

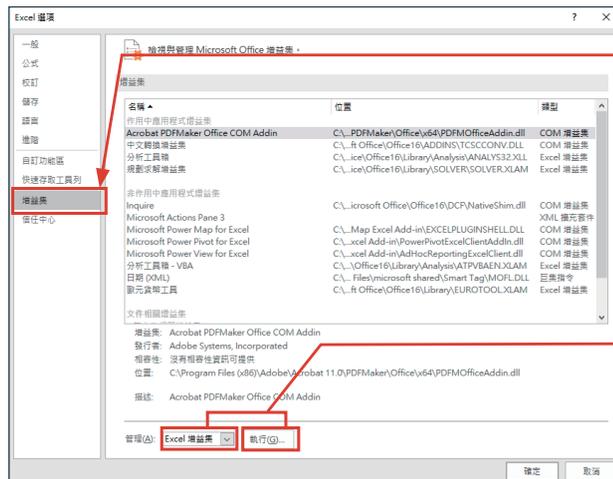
閱讀本書的準備

本書需要使用「分析工具箱」與「規劃求解增益集」。預設是無法使用的，所以請透過下列的操作新增「分析工具箱」與「規劃求解增益集」。只要完成設定，「分析工具箱」與「規劃求解增益集」就會常駐於工具列，不需要每次啟動 Excel 都重新設定一次。

請先啟動 Excel。Excel 2013/2016 需先點選「空白活頁簿」之後開始操作。

Excel 2007

▶ 步驟 1 先點選「Office 按鈕」，再從選單點選「Excel 的選項」。



3 點選「▼」再選擇「Excel 增益集」，然後點選「執行」。

02

計算以超群平衡感自誇的平均值

即使是討厭計算的人，平常也會計算平均值吧！比方說，優惠包裝與一般包裝的每克價格計算、孩子們的五科成績平均分數、各付各的時候，每個人要付多少的計算，總之，計算平均值的機會真的很多。在職場也是一樣，需要算出大略的業績規模時，平均值是個很好用的數值。這節就帶大家重新思考一下熟悉的平均值。

導入

例題

「無法掌握業績規模」

年營業額2億元的A公司希望在東京銷售關西限定商品K，但是銷售時，想將商品改成商品T這個名稱。原本不是負責這項商品的X先生突然接到一個小時後要在會議報告東京的業績預測值的指示。只拿到紙本業績記錄與文件的X先生正在一籌莫展時，前輩Y先生與後進的Z先生幫忙，整理出以下的資料。

業績記錄包含銷售時間、銷售人員與售價這些內容，每件銷售數量都是一個。因此，隨機擷取出200筆商品K的售價，並且整理成適當的資料格式。此外，確認文件內容後，發現文件記載了候補店面的目標人口與預測的市佔率。

● 商品 K 的售價摘要

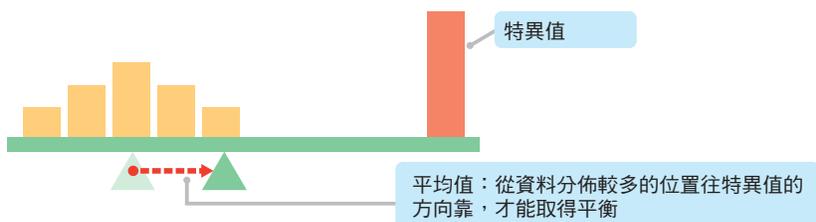
	A	B	C	D	E
1	從業績記錄隨機擷取的銷售價格				
2	400	500	410	410	370
3	500	380	400	410	400
4	500	370	410	370	370
5	500	370	410	400	620
6	500	480	370	450	440
38	440	440	400	440	480
39	410	370	400	500	480
40	480	440	500	550	550
41	370	480	480	400	480
42					

