3667	127.1544	-1277.3570	0.02564	-1258.06735	x1 x4 x5 x8
4	<b>0.3715</b>	0.3642	129.0133	-1275.9671	0.02574	-1256.67739	x2 x4 x7 x8
4	<b>0.3680</b>	0.3606	131.6436	-1274.0097	0.02588	-1254.72002	x1 x2 x7 x8
4	<b>0.3677</b>	0.3604	131.8377	-1273.8657	0.02589	-1254.57602	x1 x2 x4 x7
4	<b>0.3398</b>	0.3322	152.6342	-1258.7695	0.02703	-1239.47988	x1 x6 x7 x8
4	<b>0.3390</b>	0.3313	153.2586	-1258.3262	0.02707	-1239.03651	x1 x4 x6 x7
4	<b>0.3313</b>	0.3235	159.0455	-1254.2439	0.02739	-1234.95424	x4 x6 x7 x8
4	<b>0.3174</b>	0.3095	169.3528	-1247.0886	0.02795	-1227.79898	x1 x4 x7 x8
4	<b>0.3137</b>	0.3058	172.1129	-1245.1971	0.02810	-1225.90745	x1 x2 x4 x6
4	<b>0.3129</b>	0.3050	172.7261	-1244.7783	0.02814	-1225.48866	x2 x4 x6 x8
4	<b>0.3111</b>	0.3031	174.0756	-1243.8583	0.02821	-1224.56867	x1 x2 x6 x8
4	<b>0.2574</b>	0.2488	214.1308	-1217.5977	0.03041	-1198.30802	x1 x2 x4 x8
4	<b>0.1057</b>	0.0954	327.3391	-1152.5335	0.03662	-1133.24384	x1 x4 x6 x8
5	<b>0.5315</b>	0.5247	11.5973	-1376.8142	0.01924	-1353.66665	x1 x2 x3 x6 x7
5	<b>0.5286</b>	0.5217	13.7785	-1374.6374	0.01936	-1351.48975	x1 x3 x4 x6 x7
5	<b>0.5266</b>	0.5198	15.2427	-1373.1836	0.01944	-1350.03598	x1 x3 x6 x7 x8
5	<b>0.5257</b>	0.5188	15.9434	-1372.4900	0.01948	-1349.34237	x1 x2 x3 x5 x6
5	<b>0.5230</b>	0.5161	17.9581	-1370.5034	0.01959	-1347.35578	x1 x3 x5 x6 x7

數目	$R^2$	調整 $R^2$	C(p)	AIC	MSE	SBC	模型中的變數
5	<b>0.5226</b>	0.5156	18.2720	-1370.1949	0.01961	-1347.04726	x1 x2 x3 x4 x6
5	<b>0.5207</b>	0.5137	19.6876	-1368.8070	0.01969	-1345.65939	x1 x3 x4 x5 x6
5	<b>0.5204</b>	0.5134	19.8936	-1368.6054	0.01970	-1345.45780	x1 x2 x3 x6 x8
5	<b>0.5188</b>	0.5118	21.0952	-1367.4323	0.01976	-1344.28467	x2 x3 x5 x6 x7
5	<b>0.5187</b>	0.5117	21.1989	-1367.3312	0.01977	-1344.18362	x1 x3 x5 x6 x8
5	<b>0.5178</b>	0.5108	21.8184	-1366.7281	0.01980	-1343.58054	x2 x3 x6 x7 x8
5	<b>0.5176</b>	0.5106	21.9877	-1366.5635	0.01981	-1343.41588	x2 x3 x4 x6 x7
5	<b>0.5147</b>	0.5077	24.1196	-1364.4968	0.01993	-1341.34922	x1 x3 x4 x6 x8
5	<b>0.5138</b>	0.5067	24.8389	-1363.8023	0.01997	-1340.65468	x3 x4 x5 x6 x7
5	<b>0.5136</b>	0.5065	24.9837	-1363.6626	0.01998	-1340.51504	x3 x5 x6 x7 x8
5	<b>0.5109</b>	0.5038	26.9580	-1361.7642	0.02009	-1338.61658	x3 x4 x6 x7 x8
5	<b>0.5083</b>	0.5012	28.9084	-1359.8987	0.02019	-1336.75107	x2 x3 x4 x5 x6
5	<b>0.5082</b>	0.5010	29.0064	-1359.8052	0.02020	-1336.65757	x2 x3 x5 x6 x8
5	<b>0.5073</b>	0.5002	29.6543	-1359.1879	0.02023	-1336.04030	x1 x2 x5 x6 x7
5	<b>0.5021</b>	0.4948	33.5823	-1355.4683	0.02045	-1332.32066	x3 x4 x5 x6 x8
5	<b>0.4998</b>	0.4925	35.2789	-1353.8739	0.02054	-1330.72632	x2 x3 x4 x6 x8
5	<b>0.4995</b>	0.4922	35.5314	-1353.6372	0.02056	-1330.48960	x2 x5 x6 x7 x8
5	<b>0.4984</b>	0.4911	36.3449	-1352.8758	0.02060	-1329.72819	x2 x4 x5 x6 x7
5	<b>0.4860</b>	0.4785	45.5619	-1344.3626	0.02111	-1321.21496	x1 x3 x4 x7 x8
5	<b>0.4856</b>	0.4782	45.8432	-1344.1060	0.02113	-1320.95835	x1 x3 x4 x5 x7
5	<b>0.4855</b>	0.4780	45.9329	-1344.0242	0.02113	-1320.87663	x1 x2 x3 x4 x7



數目	$R^2$	調整 $R^2$	C(p)	AIC	MSE	SBC	模型中的變數
5	<b>0.4852</b>	0.4778	46.1342	-1343.8407	0.02114	-1320.69312	x1 x2 x3 x7 x8
5	<b>0.4851</b>	0.4776	46.2249	-1343.7581	0.02115	-1320.61048	x1 x3 x5 x7 x8
5	<b>0.4817</b>	0.4741	48.7910	-1341.4284	0.02129	-1318.28076	x1 x2 x4 x5 x6
5	<b>0.4816</b>	0.4741	48.8541	-1341.3713	0.02129	-1318.22370	x1 x2 x5 x6 x8
5	<b>0.4789</b>	0.4714	50.8456	-1339.5741	0.02140	-1316.42652	x1 x2 x3 x5 x7
5	<b>0.4778</b>	0.4702	51.7257	-1338.7828	0.02145	-1315.63524	x1 x3 x4 x5 x8
5	<b>0.4777</b>	0.4701	51.7947	-1338.7209	0.02145	-1315.57325	x1 x2 x3 x4 x5
5	<b>0.4776</b>	0.4700	51.8637	-1338.6589	0.02146	-1315.51130	x1 x2 x3 x4 x8
5	<b>0.4763</b>	0.4687	52.7784	-1337.8387	0.02151	-1314.69108	x1 x2 x3 x5 x8
5	<b>0.4748</b>	0.4671	53.9493	-1336.7915	0.02157	-1313.64392	x2 x4 x5 x6 x8
5	<b>0.4687</b>	0.4610	58.4648	-1332.7824	0.02182	-1309.63481	x2 x3 x4 x7 x8
5	<b>0.4687</b>	0.4610	58.4960	-1332.7549	0.02182	-1309.60726	x3 x4 x5 x7 x8
5	<b>0.4684</b>	0.4606	58.7372	-1332.5420	0.02183	-1309.39444	x2 x3 x5 x7 x8
5	<b>0.4682</b>	0.4605	58.8253	-1332.4643	0.02184	-1309.31670	x2 x3 x4 x5 x7
5	<b>0.4624</b>	0.4546	63.1777	-1328.6463	0.02208	-1305.49873	x1 x5 x6 x7 x8
5	<b>0.4621</b>	0.4543	63.4232	-1328.4322	0.02209	-1305.28465	x4 x5 x6 x7 x8
5	<b>0.4612</b>	0.4534	64.0793	-1327.8606	0.02213	-1304.71305	x1 x4 x5 x6 x7
5	<b>0.4593</b>	0.4515	65.4740	-1326.6487	0.02221	-1303.50112	x2 x3 x4 x5 x8
5	<b>0.4459</b>	0.4378	75.5073	-1318.0516	0.02276	-1294.90399	x2 x4 x6 x7 x8
5	<b>0.4452</b>	0.4372	75.9877	-1317.6453	0.02278	-1294.49766	x1 x2 x6 x7 x8
5	<b>0.4441</b>	0.4360	76.8564	-1316.9116	0.02283	-1293.76398	x1 x2 x4 x6 x7

數目	$R^2$	調整 $R^2$	C(p)	AIC	MSE	SBC	模型中的變數
5	0.4331	0.4249	85.0493	-1310.0669	0.02328	-1286.91928	x1 x4 x5 x6 x8
5	0.4135	0.4049	99.7006	-1298.1506	0.02409	-1275.00302	x1 x2 x5 x7 x8
5	0.4134	0.4049	99.7313	-1298.1261	0.02409	-1274.97847	x2 x4 x5 x7 x8
5	0.4129	0.4044	100.1177	-1297.8173	0.02411	-1274.66969	x1 x2 x4 x5 x7
5	0.4018	0.3931	108.4337	-1291.2361	0.02457	-1268.08852	x1 x4 x5 x7 x8
5	0.3904	0.3815	116.9147	-1284.6495	0.02504	-1261.50187	x1 x2 x4 x5 x8
5	0.3716	0.3625	130.9483	-1274.0156	0.02581	-1250.86796	x1 x2 x4 x7 x8
5	0.3413	0.3318	153.5243	-1257.5589	0.02705	-1234.41132	x1 x4 x6 x7 x8
5	0.3140	0.3040	173.9142	-1243.3330	0.02817	-1220.18538	x1 x2 x4 x6 x8
6	0.5387	0.5306	8.2368	-1380.2225	0.01900	-1353.21698	x1 x2 x3 x4 x6 x7
6	0.5385	0.5304	8.4168	-1380.0396	0.01901	-1353.03404	x1 x2 x3 x6 x7 x8
6	0.5371	0.5290	9.4203	-1379.0212	0.01907	-1352.01570	x1 x2 x3 x5 x6 x7
6	0.5311	0.5229	13.8767	-1374.5347	0.01931	-1347.52915	x1 x3 x4 x5 x6 x7
6	0.5301	0.5219	14.6596	-1373.7524	0.01936	-1346.74686	x1 x3 x5 x6 x7 x8
6	0.5301	0.5219	14.6685	-1373.7435	0.01936	-1346.73798	x1 x2 x3 x4 x5 x6
6	0.5295	0.5213	15.1035	-1373.3096	0.01938	-1346.30404	x1 x2 x3 x5 x6 x8
6	0.5288	0.5206	15.6031	-1372.8119	0.01941	-1345.80641	x1 x3 x4 x6 x7 x8
6	0.5226	0.5142	20.2610	-1368.2057	0.01966	-1341.20019	x1 x2 x3 x4 x6

