



臺南市實價登錄大數據輔助區段地價平衡（含整體開發區）之研究（第四期）

電腦大量估價模型及基礎統計概論 教育訓練

Presented by

Karen Teng 鄧筱蓉 博士

Vice President 副總經理

© Anfu Solutions Inc.

Nov 11, 2024

臺南市實價登錄大數據輔助區段地價平衡（含整體開發區）之研究（第四期）

01

基礎統計介紹

解讀資料，辨識資訊意涵

02

認識電腦大量估價

AI世代來臨改變，提供創新服務

03

線性迴歸

特徵價格模型

04

地產數位轉型

數據分析做為決策應用

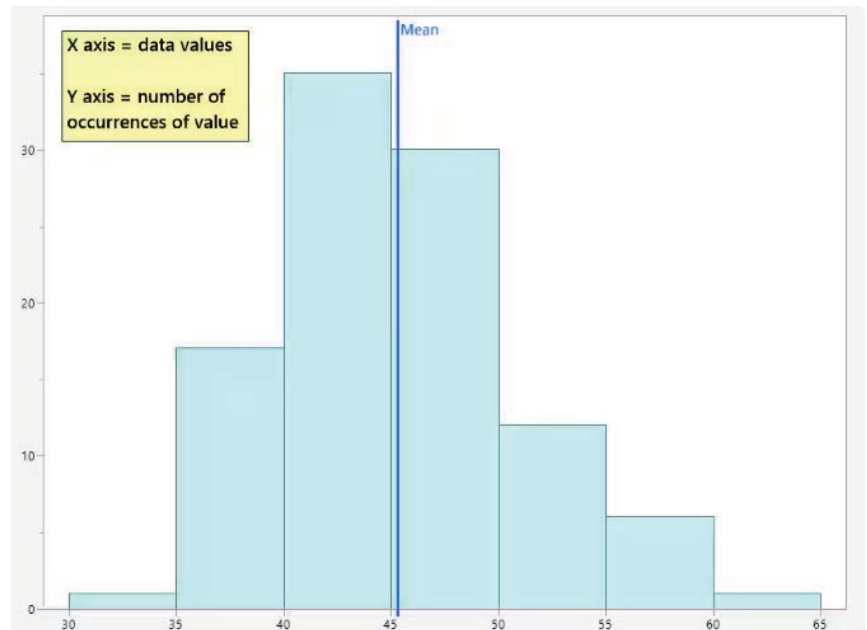
平均數

平均數是估計資料的集中趨勢

■ 母體平均數

- 母體平均數是理論母體的集中趨勢，通常未知。
- 藉由收集資料樣本，然後計算該樣本的平均數。作為母體平均數的估計值。

■ 樣本平均數、算術平均數 和樣本平均值之間有什麼 區別？



圖片來源：統計學知識入口網站

中位數

中位數是樣本資料的第 50 個百分位數

- 中位數是樣本資料的第 50 個百分位數。50% 的資料值高於中位數，50% 低於中位數
- 將樣本資料值從低到高排序，然後找到中間值。
- 平均數和中位數都是用以估計資料集中趨勢的指標，通常兩者會一起報告。
- 中位數受極端資料值或不對稱資料的影響較小。

假設資料值又是 4、5 和 6。

首先，將值從低到高排序：4 — 5 — 6。

中間值（在此範例中為 5）是中位數。

極端資料值如何影響樣本平均數和樣本中位數

假設樣本資料值現在是 4、5 和 6。

- 樣本平均值為 $(4+5+6)/3=5$
- 樣本中位數是排序資料值 4 – 5 – 6 中的中間值，即 5

假設樣本資料值現在是 4、5 和 12。

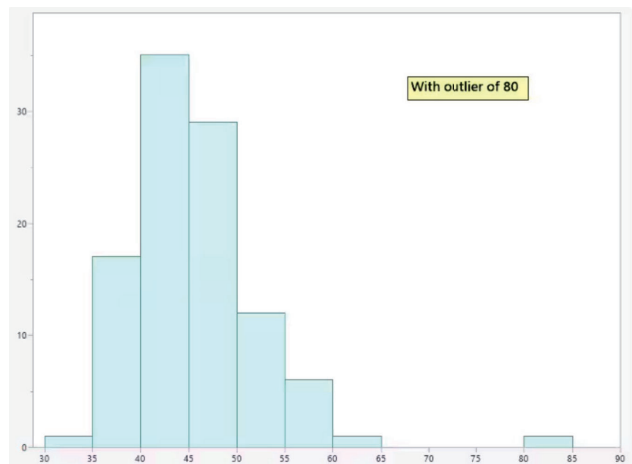
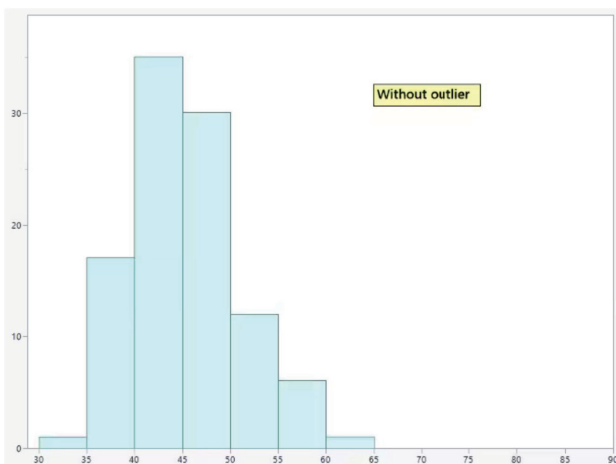
- 樣本平均值為 $(4+5+12)/3=7$
- 樣本中位數是排序資料值 4 – 5 – 12 中的中間值，即 5