● 開店候選資料

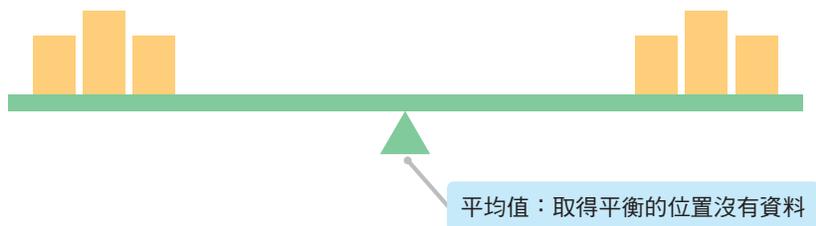
	A	B
1	開店候選資料	
2	目標人口合計(人)	150,000
3	市佔率	5%
4	年平均預測購買個數/人	5.5
5		

尤其當資料的分佈呈現兩極化時，這問題就越加嚴重，因為為了取得平衡，平均值很可能落在沒有資料的區間裡，如此一來，平均值就不太能代表這些資料。

● 平均值被特異值影響



● 在兩極化的資料，平均值的周圍沒有資料



現在已是多元化的時代，計算平均值的時候，也常常會算出上述這種失準的平均值。不過，不管以前還是現在，平均值一定位於資料的平衡之處。之所以會讓人覺得失準，是因為我們總以為「平均值位於多數資料的中心點」。

今後如果算出失準的平均值，不妨先懷疑資料的分佈是否呈兩極化，又或者是不是挾雜的特異值。

► 特異值的處理

特異值有兩種，一種是不小心輸入錯誤，一種是因為某些原因而變得特異的值。例題的開店活動就屬於因為某些原因而變得特異的情況。輸入錯誤的值可排除在平均值的計算對象之外，但是因為某些原因而造成偏差的話，就必須視情況決定。要注意的是，就算是偏差的值，也不一定就能無條件排除。

● 特異值與偏差值

所謂的偏差值指的是極端遠離資料群的數值，算是特異值的一種。不過，特異的資料不一定就等於偏差值。偏差值的標準是位於 99.7% 資料範圍外側的值。這部分會在第 5 章解說，但所謂的「99.7% 資料範圍」就是「平均值 $\pm 3 \times$ 標準差」的意思。不過這終究只是參考，要不要排除偏差值，還是得依照要計算的值（目的）決定。

Column 被當成壞人看待的平均值

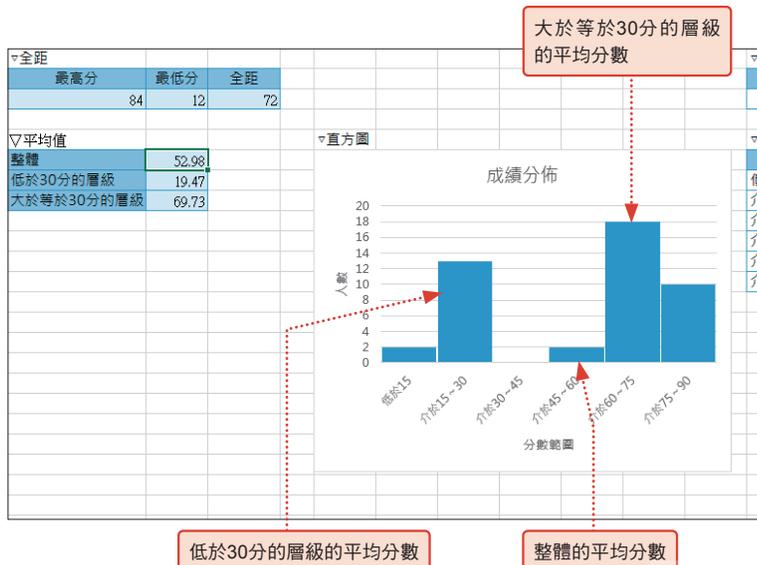
最近平均值似乎很討人厭，很多人會說，平均值沒什麼用，算出平均值也無法得到任何資訊，但是，一如前述，平均值本身沒有不好。話雖如此，平均值旁邊沒有資料的現象越來越多，會覺得平均值越來越無法說明資料也是情有可原。問題在於，現在還是習慣只憑平均值解釋資料的全貌。下一節會提到，要解釋資料的全貌，需要中位數、眾數，最好還有標準差。當然，也少不了平均值。觀察多個代表值才能想像資料的特徵與分佈情況。偶爾會有人提到平均值、中位數與眾數，但現在平均值出現的頻率還是最高。等到大家覺得其他的代表值一起公佈才正常，平均值才有可能不再被當成壞人。

