

相關與迴歸函數

相關係數

- **CORREL(範圍1,範圍2)**
- 計算兩組數字範圍之相關係數，兩組數字範圍之資料點必須相同
- 所謂相關是指兩變項（X、Y）間相互發生之關聯
- 要瞭解相關，通常有二種方式
 - 繪製資料散佈圖（即Excel之XY圖）
 - 計算相關係數（表示相關程度強弱、相關方向異同之量數）

相關係數

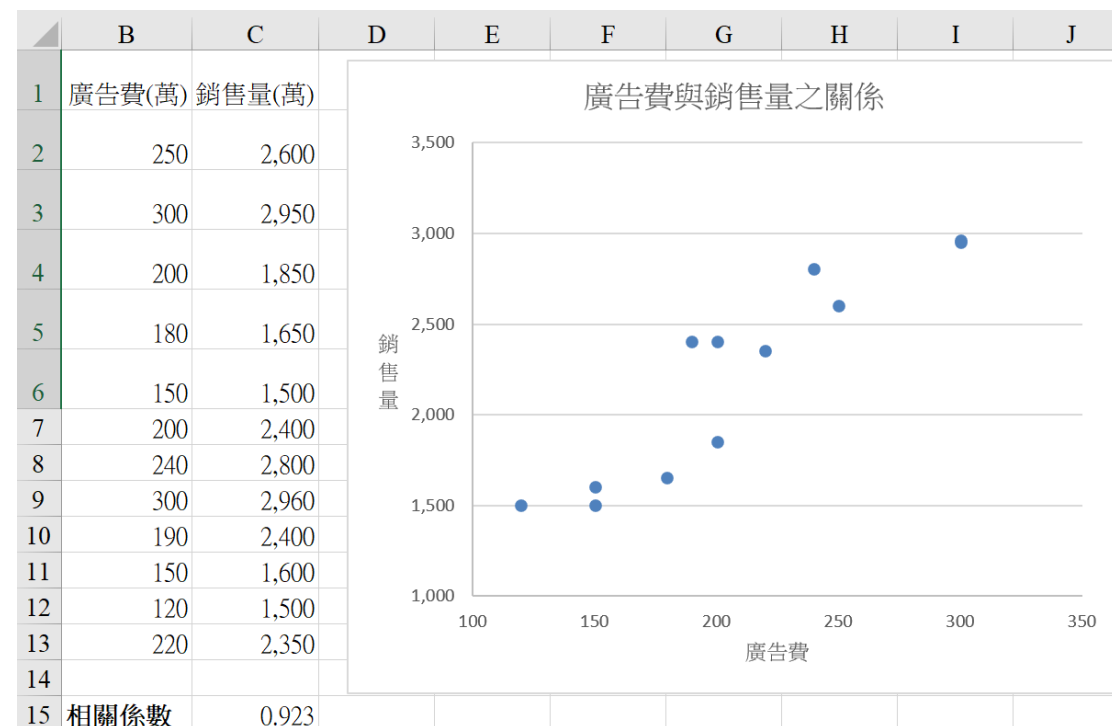
- 相關係數之計算公式為：

$$\rho_{x,y} = \frac{\frac{1}{n} \sum_{j=1}^n (x_j - \mu_x)(x_j - \mu_y)}{\sigma_x \cdot \sigma_y}$$

- 相關係數係一介於-1到+1之數字，其情況可有下列三種：
- =0 無關
- >0 正相關
- <0 負相關

相關係數

- 當相關係數之絕對值
 - 達0.4~0.6時，即為中度相關
 - 達0.6~0.8時，即為高度相關
 - 若達0.8以上時，即為非常高度相關
- 年度每月份之廣告費與銷售量之數字繪製XY散佈圖；
 - **=CORREL(B2:B13,C2:C13)**
 - 算出其相關係數為0.923
- 表銷售量與廣告費間存有極高度之正相關，銷售量會隨廣告費遞增而明顯增加。



相關係數

- 查看學童之仰臥起坐與伏地挺身個數之相關情況

	A	B	C	D
1	學童之仰臥起坐與伏地挺身個數			
2				
3		仰臥起坐	伏地挺身	
4		9	12	
5		30	40	
6		26	32	
7		25	30	
8		40	36	
9		18	21	
10		15	14	
11		20	16	
12		8	7	
13				
14		相關係數	0.91486	
15				

使用『資料分析』求相關矩陣

- CORREL()函數，僅能求算兩組資料間之相關係數。
- 使用「資料/資料分析...」
 - （得先安裝『增益集』之『分析工具箱』）
- 還可計算出多組資料間之相關係數，組成一個相關係數表

使用『資料分析』求相關矩陣

	A	B	C	D	E	F	G
1	汽車之鈹金，省油與價格之滿意度(5-很滿意，1-很不滿意)						
2							
3	鈹金	省油	價格				
4	4	3	2				
5	5	2	1				
6	4	3	3				
7	3	4	3				
8	3	3	4				
9	4	3	3				
10	1	5	5				
11	4	3	3				
12	3	3	4				
13	4	3	3				
14	5	2	2				
15	3	4	3				
16	2	5	5				
17	1	5	5				
18							

資料分析

分析工具(A)

- 單因子變異數分析
- 雙因子變異數分析：重複試驗
- 雙因子變異數分析：無重複試驗
- 相關係數**
- 共變數
- 敘述統計
- 指數平滑法
- F-檢定：兩個常態母體變異數的檢定
- 傅立葉分析
- 直方圖

確定 取消 說明(H)

相關係數

輸入

輸入範圍(I):

分組方式:

☒ 逐欄(C) ☐ 逐列(R)

☒ 類別軸標記是在第一列上(L)

輸出選項

☒ 輸出範圍(O):

☐ 新工作表(P):

☐ 新活頁簿(W)

確定 取消 說明(H)

使用『資料分析』求相關矩陣

- 點選「類別軸標記在第一列上(L)」（因各組資料均含標題之字串標記）
- 設定輸出範圍：目前工作表之E3
- 按〔確定〕鈕結束，即可獲致多組資料之相關係數表（因為對稱矩陣，故顯示一半即可）

	A	B	C	D	E	F	G	H	I
1	汽車之鉅金，省油與價格之滿意度(5-很滿意，1-很不滿意)								
2									
3	鉅金	省油	價格			鉅金	省油	價格	
4	4	3	2		鉅金	1			
5	5	2	1		省油	-0.93897	1		
6	4	3	3		價格	-0.91483	0.834895	1	
7	3	4	3						
8	3	3	4						
9	4	3	3						
10	1	5	5						
11	4	3	3						
12	3	3	4						
13	4	3	3						
14	5	2	2						
15	3	4	3						
16	2	5	5						
17	1	5	5						
18									

相關係數

輸入

輸入範圍(I):

分組方式:

☒ 逐欄(C)

☐ 逐列(R)

☒ 類別軸標記是在第一列上(L)

輸出選項

☒ 輸出範圍(O):

☐ 新工作表(P):

☐ 新活頁簿(W)

使用『資料分析』求相關矩陣

- 依此結果，顯示『鈹金與省油』及『鈹金與價格』之滿意度間均呈高度負相關，對鈹金越滿意對其省油與價格將越不滿意。鈹金好的車身重量大，當然較不省油，且其售價一般也比較高。
- 另外，『省油與價格』之滿意度間則呈高度正相關，因省油的車一般價位比較低之故。

	A	B	C	D	E	F	G	H	
1	汽車之鈹金，省油與價格之滿意度(5-很滿意，1-很不滿意)								
2									
3	鈹金	省油	價格			鈹金	省油	價格	
4	4	3	2		鈹金	1			
5	5	2	1		省油	-0.93897	1		
6	4	3	3		價格	-0.91483	0.834895	1	
7	3	4	3						

使用『資料分析』求相關矩陣

- 求本班學生上學期之總平均成績、出席率、選修學分數與每週打工時數間之相關係數表：

	F	G	H	I	J	K
1		總平均成績	出席率	選修學分數	每週打工時數	
2	總平均成績	1				
3	出席率	0.4848	1.0000			
4	選修學分數	0.5604	0.6045	1.0000		
5	每週打工時數	-0.7136	-0.2139	-0.1576	1	
6						

	A	B	C	D	E
1	總平均成績、出席率、選修學分數與每週打工時數				
2					
3	總平均成績	出席率	選修學分數	每週打工時數	
4	82	96%	14	4	
5	75	80%	16	8	
6	68	70%	10	10	
7	88	82%	12	0	
8	84	90%	14	6	
9	71	75%	10	0	
10	66	80%	12	15	
11	90	85%	16	0	
12	83	90%	18	4	
13	80	100%	15	4	
14	75	95%	14	6	
15					

繪圖中加入趨勢線

- 繪圖中加入趨勢線，一併求算迴歸公式是最簡便之方式。
- 且其可求算之迴歸種類也最多，包括：
 - 直線（一次式）
 - 多次式
 - 指數
 - 對數...等

直線迴歸

- 收集了某一廠牌同一車型中古車之車齡及其售價資料
- 擬繪製其資料散佈圖並求車齡對售價之迴歸方程式
- 選取**A1:B11**之範圍
- 插入『圖表』/『**XY**散佈圖』
- 執行「圖表(C)/加上趨勢線(R)...」
- 散佈圖顯示各圖點之分佈接近直線=『類型』/選「線性(L)」
- 『選項』標籤，選「圖表上顯示公式」與「圖表上顯示R-squared值(R)」

直線迴歸

	A	B	C
1	車齡	價格(萬)	
2	1	56.0	
3	2	48.5	
4	3	42.0	
5	4	37.6	
6	5	32.5	
7	6	28.7	
8	7	22.2	
9	8	18.5	
10	9	15.0	
11	10	12.5	
12			
13	6.5	26.54	
14			

趨勢線格式

趨勢線選項



趨勢線選項



☐ 指數(X)



☒ 線性(L)



☐ 對數(O)



☐ 多項式(P)

冪次(D)

2



☐ 乘冪(W)



☐ 移動平均(M)

