

用Excel來做迴歸分析 STEP BY STEP

花蓮女中◆歐漢文 老師

高三的地理課程中，計量地理部分有講授迴歸分析及趨勢線的計算，在地理課程中算是比較困難的部分，在課堂上用一堆數字套上統計的公式，而社會組的學生通常都對數學有恐懼感，所以教起來事倍功半，通常在講授前會請學生帶計算機，再來課堂中的一堆數字，在演算練習過程中有平方相加，再開根號，到答案出來後，學生可能已不清楚數字所代表的地理意義了。如果能帶學生到電腦教室，利用Excel來計算，又或者利用手提電腦接單槍投影機，在教室直接顯示出來，可以縮短演算的時間，明白本單元的目標並非著重繁瑣的計算過程，重要的是數字所表示的地理意義。

地理學計量分析中的科學化推論，建立因果關係模式的三個重要主題是：散佈圖、相關分析及迴歸分析。我們就用下列例子來說明：

在93年指定科目考試中有一題是測驗考生相關係數的觀念，題目是：變數甲和變數丙間的相關係數(r)值為何？答案有四個選項，分別是：(A) -0.74 (B) -0.12 (C) 0.12 (D) 0.74。題目附有一個公式：

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{nS_x S_y}$$

雖然附有一個公式，但考試時大概不太可能代入公式來計算，不要說是否有帶或能否

帶計算機，如果用手來算，光是把一堆數字，計算標準差已經花太多時間了，而且又要開根號，又要平方，所以其實是考概念，變數甲由200到950是變大，變數丙由兩位數到個位數，明顯地可以看出變數丙隨著變數甲的增加而減少，兩者呈現負相關。(B)選項和(C)選項中的相關程度太低，(D)選項是高度正相關，所以正確答案是(A)選項，因此如果學生有正確的相關係數觀念，其實是很容易就知道要選(A)的。

該題目的三組變數如下：

甲	乙	丙
200	0.9	11
280	1.1	12
290	1.9	10
310	1.9	11
340	2.6	3
560	2.7	11
570	3.2	5
590	3.8	4
630	2.8	3
690	2.4	3
820	9.5	4
840	9	1
890	4.4	5
950	11.2	2

我們先做甲變數和乙變數的迴歸分析。

	A	B	C	D
1	甲	乙	丙	
2	200	0.9	11	
3	280	1.1	12	
4	290	1.9	10	
5	310	1.9	11	
6	340	2.6	3	
7	560	2.7	11	
8	570	3.2	5	
9	590	3.8	4	
10	630	2.8	3	
11	690	2.4	3	
12	820	9.5	4	
13	840	9	1	
14	890	4.4	5	
15	950	11.2	2	
16				

步驟 1：首先把指考題目中的三項變數放到試算表中。

	A	B	C	D	E	F
1	甲	乙	丙			
2	200	0.9	11			
3	280	1.1	12			
4	290	1.9	10			
5	310	1.9	11			
6	340	2.6	3			
7	560	2.7	11			
8	570	3.2	5			
9	590	3.8	4			
10	630	2.8	3			
11	690	2.4	3			
12	820	9.5	4			
13	840	9	1			
14	890	4.4	5			
15	950	11.2	2			
16						

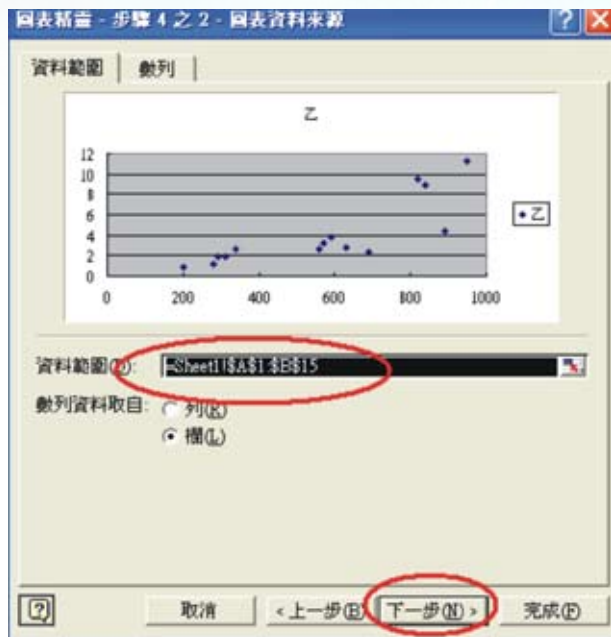
步驟 3：選取圖表的符號。

	A	B	C	D
1	甲	乙	丙	
2	200	0.9	11	
3	280	1.1	12	
4	290	1.9	10	
5	310	1.9	11	
6	340	2.6	3	
7	560	2.7	11	
8	570	3.2	5	
9	590	3.8	4	
10	630	2.8	3	
11	690	2.4	3	
12	820	9.5	4	
13	840	9	1	
14	890	4.4	5	
15	950	11.2	2	
16				