■ 單一極限資料值對樣本平均數的影響較大，但對樣本中位數的影響較小。

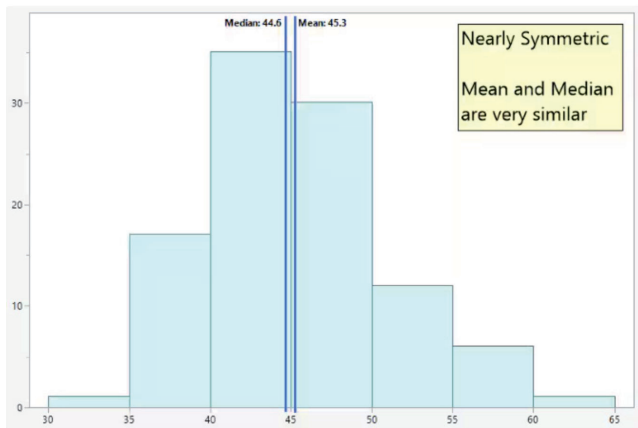
■ 中位數對離群值或極端資料值具有穩健性。

極端資料值如何影響樣本平均數和樣本中位數

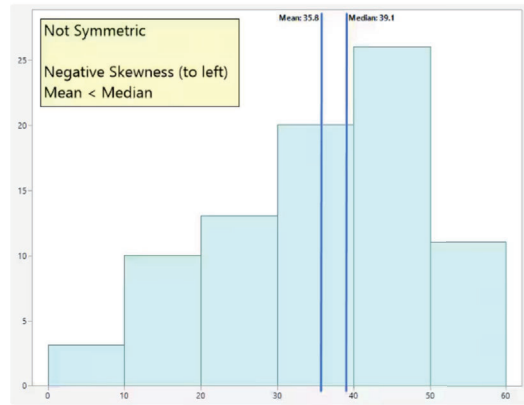
兩組資料的中位數均為 44.6。無離群值的資料的平均數為 45.3，有離群值的資料的平均數為 45.6