數目	$R^2$	調整 $R^2$	C(p)	AIC	MSE	SBC	模型中的變數
							x8
6	0.5224	0.5140	20.4103	-1368.0590	0.01967	-1341.05345	x2 x3 x5 x6 x7 x8
6	0.5216	0.5132	21.0114	-1367.4692	0.01971	-1340.46367	x2 x3 x4 x5 x6 x7
6	0.5207	0.5123	21.6679	-1366.8262	0.01974	-1339.82065	x1 x3 x4 x5 x6 x8
6	0.5182	0.5098	23.5510	-1364.9883	0.01985	-1337.98275	x2 x3 x4 x6 x7 x8
6	0.5142	0.5057	26.5316	-1362.0989	0.02001	-1335.09334	x3 x4 x5 x6 x7 x8
6	0.5088	0.5002	30.5280	-1358.2617	0.02023	-1331.25621	x1 x2 x5 x6 x7 x8
6	0.5085	0.4999	30.7878	-1358.0137	0.02025	-1331.00822	x2 x3 x4 x5 x6 x8
6	0.5074	0.4988	31.5945	-1357.2448	0.02029	-1330.23928	x1 x2 x4 x5 x6 x7
6	0.5009	0.4922	36.4329	-1352.6680	0.02056	-1325.66246	x2 x4 x5 x6 x7 x8
6	0.4861	0.4771	47.4875	-1342.4305	0.02117	-1315.42492	x1 x3 x4 x5 x7 x8
6	0.4861	0.4771	47.5316	-1342.3902	0.02117	-1315.38467	x1 x2 x3 x4 x7 x8
6	0.4857	0.4767	47.8268	-1342.1210	0.02119	-1315.11542	x1 x2 x3 x4 x5 x7
6	0.4853	0.4763	48.1143	-1341.8589	0.02120	-1314.85334	x1 x2 x3 x5 x7 x8
6	0.4830	0.4740	49.7911	-1340.3343	0.02129	-1313.32879	x1 x2 x4 x5 x6 x8

數目	$R^2$	調整 $R^2$	C(p)	AIC	MSE	SBC	模型中的變數
6	0.4778	0.4686	53.7234	-1336.7849	0.02151	-1309.77936	x1 x2 x3 x4 x5 x8
6	0.4687	0.4594	60.4515	-1330.7942	0.02188	-1303.78864	x2 x3 x4 x5 x7 x8
6	0.4633	0.4539	64.5039	-1327.2347	0.02211	-1300.22917	x1 x4 x5 x6 x7 x8
6	0.4476	0.4380	76.2022	-1317.1580	0.02275	-1290.15251	x1 x2 x4 x6 x7 x8
6	0.4163	0.4061	99.5847	-1297.8467	0.02404	-1270.84114	x1 x2 x4 x5 x7 x8
7	0.5427	0.5334	7.2332	-1381.2897	0.01889	-1350.42628	x1 x2 x3 x5 x6 x7 x8
7	0.5422	0.5328	7.6460	-1380.8667	0.01891	-1350.00322	x1 x2 x3 x4 x5 x6 x7
7	0.5393	0.5299	9.7803	-1378.6869	0.01903	-1347.82347	x1 x2 x3 x4 x6 x7 x8
7	0.5315	0.5219	15.6099	-1372.8016	0.01935	-1341.93816	x1 x3 x4 x5 x6 x7 x8
7	0.5303	0.5207	16.5290	-1371.8827	0.01940	-1341.01922	x1 x2 x3 x4 x5 x6 x8
7	0.5224	0.5127	22.3746	-1366.0941	0.01973	-1335.23064	x2 x3 x4 x5 x6 x7 x8
7	0.5104	0.5004	31.3710	-1357.3683	0.02023	-1326.50480	x1 x2 x4 x5 x6 x7 x8
7	0.4862	0.4756	49.4530	-1340.4620	0.02123	-1309.59853	x1 x2 x3 x4 x5 x7 x8
8	0.5430	0.5323	9.0000	-1379.5291	0.01893	-1344.80767	x1 x2 x3 x4 x5 x6 x7 x8

## 一、複判定係數法( $R^2$ )

選取最大的 $R^2$ 為最有效的迴歸模型，但根據差異判斷最有效模型。

$$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}$$

$R^2=0.5387$  時，SAS 系統投入的變數為  $X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_6$ 、 $X_7$

$R^2=0.5328$  時，SAS 系統投入的變數為  $X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_5$ 、 $X_6$ 、 $X_7$

$R^2=0.5299$  時，SAS 系統投入的變數為  $X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_6$ 、 $X_7$ 、 $X_8$

整體 $R^2$ 皆為中度相關，但彼此差異不大。當系統投入  $X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_6$ 、 $X_7$  時，其 $R^2=0.5387$ 。而系統投入  $X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_5$ 、 $X_6$ 、 $X_7$  時，其 $R^2=0.5328$ 。差異只有 0.0059，有此可知有沒有投入  $X_5$  影響並不大。而系統投入  $X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_6$ 、 $X_7$ 、 $X_8$  時，其 $R^2=0.5299$ 。差異只有 0.0388，有此可知有沒有投入  $X_8$  影響並不大。所以最後選擇投入  $X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_6$ 、 $X_7$  等六個解釋變數建立最佳迴歸模型。

## 二、 $C_p$ 準則

$$C_p = \frac{SSE_p}{MSE} - (n - 2p), \text{ 其中 } P \text{ 為參數 } \beta \text{ 之個數。}$$

$C_p$  值較小時，總均方誤差會很小，迴歸模型的偏誤亦會較小，且當 $C_p$ 值小於且最接近的 $p$ 時，即為最佳迴歸模型組合。

$C_p=7.6460 < P=8$ ，因此選擇  $X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_5$ 、 $X_6$ 、 $X_7$  等七個解釋變數建立最佳迴歸模型。

### 三、SBC、AIC 法

選取 SBC、AIC 值最小者為最佳的迴歸模型。

$$AIC = n \ln(SSE) - n \ln n + 2p$$

$$SBC = n \ln(SSE) - n \ln n + [\ln \cdot n]p$$

AIC=-1381.2897 時為最小值，SAS 系統選擇  $X_1$ 、 $X_2$ 、 $X_3$ 、 $X_5$ 、 $X_6$ 、 $X_7$ 、 $X_8$  等七個變數來建立最佳模型組合。

SBC=-1353.21698 時為最小值，SAS 系統選擇  $X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_6$ 、 $X_7$  等六個變數來建立最佳模型組合。

### 第五節 結論

表 4-8 模型選取方法與結論

方法	選擇之變數
前進選取法	$X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_6$ 、 $X_7$
後退刪去法	$X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_6$ 、 $X_7$
逐步回歸法	$X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_6$ 、 $X_7$
複判定係數法	$X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_6$ 、 $X_7$
Cp 準則	$X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_5$ 、 $X_6$ 、 $X_7$
SBC 分析法	$X_1$ 、 $X_2$ 、 $X_3$ 、 $X_5$ 、 $X_6$ 、 $X_7$ 、 $X_8$
AIC 分析法	$X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_6$ 、 $X_7$

綜合以上各種檢定方法，可知我們應剔除 $X_5$ 、 $X_8$ 來進行迴歸分析，以求得最佳迴歸模型：

$$\hat{Y} = 11.80876 - 0.03934 X_1 + 0.04462 X_2 + 0.00013531 X_3 - 5.3687E-7 X_4 \\ - 0.00182 X_6 + 0.00005548 X_7$$

## 第伍章 離群值及影響點之檢定

當離群值或具影響力的觀察值出現時，常常是造成迴歸分析中違反常態性或是變異數齊一性的重要原因。所以我們分別找出屬於離群值跟具有影響力的資料，若同時具有離群值跟影響力的特性時，我們將予以刪除。

### 第一節 離群值

最初模型：

$$\hat{Y} = 11.80876 - 0.03934 X_1 + 0.04462 X_2 + 0.00013531 X_3 - 5.3687E-7 X_4 \\ - 0.00182 X_6 + 0.00005548 X_7$$

#### 一、標準化殘差值

通常標準化殘差在正負兩個標準差之外即為離群值。本資料共選出 15 筆離群值，分別為第 8、18、47、72、104、140、148、150、158、167、174、241、263、273、346 筆觀察值。

表 5-1 離群值分析

觀測值	應變數	預測值	標準誤差 平均值 預測	殘差	標準誤 殘差	Student 殘差
8	8.92	8.6436	0.0216	0.2734	0.136	2.008
18	9.09	8.6714	0.0130	0.4216	0.137	3.072
47	8.32	8.6401	0.0130	-0.3227	0.137	-2.352
72	9.27	8.9414	0.0283	0.3288	0.135	2.437
104	9.03	8.7591	0.0252	0.2756	0.136	2.034
140	8.14	8.4534	0.0105	-0.3162	0.137	-2.301
148	8.29	8.5916	0.0188	-0.2991	0.137	-2.191

150	8.74	9.1637	0.0348	-0.4250	0.133	-3.187
158	8.40	8.6825	0.0155	-0.2790	0.137	-2.037
167	8.24	8.5313	0.0114	-0.2939	0.137	-2.139
174	8.07	8.4367	0.0166	-0.3654	0.137	-2.670
241	8.38	8.7850	0.0161	-0.4095	0.137	-2.991
263	8.85	8.5227	0.0138	0.3318	0.137	2.419
273	8.08	8.3786	0.0161	-0.2955	0.137	-2.158
346	8.86	8.5524	0.0122	0.3058	0.137	2.227

## 二、Y outlier

根據 student residual，當  $|t_i| \geq t_{(1-\frac{\alpha}{2n}, n-p-1)}$ ，則第 i 個點即為 Y outlier，p 為參數  $\beta$  的個數。根據此資料在顯著水準  $\alpha=0.05$  之下，且  $p=7$ ， $n=350$ ，當  $|t_i| \geq t_{(1-\frac{0.05}{2*350}, 350-7-1)} \doteq 3.1142$  時即為離群值。研究資料中無 Y 的離群值存在。

## 三、X outlier

利用帽子矩陣(Hat Diag H)，當  $H > \frac{2p}{n} = \frac{2*7}{350} = 0.04$ ，此資料為離群值，經由 sas 程式得出下表結果，可知觀測值 2、3、15、21、30、33、70、72、97、98、114、135、149、150、154、181、203、205、230、249、256、275、279、280、297、300、301、317、335、337、339，共 31 筆為離群值。

表 5-2 帽子矩陣分析

obs	Hat Diag H
2	0.0596
3	0.0412
15	0.0425
21	0.0447
30	0.1006
33	0.0459
70	0.1299
72	0.0422



97	0.0691
98	0.0836
114	0.0469
135	0.0457
149	0.0812
150	0.0639
154	0.0602
181	0.0451
203	0.0769
205	0.0446
230	0.1916
249	0.0501
256	0.1125
275	0.055
279	0.0748
280	0.0766
297	0.2424
300	0.0901
301	0.0539
317	0.0631
335	0.0676
337	0.0506
339	0.067

## 第二節 影響點

### 一、DFFITS

$$DFFITS_i = \frac{\hat{Y}_1 - \hat{Y}_{I(I)}}{\sqrt{MSE_{(i)}h_{ii}}}$$

此為個案i對配適值 $\hat{Y}_1$ 的影響力（DFFITS），當 $|DFFITS| > 2\sqrt{\frac{P}{N}} = 2\sqrt{\frac{7}{350}} = 0.28281$ 時，即為影響點。在此資料中共23筆資料為影響點，分別為 8、18、30、33、72、97、98、104、113、147、148、150、174、194、197、201、230、241、249、256、301、317、339。

表 5-3 DFFITS 影響力分析

obs	dffits
8	0.31970
18	0.29451

obs	dffits
30	-0.41013
33	0.29465
72	0.51512
97	0.37319
98	0.33739
104	0.38003
113	0.29103
147	0.30136
148	-0.30373
150	-0.84384
174	-0.32704
194	0.30737
197	-0.33221
201	0.37910
230	-0.42931
241	-0.35686
249	0.28654
256	0.39638
301	0.35627

obs	dffits
317	0.32623
339	0.36075

## 二、COOK'S D

此為個案  $i$  對所有配適值  $\hat{Y}_1$  的影響力 (COOK'S D)，當  $\text{COOK'S D} > F_{0.5}(p, n-p)$  時，此筆資料即為影響點。在本研究中，當  $\text{COOK'S D} > F_{0.5}(7, 342) = 0.908$  時為影響點，資料中並無影響點存在。