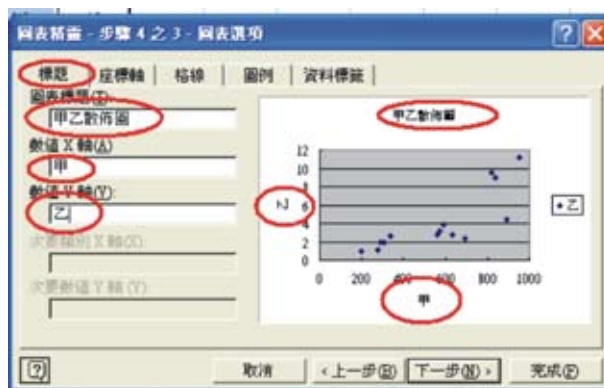
步驟 2：把甲變數及乙變數反白。由 A1 到 B15。



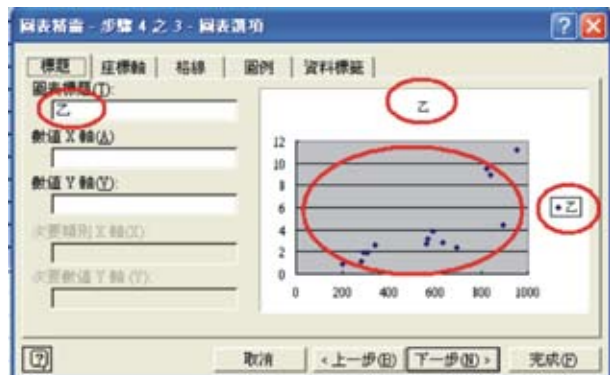
步驟 4：出現圖表精靈的步驟 4 之 1（圖表類型），要選 XY 散佈圖，副圖表類型選第一種即可，因為我們不必用到其他的副圖表。然後下一步。



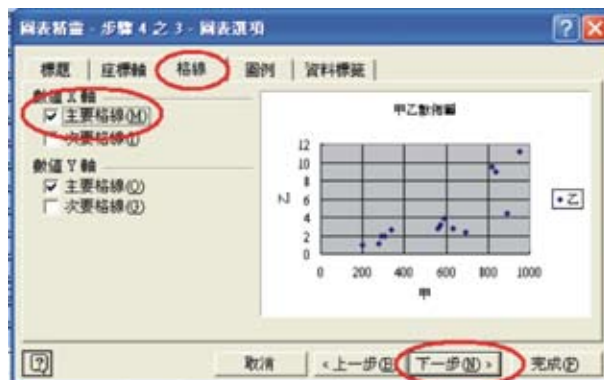
步驟 5：出現圖表精靈的步驟 4 之 2（圖表資料來源），資料範圍已自動設為 A1：B15，然後下一步。如果你直接選完成，簡單的散佈圖會直接跑出來，跟上面的長得一樣，但是我們還要調整其他的說明，所以一定要選下一步。



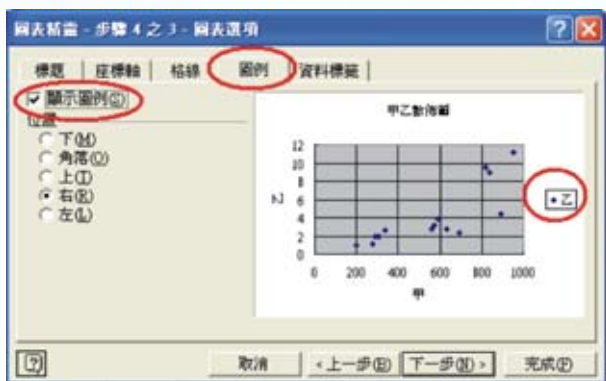
步驟 7：選取子選項：標題\圖表標題，鍵入甲乙散佈圖；數值 X 軸鍵入甲，數值 Y 軸鍵入乙。



步驟 6：出現圖表精靈的步驟 4 之 3（圖表選項），因為散佈圖上的圖表標題及旁邊的圖例都不是我們想要的，散佈圖的底色不要灰色，因為如果印成講義，灰色會不清楚，另外 X 軸及 Y 軸也要註明項目。