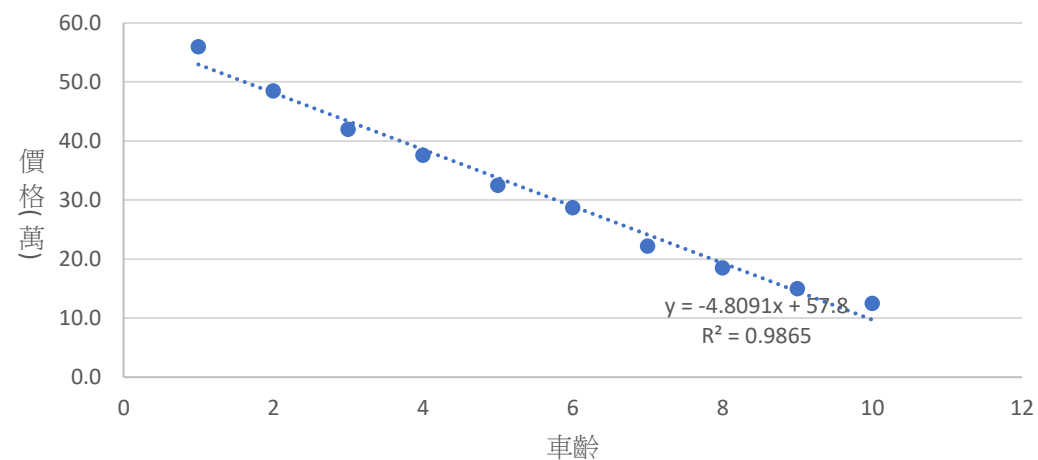
週期(E)

2

☒ 在圖表上顯示方程式(E)

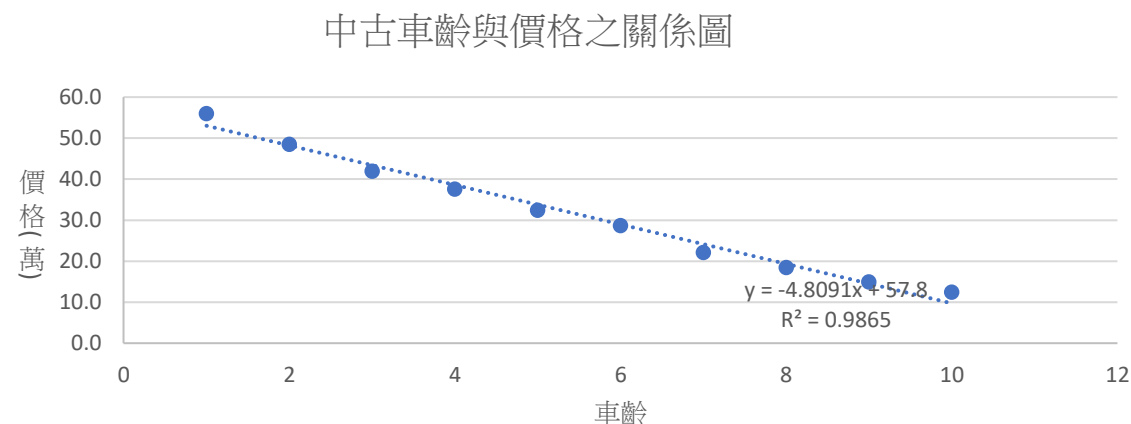
☒ 圖表上顯示 R 平方值(R)

中古車齡與價格之關係圖



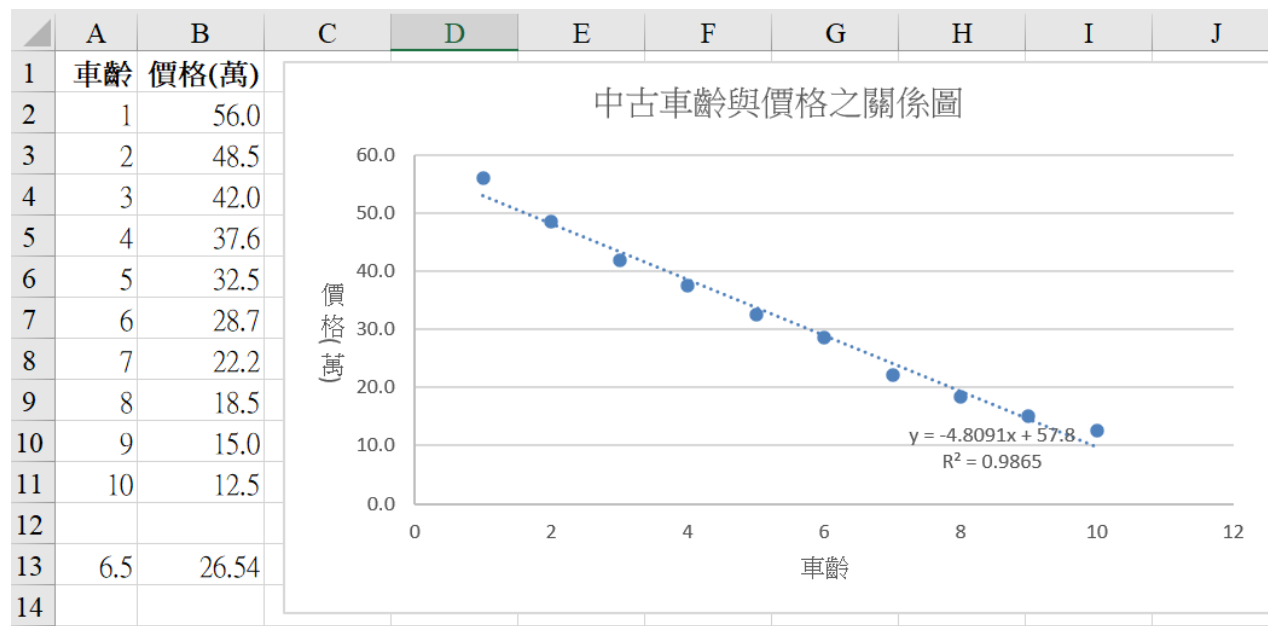
直線迴歸

- 可於圖表上獲致迴歸方程式及其判定係數（R-squared值，R2值）
- 其迴歸方程式為
 - $y = -4.8091x + 57.8$
 - 中古車車價 = $-4.8091 \times \text{車齡} + 57.8$
- 其判定係數0.9865，表整個迴歸模式之解釋力很強，即車齡的變異可解釋98.65%的售價差異。



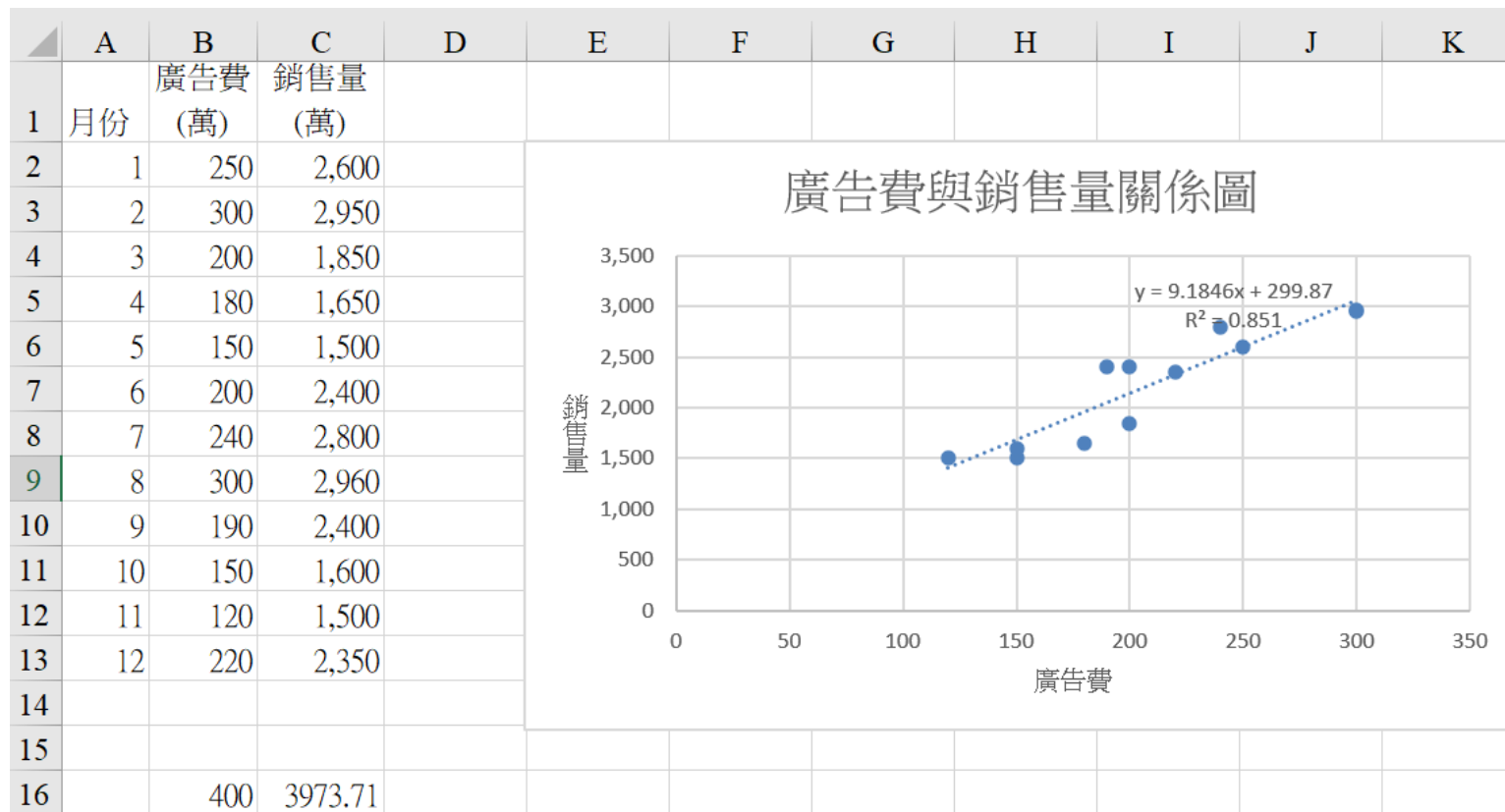
直線迴歸

- 取得迴歸方程式後，即可用以預測不同車齡之售價。
- 假定，要求當車齡為6.5年時，其售價應為多少？
- 僅須將6.5代入其迴歸方程式之x：
 - $y = -4.8091 \times (6.5) + 57.8$ 即
 - 中古車車價 = $-4.8091 \times 6.5 + 57.8 = 26.54$
 - 可求得其中古車車價為26.54萬