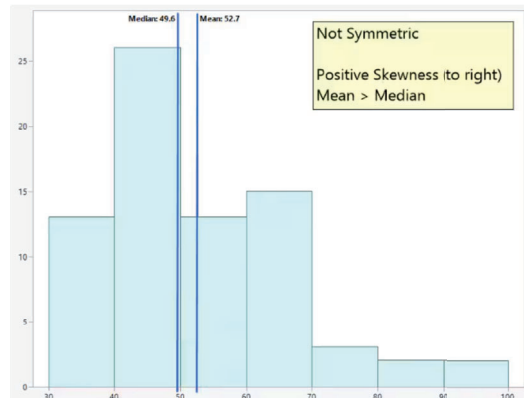
資料對稱性如何影響 樣本平均數和樣本中 位數



圖片來源：統計學知識入口網站



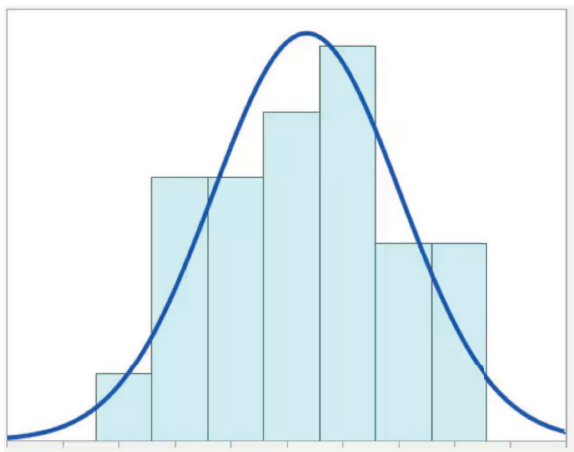
資料會受到低值影響，其呈左偏分佈。偏度統計量為負值，平均數小於中位數



該資料會受到「重高值權重」影響，其呈右偏分佈。偏度為正值，平均數大於中位數。

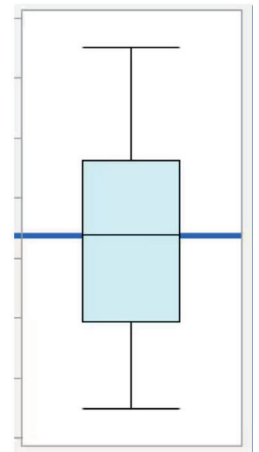
常態分佈

常態分佈的資料傾向於圍繞中心值累積；中心上方和下方之值的出現頻率呈對稱下降。



圖片來源：統計學知識入口網站

- 它完全由平均數和標準差定義。
- 平均數、中位數、眾數都相同。
- 對稱
- 呈鐘形曲線。

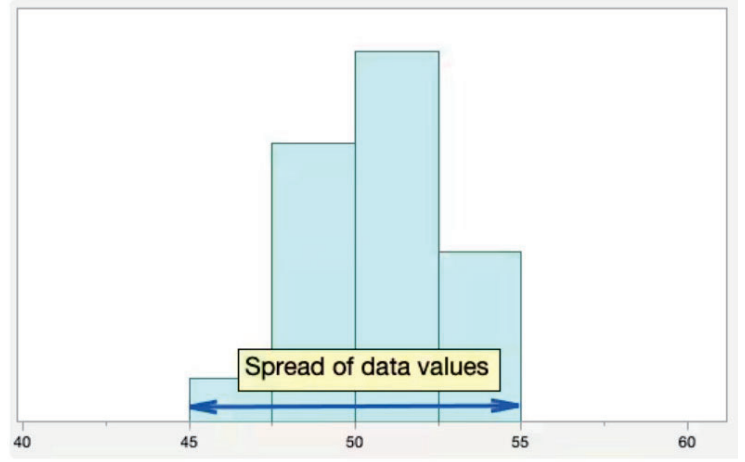


標準差

- 衡量一組資料值的分佈。
- 標準差高表示資料值分佈範圍較廣
- 標準差低表示資料值圍繞資料集的平均值集中分佈範圍較窄。

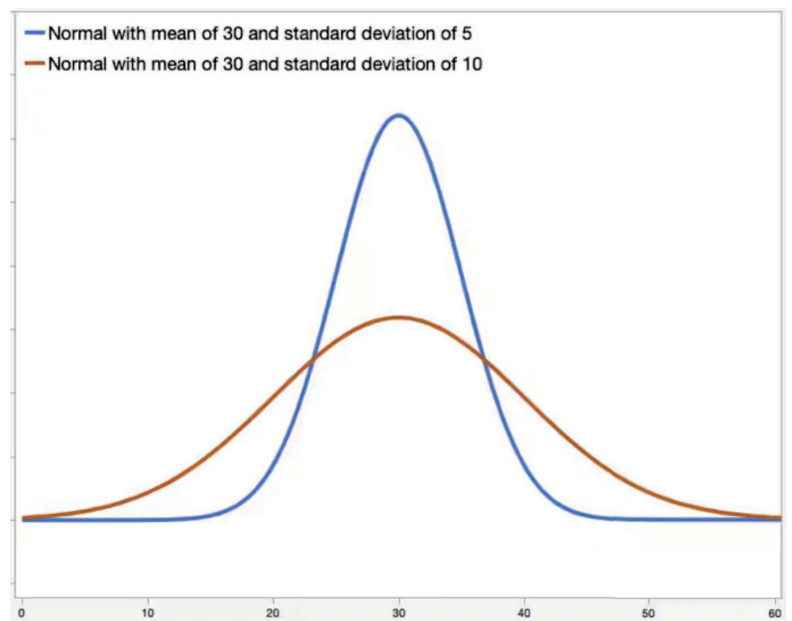


圖片來源：統計學知識入口網站



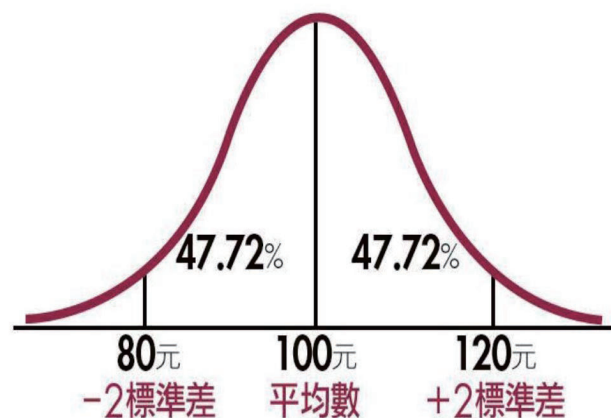
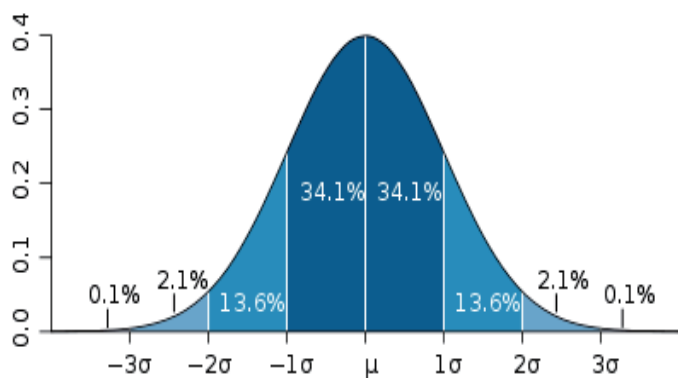
標準差

- 為兩個常態分佈，每個平均數都相同，均為 30。較窄、較高的藍色分佈的標準差為 5。較寬、較短的橙色分佈的標準差為 10。
- 標準差越小，表示資料值越接近樣本平均數。標準差越大，表示資料值分佈越廣，某些值離樣本平均數越遠。
- 標準差幾乎總是正值，但有一個例外：如果資料集中的所有值都相同，則標準差為零。資料不存在變異性或分佈。



圖片來源：統計學知識入口網站

平均數與標準差之關係



圖片來源：<https://smart.businessweekly.com.tw/Books/Special2.aspx?id=60065&type=1&s=books>