► 兩極化資料的平均值

面對兩極化的資料時，將資料分成兩群看待，就能避免平均值旁邊沒有資料的情況。資料之所以會變得兩極化，有可能是因為收集資料時，混入太多性質不同的資料。只要了解性質的差異，就能依照性質分類資料，再計算平均值，也通常能算出具代表性的平均值。此外，以性質分類資料的過程稱為分層。分層的常見範例有性別、年紀或地區。

下列的圖是 2-01 節的 B 組成績資料。B 組的平均分數附近幾乎沒有人，但是，若將資料分成低於 30 分以及大於等於 30 分的兩層，然後再計算平均分數，就會發現，平均值落在次數較高的組別裡。

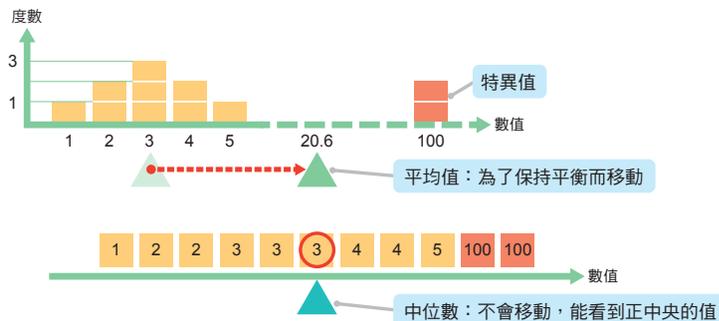
● 各層的平均值



► 不為特異值所感的中位數

中位數是先由小至大，根據數值排列資料，剛好落在正中央位置的資料。假設資料是偶數，就取正中央兩個的平均值。為了取得整體資料的平衡，平均值會移動，但中位數不會移動，一開始就待在資料的正中央，只能看到位於正中央的值。再者，由小至大排列數值之後，特異值會被趕到邊緣，所以絕對無法成為中位數。所以，就算資料摻雜了特異值，中位數也不會受到影響。

● 平均值與中位數



「不會被特異質干擾的中位數才是最適合代表資料的值。為什麼還要一味地計算平均值呢？」有些讀者會有這樣的疑問，但計算平均值的理由之一就是方便性。平均值只要加總所有數值，再除以資料筆數就能算出，所以若是對數字敏感的人，心算也能算出平均值。中位數的原理雖然只是「先排序，再觀察正中央的數值」，但是「排序」這個步驟無法用心算或電子計算機完成。

請試著以中位數代替餐費平付時使用的平均值。酒喝得不多、飯吃得很少的人或許可以接受這種算法，但結帳時錢一定會不夠，之後有可能會為了誰該付不足的部分而吵架。

平均值是「合計金額 = 平均金額 × 人數 = 個人餐費的總和」，所以就算有些人對平付的金額不滿意，但一定能結帳。

● 以中位數計算分攤的餐費

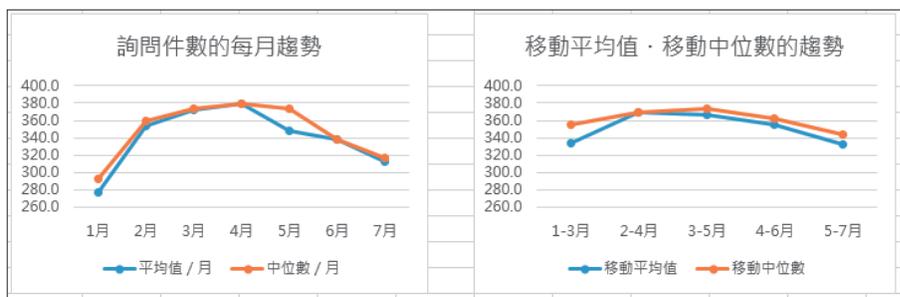
	A	B	C	D	E
1	餐費資料				
2	參加者	餐費明細		總計金額	32,000
3	D先生	2,200		參加人數	8
4	F先生	2,500		平均值	4,000
5	B先生	2,900		中位數	3,300
6	G先生	3,100			
7	A先生	3,500			
8	C先生	3,800			
9	E先生	6,800			
10	H先生	7,200			
11					