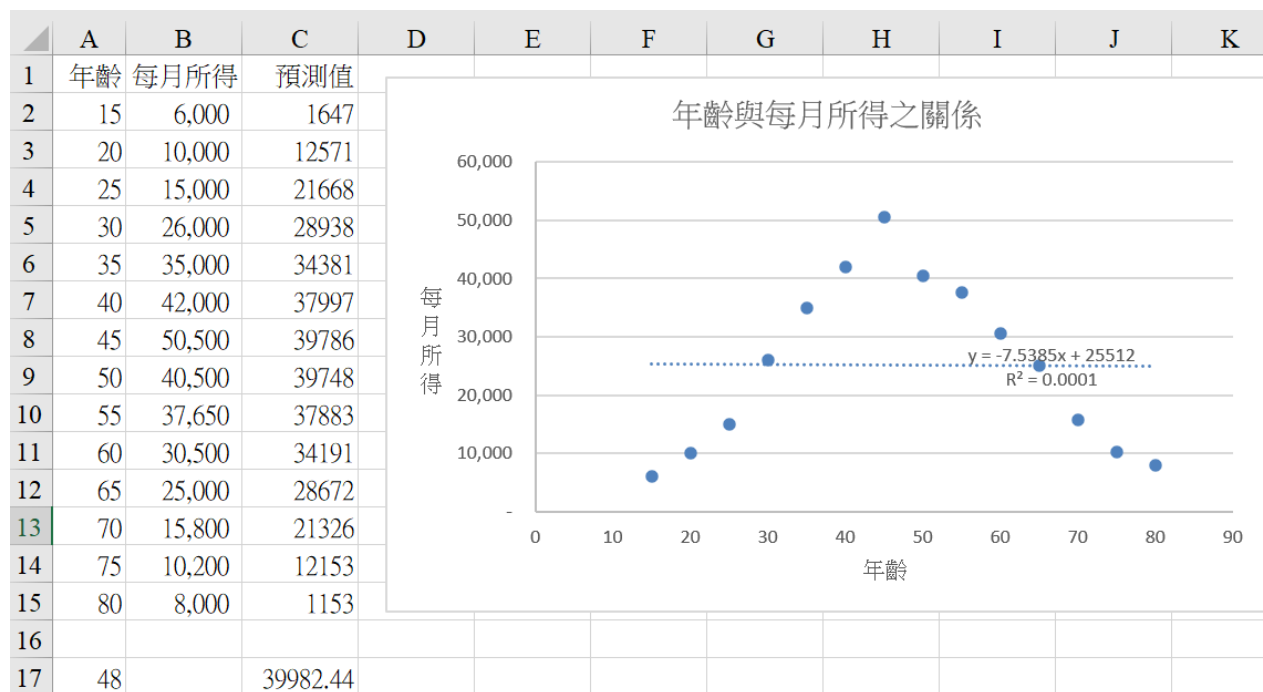
直線迴歸

- 依廣告費與銷售量之資料，繪製其資料散佈圖，求廣告費對銷售量之迴歸方程式，並求算當廣告費為400萬時，其預測之銷售量為多少？



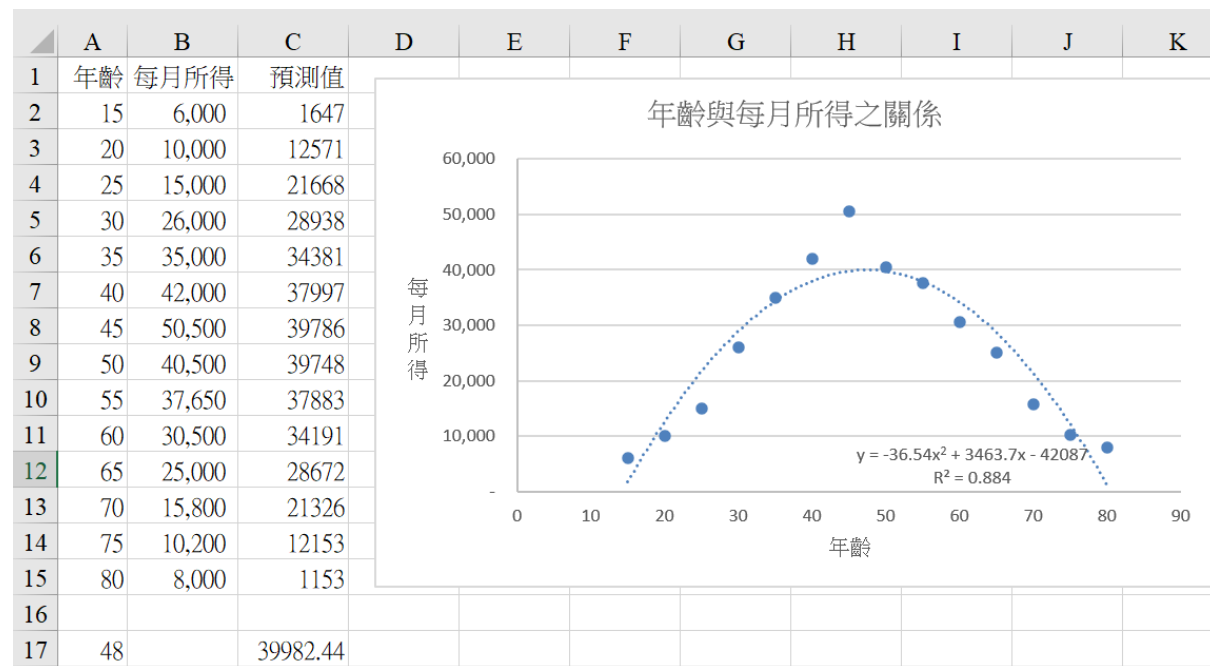
非線性迴歸

- 有些資料間並不是單純的直線關係，如下例之『年齡與所得之關係圖』資料，以「**線性(L)**」模式求其迴歸方程式，其判定係數（ R^2 ）僅為**0.0001**，根本不具任何解釋力。【**所得1**】工作表



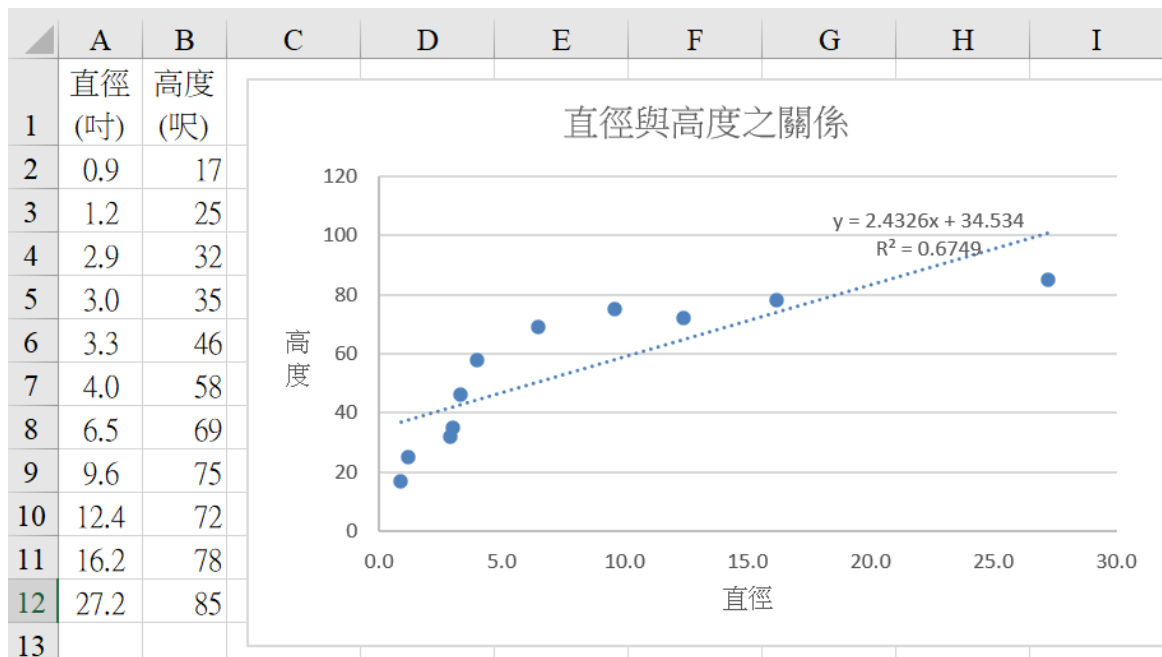
非線性迴歸

- 將其改為幕次為「2」之「多項式(P)」
- 其迴歸方程式為
 - $y = -36.54x^2 + 3463.7x - 42087$
 - 判定係數（R²）可高達0.884，就明顯較具解釋能力。



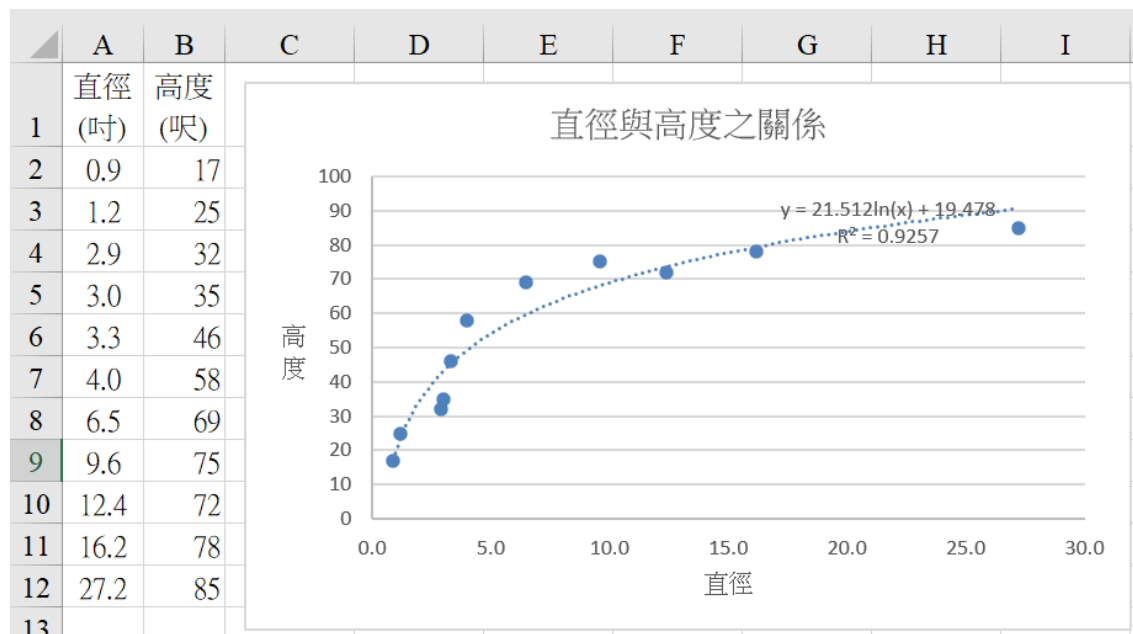
非線性迴歸

- 於繪圖中，利用加入趨勢線可求算之迴歸種類最多，包括：直線、多次式、指數、對數...等。
- 圖之樹木直徑與其高度之資料，以「線性(L)」進行迴歸，其判定係數（R²）僅為0.6749：