股價的波動



資料來源：XQ全球贏家 整理：呂郁青

極端資料值會影響樣本標準差？

極端資料值會對樣本標準差產生重大影響

樣本	1	2	3	4	5	6
價格	55	60	65	75	80	85



平均數：70
標準差：11.8

樣本	1	2	3	4	5	6	7
價格	55	60	65	75	80	85	140



平均數：80
標準差：28.6

13

臺南市實價登錄大數據輔助區段地價平衡（含整體開發區）之研究（第四期）

01

基礎統計介紹

解讀資料，辨識資訊意涵

02

認識電腦大量估價

AI世代來臨改變，提供創新服務

03

線性迴歸

特徵價格模型

04

地產數位轉型

數據分析做為決策應用

Artificial Intelligence

人工智慧

現有AI大都是建立在「機器學習」，或更精確地說「深度學習」的架構上，由巨量數據中摸索出特定的規律與模式，並以此來做出預測。

AI不等於機器人



大量估價模型

模型的理論基礎源自於迴歸分析

迴歸分析的方法是，依據相關理論建立依變數為自變數的函數，然後依樣本資料去估計模型中的參數並做預測。

人們需要研究的是那些不確定的關係

研究變數間關係的統計方法：

- 變異數分析
- 迴歸分析
- 相關分析



臺南市實價登錄大數據輔助區段地價平衡（含整體開發區）之研究（第四期）

01 基礎統計介紹

解讀資料，辨識資訊意涵

02 認識電腦大量估價

AI世代來臨改變，提供創新服務

03 線性迴歸

特徵價格模型

04 地產數位轉型

數據分析做為決策應用

簡單線性迴歸分析

迴歸分析是用來分析一個或一個以上自變數與依變數的數量關係，以了解當自變數（ X_i ）為某一水準或數量時，依變數（ Y_i ）反應的數量或水準。

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad i = 1, 2, \dots, n$$

α 稱之為截距， β 是迴歸係數， ε_i 是隨機誤差

$$Y = -3.20 + 0.897X \quad R^2 = 0.94$$

Y：台股現貨收盤指數

X：台股期貨收盤指數

台股期貨指數每上漲 1 點，現貨指數上漲 0.897 點

驗證模型準確度

依據訓練樣本和測試樣本進行交叉驗證來調校模型穩定度與準確率，最終模型需符合大量估價模型之國際標準

- $MAPE < 15\%$
- $10\%PPE > 40\%$
- $20\%PPE > 70\%$

01 MAPE

平均絕對百分比誤差

- 誤差 = (預測值 / 真實值) - 1
- 誤差取絕對值
- 計算誤差之平均值

02 Percent Predicted Error (PPE)

預測誤差百分比

- 篩選誤差 $< 10\%$ 的樣本
- 篩選誤差 $< 20\%$ 的樣本
- 計算誤差 $< 10\%$ 的樣本佔全體比例
- 計算誤差 $< 20\%$ 的樣本佔全體比例

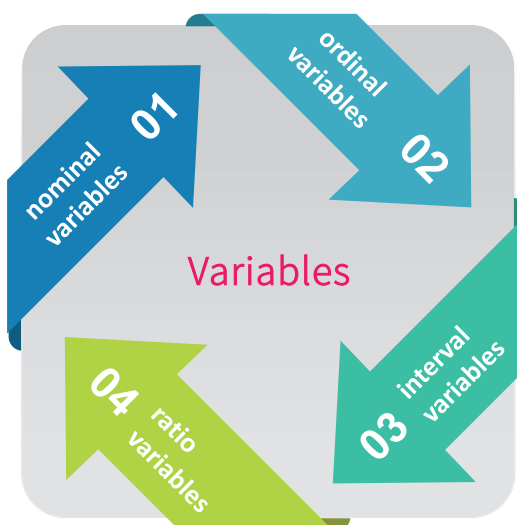
變數分類

類別變數

用來分別事物之不同
例如：有無管理組織

比率變數

依照順序、大小，來分別事物之不同，數值之間有相等的差距。另加上有絕對零點的特性。
例如：臨路路寬、身高、體重



次序變數

依照順序、大小，來分別事物之不同。
例如：名次排名、百分等級

等距變數

依照順序、大小，來分別事物之不同，數值之間有相等的差距。(可加減，但無法乘除)
例如：智商、溫度

資料清理流程

GARBAGE IN GARBAGE OUT



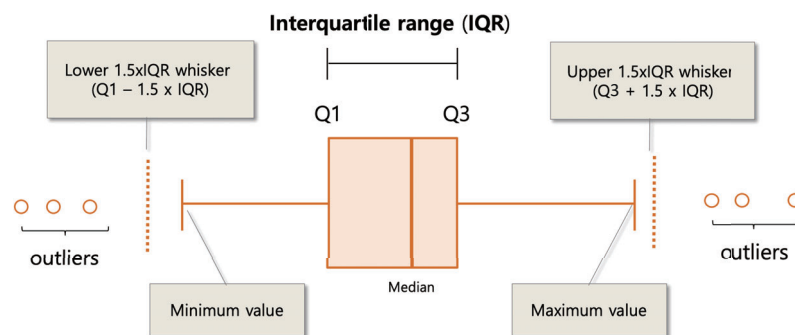
資料

資料分析

敘述統計

ID	申報書序	房地總價	交易日期	交易年	交易季	土地移轉	建物移轉	總樓層數	道路寬	土地使用	部	
1	9	A1DK1081	4950000	1081015	108	108Q4	58.84	64.8	2	700709	38.29	6都_住
2	14	A1DK1080	4600000	1080605						0131	33.43	6都_住
3	16	A1DK1080	4500000	1080629						0323	12.65	11都_住
4	29	A1DK1081	11000000	1081112						0727	14.96	8都_住
5	39	A1DK1080	11500000	1080708						0820	32.08	8都_住
6	49	A1DK1081	5900000	1080909						1212	17.36	4都_住
7	53	A1DK1080	8500000	1080420						0619	33.93	10都_住
8	77	A1DK1080	6500000	1080516						0708	20.79	9都_住
9	85	A1DK1080	9880000	1080419						0130	26.6	9都_住
10	86	A1DK1081	9100000	1080830						0627	37.96	6都_住
11				1080603						1007	25.02	8都_住
12				1081010								
13				1080921	108	108Q3	97.52	259.54	4	940525	14.33	4都_住
14				1080624	108	108Q2	93.88	195.73	4	950306	13.31	7都_住
15	120	A1DK1080	7600000	1080821	108	108Q3	82.22	180.88	4	821221	25.68	8都_住
16	127	A1DK1090	12350000	1081224	108	108Q4	84.62	209.03	4	811117	27.12	14都_住