以3,300除的話，會不足5,600元。

排序餐費的明細，中間的兩個值的平均就是中位數。

▶觀察大小的長條圖將原點設定為「0」，必須輸入省略的波浪線，但是用來觀察變化（斜率）的折線圖，就不一定得將原點設定為「0」。

●詢問件數的趨勢

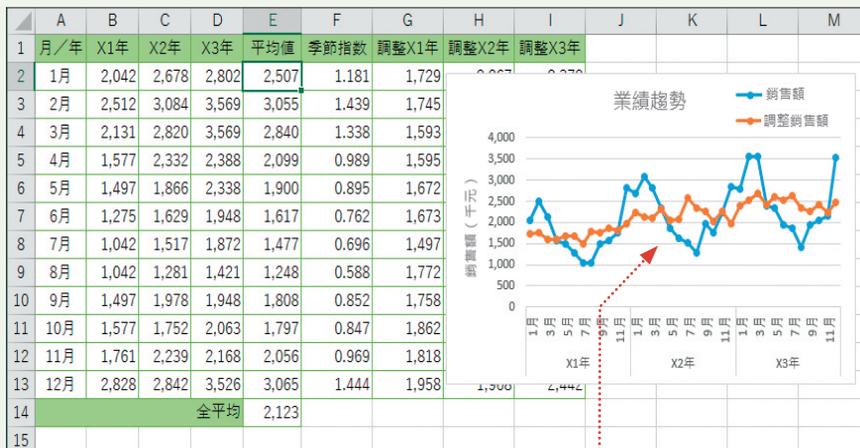


此外，實施方案前後的詢問件數平均值是否真的有落差，換言之，方案是否真有效果，必須透過 t 檢定驗證。有關檢定的部分將於第 7 章解說。

Column 以平均值與中位數了解業績動向

季節性明顯、難以掌握業績動向時，可利用平均值或中位數消除浮動。下圖為三年內的每月業績以及消除浮動之後的業績。其中計算了每個月平均業績佔全體平均值的比例（季節指數），再依照比例調整銷售額，藉此消除浮動。也可利用中位數取代每月的業績平均。

●會隨季節變動的業績走向



使用平均值消除浮動，會比較容易掌握業績動向。根據這張圖可推測業績是上揚的趨勢。

▶使用中位數時，至少需要三年份的資料。因外在因素而摻雜了較往年明顯不同的業績時，可使用這種算法。

05

了解資料的多數派

乍看很凌亂的資料，有可能會含有許多相同的資料，而最常出現的資料稱為眾數。眾數與平均值、中位數都是代表資料的數值之一，這節要一邊與平均值、中位數比較，一邊解說眾數。

導入 ▶▶▶

例題 「想了解最常出現的詢問件數」

為了參考必要的技術，求助部門的人員配置方式而整理了詢問件數。每天的詢問件數雖然不同，但還是想知道每天大概的件數，該怎麼做才好呢？

● 詢問件數

	A	B	C	D	E	F	G	H	I
1	▼詢問件數								
2	日期	1月	2月	3月	4月	5月	6月	7月	
3	1	55	322	354	475	155	336	368	
4	2	64	373	382	465	198	318	301	
5	3	88	397	350	462	188	305	352	
6	4	256	337	379	461	156	305	317	
7	5	324	379	399	422	188	345	318	
8	6	258	359	357	479	399	349	356	
9	7	330	360	398	420	351	322	308	
31	29	255		359	155	400	308	282	
32	30	260		381	202	356	341	263	
33	31	265		392		374		255	
34									