步驟 8：點選另一個子選項：格線。原來預設是只有 Y 軸有主要格線，要再勾選數值 X 軸的主要格線，這時散佈圖出現了數值 X 軸的格線。然後再下一步。



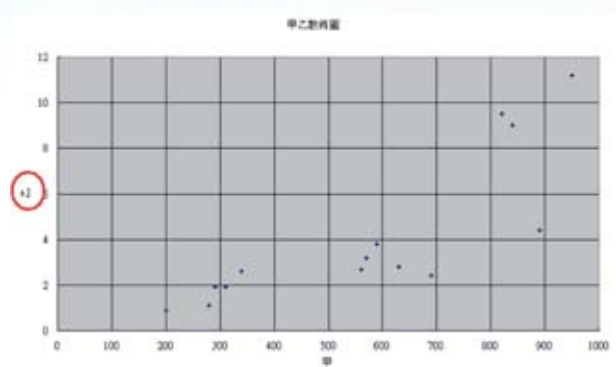
步驟 9：點選另一個子選項：圖例。把顯示圖例的選項去除。散佈圖右邊的一個圖例的方框會去掉。



步驟 10：圖例的方框不見了。再下一步。



步驟 11：出現圖表精靈的步驟 4 之 4（圖表位置），要點選新的工作表，如果選工作中的物件，會變成一個小圖，出現在工作表上，我們要示範給學生看，所以要選新的工作表，才會呈現另外一張的大圖。



步驟 12：這是甲乙的散佈圖，我們要把 Y 軸的文字右轉 90 度。



步驟 13：把指標移到乙字的上面，點一下左鍵，使乙字變成可修改的模式。



步驟 14：接下來點一下右鍵，會出現選單，再用左鍵點選座標軸標題格式。

注意：步驟 13 和步驟 14 是連續動作，也就是點一下左鍵再點右鍵，其間指標不要移動，如果離開乙字的「身上」，會出現另外一個不同的選單。



步驟 15：點選座標軸標題格式後會出現這一個對話框，選對齊方式，再在直立方向的文字上點一下，底下的角度會由原來的 90 變成 0，再確定。



步驟 16：乙字變直立了

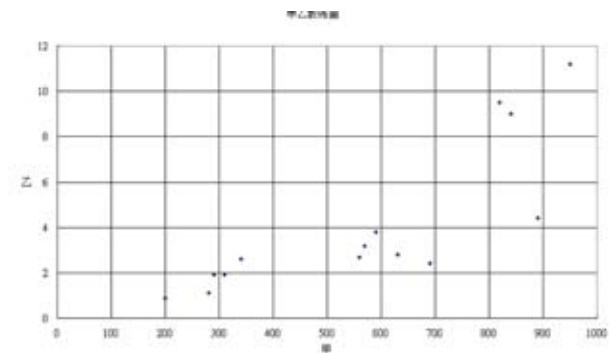


因為底色是灰色，如果想要印成講義會不明顯，所以要變成白色。

步驟 17：把指標放在灰色的區域中（只要是灰色即可，因為灰色的區域叫繪圖區，白色的外框叫圖表區，現在要改變繪圖區的底色）點選右鍵，會出現一張選單，選取繪圖區格式。



步驟 18：選取繪圖區格式後會出現對話框，在填滿的子選項上勾選無，然後再確定。



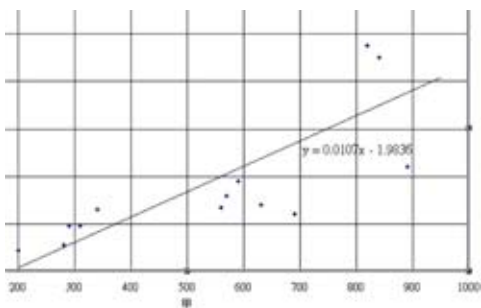
步驟 19：散佈圖完成了，明顯可以看出是正相關，我們再繼續做下去。



步驟 20：準確地把指標放在散佈圖上的任何一個資料點上（14個點其中一個點），然後點右鍵，會出現一張選單（一定要放準確，否則會出現另一張選單），再選取加上趨勢線。



步驟 21：出現加上趨勢線的選單，點選項，把圖表上顯示公式的項目勾選，再確定。



完成：出現的迴歸模式為： $y=0.0107x-1.9836$

該題的變數甲是海拔高度、變數乙是溪流坡度，變數丙是魚的種類數，因此表示溪流坡度會隨海拔高度的增加而增加，是正相關。

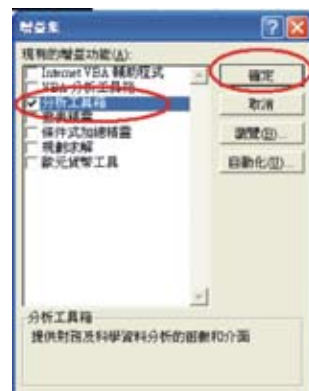
本題目的變數甲和變數丙的相關係數是 -0.74 ，我們用如下方法來做：

	甲	丙
200	11	
280	12	
290	10	
310	11	
340	3	
560	11	
570	5	
590	4	
630	3	
690	3	
820	4	
840	1	
890	5	
950	2	