變數名稱	樣本數	平均值	標準差	最小值	中位數	最大值
屋齡(年)	64,286	25.58	16.13	0.00	29.96	59.96
總樓層數(層)	64,286	9.23	5.25	1.00	7.00	42.00
土地移轉總面積(平方公尺)	64,286	27.85	41.02	1.01	24.68	4,677.51
建物移轉面積不含車位(平方公尺)	64,286	103.49	57.66	14.03	96.11	1,346.49
單價(萬元/平方公尺)	64,286	20.97	9.99	3.71	18.86	171.99
總價不含車位價(萬元)	64,286	2,241.53	2,420.41	172.00	1,665.00	61,000.00



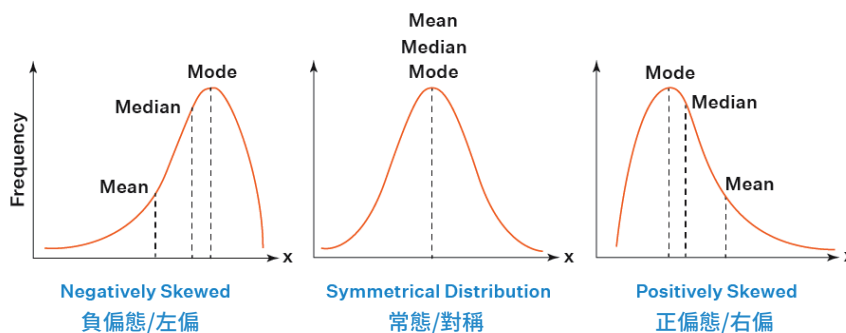
Source: EZBioCloud

變數名稱	樣本數	平均值	標準差	最小值	中位數	最大值
屋齡(年)	64,286	25.58	16.13	0.00	29.96	59.96
總樓層數(層)	64,286	9.23	5.25	1.00	7.00	42.00
土地移轉總面積(平方公尺)	64,286	27.85	41.02	1.01	24.68	4,677.51
建物移轉面積不含車位(平方公尺)	64,286	103.49	57.66	14.03	96.11	1,346.49
單價(萬元/平方公尺)	64,286	20.97	9.99	3.71	18.86	171.99
總價不含車位價(萬元)	64,286	2,241.53	2,420.41	172.00	1,665.00	61,000.00

mean < median < mode

mean = median = mode

mean > median > mode

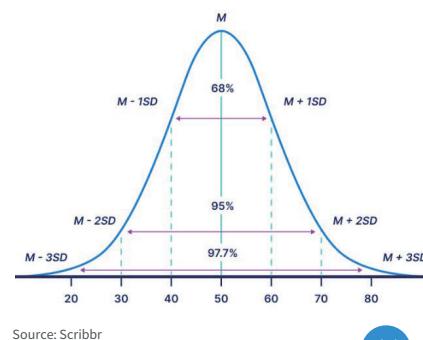
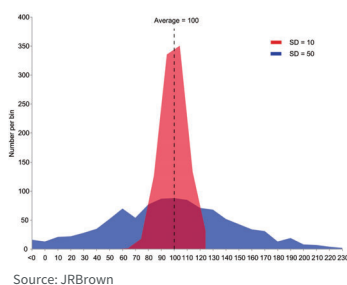
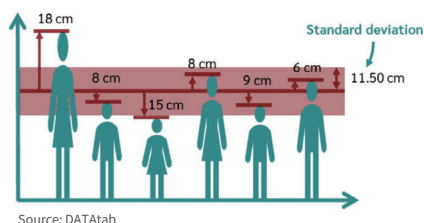


Source: CUEMATH

變數名稱	樣本數	平均值	標準差	最小值	中位數	最大值
屋齡(年)	64,286	25.58	16.13	0.00	29.96	59.96
總樓層數(層)	64,286	9.23	5.25	1.00	7.00	42.00
土地移轉總面積(平方公尺)	64,286	27.85	41.02	1.01	24.68	4,677.51
建物移轉面積不含車位(平方公尺)	64,286	103.49	57.66	14.03	96.11	1,346.49
單價(萬元/平方公尺)	64,286	20.97	9.99	3.71	18.86	171.99
總價不含車位價(萬元)	64,286	2,241.53	2,420.41	172.00	1,665.00	61,000.00

標準差 Standard Deviation
每個數據點跟平均值的差異程度

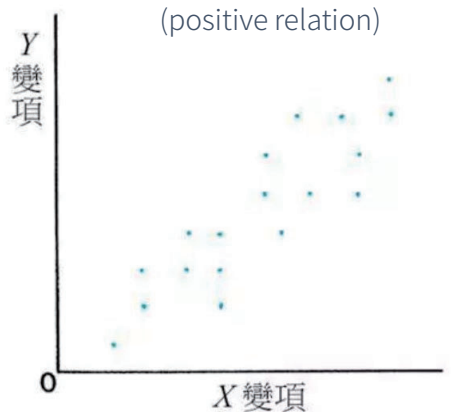
$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$$



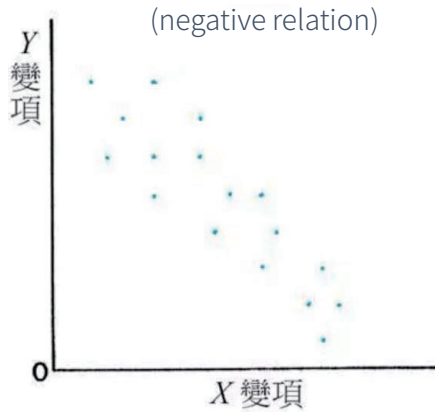
相關係數

正相關、負相關、無關係

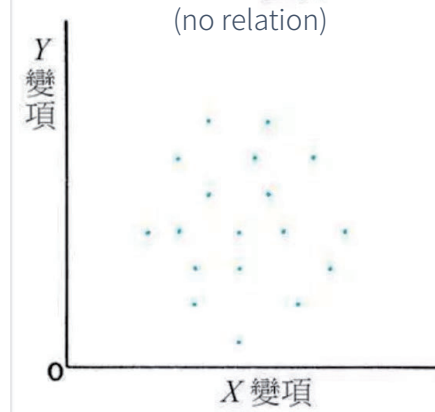
正相關
(positive relation)