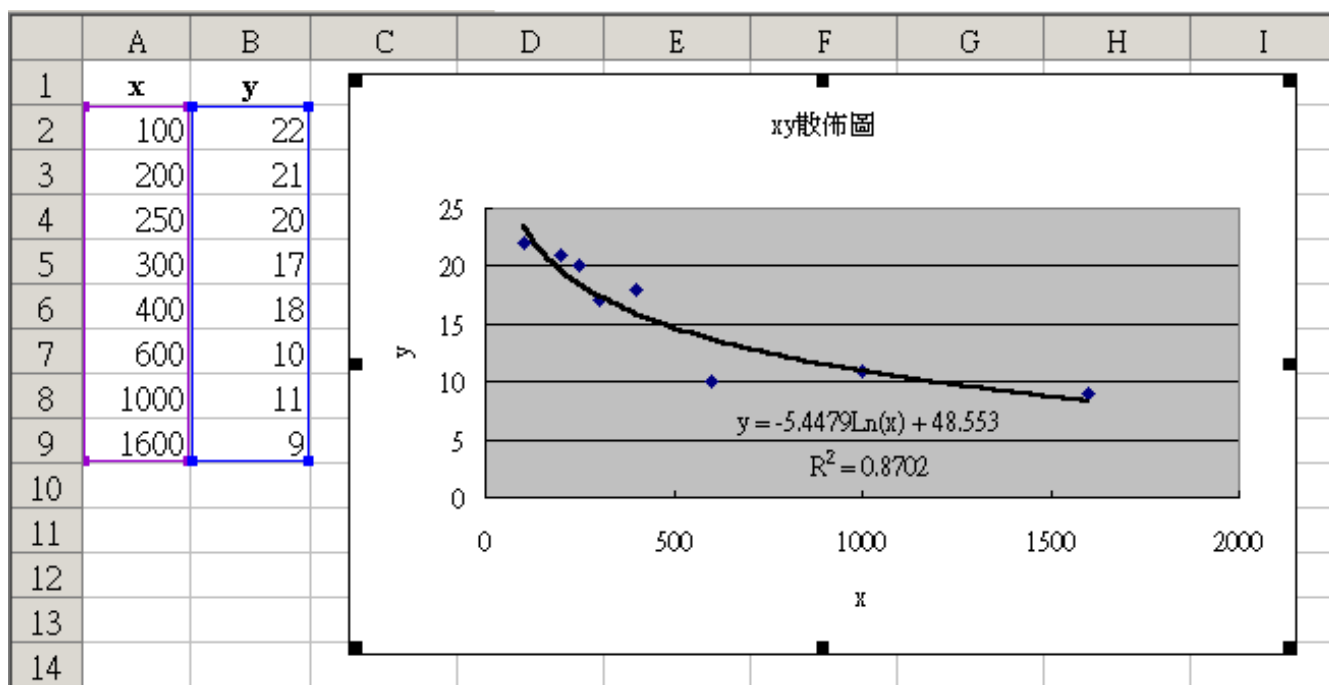
非線性迴歸

- 改為使用「對數(O)」模式：
- 其迴歸方程式為
 - $y = 21.512\ln(x) + 19.478$
 - 判定係數（R²）可高達0.9257，就很明顯的較直線模式更具解釋力



非線性迴歸

- 依資料，繪製XY散佈圖並求其對數迴歸方程式
- 【對數】



使用『資料分析』進行直線迴歸

- 於『輸入Y範圍』處，選取銷售額之範圍B1:B13
- 於『輸入X範圍』處，選取廣告費之範圍A1:A13
- 由於上述兩範圍均含標記，故點選「標記(L)」
- 點選「信賴度(O)」，可計算迴歸係數之信賴區間（預設值為95%）
- 於『輸出選項』/「輸出範圍(O) :D1
- 分析殘差：選「殘差(R)」或「標準化殘差(T)」（本例選前者）
- 繪圖：選「殘差圖(D)」或「樣本迴歸線圖(I)」（本例選後者）

迴歸

輸入

輸入 Y 範圍(Y): \$B\$2:\$B\$11

輸入 X 範圍(X): \$A\$2:\$A\$11

☒ 標記(L) ☒ 常數為零(Z)

☒ 信賴度(O) 95 %

輸出選項

☒ 輸出範圍(O): \$D\$1

☐ 新工作表(P):

☐ 新活頁簿(W)

殘差

☒ 殘差(R) ☐ 殘差圖(D)

☐ 標準化殘差(T) ☒ 樣本迴歸線圖(I)

常態機率

☐ 常態機率圖(N)

確定 取消 說明(H)

使用『資料分析』進行直線迴歸

- 摘要輸出結果一迴歸統計：
- 此部份在求算簡單相關係數（R，寫成『R的倍數』應為將 **coefficient of multiple correlation** 翻譯錯了，在複迴歸模式＝複相關係數）、判定係數（R平方）、調整後的R平方標準誤與觀察值個數。

D	E
摘要輸出	
迴歸統計	
R 的倍數	0.99653347
R 平方	0.993078956
調整的 R 平方	0.881967845
標準誤	203.4572236
觀察值個數	10

使用『資料分析』進行直線迴歸

- 此部份以t檢定，判斷迴歸係數與常數項是否為0（為0即無直線關係存在）？並求其信賴區間。
- 其虛無假設為迴歸係數與常數項為0，判斷是否顯著，只須看顯著水準（P-值）是否小於我們所指定之 α 值即可，如本例之常數項（截距）為1545.832，其t統計量為4.905255，顯著水準（P-值） $0.000618 < \alpha = 0.05$ ，故棄卻其為0之虛無假設，迴歸方程式之常數項不應為0，故不可將其省略。

	D	E	F	G	H	I	J	K	L
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	1545.831536	315.13783	4.9052554	0.0006183	843.66058	2248.0025	843.6605796	2248.00249
18	廣告費(萬)	3.14338351	1.2120872	2.5933641	0.0267949	0.4426844	5.8440827	0.442684371	5.84408265

使用『資料分析』進行直線迴歸

- 另，自變數X（廣告費）的迴歸係數為3.143384，其t統計量為2.593364，顯著水準（P-值） $0.026795 < \alpha = 0.05$ ，故棄卻其為0之虛無假設，迴歸方程式之自變數X的係數不為0，自變數與因變數間存有直線關係。
- 所以，迴歸方程式應為

$$y = 3.143384x + 1545.832$$

	D	E	F	G	H	I	J	K	L
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	1545.831536	315.13783	4.9052554	0.0006183	843.66058	2248.0025	843.6605796	2248.00249
18	廣告費(萬)	3.14338351	1.2120872	2.5933641	0.0267949	0.4426844	5.8440827	0.442684371	5.84408265

使用『資料分析』進行直線迴歸

- 此部份為於求得迴歸方程式後，將各觀察值之X（廣告費）代入方程式以求其預測之銷售量（萬），並計算預測結果與原實際銷售量間之誤差。
- 將兩者相減即可求得殘差，如觀察值1之廣告費為245萬，代入方程式所求得之預測銷售量為2315.960496萬，以原實際銷售量3000萬減去預測結果即為殘差684.0395萬。

	B	C	D	E	F	G
19						
20						
21						
22			殘差輸出			
23						
24			觀察值	則為 銷售量(	殘差	標準化殘差
25			1	2616.82722	-16.827219	-0.087180222
26			2	3140.19266	-190.19266	-0.985370122
27			3	2093.46177	-243.46177	-1.261352335
28			4	1570.09633	-70.096331	-0.363162435
29			5	2093.46177	306.53823	1.588145432
30			6	2512.15413	287.84587	1.491302116
31			7	3140.19266	-180.19266	-0.933561071
32			8	1570.09633	29.903669	0.154928068
33			9	1256.07706	243.92294	1.263741561
34			10	2302.80795	47.192048	0.244497517

使用『資料分析』進行複迴歸

- 現實中，很多狀況並非簡單之單一變數即可以解釋清楚。如；銀行計算客戶之信用分數，亦不會只決定於其每月所得而已，其動產、不動產甚或年齡、性別、教育程度、...等，亦均有可能影響其信用分數。
- 故於迴歸中，同時使用多個自變數以預測某一因變數的情況已越來越多。
- 這種同時使用多個自變數之迴歸，即稱為複迴歸（multiple regression）或多元迴歸。

使用『資料分析』進行複迴歸

- 複相關係數（R）為0.971605，判定係數（R平方）為0.944016、調整後的R平方為0.928021。
- 顯示整組迴歸方程式可解釋價格差異之程度相當高

	F	G	H	I	J	K
1	摘要輸出					
2						
3	迴歸統計					
4	R 的倍數	0.971605				
5	R 平方	0.944016				
6	調整的 R 平方	0.928021				
7	標準誤	4.692793				
8	觀察值個數	10				
9						
10	ANOVA					
11		自由度	SS	MS	F	顯著值
12	迴歸	2	2599.444	1299.722	59.01842	4.15E-05
13	殘差	7	154.1562	22.02231		
14	總和	9	2753.6			

使用『資料分析』進行複迴歸

- 最後之t檢定結果中，常數項（截距）為62.6468，其顯著水準（P-值） $2.3E-7 < \alpha = 0.05$ ，故棄卻其為0之虛無假設，迴歸方程式之常數項不應為0
- 兩個自變數中之車齡的迴歸係數為-5.37885，其顯著水準（P-值） $0.003084 < \alpha = 0.05$ ，故棄卻其為0之虛無假設，車齡與價格間存有直線關係。由其係數為負值，顯示車齡與價格間之關係為一負相關，車齡愈大售價愈低。

	F	G	H	I	J	K	L	M	N
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	62.6468	3.207102	19.53377	2.3E-07	55.06321	70.23039	55.063214	70.230385
18	車齡	-5.37385	1.216041	-4.41914	0.003084	-8.24933	-2.49838	-8.24933	-2.498377
19	里程數 (萬公里)	-0.22924	1.059104	-0.21644	0.834814	-2.73362	2.275143	-2.733617	2.2751431

使用『資料分析』進行複迴歸

- 另一個自變數里程數的迴歸係數為-0.22924，其P-值0.834814> $\alpha=0.05$ ，故無法棄卻其為0之虛無假設，里程數與價格間並無直線關係。故可將此一係數自迴歸方程式中排除掉。（少掉一個變數，即可省去蒐集其資料之時間與成本）
- 所以，最後之迴歸方程式應為

$$y = -5.37885X_1 + 62.6468$$

（價格 = -5.37885×車齡 + 62.6468）

	F	G	H	I	J	K	L	M	N
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	62.6468	3.207102	19.53377	2.3E-07	55.06321	70.23039	55.063214	70.230385
18	車齡	-5.37385	1.216041	-4.41914	0.003084	-8.24933	-2.49838	-8.24933	-2.498377
19	里程數 (萬公里)	-0.22924	1.059104	-0.21644	0.834814	-2.73362	2.275143	-2.733617	2.2751431

信用分數之實例

- 假定，銀行為核發信用卡，而蒐集了申請人之每月總收入、不動產、動產、每月房貸與扶養支出費用等資料，並以主管之經驗，主觀的給予一信用分數：（【[信用分數](#)】工作表）
- 為使評估信用分數能有一套公式，免得老是要主管抽空評分。擬以複迴歸來求得一迴歸方程式，其處理步驟為：