想知道每月的詢問件數之中，佔多數的詢問件數與日數。

▶ 搜尋頻繁出現的值

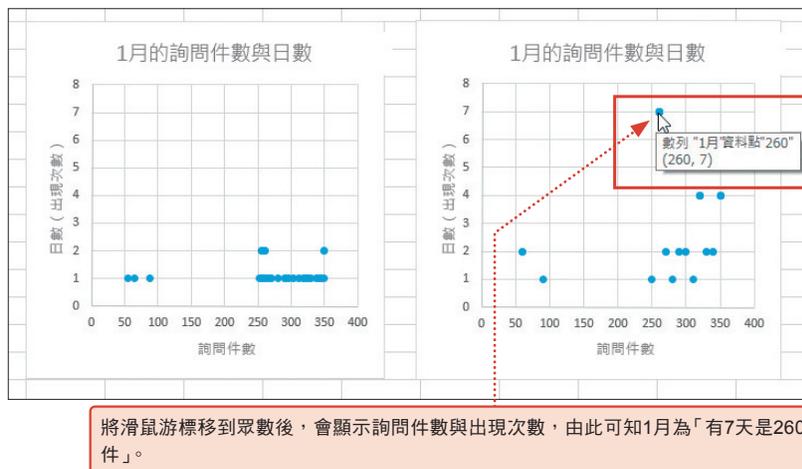
眾數就是資料之中，最頻繁出現的值，也就是資料裡的多數派。繪製直方圖或是將資料與出現次數繪製成圖表時，眾數會出現在次數最多或是出現次數最多的位置，也可說是資料最集中的部分。少數派的資料會被排除在眾數的候選之外，所以眾數不會受到特異值或偏差值的影響，而且也因為是多數派，所以眾數比平均值或中位數更能代表資料。

► 眾數的問題

眾數的問題在於資料的數值若是太瑣碎，就很難找到相同的值。解決方案之一就是對資料進行四捨五入的計算，讓資料限縮在方便操作的範圍之內。

下列是一月的詢問件數與日數（出現次數）的散佈圖與將詢問件數四捨五入至十位數的散佈圖。直接使用資料時，出現相同詢問件數的日子只有兩天，但是統整資料之後，日數就增加了。

● 以原始資料計算的眾數（左）與資料經過四捨五入之後的眾數（右）



► 資料的兩極化指的是以同一目的收集的資料分成兩種資料層的狀態。

► 兩極化資料的眾數

眾數雖然是最常出現的值，但不一定只有一個。如果有其他出現次數相同的資料，這些資料就全部都是眾數。當資料呈現兩極化的分佈，兩邊有出現次數相同的資料時，就會出現兩個以上的眾數。一如平均值與中位數的處理，在計算眾數時，如果資料也呈現兩極化的分佈，最好先替資料分層，才會比較方便判讀。

實踐

範例

2-05「操作2」工作表

► Excel 的操作①：計算每月的眾數

這次要計算資料統整前與統整後的眾數。這次利用 ROUND 函數將資料四捨五入至十位數，眾數則是利用 MODE.MULT 函數計算。

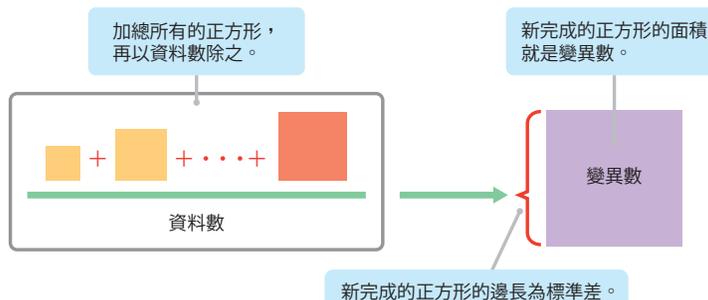
從上圖來看，變異數就是加總所有正方形面積，再除以資料筆數的值。雖然我們不知道邊長，但是變異數也會是正方形。由於線條變成面積，所以變異數會比偏差值大也不意外。此外，變成面積後，變異數的單位也變成原本的單位的平方。要還原成原本的單位必須算出變異數的正方形的邊長，也就是開根號或計算平方根。變異數的正平方根就是所謂的標準差。標準差是正方形的邊長，單位也是原始資料的單位。