## 三、DFBETAS

此為個案  $i$  對第  $k$  個迴歸係數的影響力，當  $|\text{DFBETAS}_{k(i)}| > 1$  (or  $n \rightarrow \infty \text{ DFBETAS} > \frac{2}{\sqrt{n}}$ )，表示第  $i$  個案對迴歸係數  $b_k$  具影響力。本研究資料中無影響點存在。

## 第三節 結論

表 5-4 離群值與影響點分析之結論

離群值	2、3、8、15、18、21、30、33、47、70、72、97、98、104、114、135、140、148、149、150、154、158、167、174、181、203、205、230、241、249、256、263、273、275、279、280、297、300、301、317、335、337、339、346、
影響點	8、18、30、33、72、97、98、104、113、147、148、150、174、194、197、201、230、241、249、256、301、317、339

當一筆資料同時為離群值和影響點時予以刪除，兩者的交集分別為觀察值 8、18、30、33、72、97、98、104、148、150、174、230、241、249、256、301、317、339，我們將這 18 筆資料從中移除。

表 5-5 離群值與影響點分析之結論

obs	student	cookd	Hat Diag H	Rstudent	DFFITS
8	2.00804	0.014471	0.02451	2.01700	0.31970
18	3.07236	0.012085	0.00888	3.11098	0.29451
30	-1.22548	0.023994	0.10059	-1.22638	-0.41013
33	1.34120	0.012374	0.04594	1.34276	0.29465
72	2.43742	0.037360	0.04216	2.45522	0.51512
97	1.36749	0.019845	0.06915	1.36924	0.37319
98	1.11669	0.016250	0.08359	1.11709	0.33739
104	2.03401	0.020442	0.03343	2.04340	0.38003
113	1.51811	0.012053	0.03532	1.52101	0.29103
147	1.64189	0.012910	0.03243	1.64598	0.30136
148	-2.19069	0.013032	0.01865	-2.20296	-0.30373
150	-3.18667	0.099000	0.06388	-3.23020	-0.84384
174	-2.67023	0.015006	0.01452	-2.69449	-0.32704
194	1.78604	0.013410	0.02859	1.79179	0.30737
197	-1.64658	0.015687	0.03893	-1.65072	-0.33221
201	1.94252	0.020365	0.03640	1.95044	0.37910
230	-0.88218	0.026347	0.19158	-0.88190	-0.42931
241	-2.99120	0.017770	0.01371	-3.02657	-0.35686
249	1.24660	0.011710	0.05010	1.24761	0.28654
256	1.11275	0.022430	0.11253	1.11313	0.39638

obs	student	cookd	Hat Diag H	Rstudent	DFFITS
301	1.48969	0.018068	0.05392	1.49236	0.35627
317	1.25640	0.015178	0.06306	1.25747	0.32623
339	1.34498	0.018548	0.06697	1.34657	0.36075

經過進一步的校正後，建立最適迴歸模型：

$$\hat{Y} = 12.21085 - 0.04529 X_1 + 0.04553 X_2 + 0.0001417 X_3 - 7.05091E-7 X_4 + 0.00202 X_6 + 0.00005175 X_7$$

## 第陸章 殘差檢定

### 第一節 均質性

模型為：

$$\hat{Y} = 12.21085 - 0.04529 X_1 + 0.04553 X_2 + 0.0001417 X_3 - 7.05091E-7 X_4 + 0.00202 X_6 + 0.00005175 X_7$$

### 一、White test

表 6-1 均質性檢定

第一個和第二個動差規格的檢定		
DF	卡方	Pr > ChiSq
27	39.13	0.0616

統計假設如下：

$H_0$ : 誤差項無異質變異

$H_1$ : 誤差項有異質變異

從SAS結果可知，p-value=0.0616>0.05=α，所以不拒絕 $H_0$ ，故沒有足夠的證據證明誤差項有異質變異，即符合均質性假設。

## 二、殘差分析圖

刪除資料前的原始模型：

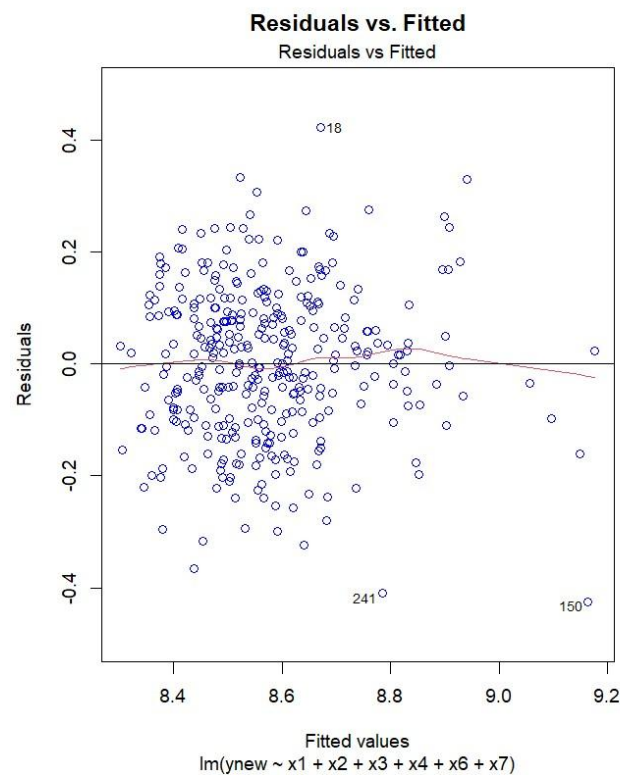


圖 6-1 原始模型 Residuals vs. Fitted plot

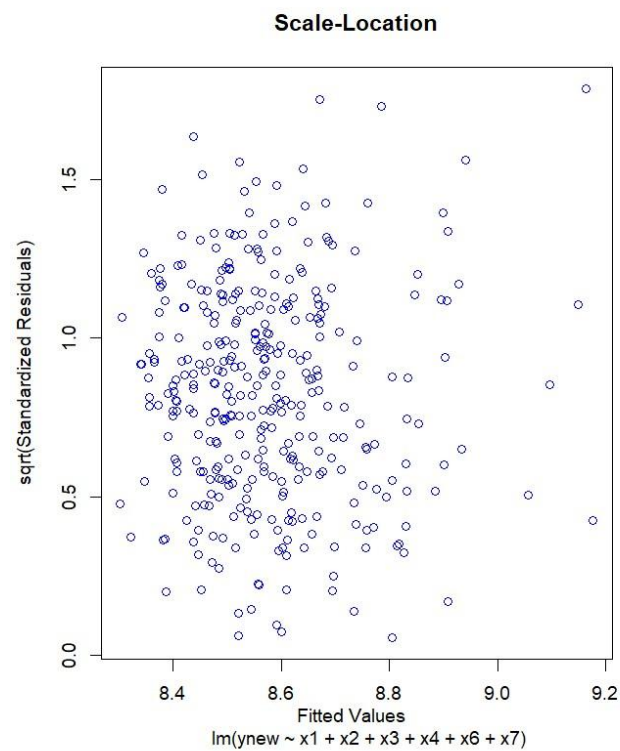


圖 6-2 原始模型 Standardized Residuals vs. Fitted plot

刪除資料後的配適模型：

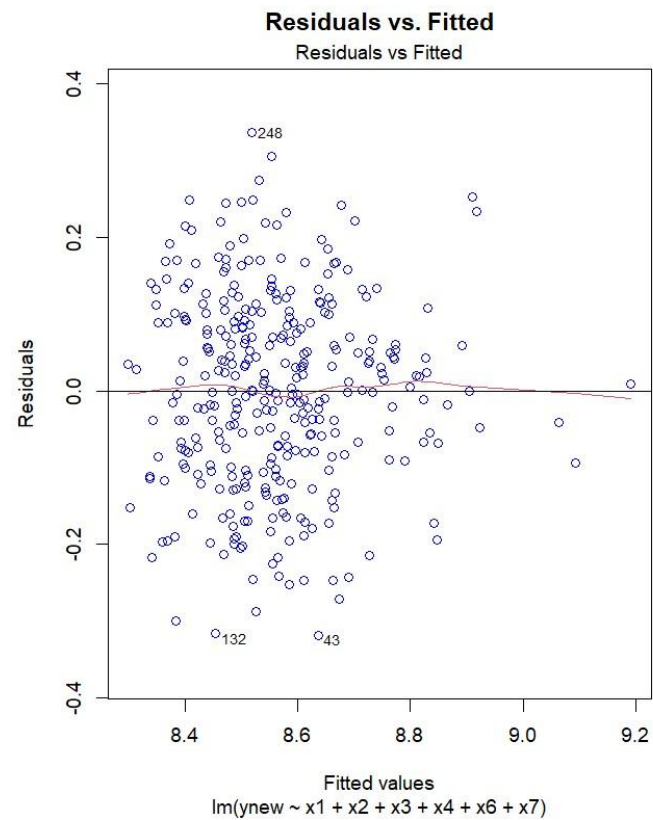


圖 6-3 配適模型 Residuals vs. Fitted plot

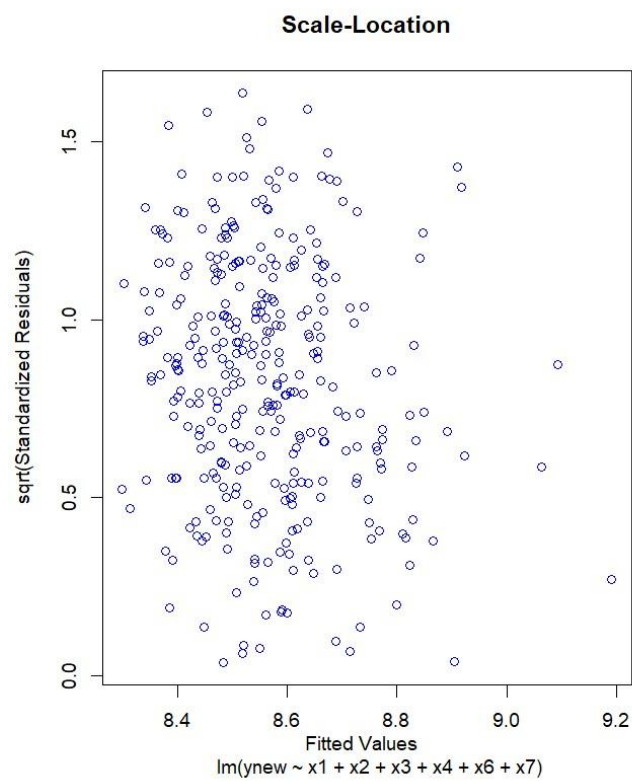


圖 6-4 配適模型 Standardized Residuals vs. Fitted plot

從上圖可觀察出，殘差值無明顯隨著配適值變大而有變大或變小的趨勢，而是都在以 0 為基準的橫線上，上下隨機分佈，符合均質性假設。

表 6-2 分位數殘差值

分位數 (定義 5)	
層級	分位數
100% Max	0.331781488
99%	0.262842419
95%	0.205530669
90%	0.159575923
75% Q3	0.087323858
50% 中位數	0.000164516
25% Q1	-0.095995652
10%	-0.176455871
5%	-0.210473599
1%	-0.293866460
0% 最小值	-0.322743956

由表 6-2 可知，有 90% 的殘差值小於 0.159575923，大多落在零附近。



## 第二節 常態性

### 一、常態性檢定

表 6-3 常態性檢定

常態性檢定				
檢定	統計值		p 值	
Shapiro-Wilk	W	0.995032	Pr < W	0.3646
Kolmogorov-Smirnov	D	0.039662	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.068708	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.424924	Pr > A-Sq	>0.2500

統計假設如下：

$H_0$ : 誤差項服從常態分配

$H_1$ : 誤差項不服從常態分配

取顯著水準  $\alpha=0.05$

#### Shapiro-Wilk test

因為  $p\text{-value}=0.3646>0.05=\alpha$ ，所以不拒絕  $H_0$ ，故沒有足夠的證據證明誤差項不服從常態分配，即符合常態性假設。