	A	B	C	D	E	F	G
1	每月總收入 (萬)	不動產 (百萬)	動產 (百萬)	每月房貸 (萬)	扶養支出 (萬)	信用分數	
2	6.5	12.0	3.0	2.0	2.0	82	
3	7.2	8.0	2.0	0.0	2.0	86	
4	3.8	0.0	1.0	0.0	1.0	70	
5	8.5	15.0	3.5	2.8	2.0	90	
6	4.2	0.0	0.5	0.0	1.5	75	
7	6.3	6.0	2.0	1.8	1.5	80	
8	3.0	0.0	0.6	0.0	0.0	70	
9	3.2	3.0	1.0	0.0	2.0	65	
10							

信用分數之實例

- 複相關係數（R）為0.990989，判定係數（R平方）為0.98206、調整後的R平方為0.93721。
- 顯示整組迴歸方程式可解釋信用分數差異之程度相當高。

	H	I	J	K	L	M
1	摘要輸出					
2						
3	迴歸統計					
4	R 的倍數	0.990989				
5	R 平方	0.98206				
6	調整的 R 平方	0.93721				
7	標準誤	2.179361				
8	觀察值個數	8				
9						
10	ANOVA					
11		自由度	SS	MS	F	顯著值
12	迴歸	5	520.0008	104.0002	21.89655	0.044248
13	殘差	2	9.499228	4.749614		
14	總和	7	529.5			

信用分數之實例

- ANOVA表中之F檢定的顯著水準 $0.044248 < \alpha = 0.05$ ，故其結果為棄卻因變數與自變數間無迴歸關係存在之虛無假設。
- 顯示每月總收入、不動產、動產、每月房貸、扶養支出與信用分數整體間有明顯迴歸關係存在。

	H	I	J	K	L	M
1	摘要輸出					
2						
3	迴歸統計					
4	R 的倍數	0.990989				
5	R 平方	0.98206				
6	調整的 R 平方	0.93721				
7	標準誤	2.179361				
8	觀察值個數	8				
9						
10	ANOVA					
11		自由度	SS	MS	F	顯著值
12	迴歸	5	520.0008	104.0002	21.89655	0.044248
13	殘差	2	9.499228	4.749614		
14	總和	7	529.5			

信用分數之實例

- 所有五個自變數中，僅每月總收入之顯著水準（P-值）為 $0.032912 < \alpha = 0.05$ ，可棄卻其為0之虛無假設，表示每月總收入與信用分數間存有直線關係。其係數為5.350913，顯示每月總收入與信用分數間之關係為正相關，收入愈高信用分數愈高。

	H	I	J	K	L	M	N	O	P
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	57.0761	4.950432	11.52952	0.007439	35.7761	78.37611	35.7761	78.37611
18	每月總收入 (萬)	5.350913	0.995484	5.375187	0.032912	1.067688	9.634138	1.067688	9.634138
19	不動產 (百萬)	0.703921	0.930382	0.756593	0.528274	-3.29919	4.707036	-3.29919	4.707036
20	動產 (百萬)	-4.96189	5.445107	-0.91126	0.458351	-28.3903	18.46653	-28.3903	18.46653
21	每月房貸 (萬)	-0.08989	1.715809	-0.05239	0.96298	-7.47243	7.292645	-7.47243	7.292645
22	扶養支出 (萬)	-2.49919	1.704976	-1.46582	0.280338	-9.83511	4.836735	-9.83511	4.836735

信用分數之實例

- 其餘之不動產、動產、每月房貸與扶養支出等四個變數之顯著水準（P-值）均大於 $\alpha=0.05$ ，故無法棄卻其為0之虛無假設，顯示信用分數與這些變數間並無顯著之直線關係。故可將這些變數之係數自迴歸方程式中排除掉。
 - 最後之迴歸方程式： $y = 5.350913X_1 + 57.0761$
- 信用分數 = $5.350913 \times \text{每月總收入} + 57.0761$

	H	I	J	K	L	M	N	O	P
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	57.0761	4.950432	11.52952	0.007439	35.7761	78.37611	35.7761	78.37611
18	每月總收入 (萬)	5.350913	0.995484	5.375187	0.032912	1.067688	9.634138	1.067688	9.634138
19	不動產 (百萬)	0.703921	0.930382	0.756593	0.528274	-3.29919	4.707036	-3.29919	4.707036
20	動產 (百萬)	-4.96189	5.445107	-0.91126	0.458351	-28.3903	18.46653	-28.3903	18.46653
21	每月房貸 (萬)	-0.08989	1.715809	-0.05239	0.96298	-7.47243	7.292645	-7.47243	7.292645
22	扶養支出 (萬)	-2.49919	1.704976	-1.46582	0.280338	-9.83511	4.836735	-9.83511	4.836735

信用分數之實例

練習：老師為找出學生出席率高低之主要原因，以問卷調查蒐集了下列資料，試以複迴歸求出席率高低之迴歸方程式：(【出席率】工作表)

$$\text{出席率} = 0.47928 + 0.1000 * \text{上課內容} - 0.0574 * \text{上課時段}$$

	A	B	C	D	E	F
1	受測者對影響出席率之因素的同意程度					
2	是否點名	成績高低	上課內容	上課時段	出席率	
3	2	3	5	2	95%	
4	1	5	3	4	65%	
5	3	3	5	2	100%	
6	2	2	5	3	90%	
7	2	2	3	4	60%	
8	2	3	2	2	65%	
9	1	2	5	1	95%	
10	2	2	4	2	90%	
11	3	3	5	2	95%	
12						
13	5-非常同意，1-非常不同意					
14						

	H	I	J	K	L	M	N
10	ANOVA						
11		自由度	SS	MS	F	顯著值	
12	迴歸	4	0.192545	0.048136	30.35348	0.002987	
13	殘差	4	0.006343	0.001586			
14	總和	8	0.198889				
15							
16		係數	標準誤	t 統計	P-值	下限 95%	上限 95%
17	截距	0.479278	0.098561	4.862774	0.008263	0.205629	0.752927
18	是否點名	0.026372	0.020997	1.255964	0.277485	-0.03193	0.084669
19	成績高低	0.012951	0.015849	0.817132	0.459726	-0.03105	0.056956
20	上課內容	0.100028	0.014268	7.010792	0.00218	0.060414	0.139641
21	上課時段	-0.05741	0.016316	-3.51863	0.02448	-0.10271	-0.01211
22							
23	出席率與上課內容呈正相關；與上課時段呈負相關						
24							

含二次式之複迴歸

	A2		fx	=B2^2
	A	B	C	D
1	年齡平方	年齡	每月所得	
2	225	15	6,000	
3	400	20	10,000	
4	625	25	15,000	
5	900	30	26,000	
6	1225	35	35,000	
7	1600	40	42,000	
8	2025	45	50,500	
9	2500	50	40,500	
10	3025	55	37,650	
11	3600	60	30,500	
12	4225	65	25,000	
13	4900	70	15,800	
14	5625	75	10,200	
15	6400	80	8,000	
16				

迴歸

輸入

輸入 Y 範圍(Y):

輸入 X 範圍(X):

☒ 標記(L) ☐ 常數為零(Z)

☒ 信賴度(O) %

輸出選項

☒ 輸出範圍(O):

☐ 新工作表(P):

☐ 新活頁簿(W)

殘差

☐ 殘差(R) ☐ 殘差圖(D)

☐ 標準化殘差(T) ☐ 樣本迴歸線圖(I)

常態機率

☐ 常態機率圖(N)

確定 取消 說明(H)

含二次式之複迴歸

7. 按〔確定〕鈕，即可獲致迴歸結果

- 判定係數（R平方）為**0.884033**，ANOVA表中之F檢定的顯著水準 **$7.14E-06 < \alpha = 0.05$** ，故其結果為棄卻因變數與自變數間無迴歸關係存在之虛無假設。

	E	F	G	H	I	J
1	摘要輸出					
2						
3	迴歸統計					
4	R 的倍數	0.94023				
5	R 平方	0.884033				
6	調整的 R 平方	0.862948				
7	標準誤	5383.55				
8	觀察值個數	14				
9						
10	ANOVA					
11		自由度	SS	MS	F	顯著值
12	迴歸	2	2.43E+09	1.22E+09	41.9271	7.14E-06
13	殘差	11	3.19E+08	28982611		
14	總和	13	2.75E+09			

含二次式之複迴歸

- 最後之t檢定結果中，常數項（截距）、年齡平方與年齡等之顯著水準（P-值）均小於 $\alpha=0.05$ ，故棄卻其為0之虛無假設，故均不可將其省略。
- 迴歸方程式： $y = -36.53984x^2 + 3463.746x - 42087$
- 每月所得 = $-36.53984 \times \text{年齡平方} + 3463.746 \times \text{年齡} - 42087$

	E	F	G	H	I	J	K	L	M
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	-42087	8250.414	-5.1012	0.000343	-60246.1	-23928	-60246.1	-23928
18	年齡平方	-36.5398	3.990553	-9.15658	1.77E-06	-45.323	-27.7567	-45.323	-27.7567
19	年齡	3463.746	385.7649	8.978904	2.14E-06	2614.683	4312.809	2614.683	4312.809

含二次式之複迴歸

練習：下示資料分佈情況接近對數圖形，試新增一欄 $\ln(x)$ 資料，並以「資料/資料分析...」求其迴歸方程式：【對數】工作表

$$y = -6.9593\ln(x) + 57.834$$

	A	B	C
1	x	y	
2	100	25	
3	200	22	
4	250	21	
5	300	17	
6	400	18	
7	600	10	
8	1000	9	
9	1600	8	
10			

	C2		f _x	=LN(A2)
	A	B	C	D
1	x	y	Ln(x)	
2	100	25	4.605	
3	200	22	5.298	
4	250	21	5.521	
5	300	17	5.704	
6	400	18	5.991	
7	600	10	6.397	
8	1000	9	6.908	
9	1600	8	7.378	
10				

	E	F	G	H	I	J	K
17	迴歸統計						
18	R 的倍數	0.96056					
19	R 平方	0.922675					
20	調整的 R 平方	0.909788					
21	標準誤	1.951472					
22	觀察值個數	8					
23							
24	ANOVA						
25		自由度	SS	MS	F	顯著值	
26	迴歸	1	272.6505	272.6505	71.59481	0.000149	
27	殘差	6	22.84947	3.808244			
28	總和	7	295.5				
29							
30		係數	標準誤	t 統計	P-值	下限 95%	上限 95%
31	截距	57.83382	4.962743	11.6536	2.41E-05	45.69042	69.97722
32	Ln(x)	-6.95925	0.822473	-8.46137	0.000149	-8.97177	-4.94673
33							