- 變異數的單位是資料的平方，所以算出正的平方根，就能還原次數資料的單位。
- 變異數的正平方根就是標準差。標準差的單位就是資料的單位。

之所以會算出變異數與標準差，主要是為了避免偏差值的合計為 0，而刻意乘以平方，然後為了還原次數原本的單位，才計算平方根。值得繞遠路的變異數與標準差不會忽略任何一筆資料，是代表所有資料分佈程度的代表值。

此外，變異數就是偏差平方和的「平均值」，所以開根號之後的標準差就是資料整體分佈的平均值。

● 變異數與標準差



實踐

範例

2-07 「操作 1」工作表



► Excel 的操作①：計算偏差值與偏差值的合計

計算 A 店與 B 店的偏差值與偏差值的合計，確認偏差值的合計會是 0。

計算偏差值

- 在儲存格「K1」、「I3」輸入的公式

K1	=AVERAGE(A3:C32)	I3	=A3-\$K\$1
O1	=AVERAGE(E3:G32)	M3	=E3-\$O\$1

01

兩種資料之間的意外關係

「尿布的業績提升，啤酒的業績也會跟著提升」是找出看似毫無關聯的資料的關聯性，讓業績因此提升的經典案例。這節將為大家解說兩種資料的關聯性與關聯性的強度。

導入

例題 「想了解橘子跟暖桌的關係、售價與銷售數量的關係」

調查橘子與暖桌的出貨量之後，整理出下列的資料。下列的資料之間存在著何種關聯性呢？

● 橘子與暖桌的年度資料

	A	B	C	D	E
1	年度	橘子的10a的平均出貨量(t)	年度	暖桌	
2	民國99年產	1,005,000	民國99年度		713,000
3	民國100年產	743,200	民國100年度		511,596
4	民國101年產	950,500	民國101年度		268,498
5	民國102年產	807,800	民國102年度		322,292
6	民國103年產	893,400	民國103年度		296,456
7	民國104年產	700,100	民國104年度		283,087
8	民國105年產	828,600	民國105年度		241,870
9	民國106年產	757,300	民國106年度		195,931
10	民國107年產	804,400	民國107年度		174,622
11	民國108年產	782,000	民國108年度		201,800
12					
13	出處〈左〉：從農林水產省農作調查(果樹)節錄「橘子」的資料再加工				
14	出處〈右〉：經濟產業省生產動態統計年報 纖維、生活用品統計篇				
15	節錄「暖桌」資料再加工				

接著是將商品 A 與商品 B 的二十天份售價、銷售數量的資料整理成表格。此外，商品 A 是老字號 A 社的商品，商品 B 是與商品 A 同類的自選商品。商品 A 與商品 B 的價格與數量有何關係存在呢？

02

利用手邊的資料計算預測值

兩種資料之間呈現線性的相關關係時，可繪製一條逼近資料的線條，根據繪製線條的公式計算沒有資料的位置。這節挑選了兩種透過企業活動收集，能說明線性相關關係的資料，介紹預測業績的方法。