負相關
(negative relation)

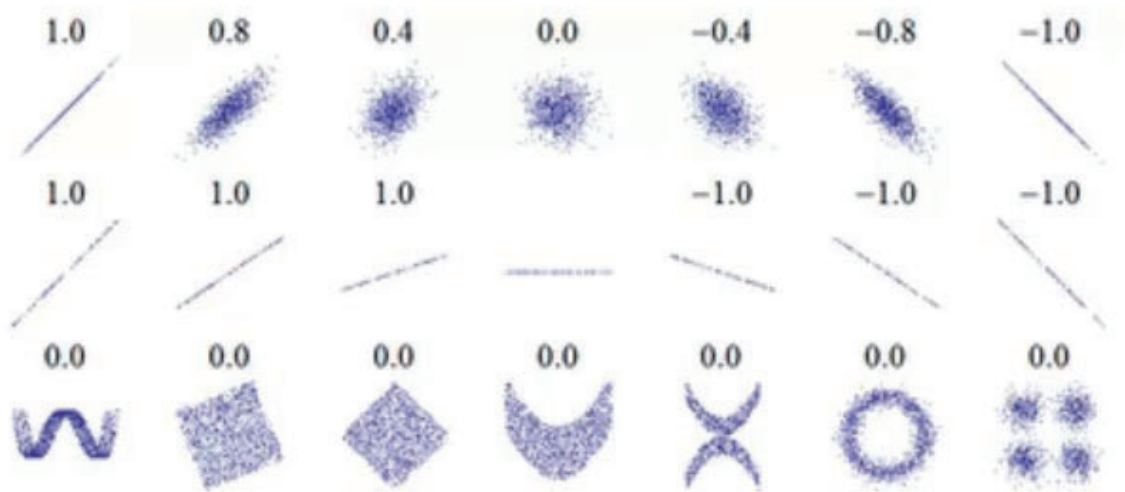


無關係
(no relation)



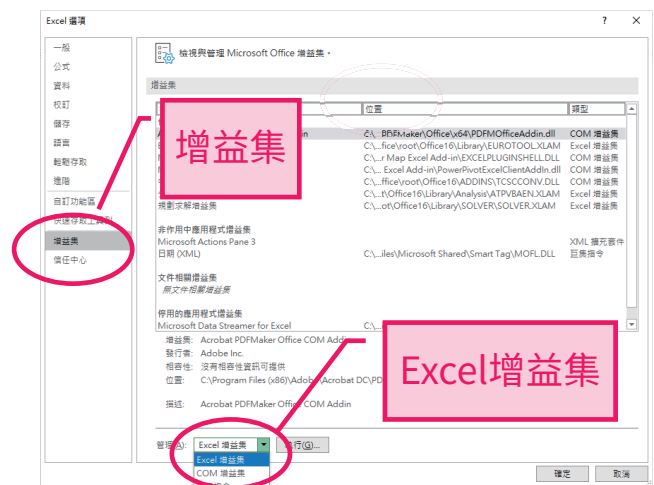
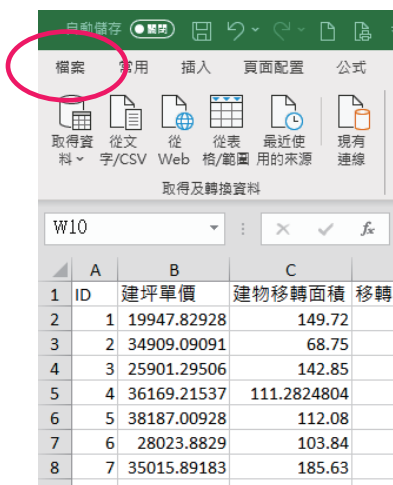
相關係數

無相關不等於無關係

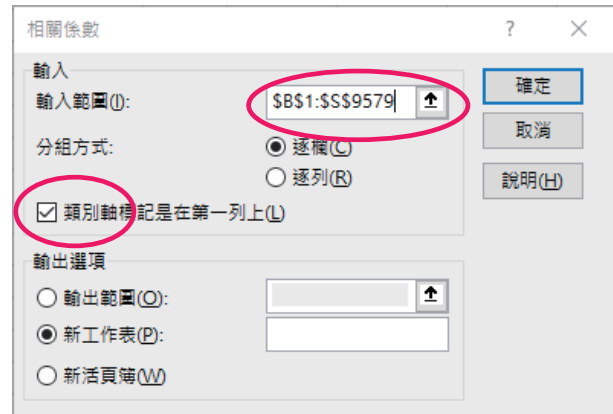
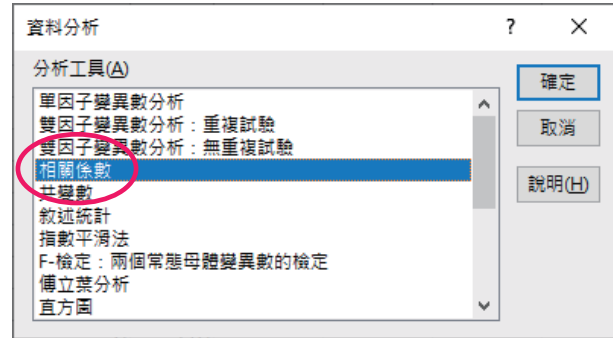
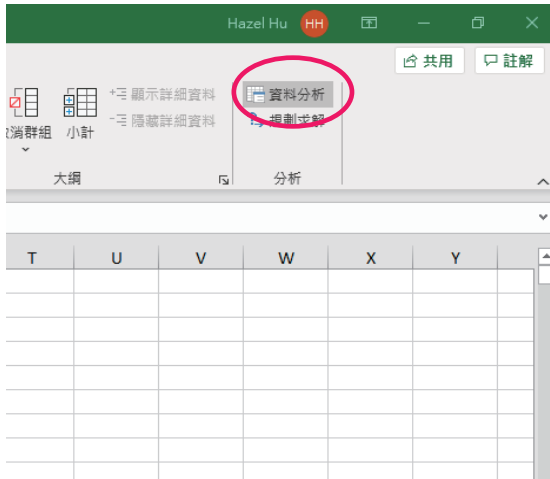