#### Kolmogorov-Smirnov test

因為  $p\text{-value}=0.15>0.05=\alpha$ ，所以不拒絕  $H_0$ ，故沒有足夠的證據證明誤差項不服從常態分配，即符合常態性假設。

## 二、常態圖

刪除資料前的原始模型：

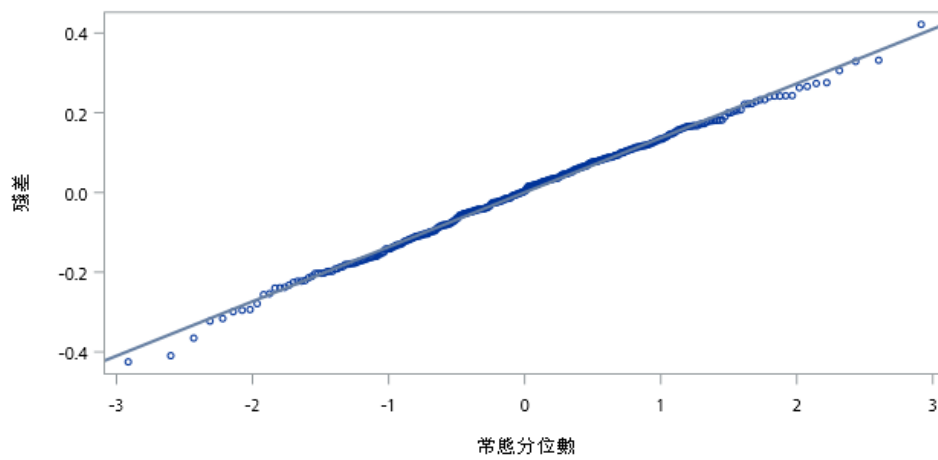


圖 6-5 Q-Q plot-1

刪除資料後的配適模型：

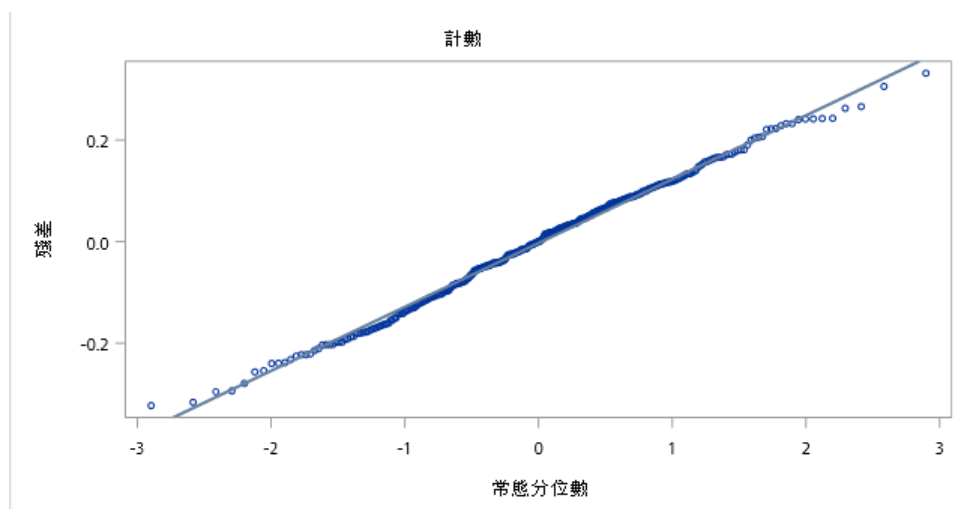


圖 6-6 Q-Q plot-2

由刪除離群值與影響點後的模型，可看出其殘差直線的斜率較靠近1，表示較接近常態分配。

### 第三節 獨立性

#### 一、一階自我相關檢定

表 6-4 獨立性檢定

Durbin-Watson D	1.886
觀測值數目	332
一階自相關	0.055

統計假設如下：

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

#### Durbin-Watson test

判斷方法：當檢定統計量D越趨近於2，則代表樣本間沒有一階自我相關性存在。

由表6-4可知， $D=1.886 \approx 2$ ，所以不拒絕 $H_0$ ，故沒有足夠證據顯示，誤差項有一階自我相關性存在，即表示誤差項符合獨立性假設。

## 第柒章 模型確認

配適模型：

$$\hat{Y} = 12.21085 - 0.04529 X_1 + 0.04553 X_2 + 0.0001417 X_3 - 7.05091E-7 X_4 \\ + 0.00202 X_6 + 0.00005175 X_7$$

### 第一節 最終模型解釋能力

表 7-1 配適模型解釋能力分析

根 MSE	0.12661	$R^2$	0.5425
應變平均值	8.56093	調整 $R^2$	0.5341
變異係數	1.47896		

配適模型：

$$R^2 = 0.5425$$

$$\text{調整} R^2 = 0.5341$$

配適模型的  $R^2$ ，表示此迴歸能解釋 54.25% 的 Y（房價）變異。校正過後的調整  $R^2$  些微降低，為 53.92%，表示配適模型對 Y（房價）具有相當程度的解釋力。

表 7-2 原始模型與配適模型解釋能力之比較

模型	$R^2$	調整 $R^2$
原始模型	0.5387	0.5306
配適模型	0.5425	0.5341

表 7-2 可知，配適模型  $R^2$  與調整後的  $R^2$  皆比原始模型高一點，模型也較精簡，故我們選擇配適模型作為最終模型。

## 第二節 最終模型預測能力

### 一、MAPE

為證明此模型的準確性，我們從原始資料抽取另外 150 筆資料建立測試集，檢測此模型的平均絕對誤差(MAPE)，確認模型是否具有預測能力。

Step 1:

表 7-3 最終模型 ANOVA 表

來源	DF	平方和	均方	F 值	Pr > F
模型	6	6.17835	1.02973	64.23	<.0001
誤差	325	5.20999	0.01603		
已校正的總計	331	11.38834			

Step 2:

表 7-4 確認資料集 ANOVA 表

來源	DF	平方和	均方	F 值	Pr > F
模型	6	4.30324	0.71721	28.66	<.0001
誤差	143	3.57856	0.02502		
已校正的總計	149	7.88179			

Step 3: MAPE(平均絕對預測誤差)

$$MAPE = \frac{1}{n} \sum \left| \frac{y_i - \hat{y}_i}{y_i} \right| = 0.014922$$

$y_i$  = 新資料的反應變數     $\hat{y}_i$  =  $y_i$  的預測值     $n$  = 確認資料集筆數

經運算後， $MAPE = 0.014922$ ，表示所選的最終迴歸模型預測偏離的程度很小，表示此模型在預測上有較高的精確性，且可能有建模資料以外的應用能力。

## 二、最終模型的應用

從原始資料中另外抽取一筆數據（臥室 3 間、浴室 2 間、居住面積 1880 平方英尺、土地面積 3048 平方英尺、建築年份 1880 平方英尺、最近 15 棟房屋的內部居住空間 1680 平方英尺），代入最終模型作測試，得到轉換過後的房價為 8.6 美元。再將  $Y_{new} \rightarrow Y$ ，得預測房價為 519943.5 美元。相較原始資料的 449950 美元有些許誤差。

## 第捌章 結論

本研究主要是想探討藉由美國西雅圖金郡房屋的八個變數：「臥室數量」、「浴室數量」、「居住面積」、「土地面積」、「地上居住面積」、「建築年份」、「最近 15 棟房屋的內部居住空間」、「最近 15 棟房屋的土地平均大小」，了解影響房價的主要因素。我們使用向前選取法、後退刪去法、逐步迴歸法、其他選取法( $R^2$ 、 $C_p$  準則法、AIC 法和 SBC 法)，來篩選影響力較小的變數以得到最適合的迴歸模型。檢定過後，我們選擇刪除「土地面積」及「最近 15 棟房屋的土地平均大小」兩項不適合的預測變數。

利用剩下六個預測變數建立迴歸模型為：

$$\hat{Y} = 11.80876 - 0.03934 X_1 + 0.04462 X_2 + 0.00013531 X_3 - 5.3687E-7 X_4 - 0.00182 X_6 + 0.00005548 X_7$$

接著檢測資料是否存在離群值和影響點，我們發現 18 筆同為離群值及影響點的資料，將其刪除校正後，取得一個新的迴歸模型為：

$$\hat{Y} = 12.21085 - 0.04529 X_1 + 0.04553 X_2 + 0.0001417 X_3 - 7.05091E-7 X_4 + 0.00202 X_6 + 0.00005175 X_7$$

之後以此迴歸模型進行殘差檢定，結果符合三大假設：同質性、常態性、獨立性，故選擇此迴歸模型作為最佳迴歸模型。在最佳迴歸模型中，各參數代表的意義為：

$\beta_0 = 12.21085$ ，代表在不考慮有其他變數影響的情況下，即  $X_1 = X_2 = X_3 = X_4 = X_6 = X_7 = 0$  時， $E(Y) = 12.21085$ 。

$\beta_1 = -0.04529$ ，代表在其他變數不變的情況下，房子每增加 1 間臥室，房價將會降低 0.04529 美元。

$\beta_2 = 0.04553$ ，代表在其他變數不變的情況下，房子每增加 1 間浴室，房價將會上升 0.04553 美元。

$\beta_3 = 0.0001417$ ，代表在其他變數不變的情況下，居住面積每增加 1 平方英尺，房價將會上升 0.0001417 美元。

$\beta_4 = -5.3687E-7$ ，代表在其他變數不變的情況下，土地面積每增加 1 平方英尺，房價將會下降 5.3687E-7 美元。

$\beta_6 = 0.00202$ ，代表在其他變數不變的情況下，建築年份每晚 1 年，房價將會上升 0.00202 美元。

$\beta_7 = 0.00005175$ ，代表在其他變數不變的情況下，最近 15 棟房屋的內部居住空間每增加 1 平方英尺，房價將會上升 0.00005175 美元。

在此報告中，配適模型的  $R^2 = 0.5425$ ，表示此模型可解釋 54.25% 的 Y(房價) 變異，有一定程度的解釋能力。

由上述結論，雖然研究結果顯示配適模型具有解釋能力，但由最終迴歸模型參數中，卻發現各個因素對房價影響不大。我們推測其原因為本研究受限於線性迴歸模型僅能處理數值型資料的限制，未納入類別型資料進行分析，然而房價的複雜性涉及多方面因素，例如政府政策、建築成本上升、地區、交通等，模型未全面考慮這些重要因素。若能使用邏輯迴歸、機器學習等其他具有更大的靈活性、能夠處理不同型別特徵的非線性方法，可能有助於更全面地理解影響房價變動的因素，優化模型以降低預測誤差。