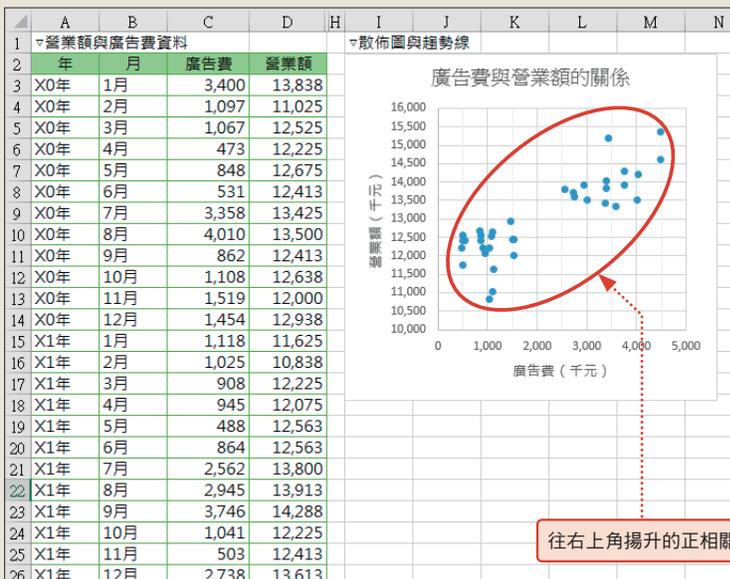
導入 ▶▶▶

例題 「想根據廣告費預測營業額」

下圖是過去三年內的每月廣告費與營業額資料。根據這些資料繪製散佈圖之後，發現廣告費的多寡與營業額呈正相關。

該如何根據廣告費預測營業額呢？

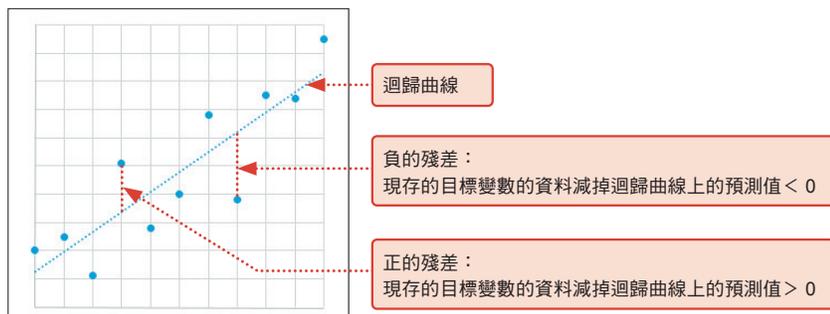
● 每月廣告費與營業額資料



往右上角揚升的正相關

- ① 迴歸曲線必須通過平均值。
 以例題而言，線性的迴歸曲線會通過營業額平均值與廣告費平均值。
- ② 繪製迴歸曲線時，必須讓散佈圖裡的每項資料與迴歸曲線之間的距離縮至最小。
 資料與迴歸曲線之間的距離稱為殘差。殘差就是直軸資料，也就是目標變數的資料與迴歸曲線上的預測值「y」之間的差距。

● 迴歸曲線與殘差



▶ 判斷迴歸曲線能不能用於預測

即使遵守上述規則，繪製了殘差為最小值的迴歸曲線，迴歸曲線的預測精確度仍會受到相關性的強弱而影響。直覺上，相關係數的絕對值（無符號的大小）越接近 1，精確度有可能越高。這個直覺基本上是正確的，在迴歸分析裡，會利用相關係數乘以平方的決定係數判斷。相關係數的範圍介於 -1 ~ 1，而乘以平方代表不再會有負數，所以決定係數的範圍會在 0 ~ 1 之間。越接近 1 代表預測的精確度越高，但只要超過 0.5，就代表迴歸公式可用於預測。決定係數「0.5」代表迴歸公式可說明 50% 的資料。此外，決定係數是相關係數的平方，所以又寫成「R 平方值」或「 R^2 」。

▶ 嚴格來說，繪製曲線時，必須讓殘差的平方和縮至最小。

實踐 ▶▶▶

範例

3-02



▶ Excel 的操作①：在散佈圖裡繪製趨勢線

Excel 可在代表兩種資料相關性的圖表裡繪製各種趨勢線、代表趨勢線的公式以及追加決定係數。這次使用的是迴歸分析，所以趨勢線也選用能說明直線關係的「線性」。此外，可確認迴歸曲線通過資料的平均值。

繪製散佈圖

請參考 P.92 的步驟，繪製廣告費為橫軸、營業額為直軸的散佈圖。此外，請在儲存格「G2」輸入 CORREL 函數，算出相關係數。