計算相關係數

運用Excel分析工具



計算相關係數



48

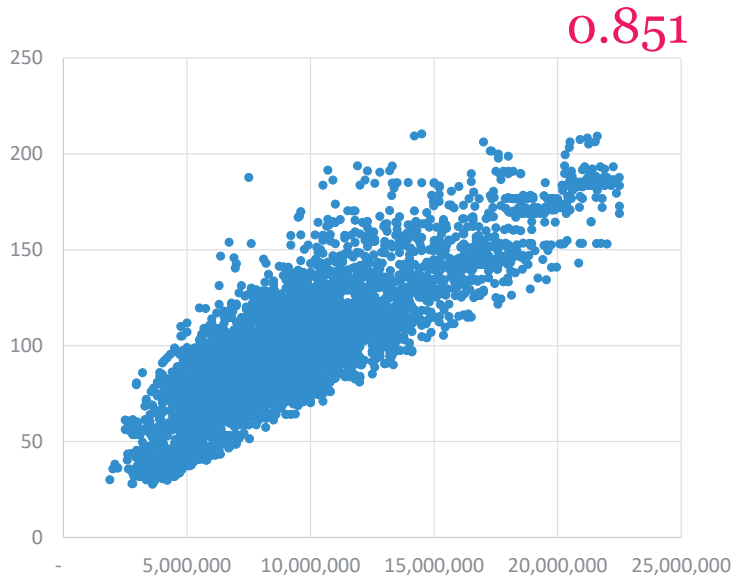
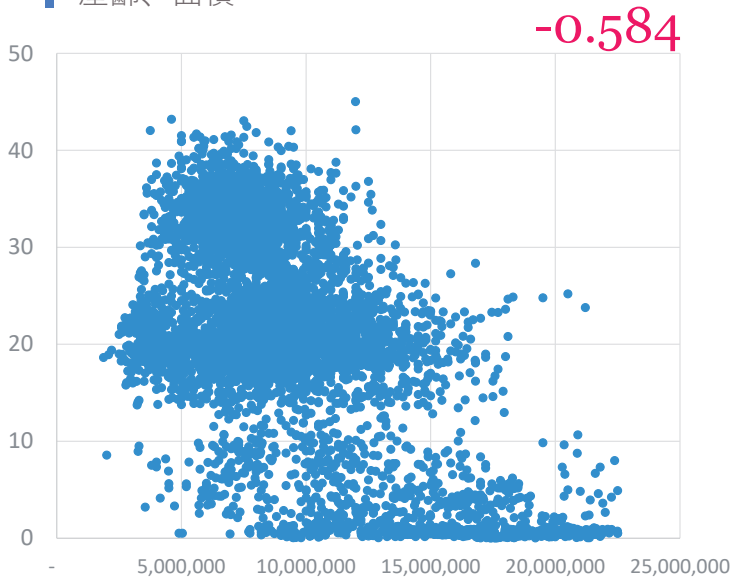
計算相關係數

	房地總價	所在樓層	總樓層	面積	屋齡	路寬	大樓	華廈	公寓	套房
房地總價	1.00									
所在樓層	0.32	1.00								
總樓層	0.50	0.68	1.00							
面積	0.85	0.23	0.36	1.00						
屋齡	-0.58	-0.49	-0.73	-0.48	1.00					
路寬	0.18	0.20	0.26	0.12	-0.15	1.00				
大樓	0.46	0.55	0.79	0.39	-0.60	0.22	1.00			
華廈	0.00	-0.19	-0.25	0.04	-0.10	-0.08	-0.40	1.00		
公寓	-0.34	-0.48	-0.68	-0.24	0.74	-0.18	-0.63	-0.32	1.00	
套房	-0.31	0.09	0.11	-0.41	-0.05	0.02	-0.22	-0.11	-0.17	1.00

49

變數與房價之關係

屋齡、面積



$$Y_i = \alpha + \beta_1 \text{是否為公寓}_i + \beta_2 \text{屋齡}_i + \beta_3 \text{交易樓層}_i + \beta_4 \text{總樓層}_i + \beta_5 \text{建物面積}_i + \beta_6 \text{土地面積}_i + \varepsilon_i$$

$$Y_i = 2,066,559 + (-3,733,443) \text{是否為公寓}_i + (-168,351) \text{屋齡}_i + 54,096 \text{交易樓層}_i + 92,172 \text{總樓層}_i + 285,096 \text{建物面積}_i + (-89,122) \text{土地面積}_i + \varepsilon_i$$