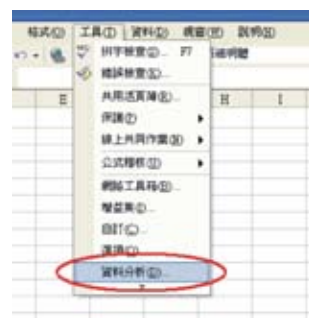
步驟 1：先把變數甲及變數丙兩個欄位copy在一起。



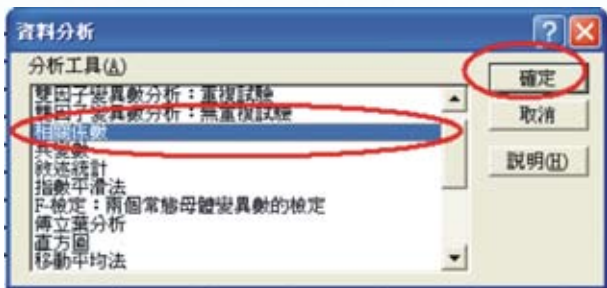
步驟 2：選工具，再選增益集。



步驟 3：出現增益集對話框，勾選分析工具箱，再確定。對話框會消失。



步驟 4：再選工具，這時多了一個資料分析的選項。



步驟 5：出現資料分析的對話框，選相關係數，再按確定。



步驟 6：相關係數的對話框出現，因為變數甲和變數丙是放在E2到F15的位址（E1及F1的文字千萬不要列進去）所以輸入範圍是E2到F15；輸出範圍隨便找一個空白的地方，就選F18好了，然後按確定。

840	1
890	5
950	2

	欄 1	欄 2
欄 1	1	
欄 2	-0.743281	1

完成：確定後，F18的位址出現相關係數為 -0.743281 。

以上兩個方法，筆者認為描繪散佈圖是很好的教學模式，至於計算相關係數的公式，的確可以簡化計算過程，但如果要講解，筆者是利用一張試算表來講授，可以逐步解釋，連同標準差的過程，一併講授。學生可以也在過程中熟悉公式的演算，下表提供參考。

	A	B	C	D	E	F	G
1	甲			丙			
2	200	368.57	135844.9	11	-4.92857	24.29082	-1816.53
3	280	288.57	83273.47	12	-5.92857	35.14796	-1710.82
4	290	278.57	77602.04	10	-3.92857	15.43367	-1094.39
5	310	258.57	66859.18	11	-4.92857	24.29082	-1274.39
6	340	228.57	52244.9	3	3.071429	9.433673	702.0408
7	560	8.57	73.46939	11	-4.92857	24.29082	-42.2449
8	570	-1.43	2.040816	5	1.071429	1.147959	-1.53061
9	590	-21.43	459.1837	4	2.071429	4.290816	-44.3878
10	630	-61.43	3773.469	3	3.071429	9.433673	-188.673
11	690	-121.43	14744.9	3	3.071429	9.433673	-372.959
12	820	-251.43	63216.33	4	2.071429	4.290816	-520.816
13	840	-271.43	73673.47	1	5.071429	25.71939	-1376.53
14	890	-321.43	103316.3	5	1.071429	1.147959	-344.388
15	950	-381.43	145487.8	2	4.071429	16.57653	-1552.96
16	568.57		820571.4	6.071429		204.9286	
17			905.85			14.3153	-9638.57

1. A16 是變數甲的平均值；D16 是變數丙的平均值。
2. B 欄是 $(X_i - \bar{X})$ ；C 欄是 $(X_i - \bar{X})^2$ ；C16 是 $\Sigma (X_i - \bar{X})^2$ 。
3. E 欄是 $(Y_i - \bar{Y})$ ；F 欄是 $(Y_i - \bar{Y})^2$ ；F16 是 $\Sigma (Y_i - \bar{Y})^2$ 。
4. C17 是 C16 開平方；F17 是 F16 開平方。
5. G17 是 $\Sigma (X_i - \bar{X})(Y_i - \bar{Y})$ 。
6. 代入相關係數的公式：
$$\frac{-9638.57}{905.85 \times 14.32} = -0.74$$