直線迴歸LINEST()函數

LINEST(因變數範圍, 自變數範圍, 常數項是否不為0, 統計值)

- 本函數使用最小平方法計算最適合於因變數範圍之迴歸直線公式，並傳回該直線公式的陣列。
- 陣列公式：
 - 先選取輸出範圍
 - 於輸入公式後，以Ctrl+Shift+Enter完成輸入
 - 才可傳回完整之陣列公式

直線迴歸LINEST()函數

- 因變數範圍與自變數範圍之儲存格個數應一致。
- 常數項是否不為0係一邏輯值，省略或TRUE，表迴歸公式中應計算出常數項，公式變為 $y = a + bx$ ；FALSE表將常數項安排為0，公式變為 $y = bx$ 。
- 統計值也是一邏輯值，用以設定是否要傳回額外的迴歸直線統計值。為TRUE時，將依對應位置，傳回所有統計值

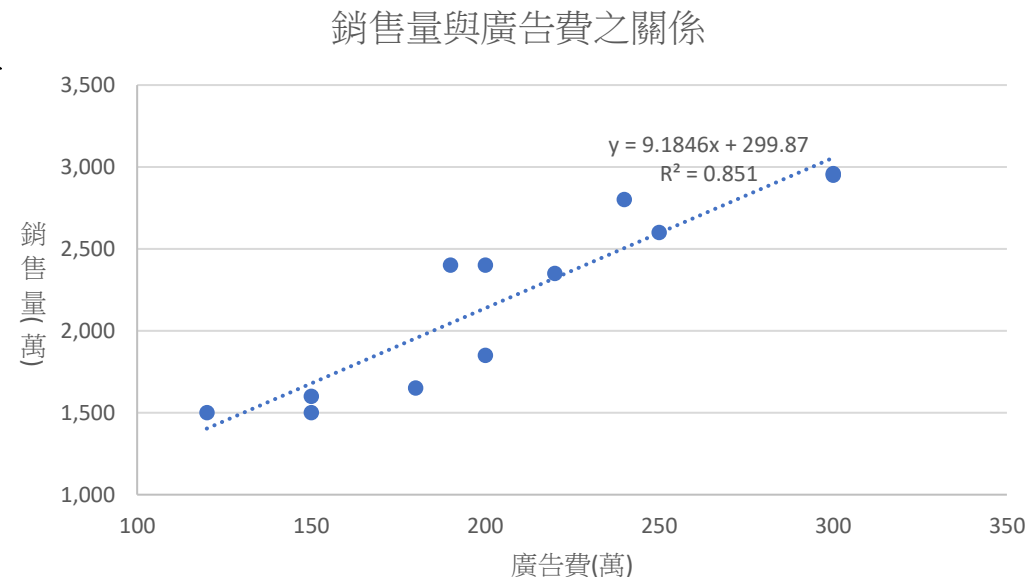
	A	B	C	D	E
1	單一變量時				
2	係數(b)	常數(a)			
3	標準誤(b)	標準誤(a)			
4	判定係數(r^2)	對y 估計值的標準誤差			
5	F 統計值	F檢定之自由度			
6	迴歸平方	殘差平方			
7					
8	多變量時				
9	係數(b_n)	係數 b_{n-1}	...	係數1(b_1)	常數(a)
10	標準誤(b_n)	標準誤(b_{n-1})	...	標準誤(b_1)	標準誤(a)
11	判定係數(r^2)	對y 估計值的標準誤差			
12	F 統計值	F檢定之自由度			
13	迴歸平方	殘差平方			

直線迴歸LINEST()函數

- 統計值若省略或安排為**FALSE**時，將只傳回常數（**a**）與迴歸係數（**b**，即斜率）。**F**統計值是用來判斷自變數和因變數間的關係是否是巧合？
- 這應該是最難懂的迴歸方式了，因為使用者常不清楚要選取多少範圍才算正確，往往多選而導致一堆的**#N/A**值。
- 且其結果也無字串說明，一下丟給您一堆數字，那個數字代表什麼？非仔細對照，很難知曉其作用！不小心，對錯了，也是在所難免。

單一變量資料

- 廣告費與銷售量之單一變量資料為例：【[銷售1](#)】工作表
- 其迴歸方程式與判定係數（ R^2 ）分別為
 - $y = 9.1846x + 299.87$
 - $R^2 = 0.851$
- 若要以LINEST()來求得這些數字，得以下示步驟求得：



單一變量資料

5. 依前述對應位置，找出各統計數字
其迴歸方程式與判定係數（ R^2 ）分別為

$$y = 9.1846298x + 299.8688$$

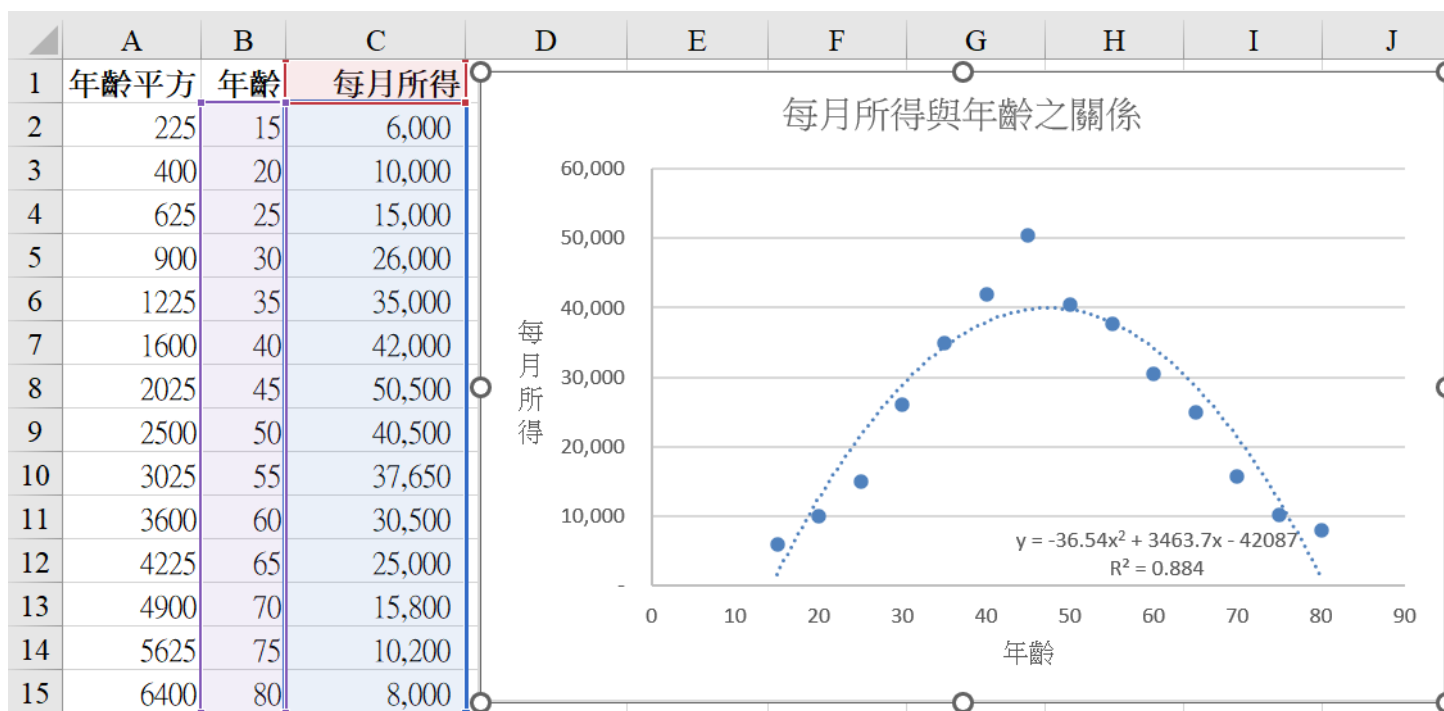
$$R^2 = 0.8510398$$

	B17	fx {=LINEST(B3:B14,A3:A14,TRUE,TRUE)}					
	A	B	C	D	E	F	G
16							
17		9.1846298	299.8688				
18		1.2151275	261.6524				
19		0.8510398	229.1624				
20		57.13202	10				
21		3000312.4	525154.3				
22							

	B17	fx {=LINEST(B3:B14,A3:A14,TRUE,TRUE)}					
	A	B	C	D	E	F	G
17	係數(b) →	9.1846298	299.8688	← 常數(a)			
18	標準誤(b) →	1.2151275	261.6524	← 標準誤(a)			
19	判定係數(r^2)	0.8510398	229.1624	← 對y 估計值的標準誤差			
20	F 統計值 →	57.13202	10	← F檢定之自由度			
21	迴歸平方 →	3000312.4	525154.3	← 殘差平方			
22							

多變量資料

- 若處理對象係多變量資料，則安排給LINEST()之陣列範圍就要變大一些，因使用者常不清楚要選取多少範圍才算正確，往往多選而導致一堆的#N/A值。
（就算範圍完全正確，也會有一些無作用的#N/A值）
- 年齡與所得資料：【[所得](#)】工作表



多變量資料

- 其迴歸方程式與判定係數（ R^2 ）分別為

$$y = -36.54x^2 + 3463.7x - 42087$$

$$R^2 = 0.884$$

- 若要以LINEST()來求得這些數字，得以下示步驟求得：

- 由於所求對象係單兩欄變量，故LINEST()之結果將為一5x3之陣列，所以於輸入公式之前，先選取五列三欄之儲存格範圍

	A	B	C	D	E	F
16						
17						
18						
19						
20						
21						
22						
23						

多變量資料

2. 輸入下示公式

=LINEST(C2:C15,A2:B15,TRUE,TRUE)

各引數之意義，依序為因變數範圍、自變數範圍、常數項是否不為0與是否求算統計值。

3. 按Ctrl+Shift+Enter完成輸入，獲致陣列內容

4. 依前述對應位置，找出各統計數字

PMT		✖ ✓ ✎		=LINEST(C2:C15,A2:B15,TRUE,TRUE)			
	A	B	C	D	E	F	G
16							
17							
18			UE,TRUE)				
19							
20							
21							
22							
23							

C18		✖ ✎		{=LINEST(C2:C15,A2:B15,TRUE,TRUE)}			
	A	B	C	D	E	F	G
17							
18			3463.745879	-36.539835	-42087		
19			385.7649034	3.9905528	8250.414		
20			0.884032544	5383.55	#N/A		
21			41.92709886	11	#N/A		
22			2430313600	318808722	#N/A		
23							