G	H	I	J	K	L	M
總價	是否為公寓	交易樓層	總樓層	建物面積	土地面積	
5080080	1	1	2	13	21.1	3.2
14502528	1	25	12	12	71.6	9.7
8200380	1	24	5	7	41.8	6.5

$$2,066,559 + (-3,733,443) * 1 + (-168,351) * 1 + 54,096 * 2 + 92,172 * 13 + 285,096 * 21.1 + (-89,122) * 3.2 = 5,191,272$$

	A	B	C	D	E	F	G	H	I
1	摘要輸出								
2									
3	迴歸統計								
4	R 的倍數	0.93							
5	R 平方	0.86							
6	調整的 R 平方	0.86							
7	標準誤	2128926.66							
8	觀察值個數	444							
9									
10	ANOVA								
11		自由度	SS	MS	F	顯著值			
12	迴歸	6	1.2163E+16	2.02717E+15	447.269222	5.1E-183			
13	殘差	437	1.98063E+15	4.53233E+12					
14	總和	443	1.41437E+16						
15									
16		係數	標準誤	t 統計	P-值	下限 95%	上限 95%	下限 95.0%	上限 95.0%
17	截距	2066559	848385.79	2.44	0.02	399135.5	3733983	399135.5	3733982.79
18	是否為公寓	-3733443	591658.38	-6.31	0.00	-4896293	-2570594	-4896293	-2570593.8
19	屋齡	-168351	11890.21	-14.16	0.00	-191720	-144982	-191720	-144982
20	交易樓層	54096	25810.19	2.10	0.04	3368.423	104823.5	3368.423	104823.495
21	總樓層	92172	34218.02	2.69	0.01	24919.69	159424.4	24919.69	159424.373
22	建物面積	285096	10914.22	26.12	0.00	263644.8	306546.6	263644.8	306546.58
23	土地面積	-89122	64555.10	-1.38	0.17	-215999	37755.49	-215999	37755.4901

	A	B	C	D	E	F	G
觀察值	預測為 總價	殘差	標準化殘差	總價	APE	10%HR	20HR
1	5,191,272	- 111,192	-0.05	5,080,080	0.021419	1	1
2	15,431,637	- 929,109	-0.44	14,502,528	0.060208	1	1
3	6,529,337	1,671,043	0.79	8,200,380	0.255928	0	0
4	15,791,089	2,294,918	1.09	18,086,007	0.14533	0	1
5	15,791,089	2,294,918	1.09	18,086,007	0.14533	0	1
6	10,970,555	530,649	0.25	11,501,204	0.04837	1	1
7	3,121,898	1,763,766	0.83	4,885,664	0.564966	0	0
8	26,535,235	- 2,443,749	-1.16	24,091,486	0.092094	1	1
9	2,588,293	2,351,029	1.11	4,939,322	0.908332	0	0
10	12,962,636	137,153	0.06	13,099,789	0.010581	1	1
11	4,493,821	- 445,306	-0.21	4,048,515	0.099093	1	1
12	13,295,926	6,621,153	3.13	19,917,079	0.497984	0	0
13	11,502,680	- 5,132,520	-2.43	6,370,160	0.446202	0	0
14	9,893,890	- 2,750,022	-1.30	7,143,868	0.277952	0	0

模型表現

01 MAPE

0.198

02 10% PPE

0.336

03 20% PPE

0.599

	MAPE	PPE10	PPE20	計次
公寓	0.375	2	3	16
住宅大樓	0.172	128	232	354
套房(1房(1廳)1衛)	0.622	0	0	3
華廈	0.271	19	31	71
總計	19.8%	33.6%	59.9%	444

臺南市實價登錄大數據輔助區段地價平衡（含整體開發區）之研究（第四期）

01

基礎統計介紹

解讀資料，辨識資訊意涵

02

認識電腦大量估價

AI世代來臨改變，提供創新服務

03

線性迴歸

特徵價格模型

04

地產數位轉型

數據分析做為決策應用

Digital Transfer

數位轉型

數位轉型是改變機構運作方式、最佳化內部資源。使用現代數位技術，打造或調整業務流程、文化和客戶體驗，以因應商業環境和市場局勢變化。

資料存進資料庫 $\neq$ 數位轉型



數位轉型的優勢



提升效率

利用新工具和功能，以更有效的方式管理來自不同裝置、來源和系統的龐大資料。

管理資料



應用資訊

利用最先進的工具執行更具智慧的業務分析作業，進而獲得更加精確的分析資料。

獲取深入分析資料



強化與客戶關係

導入新數位技術並建立現代化流程和應用程式，以快速找出並解決與企業營運和客戶互動方式相關問題。

解決業務問題



成本效益最佳化

善用數位工具來簡化流程，並找出業務中績效最高和績效較低的部分。

實現成本節約效益



Contact Info:

KarenTeng 鄧筱蓉

Vice President 副總經理

+886 2 2397 9993

Karen.teng@anfusolutions.com
Anfu Solutions Inc.

©安富財經科技股份有限公司版權所有。未經許可，不得轉載。本簡報刊載的一切資料，雖力求精確，但僅作參考之用，並非對報告中所載之市場趨勢或行情狀況做全面描述。簡報中所引用的數據資料來自中華民國政府之公開資料以及實價登錄，儘管安富財經科技相信其可靠性，也盡其所能之核實，但安富財經科技對於簡報所載資訊的準確性和完整性，不做任何明示或默示的擔保，也不承擔任何責任。安富財經科技在報告中所述的任何觀點僅供參考，並不對依賴該觀點而採取的任何措施或行動、以及由此引起的任何風險承擔任何責任。