## 第玖章 附錄

### 第一節 資料

#### 一、原始資料(前二十筆)

1	id	date	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	condition	grade	sqft_above	sqft_basement	yr_built	yr_renovated	zipcode	lat	long	sqft_living	sqft_lot
2	7.1E+09	20141013	221900	3	1	1180	5650	1	0	0	3	7	1180	0	1955	0	98178	47.5112	-122.257	1340	5650
3	6.4E+09	20141209	538000	3	2.25	2570	7242	2	0	0	3	7	2170	400	1951	1991	98125	47.721	-122.319	1690	7639
4	5.6E+09	20150225	180000	2	1	770	10000	1	0	0	3	6	770	0	1933	0	98028	47.7379	-122.233	2720	8062
5	2.5E+09	20141209	604000	4	3	1960	5000	1	0	0	5	7	1050	910	1965	0	98136	47.5208	-122.393	1360	5000
6	2E+09	20150218	510000	3	2	1680	8080	1	0	0	3	8	1680	0	1987	0	98074	47.6168	-122.045	1800	7503
7	7.2E+09	20140512	1.23E+06	4	4.5	5420	101930	1	0	0	3	11	3890	1530	2001	0	98053	47.6561	-122.005	4760	101930
8	1.3E+09	20140627	257500	3	2.25	1715	6819	2	0	0	3	7	1715	0	1995	0	98003	47.3097	-122.327	2238	6819
9	2E+09	20150115	291850	3	1.5	1060	9711	1	0	0	3	7	1060	0	1963	0	98198	47.4095	-122.315	1650	9711
10	2.4E+09	20150415	229500	3	1	1780	7470	1	0	0	3	7	1050	730	1960	0	98146	47.5123	-122.337	1780	8113
11	3.8E+09	20150312	323000	3	2.5	1890	6560	2	0	0	3	7	1890	0	2003	0	98038	47.3684	-122.031	2390	7570
12	1.7E+09	20150403	662500	3	2.5	3560	9796	1	0	0	3	8	1860	1700	1965	0	98007	47.6007	-122.145	2210	8925
13	9.2E+09	20140527	468000	2	1	1160	6000	1	0	0	4	7	860	300	1942	0	98115	47.69	-122.292	1330	6000
14	1.1E+08	20140528	310000	3	1	1430	19901	1.5	0	0	4	7	1430	0	1927	0	98028	47.7558	-122.229	1780	12697
15	6.1E+09	20141007	400000	3	1.75	1370	9680	1	0	0	4	7	1370	0	1977	0	98074	47.6127	-122.045	1370	10208
16	1.2E+09	20150312	530000	5	2	1810	4850	1.5	0	0	3	7	1810	0	1900	0	98107	47.67	-122.394	1360	4850
17	9.3E+09	20150124	650000	4	3	2950	5000	2	0	3	3	9	1980	970	1979	0	98126	47.5714	-122.375	2140	4000
18	1.9E+09	20140731	395000	3	2	1890	14040	2	0	0	3	7	1890	0	1994	0	98019	47.7277	-121.962	1890	14018
19	6.9E+09	20140529	485000	4	1	1600	4300	1.5	0	0	4	7	1600	0	1916	0	98103	47.6648	-122.343	1610	4300
20	1.6E+07	20141205	189000	2	1	1200	9850	1	0	0	4	7	1200	0	1921	0	98002	47.3089	-122.21	1060	5095
21	8E+09	20150424	230000	3	1	1250	9774	1	0	0	4	7	1250	0	1969	0	98003	47.3343	-122.306	1280	8850

#### 二、選取資料(前二十筆)

id	y	x1	x2	x3	x4	x5	x6	x7	x8
1	221900	3	1	1180	5650	1180	1955	1340	5650
2	538000	3	2.25	2570	7242	2170	1951	1690	7639
3	180000	2	1	770	10000	770	1933	2720	8062
4	604000	4	3	1960	5000	1050	1965	1360	5000
5	510000	3	2	1680	8080	1680	1987	1800	7503
6	1.23E+06	4	4.5	5420	101930	3890	2001	4760	101930
7	257500	3	2.25	1715	6819	1715	1995	2238	6819
8	291850	3	1.5	1060	9711	1060	1963	1650	9711
9	229500	3	1	1780	7470	1050	1960	1780	8113
10	323000	3	2.5	1890	6560	1890	2003	2390	7570
11	662500	3	2.5	3560	9796	1860	1965	2210	8925
12	468000	2	1	1160	6000	860	1942	1330	6000
13	310000	3	1	1430	19901	1430	1927	1780	12697
14	400000	3	1.75	1370	9680	1370	1977	1370	10208
15	530000	5	2	1810	4850	1810	1900	1360	4850
16	650000	4	3	2950	5000	1980	1979	2140	4000
17	395000	3	2	1890	14040	1890	1994	1890	14018
18	485000	4	1	1600	4300	1600	1916	1610	4300
19	189000	2	1	1200	9850	1200	1921	1060	5095
20	230000	3	1	1250	9774	1250	1969	1280	8850



### 三、訓練資料

y	x1	x2	x3	x4	x5	x6	x7	x8
450000	3	1.5	1520	7903	1000	1955	1520	9830
1470000	4	3.25	5010	34460	5010	1988	4760	34460
439000	2	1.75	1620	113862	1620	1995	1560	54806
252500	2	1.5	1110	986	950	2009	1110	3515
822000	3	2.5	2290	9158	2290	1984	2210	9588
937500	4	2.5	3130	21100	2530	1956	3680	23000
165000	3	1	970	7503	970	1967	1230	9504
1180000	2	2.5	1770	7155	1770	1957	2410	10476
960000	3	2.5	2010	6857	1450	1955	2380	6370
954500	3	2.25	2440	9689	1830	1974	2730	9689
639500	4	2.25	2330	8994	2330	1986	2330	8396
330000	4	2.25	2440	8098	2440	1983	2110	7911
585000	4	3.25	2400	29252	2400	1982	2280	45000
275000	2	1	1180	6552	1180	1949	1070	7200
895000	6	2.5	3550	6533	3550	1925	3140	6234
418000	3	2	1250	1306	1250	2001	1320	1250
480000	2	1	1030	3060	790	1918	1390	3060
1900000	3	2.5	2660	13367	2660	1992	2660	13367
410000	2	2.25	1660	2128	1660	1974	1640	2128
180000	3	1	1020	5500	1020	1961	1160	5500
1060000	5	4.5	4140	7924	4140	2005	3960	8410
580000	2	1	860	4013	860	1925	1490	4013
978000	3	1.5	2390	4000	1690	1936	2350	4000
294700	3	2	1970	9600	1300	1967	1710	7703
425000	3	2.25	1790	10209	1290	1967	1840	9900
1140000	4	3.25	4130	11444	4130	2001	3720	11431
486940	4	2.5	3250	13360	3250	2014	2612	14448
470000	3	2.5	2280	6134	2280	2004	2640	6167
572000	5	2.25	2340	9225	2340	1973	2140	9348
1430000	4	4.25	4960	6000	3680	1909	2160	4080
502000	3	2	1300	14350	1300	1955	1520	10670
480000	4	1.75	1920	9380	1920	1964	1580	8580
1050000	4	3	3560	4000	2970	1996	1190	4000
403000	3	2	1960	13100	1650	1957	1960	10518
461000	3	3.25	2770	6278	1980	2006	1900	7349
757000	3	3.25	3190	5283	3190	2007	2950	5198
355000	3	2.25	2550	9674	1850	1959	2240	9674
220000	2	1	890	4804	890	1928	1010	3844
537000	3	2.5	1400	4800	1200	1921	1440	3840
320000	3	1	860	5060	860	1927	880	5060
515000	3	2.5	2360	11254	2360	1990	2390	11254
475000	2	2.25	1060	925	980	2006	1020	4000

303700	3	2.5	1981	5700	1981	2010	1981	5894
435000	4	2.25	1990	8548	1990	1973	2320	8926
763776	4	2.5	2750	16139	2750	1989	2810	13093
432000	3	1.75	1550	8134	1550	1959	1360	8000
260000	4	2.5	2360	9647	1530	1964	2580	9680
373000	3	1	1770	5720	1140	1926	1500	4406
575000	3	2.25	2400	5000	1440	1926	1630	5000
185000	4	1	1490	6600	1490	1969	1280	6600
380000	2	1	1040	7372	840	1939	1930	5150
775000	5	2	3540	9970	3540	1970	2280	7195
725000	3	2.5	3410	41022	3410	1990	2150	21429
522000	4	2.5	2835	6598	2835	2002	2770	6969
1190000	5	2.5	2710	14989	1720	1959	3180	16624
700000	4	2.5	2490	7694	2490	1987	2140	8126
499950	3	3.5	1820	1501	1430	2014	1550	1501
575000	3	1.75	2130	6500	1170	1954	2020	5000
388500	3	2.25	1350	944	1050	2007	1350	1245
262000	4	2.5	2020	6236	2020	2002	1940	5076
350000	3	1	1010	7680	1010	1967	1320	7373
795000	4	2.5	3160	16564	3160	1987	3160	12415
672000	3	2.25	1130	4445	1130	1930	1330	4445
637500	3	1.75	1680	10685	1680	1966	3340	10390
425000	3	1.5	1570	12412	1570	1954	2130	12412
281000	4	2.25	1677	3600	1677	2012	1677	3600
322000	3	2	1760	43575	1160	1988	1760	46038
425000	2	1	1430	13300	1230	1921	1950	11421
504750	3	2.5	2490	21937	2490	1993	2450	21937
400000	2	1.75	1800	224769	1420	1950	1620	112384
452000	4	1.5	1580	7350	960	1963	1560	7350
3120000	3	3.5	4490	56609	4490	1993	2710	51330
675000	4	2.5	1770	9858	1770	1971	2470	9858
820000	4	2.75	2420	4635	2420	1905	1590	3150
320000	3	1	1780	6840	1780	1947	1410	7200
515000	4	2.5	2680	7178	2680	2004	2540	7133
495000	3	1.75	2080	3000	1040	1925	1000	3193
415000	3	1.75	1640	6435	1190	1972	1770	6930
550000	2	1	950	4080	950	1924	1120	4080
485000	3	2.5	2270	7887	2270	2004	2550	7133
710000	4	2.5	3060	5000	3060	2006	2870	5548
250000	3	2.25	1481	2820	1481	2012	1481	2889
552700	2	1	1100	2800	1100	1925	1110	1673
535000	4	2.5	2710	12138	1700	1968	2390	10052
556000	3	1.75	1590	2500	1190	1908	1420	3800
875000	3	2.75	3270	39586	3270	1988	3480	35998
345000	3	3.25	1600	1882	1360	2000	1390	1379
193000	2	1	680	8640	680	1951	1320	13140

398000	3	1.75	1890	16001	1890	1950	1820	11450
249000	2	2	1090	1357	1090	1990	1078	1318
1050000	4	3.5	3500	4000	2560	2000	1910	4000
605000	3	1.75	2100	5058	1340	1941	1640	5000
219500	1	1	820	1060	760	2007	1770	1924
654000	3	2.5	2220	2873	2010	2012	2290	3213
1040000	4	2.5	3770	10893	3770	1997	3710	9685
948000	3	2.5	3510	9824	3510	2002	3510	10588
1990000	5	3.5	5230	8960	4450	2014	2310	9603
808000	8	3.75	3460	4600	2860	1987	2170	3750
939000	4	2.75	3270	12880	3270	2014	2420	7505
825000	4	2.75	3990	6637	3990	2003	3500	7074
445830	3	2.5	1870	5449	1870	2003	2000	7687
749950	3	1.75	1800	5700	1000	1941	1680	5350
363500	4	1.75	2180	9702	1090	1962	2020	9792
1620000	3	3.5	3490	4000	2570	2009	1880	4000
325000	4	2.5	2230	8500	2230	1994	2270	8770
440000	4	2.5	2880	8061	2880	1988	2650	7660
265000	3	1	1070	4800	970	1947	1120	1198
450000	3	1.75	1540	9154	1540	1983	1990	10273
570000	4	2.5	2890	5801	2890	2005	2890	6286
250800	3	1.75	1290	4000	1170	1943	1140	4000
373500	4	2.5	2610	4978	2610	2004	2470	5024
554000	3	2.5	2140	4126	1960	2005	2280	2615
549000	5	1	1500	3978	1500	1929	1350	4080
985000	5	4.25	4650	108464	3260	1999	2810	155509
342000	3	2	1930	11947	1930	1966	2200	12825
209500	3	1.5	970	5488	970	1976	1040	5488
785000	4	2.75	2900	17400	2410	1978	2620	12240
425000	3	2.25	1660	6000	1110	1979	1440	4080
625000	3	2.75	2240	3600	1650	1901	1716	3120
960000	5	2.75	3040	10257	3040	1993	2860	9327
275000	5	1.75	2180	9178	1140	1963	2140	9261
354000	3	1.75	1340	6300	1340	1972	1780	7200
245000	3	2.25	1530	12000	1070	1979	2140	13636
265000	4	3	1940	8170	1940	1948	1310	8169
820000	3	1.75	2160	6272	1390	1960	1290	5376
330000	4	2.5	2240	7589	2240	1994	2250	7300
229900	3	1	1010	8848	1010	1968	1380	10650
649500	4	3	3150	6599	3150	2008	2680	9430
269950	4	2.5	2540	8400	2540	1977	1600	8050
465000	3	1	910	3880	780	1942	1220	3880
475000	2	1.75	1320	3420	1080	1955	2070	6000
275000	3	1.5	1170	1174	840	2007	1170	2537
645000	3	1.75	2070	5500	1130	1946	1800	4400
618000	4	2.25	2530	8736	1210	1958	1720	8500