多變量資料

- 可查得其迴歸方程式與判定係數（ R^2 ）分別為

$$y = -36.539835x^2 + 3463.745879x - 42087$$

$$R^2 = 0.884032544$$

C18		fx {=LINEST(C2:C15,A2:B15,TRUE,TRUE)}					
	A	B	C	D	E	F	G
17			係數 ₂ (b ₂)	係數 ₁ (b ₁)	常數(a)		
18			3463.745879	-36.539835	-42087		
19			385.7649034	3.9905528	8250.414		
20	判定係數(r ²)		0.884032544	5383.55	#N/A		
21			41.92709886	11	#N/A		
22			2430313600	318808722	#N/A		
23							

截距

INTERCEPT(因變數範圍, 自變數範圍)

- 本函數在求利用已知的因變數範圍與自變數範圍，所求出直線迴歸方程式中的截距值。
- 以下示之廣告費與銷售量資料為例，其迴歸方程式為

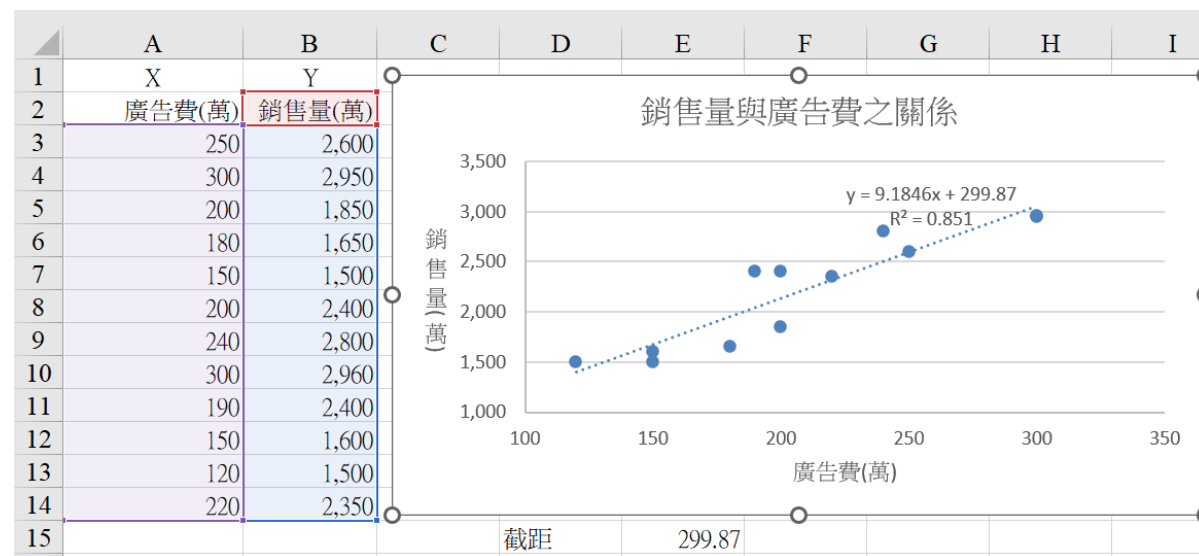
$$y = 9.1846x + 299.87$$

其截距值299.87，亦可以

=INTERCEPT(B3:B14,A3:A14)

來直接求得。

- **【截距】** 工作表



斜率

SLOPE(因變數範圍,自變數範圍)

- 在求利用已知的因變數範圍與自變數範圍，所求出直線迴歸方程式中的迴歸係數值（斜率）
- 以下示之廣告費與銷售量資料為例，其迴歸方程式為

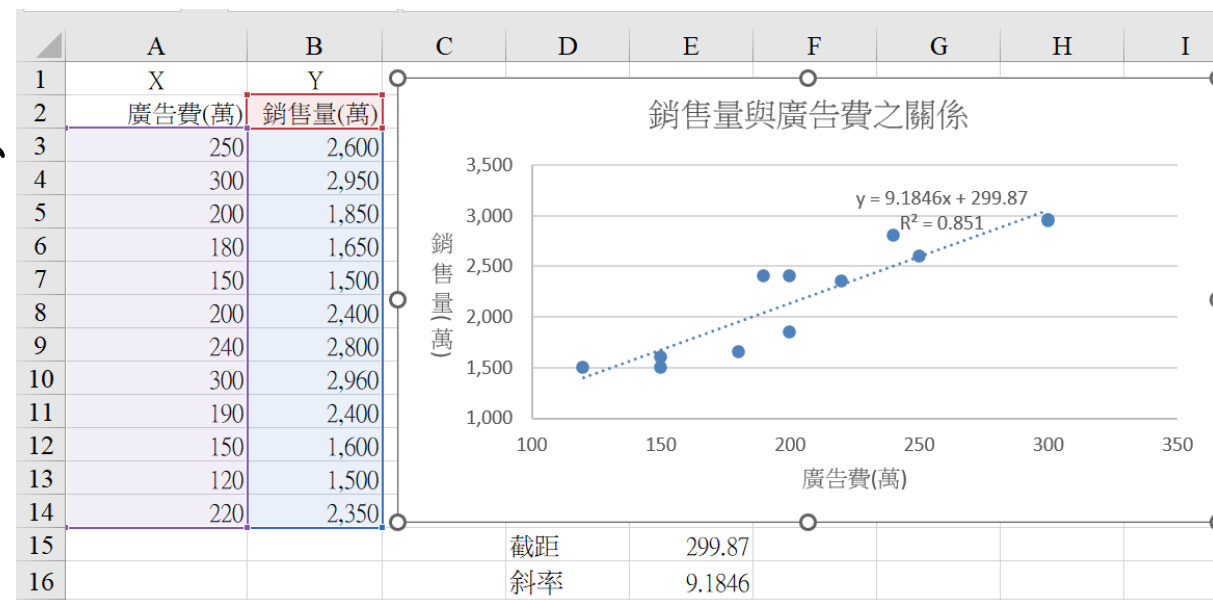
$$y = 9.1846x + 299.87$$

其迴歸係數值9.1846，亦可以

$$=SLOPE(B3:B14,A3:A14)$$

來直接求得。

- 【斜率】工作表

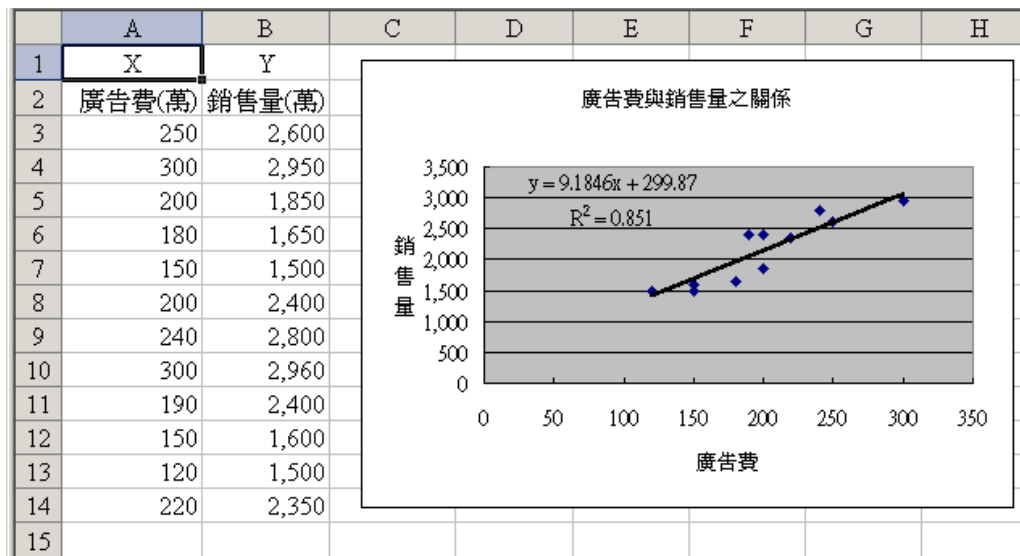


預測

FORECAST(x,因變數範圍,自變數範圍)

- 此函數利用迴歸分析，以已知之因變數範圍與自變數範圍，求算其線性迴歸方程式（但不顯示迴歸方程式內容），並將使用者所指定的一組新的x值，代入迴歸方程式求其y估計值。
- 廣告費與銷售量資料為例，其迴歸方程式為

$$y = 9.1846x + 299.87$$



預測

- 以FORECAST()直接求其預測值，可以於C3輸入下示公式：

$$=FORECAST(A3, \$B\$3:\$B\$14, \$A\$3:\$A\$14)$$
- 續將其抄給C4:C14，即可求得。
- 為驗證其結果是否等於以
 - $y = 9.1846x + 299.87$
- 之迴歸方程式所求？
 - $D3 = 9.1846 * A3 + 299.87$
 - 將A3之X值代入迴歸方程式
 - 續將其抄給D3:D14
- 可發現兩者之所求結果完全相同
- 【預測】工作表

	C3		=FORECAST(A3,\$B\$3:\$B\$14,\$A\$3:\$A\$14)				
	A	B	C	D	E	F	G
1	X	Y	以FORCATE求	以迴歸方程式			
2	廣告費(萬)	銷售量(萬)	預測銷售量(萬)	預測銷售量(萬)			
3	250	2,600	2596.0	2596.0	← =9.1846*A3 + 299.87		
4	300	2,950	3055.3	3055.3			
5	200	1,850	2136.8	2136.8			
6	180	1,650	1953.1	1953.1			
7	150	1,500	1677.6	1677.6			
8	200	2,400	2136.8	2136.8			
9	240	2,800	2504.2	2504.2			
10	300	2,960	3055.3	3055.3			
11	190	2,400	2044.9	2044.9			
12	150	1,600	1677.6	1677.6			
13	120	1,500	1402.0	1402.0			
14	220	2,350	2320.5	2320.5			
15							

線性趨勢

TREND(因變數範圍, 自變數範圍, 新x範圍, 是否要常數)

- 利用迴歸分析最小平方法，以已知之因變數範圍與自變數範圍，求算其線性迴歸方程式（但不顯示迴歸方程式內容），並將使用者所指定的一組新的x值，代入迴歸方程式求其y估計值。
- 與**FORECAST()**作用相同，只不過其輸出結果係一陣列，故於選妥範圍，輸入公式後，按**Ctrl+Shift+Enter**完成輸入。

線性趨勢

- 因變數範圍、自變數範圍與新x範圍之儲存格個數應一致。新x範圍若省略將取用自變數範圍。
- 是否要常數為一邏輯值
 - 為TRUE或省略，表要求計算常數b（即截距），其迴歸結果為 $y = b + mx$
 - 為FALSE，則將常數b設定為0，迴歸結果將為 $y = mx$
- 以廣告費與銷售量資料為例，其迴歸方程式為 $y = 9.1846x + 299.87$
- 以TREND()直接求其預測值，可以下示步驟求得：
- 【TREND】工作表