375900	6	1.5	2550	33740	1750	1958	2010	28200
510000	5	2	2740	3838	1370	1959	1660	4040
502000	4	2.5	2680	5539	2680	2013	2680	5992
770000	4	3	2320	7200	1260	1943	1760	7200
515000	3	2.5	1790	7167	1790	1989	1680	7418
170000	3	1.75	1370	10780	1370	1959	1370	10317
476800	3	1.75	1900	43700	1900	1919	2070	43995
505000	3	2.5	1820	11012	1820	1988	1860	7767
350000	3	2.75	2370	4632	2370	2002	2290	5012
388000	2	1.75	800	4800	800	1922	1090	5000
535000	4	2.75	2360	15100	2360	2014	1440	13346
288349	3	1	1250	18616	1250	1972	1920	15654
785000	2	0.75	1260	4800	1080	1942	2660	7200
245000	2	1.75	1840	7230	1570	1938	1282	6769
2560000	4	2.5	5300	26211	4570	1923	3890	19281
740000	5	5	5774	31675	4490	1984	3260	13200
250000	4	2.75	1830	6643	1830	2001	2400	6472
487000	2	1	1440	4046	960	1946	1400	4046
650000	3	2.5	2320	5284	2320	2004	2320	4383
539000	7	2.25	2620	6890	2620	1961	2070	7910
280000	3	2.25	1810	7630	1810	1959	1830	7594
284850	3	1.5	1590	8256	1090	1969	1950	7840
499950	3	3.5	1820	1991	1430	2014	1550	1460
320000	5	3	2970	7000	1810	1979	2630	8062
308900	3	2.5	2250	7294	2250	1994	2140	7363
343000	2	1	1110	9920	700	1942	2830	10091
893880	6	2.5	2820	8600	1430	1967	2070	8900
450000	5	2	1870	7344	1870	1960	1870	7650
540000	3	2.5	1380	1021	1160	2008	1440	1021
456000	4	1.75	1670	9886	1670	1947	2590	9997
945800	4	2.75	3360	9100	1760	1973	2620	8925
478000	3	2.25	1570	9500	1070	1977	2250	9583
215000	4	2.25	1900	9600	1900	1967	2040	9600
556000	3	2.5	2750	35440	2750	1994	2710	35440
641500	1	1	1000	9084	1000	1950	1090	6536
810000	4	2.5	2810	10613	2810	1989	2800	9933
640000	3	2	1380	4800	1380	1948	1740	5640
463000	2	1	1150	4400	1150	1905	1240	4400
357000	4	2.5	2380	7066	2380	1997	2310	8127
146000	2	1	780	9750	780	1937	1670	9750
657000	4	2.75	3060	35380	1810	1982	1980	10425
508800	3	2	1720	10098	1140	1981	1840	10098
459950	4	2	2760	21465	2120	1979	2550	13144
389500	5	3.5	2960	12527	1940	1986	2210	10952
994000	4	2.5	3470	20445	3470	1963	3360	21950
668000	5	3.5	3990	42436	2710	2002	3030	41684

330000	6	2.25	3040	28535	1890	1951	1360	8250
459000	4	1.75	2260	9703	1660	1978	2390	8455
365000	3	1	950	8450	950	1962	1610	10080
399500	4	1.5	1660	6617	1660	1947	950	7440
359800	4	2.5	2390	6426	2390	1999	2520	6700
402000	4	2.25	2000	3672	1650	1926	2000	5100
405000	4	2.5	2220	4652	2220	2001	1840	4500
690000	3	2.75	2480	3240	1890	1929	2160	3240
532000	3	2.25	1910	35015	1430	1977	2340	36680
695000	3	2	2500	4080	1680	1922	1550	3060
700000	3	2.5	2050	4185	2050	2011	1150	5000
415000	4	1	1620	4329	1620	1927	1220	5520
356200	3	2	1690	10062	940	1928	2390	6650
2230000	3	3	3450	16740	1960	1949	3220	12528
324500	3	2.5	1660	3990	1660	2009	1670	4050
398000	4	1	1430	6240	1430	1953	1410	6240
225000	2	1	1396	111949	1396	1940	2020	111949
425000	4	2.5	3052	12145	3052	2004	2767	8604
289999	5	2.5	2180	8240	1220	1977	2050	7590
560000	3	2.5	2280	12498	2280	1987	2330	8844
2300000	4	2.5	3280	7100	2180	1911	3240	6674
575000	2	1.5	1750	19709	1440	1978	1950	21075
625000	5	1	3240	5324	3240	2007	3240	6036
445000	3	1.75	1360	4998	1360	1968	1350	4998
610000	6	2.75	2040	8560	1100	1961	2230	8560
519900	4	2	1820	9350	1820	1967	2260	9299
424000	3	1.5	1230	7200	1230	1953	1400	6240
620000	4	3.5	3880	8244	3060	2007	3180	10947
489000	3	3.5	1500	1249	1240	2004	1440	1850
500000	2	1.5	2070	5432	1370	1951	2070	5505
317000	3	2	1790	8228	1390	1980	1880	8228
640000	4	1.75	2060	9828	2060	1960	2260	9996
565000	3	1	1260	4080	1260	1928	1130	3876
1290000	5	4	4360	8030	4360	2015	3570	6185
465000	2	1	1120	4000	1120	1926	1620	4000
435000	3	2.25	1890	7200	1230	1973	1970	8101
280000	3	2.5	1370	22326	1370	1993	1580	10920
592500	4	1.5	2080	4400	2080	1925	1680	4400
565000	3	2.5	1870	1058	1380	2007	1440	1136
418000	3	1.75	1500	19113	1500	1984	1820	18151
573500	3	1	2200	21450	1600	1952	1880	11250
296500	3	1.5	1580	10100	1580	1961	1650	10032
669000	4	2.5	2470	4945	2470	2012	2510	4988
690000	5	3.5	2720	7598	1860	1993	2430	7728
364950	4	2.5	1930	6957	1930	1995	2090	6996
350000	3	1.5	1870	9000	1120	1923	1850	6000

770000	3	2.5	2650	40705	2650	1994	2550	42625
223000	3	1.75	1360	10573	1360	1984	1580	8280
286000	2	1	610	4000	610	1918	870	5160
850000	4	2.75	5440	239580	2720	1969	1970	40392
785000	4	2.5	3440	56192	3440	1994	3150	44431
343000	2	1.5	1040	8526	1040	1953	1310	8504
352000	4	1.5	1440	8680	1440	1922	1440	9000
769950	3	2	2190	5400	1260	1952	2160	5400
415000	2	1	1460	4200	880	1914	1540	4200
272000	4	2.5	2070	6175	2070	1993	1940	6175
290500	4	3.25	2510	7686	2510	2003	2510	6732
424000	4	2.5	2650	8685	2650	2006	2650	7932
555000	3	2.5	2690	4819	2690	2004	2360	4829
418000	4	3.25	3266	5969	3266	2014	3087	5169
299000	3	2.75	3080	19635	1610	1958	2424	12410
367000	3	1	810	7000	810	1968	1820	7589
500000	2	2.25	1290	1334	1290	2007	1350	1334
260000	4	1.75	1360	7700	1360	1966	1321	7756
424000	4	2	2100	4857	2100	1965	1450	5965
474500	5	2.5	2000	3600	1150	1987	1410	3600
543000	2	1	940	3864	940	1918	1440	3956
505000	4	2	2730	12000	1410	1998	2730	9039
1800000	5	5	4490	10279	3930	2013	2490	10279
618000	5	1.75	1880	18295	1880	1955	2180	20674
290000	3	1	960	9000	960	1955	1520	9000
208000	3	1	920	7314	920	1960	1010	7420
800000	4	3.75	2540	20662	2540	1998	2490	37731
242000	3	2.25	1690	7292	1250	1985	1670	7747
175000	2	1	830	2699	830	1996	1030	3574
608000	3	2	2400	217800	1590	1975	2340	207781
585000	3	1	1870	2807	1580	1927	1640	4500
814950	4	2.75	2990	6626	2990	2014	2910	5533
337500	3	1.75	1350	5850	1050	1973	1710	7757
265000	2	1	1000	31505	1000	1960	1560	22597
385000	3	1.5	1360	2030	1360	2008	1360	1167
399950	5	2.75	1970	5400	1320	1986	1280	2150
999999	3	2.5	2100	4097	2100	2008	1780	4764
303000	2	1	1020	9200	1020	1978	1520	9600
418000	4	3	2360	6250	1460	1960	2500	6250
825000	5	3	4080	7500	2720	1961	2000	4545
540000	4	2.5	2710	9248	2710	2011	2710	5000
930800	5	2.5	4150	96574	4150	1988	3320	40803
520000	4	2.5	3290	11446	3290	1992	2050	11933
460000	2	1	1430	12092	1430	1938	2320	10800
365000	3	2.5	2190	5091	2190	2004	2200	5948
315000	5	2.5	2090	4698	2090	1998	2070	4698

150000	2	1	870	5700	870	1957	1020	5700
806000	4	2.5	3370	9629	3370	1999	3360	10335
570000	4	3	1490	6766	1490	1915	1990	6526
455000	2	2	1190	1303	800	2009	1670	2425
362000	3	1	1290	10125	1290	1956	1760	14460
710000	4	1.75	2700	7625	1450	1937	1760	3300
400000	4	2.5	3030	180263	2030	1987	2600	182509
1570000	5	4.5	5620	5800	4700	1999	2360	5800
270000	4	1	1900	8505	1200	1956	1440	8505
345950	3	2.5	2110	4118	2110	1989	2110	4044
490000	2	1	950	3060	810	1925	1510	3780
315000	3	2.5	1790	6452	1790	2004	1610	3600
385000	3	3.5	1900	4805	1560	1999	1360	4800
234550	3	1	1990	15375	1140	1946	1160	10236
249900	3	2.5	1630	7700	1120	1978	1640	8160
392500	3	1	1390	10500	1390	1940	1350	9800
268000	3	1.75	1970	10270	1970	1966	1970	8400
755000	3	2.5	2000	1950	2000	2005	1560	1340
483453	4	2.75	2790	5527	2790	2014	2620	5509
537100	3	2.5	2450	7051	1870	1990	2450	8788
270000	2	1.5	840	867	840	2005	840	1322
550000	2	1	980	3080	980	1910	1450	3333
325000	3	2.25	1440	6443	1440	1995	1700	6749
638250	4	2.5	2460	8029	2460	1989	2420	7987
800000	4	2.75	4600	322188	2400	1989	3740	114562
480000	5	2.75	2760	7200	1430	1979	2460	8750
253200	3	1	1360	5840	1360	1953	1250	6708
468000	4	2.5	2480	176418	2480	1927	1640	112384
950000	4	1.75	2500	92347	1500	1970	3750	20267
880000	4	2.5	2580	3436	2580	2013	1780	5772
440000	3	2.5	2290	6302	2290	1997	2700	7500
520000	3	2	1410	2700	1410	1902	1750	4000
775000	4	2.5	3890	34513	3890	1996	3600	28435
975000	5	2.75	3100	10014	1660	1973	3230	10279
333000	4	2.5	2400	7270	2400	1993	2150	6584
820000	3	2.5	2950	35108	2950	1995	3810	35181
1250000	4	3.25	3820	24166	3310	1990	2900	18786
246000	4	2	1400	7632	1400	1930	1400	7632
800000	4	2.25	2120	9921	2120	1981	1890	7845
440000	3	2.25	1230	1097	1230	2009	1420	1437
595000	4	3	3680	35736	2320	1970	3210	39512
1200000	2	2.25	2160	17861	2160	1956	2660	18530
345000	2	2.25	1110	1290	1110	2006	1360	1251
303500	4	2.5	1920	7345	1920	1992	1860	7364
1750000	6	3	3510	5760	3510	1906	3450	5760
323800	3	2.5	2060	7658	2060	2003	2250	7299

263000	3	1	1250	7560	1250	1959	1270	7615
599000	3	1.75	1650	1180	1650	2014	1720	1960
900000	3	2.5	2320	5000	1620	1907	2160	5000
615000	3	1.75	2350	20820	1800	1978	2040	10800
980000	5	2.5	2750	6000	1750	1904	1520	4158
450000	3	2	1430	3480	980	1947	1450	4650
335000	5	3	2660	7700	1670	1962	1610	7700
280000	4	2.5	1934	5677	1934	2013	1919	5049
898000	4	2.5	3700	63991	3700	1992	3210	47215
255000	3	1.75	1740	8800	1140	1978	1690	7920
403900	4	2.5	2050	8909	1690	1986	2190	8912
310000	3	2	2070	9195	1220	1982	2080	9551
356000	2	1	920	4095	920	1914	1460	4945
317000	3	1.5	1390	8300	1390	1974	1470	7500
400000	4	2	1350	7255	1350	1959	1050	7288
258000	3	1.75	1040	5650	1040	1951	1290	5650
385000	3	1.75	1350	155073	1350	1969	1560	50965
400000	3	1	1060	12000	1060	1952	1550	10500
880000	4	2.5	3550	8618	3550	2007	1310	9746
398500	3	2.5	1630	1971	1630	1996	1630	3451
465000	3	1	1100	145490	1100	1915	1100	11610
289950	3	2	1670	7757	1670	1992	2290	7859
671300	4	2.5	3280	5232	3280	2004	2600	5080
900000	3	2.75	2980	12600	1590	1941	1520	9660
865000	4	2	2750	5527	2130	1901	1290	1764
410000	4	2.5	1700	9000	1700	1972	1370	7592
650000	4	2.5	2290	10186	2290	1985	2150	10186
1010000	3	2.25	2160	7500	2160	1982	1550	3839
670000	3	1.75	1280	2147	1280	1910	2010	2640
481000	4	2.5	2286	8269	2286	2002	2110	4711
330000	3	2	2170	3978	1340	1919	1350	4680
307000	3	2	1790	7259	1390	1980	1790	7700

#### 四、確認資料

y	x1	x2	x3	x4	x5	x6	x7	x8
200500	3	2	1280	14972	1280	1963	1800	9698
643500	6	5.25	3600	3960	2400	1971	1450	4600
728935	4	2.5	2980	10194	2980	1993	2980	10053
150000	3	1	1010	25000	1010	1966	2020	101494
762400	3	1.75	2430	14607	1230	1949	2750	17425
255000	3	2.5	2555	5720	2555	2006	2844	5769
385000	4	2.5	2960	5054	2960	2006	2960	5000
280000	4	2	1710	5440	1030	1976	1620	6696
552625	4	2	2560	9390	1280	1957	1830	8169
1450000	4	3.5	3720	8301	2880	2008	2080	9676



537000	4	1.75	2230	7957	2230	1967	2230	8040
1220000	4	3	3470	4750	2370	2014	2420	4761
286000	2	1	1170	6543	1170	1913	1550	7225
530000	3	1	1340	4284	1080	1910	1960	4080
267000	3	2.5	1690	10336	1690	1989	1580	7700
546000	3	1.75	2000	5000	1110	1921	1750	5000
961500	3	2.5	3910	14000	3910	1999	2280	14000
995000	3	4.5	4380	47044	3720	1968	2460	18512
680000	4	2.5	2980	8770	2980	1990	2940	9238
305495	3	1.75	2110	10200	2110	1966	1800	10200
619950	2	1	1520	5400	920	1951	1600	5400
282000	4	2	1890	6302	1890	1997	1690	5369
435000	4	2.75	2230	9640	1320	1998	2100	9600
995000	3	2.5	2460	10300	2460	1992	2410	5250
449000	4	2.25	2230	8440	2230	1968	2160	9412
1000000	4	2.75	2290	6120	2170	1926	1910	4590
515000	2	1.75	1930	5570	1930	2005	1810	5178
195000	3	1.75	1070	6110	1070	1968	1350	6148
2570000	4	3.75	4475	20424	2659	1999	4340	5250
305000	4	2.5	2230	6487	2230	2003	2230	6882
337000	4	2.5	2230	5970	2230	2002	1970	4919
790000	2	1.5	1940	4400	970	1923	1480	3080
386100	3	2	1270	6760	1270	1972	1550	5734
460000	4	2.5	2080	17532	2080	1996	2550	12560
294900	3	2.25	1500	8100	1210	1968	1600	7896
1130000	6	3	2880	3192	2180	1919	1870	4533
391500	3	2.5	1920	9625	1920	1993	1920	10343
530000	4	2.5	3010	9000	3010	2008	3250	5531
475000	3	1.5	1240	8738	1240	1959	1440	9344
291750	3	2.25	1310	12825	1310	1950	1600	11250
301000	3	2.5	2420	4750	2420	2003	2690	4750
585000	3	3.25	3050	12700	2240	1990	3050	12386
599000	3	2	2560	14680	1330	1987	2390	13848
350000	2	1	1620	9205	850	1921	2460	5469
315000	4	2	1780	5336	930	1918	1910	10304
249900	3	1.75	1380	14000	1380	1939	1490	18503
840000	4	2.5	2170	9796	1650	1980	2350	9796
517000	2	1.5	1920	3408	960	1912	1130	3408
395000	4	1.75	1540	5120	770	1943	1080	5120
260000	3	1.75	1340	8000	1340	1980	1690	8000
799950	5	3.75	3760	4702	2780	2014	2950	5981
910000	3	2	2700	6120	1350	1962	1700	4590
320000	4	2.5	2110	6295	2110	2000	2720	6311
365000	3	2.25	1960	1985	1750	2003	1760	1985
413565	3	2.5	1690	1613	1150	2014	1690	1619
234000	4	2	1710	7455	1030	1975	1710	7350

735000	4	2.5	2820	6180	2050	2013	1390	4635
220000	4	1.75	1660	11664	1010	1952	1670	9975
431000	2	1	1000	6500	1000	1918	1280	6500
279000	3	2.25	2070	7800	1170	1964	1570	8400
192000	2	1	700	10540	700	1953	890	10540
315000	3	1.75	1660	8160	1660	1951	1490	8100
95000	2	1	960	7000	960	1918	1850	8120
1900000	4	2.75	4280	12668	3900	1947	3590	12670
270000	4	1	1430	5909	1070	1947	1460	8433
462550	2	1.75	1070	1276	1070	2006	1290	2057
915000	4	2.5	3210	8532	3210	1998	2950	6753
305000	4	2.25	2050	12581	2050	1978	1620	7400
370000	2	1	780	1133	780	2009	1130	1270
748000	4	2.5	3220	8379	3220	2004	2720	7635
475000	4	2.75	2200	16288	1290	1980	2650	6620
588000	5	3	2190	4900	1490	1940	1370	4606
558000	4	2	2180	3870	1020	1900	1520	2580
387865	3	2.5	2370	4200	2370	2014	2370	4200
411100	3	1.5	1040	10323	1040	1958	1580	7200
210000	5	2	2050	10200	1430	1956	1940	8625
195000	3	1.75	1570	8459	1570	1991	1650	8844
610000	4	2	2220	5821	1380	1916	1850	5000
374000	4	1.75	2220	15600	1140	1963	1670	4800
316000	2	2	860	3000	860	1906	1290	3500
622000	3	2.25	1550	1919	1550	2003	1550	2918
573300	2	1.75	1290	6600	870	1951	2380	7370
223000	3	1.75	1340	7473	1340	1973	1510	8240
560000	2	2	1680	4647	1680	2008	1680	4950
325000	3	1.5	1590	7936	1590	1956	1590	7936
280000	3	2.5	1550	4486	1550	2000	1700	4487
607010	4	2.5	2180	4000	1700	2002	2180	4000
183750	5	2.75	1650	5453	1650	1970	1670	5885
262000	2	1	1500	4120	880	1928	1300	4120
1020000	4	3.75	4060	19290	4060	2002	4020	13250
560000	3	2.5	2840	216493	2840	1991	2820	175111
363000	2	1	1120	4800	770	1926	1510	4800
380000	3	2.25	2530	12042	2530	1989	2480	10172
452000	4	2.25	2590	10002	1340	1968	1550	10436
405000	2	1	880	3000	880	1927	1300	3000
799000	3	2.5	2140	9897	2140	1959	2680	10083
937500	3	1	1320	8500	1320	1954	2790	10800
485500	2	1.5	1340	1286	1190	2006	1460	1245
626000	3	2.25	1750	1572	1470	2005	2410	3050
1080000	4	2.75	2990	7389	2090	1923	2510	6157
998000	3	3.75	3710	34412	2910	1978	2390	34412
455000	4	1.75	2050	94525	1250	1959	2270	47480

170000	2	1	810	9882	810	1944	950	7200
310000	3	1	1180	8474	1180	1956	1180	7200
525000	2	1	1000	4950	800	1948	1060	5500
215000	3	1.5	1240	9405	1240	1966	2260	7611
425000	4	2.75	2440	15349	2440	1957	2280	9250
75000	1	0	670	43377	670	1966	1160	42882
236000	4	2.25	1830	9485	1200	1989	1730	10109
520000	4	3.5	2734	210201	2734	1974	2270	187308
750000	4	2.75	3190	9023	3190	2005	3159	5615
782500	4	2.5	2930	7806	2930	2005	2600	6051
575000	3	1.75	1580	11750	1180	1951	2150	11750
791500	4	2.5	3250	8970	3250	1994	3240	8449
585000	4	2.5	2910	6250	2910	1999	2550	6250
975000	3	2.5	2530	7000	2530	1915	2380	7000
485000	3	2.5	1650	3436	1650	2003	1680	3446
309950	2	2.25	1460	1607	1460	2007	1460	1826
268000	3	1.75	1480	8009	980	1980	1790	7678
775000	4	3.5	3080	5250	3080	2003	2980	5509
690000	3	1	1090	4000	1090	1945	1520	4000
260000	3	2.5	1670	5797	1670	1988	1670	6183
336750	4	2.25	1720	7803	1350	1955	1720	7803
372500	2	1	880	10950	880	1944	1450	7560
273500	4	1	1360	6000	1020	1944	1230	6000
345000	2	1	860	8250	860	1940	1780	11200
400000	3	2.5	1910	4408	1910	2007	1910	8154
247500	2	1	700	6046	700	1950	990	6790
285000	5	2.5	2270	6300	2270	1995	2240	7005
204000	3	1	1040	7405	1040	1971	1580	7405
830000	5	2.25	2780	10192	2780	1978	2740	9900
370000	4	2.5	2720	8666	2720	1992	2410	8100
304999	4	2.25	2270	9600	1290	1976	1930	15000
520000	4	2.5	2410	6440	1550	1974	2330	6938
466200	3	1.5	1340	8856	1340	1959	1850	8740
487500	5	3.5	3530	218472	2380	1999	2110	211404
765000	3	1.75	2190	6450	1480	1957	2190	6450
240000	3	1	1230	2353	1230	1925	1280	1572
594000	4	1.75	2140	8000	1410	1965	2080	8000
580000	3	2.5	2200	11000	2200	1978	2200	12851
2890000	5	6.25	8670	64033	6120	1965	4140	81021
310000	2	1	770	8149	770	1948	820	8149
422000	4	2.5	3200	6691	3200	2002	2610	6510
1010000	3	2.5	2570	5000	1480	1940	2420	5484
245000	4	1	1530	7200	1400	1948	1530	7200
450000	2	1.75	1760	2275	1040	1912	1380	3750
697000	3	2.25	2420	5304	1640	1947	1560	5304
322000	3	1.75	1510	8400	980	1979	1500	10125