線性趨勢

1. 結果將為一陣列，所以於輸入公式之前先選取C3:C14之儲存格範圍
2. 輸入**=TREND(B3:B14,A3:A14)**，各引數之意義，依序為因變數範圍與自變數範圍
3. 按Ctrl+Shift+Enter完成輸入，可獲致預測結果陣列內容
 - D3:D14係用以驗證其結果是否等於以
$$y = 9.1846x + 299.87$$
之迴歸方程式所求？可發現兩者之所求結果完全相同。

線性趨勢

PMT ▾ ✖ ✔ ✎ =TREND(B3:B14,A3:A14)							
	A	B	C	D	E	F	G
1	X	Y	以TREND求	以迴歸方程式			
2	廣告費(萬)	銷售量(萬)	預測銷售量(萬)	預測銷售量(萬)			
3	250	2,600	B14,A3:A14)	2596.0	← =9.1846*A3 + 299.87		
4	300	2,950		3055.3			
5	200	1,850		2136.8			
6	180	1,650		1953.1			
7	150	1,500		1677.6			
8	200	2,400		2136.8			
9	240	2,800		2504.2			
10	300	2,960		3055.3			
11	190	2,400		2044.9			
12	150	1,600		1677.6			
13	120	1,500		1402.0			
14	220	2,350		2320.5			
15							

C3 ▾ ✎ {=TREND(B3:B14,A3:A14)}							
	A	B	C	D	E	F	G
1	X	Y	以TREND求	以迴歸方程式			
2	廣告費(萬)	銷售量(萬)	預測銷售量(萬)	預測銷售量(萬)			
3	250	2,600	2596.0	2596.0	← =9.1846*A3 + 299.87		
4	300	2,950	3055.3	3055.3			
5	200	1,850	2136.8	2136.8			
6	180	1,650	1953.1	1953.1			
7	150	1,500	1677.6	1677.6			
8	200	2,400	2136.8	2136.8			
9	240	2,800	2504.2	2504.2			
10	300	2,960	3055.3	3055.3			
11	190	2,400	2044.9	2044.9			
12	150	1,600	1677.6	1677.6			
13	120	1,500	1402.0	1402.0			
14	220	2,350	2320.5	2320.5			
15							

指數迴歸

LOGEST(因變數範圍, 自變數範圍, 是否要常數, 統計值)

- 此函數利用迴歸分析，計算以已知之因變數範圍與自變數範圍，所求算之指數曲線

$$y = b * m^x$$

- 並傳回描述該曲線的數值陣列
- 於輸入公式後，必須以Ctrl+Shift+Enter完成輸入，才可傳回完整之陣列公式

指數迴歸

- 因變數範圍與自變數範圍之儲存格個數應一致。
- 是否要常數為一邏輯值
 - 為TRUE或省略，表要求計算常數**b**，其迴歸結果為 $y = b * m^x$ 。
 - 為FALSE，則將常數**b**設定為1，迴歸結果將為 $y = m^x$ 。
- 統計值也是一邏輯值，用以設定是否要傳回額外的迴歸直線統計值
 - 為TRUE時，將依下表之對應位置，傳回所有統計值
 - 為FALSE或省略，LOGEST()將只傳回**m**係數和常數項**b**

	A	B	C	D	E	F
1	基底(m_n)	基底(m_{n-1})	...	基底(m_1)	常數(b)	
2	標準誤(m_n)	標準誤(m_{n-1})	...	標準誤(m_1)	標準誤(b)	
3	判定係數(r^2)	對y 估計值的標準誤差				
4	F 統計值	F檢定之自由度				
5	迴歸平方	殘差平方				
6						

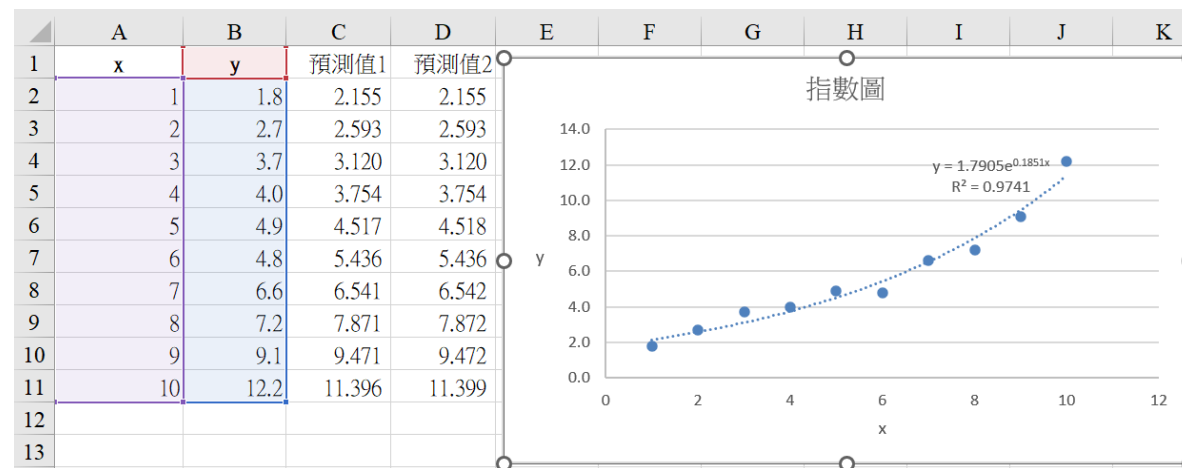
指數迴歸

- 以xy資料為例：
- 其迴歸方程式與判定係數（ R^2 ）分別為

$$y = 1.79049 \times 1.203317^x$$
$$R^2 = 0.9647$$

- 若要以LOGEST()來求得這些數字，得以下示步驟求得：

1. 由於所求對象係單一變量，故LOGEST()之結果將為一5x2之陣列，所以於輸入公式之前先選取五列二欄之儲存格範圍



	A	B	C	D
14				
15				
16				
17				
18				
19				
20				

指數迴歸

2. 輸入下示公式

=LOGEST(B2:B11,A2:A11,TRUE,TRUE)

各引數之意義，依序為因變數範圍、自變數範圍、是否要常數與是否求算統計值。

3. 按Ctrl+Shift+Enter完成輸入

4. 依前述對應位置，找出各統計數字

5. 其迴歸方程式與判定係數（ R^2 ）分別為

$$y = 1.79049 \times 1.203317^x$$

$$R^2 = 0.9647$$

14		基底(m_1)	常數(b)
15		1.2033165	1.7904902
16		0.0125189	0.0776777
17	判定係數(r^2)	0.9646911	0.1137085
18		218.57197	8
19		2.8260512	0.1034369

B15		✕ ✓ f_x		{=LOGEST(B2:B11,A2:A11,TRUE,TRUE)}			
	A	B	C	D	E	F	G
13							
14		基底(m_1)	常數(b)				
15		1.2033165	1.7904902				
16		0.0125189	0.0776777				
17	判定係數(r^2)	0.9646911	0.1137085				
18		218.57197	8				
19		2.8260512	0.1034369				

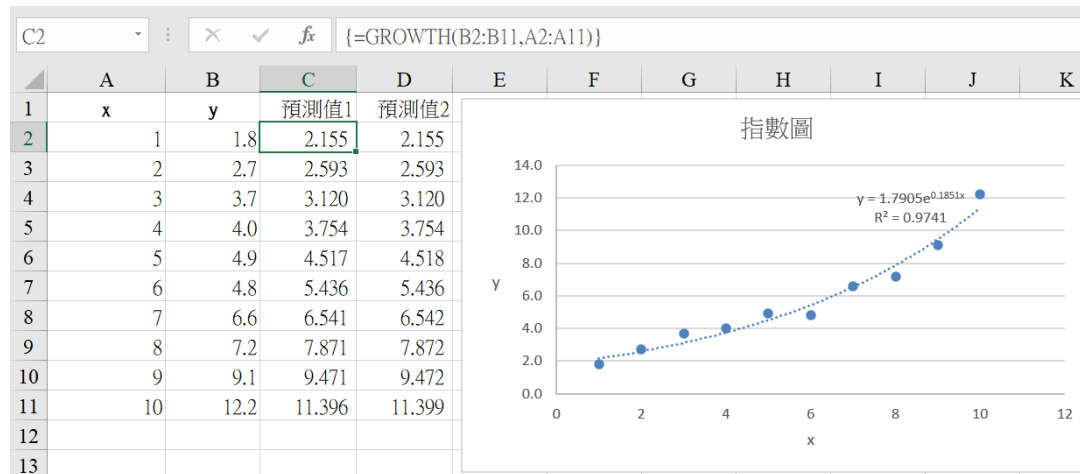
指數曲線趨勢

GROWTH(因變數範圍, 自變數範圍, 新x範圍, 是否要常數)

- 此函數利用迴歸分析，以已知之因變數範圍與自變數範圍，求算其指數曲線（但不顯示迴歸方程式內容），並將使用者所指定的一組新的x值，代入迴歸方程式求其y估計值。
- 由於其結果為一陣列，選妥範圍，輸入公式後，記得按 **Ctrl+Shift+Enter** 完成輸入。

指數曲線趨勢

- 因變數範圍、自變數範圍與新x範圍之儲存格個數應一致。新x範圍若省略將取用自變數範圍。
- 是否要常數為一邏輯值
 - 為TRUE或省略，表要求計算常數b，其迴歸結果為 $y = b * m^x$
 - 為FALSE，則將常數b設定為1，迴歸結果將為 $y = m^x$
- 茲仍以前例求指數迴歸之xy資料為例，其迴歸方程式與判定係數（ R^2 ）分別為
 - $y = 1.79049 \times 1.203317^x$
 - $R^2 = 0.9647$
- **【GROWTH】** 工作表



指數曲線趨勢

- C2:C11之預測值，可以下示步驟求得：

1. 由於所求結果將為一陣列，所以於輸入公式之前先選取C2:C11之儲存格範圍
2. 輸入公式
=GROWTH(B2:B11,A2:A11)
各引數之意義，依序為因變數範圍與自變數範圍。
3. 按Ctrl+Shift+Enter完成輸入，獲致預測結果之陣列內容

C2		fx {=GROWTH(B2:B11,A2:A11)}				
	A	B	C	D	E	F
1	x	y	預測值			
2	1	1.8	2.154526			
3	2	2.7	2.592577			
4	3	3.7	3.119691			
5	4	4.0	3.753976			
6	5	4.9	4.517221			
7	6	4.8	5.435647			
8	7	6.6	6.540804			
9	8	7.2	7.870657			
10	9	9.1	9.470892			
11	10	12.2	11.39648			
12						

PMT		fx {=GROWTH(B2:B11,A2:A11)}				
	A	B	C	D	E	F
1	x	y	預測值			
2	1	1.8	A2:A11)			
3	2	2.7				
4	3	3.7				
5	4	4.0				
6	5	4.9				
7	6	4.8				
8	7	6.6				
9	8	7.2				
10	9	9.1				
11	10	12.2				
12						