217000	3	1	1400	7800	1400	1962	1400	7800
498000	3	2.5	1620	2640	1620	1900	1370	3840

## 第二節 程式碼

### 一、SAS

```
data a1;
    set work.train;
    keep x1-x8 y;
run;

PROC reg data=a1;
MODEL Y=X1-x8 / r partial tol vif collinoint;
run;

ods graphics on;
proc transreg details data=a1;
model boxcox(y/lambda=-2 to 2 by 0.01)=
identity(x1-x8);
run;
ods graphics off;

data a2;
set work.a1;
Ynew=((y**(-0.07))-1)/(-0.07);
run;
proc print data=a2;
run;
proc means data=a2(drop=y);
LABEL ynew="房價" x1="臥室數量" x2="浴室數量" x3="居住面積"
x4="土地面積" x5="地上居住面積"
x6="建築年份" x7="最近15棟房屋的內部居住空間平均大小" x8="最近15
棟房屋的土地平均大小";
run;

data a;
set a1;
y=y/10000;
x4=log(x4);
x8=log(x8);
run;
proc means data=a1 median;
var x4 x8;
run;

proc sgplot data=a;
histogram y /binwidth=25 datalabel;
title "房價(y)";
xaxis label="房價(萬美元)" values=(12.5 to 300 by 25);
yaxis label="百分比(%)" ;
run;
```

```

proc sgplot data=a2 ;
  histogram ynew / binwidth=0.05 ;
  title "房價(ynew)";
  xaxis label="房價" values=(8 to 9.25 by 0.05) ;
  yaxis label="百分比(%)" ;
run;

proc sgplot data=a2;
  histogram x1 / binwidth=1 datalabel;
  title "臥室數量(x1)";
  xaxis label="臥室數量" values=(0 to 8 by 1);
  yaxis label="百分比(%)" ;
run;

proc sgplot data=a2;
  histogram x2 / binwidth=1 datalabel;
  title "浴室數量(x2)";
  xaxis label="浴室數量" values=(0 to 6 by 1);
  yaxis label="百分比(%)" ;
run;

proc sgplot data=a2;
  histogram x3 / datalabel;
  title "居住面積 (x3)";
  xaxis label="居住面積 (平方英尺)" values=(500 to 6000 by
500);
  yaxis label="百分比(%)" ;
run;

proc sgplot data=a2;
  histogram x4 / datalabel;
  title "土地面積 (x4)";
  xaxis label="土地面積 (平方英尺)" values=(0 to 330000 by
30000);
  yaxis label="百分比(%)" ;
run;
proc sgplot data=a;
  histogram x4/ binwidth=0.5 datalabel;
  title "土地面積 (x4)";
  xaxis label="土地面積(lnx)" values=(6.75 to 12.5 by 0.5);
  yaxis label="百分比(%)" ;
run;

proc sgplot data=a2;
  histogram x5/binwidth=500 datalabel;
  title "地上居住面積 (x5)";
  xaxis label="地上居住面積 (平方英尺)" values=(250 to 5000
by 500);

```

```

    yaxis label="百分比(%)" ;
run;

proc sgplot data=a2;
    histogram x6/binwidth=1 datalabel;
    title "建築年份(x6)";
    xaxis label="建築年份" values=(1901 to 2015 by 10);
    yaxis label="百分比(%)" ;
run;
proc freq data=a2;
    title "建築年份(x6)";
    table x6 ;
run;

proc sgplot data=a2;
    histogram x7/binwidth=500 datalabel;
    title "最近15棟房屋的內部居住空間平均大小(x7)";
    xaxis label="最近15棟房屋的內部居住空間平均大小(平方英尺)"
    values=(750 to 5000 by 500);
    yaxis label="百分比(%)" ;
run;

proc sgplot data=a2;
    histogram x8 / datalabel;
    title "最近15棟房屋的土地平均大小";
    xaxis label="最近15棟房屋的土地平均大小(平方英尺)"
    values=(1021 to 210000);
    yaxis label="百分比(%)" ;
run;
proc sgplot data=a;
    histogram x8/ binwidth=0.5 datalabel;
    title "最近15棟房屋的土地平均大小";
    xaxis label="最近15棟房屋的土地平均大小(lnx)" values=(6.75 to
    12.5 by 0.5);
    yaxis label="百分比(%)" ;
run;

proc corr data=a2;
    VAR Ynew x1-x8;
    TITLE "檢定是否有相關";
    LABEL y="房價" x1="臥室數量" x2="浴室數量" x3="居住面積" x4="
    土地面積" x5="地上居住面積"
    x6="建築年份" x7="最近15棟房屋的內部居住空間平均大小" x8="最近15
    棟房屋的土地平均大小";
run;

proc reg data=a2;
    MODEL Ynew=x1-x8/ r partial tol vif collinooint;

```

```

TITLE "檢定是否有共線性關係";
LABEL y="房價" x1="臥室數量" x2="浴室數量" x3="居住面積" x4="
土地面積" x5="地上居住面積"
x6="建築年份" x7="最近15棟房屋的內部居住空間平均大小" x8="最近15
棟房屋的土地平均大小";
run;

proc reg data=a2;
model Ynew=X1-x8 / selection=Forward sle=0.05;
title '以前進選取法來選取';
run;
proc reg data=a2;
model Ynew=X1-X8 / selection=Backward sls=0.05;
title '以後退刪去法選取';
run;
proc reg data=a2;
model Ynew=X1-x8/ selection=stepwise sls=0.05
sle=0.05;
title '以逐步篩選法選取';
run;
proc rsquare data=a2 adjrsq cp aic mse sbc;
model Ynew=X1-X8;
title '其他選取變數方法';
run;

data a3;
set a2;
drop x5 x8;
obs=_n_;
run;

proc reg data=a3;
model Ynew=X1-x4 x6 x7/ vif r influence dw;
output out=outlier r=r h=h rstudent=rs
student=student cookd=cookd dffits=dffits;
title '極端值與影響點';
run;

ods graphics on;
proc univariate data=outlier normal plot;
var r;
title '未刪除離群值的資料';
run;
ods graphics off;

data a4;
set outlier;
p=7;
n=350;

```



```

hh=2*p/n;
dif=((p/n)**0.5)*2;
c=finv(0.5,p,n-p);
t=tinv(1-0.05/(2*n),n-p-1);
if (abs(student)>2 or h>hh or abs(rs)>t) and
(cookd>c or abs(dffits)>dif )then delete;
drop p n hh dif c t;
run;

data r_outliers y_outliers x_outliers dif_outliers
cookd_outliers;
set outlier;
p=7;
n=350;
hh=2*p/n;
dif=((p/n)**0.5)*2;
c=finv(0.5,p,n-p);
t=tinv(1-0.05/(2*n),n-p-1);
if abs(student) > 2 then output r_outliers;
if abs(rs) > t then output y_outliers;
if h > hh then output x_outliers;
if abs(dffits) > dif then output dif_outliers;
if cookd > c then output cookd_outliers;
drop p n hh dif c t;
run;

proc print data=a4(drop=obs);
title '刪除離群值後的資料';
run;

proc reg data=a4(drop=obs);
model Ynew=x1-x4 x6 x7/ r ;
output out=a5 r=r p=pd ;
run;

ods graphics on;
proc univariate data=a5 normal plot;
var r;
title '殘差檢定';
run;
ods graphics off;

proc corr data=a5;
var Ynew x1-x4 x6 x7;
title '檢定是否有相關性';
run;

proc reg data=a5;
model Ynew= x1-x4 x6 x7/ r vif dw influence partial
collino int spec;

```

```

title '建立最終迴歸模型';
run;

data b1;
set WORK.test;
Ynew=((y**(-0.07))-1)/(-0.07);
Ynewhat=12.21085-0.04529*X1+0.04553*x2+0.00014170 *x3
-7.05091E-7*x4-0.00202*x6+0.00005175*x7;
mape=abs((Ynew-Ynewhat)/Ynew);
title 'mape';
run;

proc reg data=b1;
model Ynew= x1-x4 x6 x7/ r ;
run;

proc univariate data=b1;
var mape;
output out=mape1 mean=m;
run;
data b2;
set mape1;
MAPE=m;
keep mape;
title 'MAPE';
proc print;
run;

```

## 二、R

```

library(readr)
library(dplyr)
a1 <- read_csv("C:\\Users\\niuno\\Desktop\\迴歸\\kc.csv")

# 設置隨機種子
set.seed(410651243)

library(tidyverse)
train <- a1 %>%
  sample_n(350, replace = FALSE)

# 未被抽取的數據
remaining_data <- anti_join(a1, train, by='id')

# 再從未被抽取的數據中抽取 150 筆
test<- remaining_data %>% sample_n(150, replace = FALSE)

remaining <- anti_join(remaining_data, test, by='id')
final_test<- remaining %>% sample_n(1, replace = FALSE)

```

```

library(openxlsx)
# 匯出抽取的 data 到 Excel
write.xlsx(train, file = "C:\\Users\\niuno\\Desktop\\迴歸\\train.xlsx", rowNames= F)
write.xlsx(test, file = "C:\\Users\\niuno\\Desktop\\迴歸\\test.xlsx", rowNames= F)

library(readxl)
rs1 <- read_excel("C:\\Users\\niuno\\Desktop\\迴歸\\rs1.xlsx")
# Fit the linear model
train.result <- lm(ynew ~ x1 + x2 + x3 + x4 + x6 + x7, data = rs1)
# Residuals vs. Fitted plot
plot(train.result, which = 1, col = "blue")
abline(h = 0, col = "black")
# Standardize residuals
std_resid <- sqrt(abs(rstandard(train.result)))
# Residuals vs. Fitted plot with standardized residuals
plot(fitted(train.result), std_resid, col = "blue", main = "Scale-Location" , xlab = "Fitted
Values
      lm(ynew ~ x1 + x2 + x3 + x4 + x6 + x7)", ylab = "sqrt(Standardized Residuals)")

rs2 <- read_excel("C:\\Users\\niuno\\Desktop\\迴歸\\rs2.xlsx")
# Fit the linear model
train.result1 <- lm(ynew ~ x1 + x2 + x3 + x4 + x6 + x7, data = rs2)
# Residuals vs. Fitted plot
plot(train.result1, which = 1, col = "blue")
abline(h = 0, col = "black")
# Standardize residuals
std_resid1 <- sqrt(abs(rstandard(train.result1)))
# Residuals vs. Fitted plot with standardized residuals
plot(fitted(train.result1), std_resid1, col = "blue", main = "Scale-Location" , xlab = "Fitted
Values
      lm(ynew ~ x1 + x2 + x3 + x4 + x6 + x7)", ylab = "sqrt(Standardized Residuals)")

remaining <- anti_join(remaining_data, test ,by='id')
final_test<- remaining %>% sample_n(1, replace = FALSE)

```

### 第三節 資料來源

House sales in King County, WA from May 2014 to May 2015.

<https://www.kaggle.com/code/daisyamber/house-prices-seattle/input>

<https://geodacenter.github.io/data-and-lab/KingCounty-HouseSales2015/>