相關分析在試驗上的應用

李 蘭 帝

一、引 言

相關研究在農業試驗上扮演很重要的角色。其目的在求原資料所含的相互關係。所得的結果可做推測之用，或試驗結果之解釋用。其應用是多方面的，尚待研究者多開發這門。

二、單相關

單相關為大家所熟悉最簡單的相關分析法。現有兩種觀測值，互成一對時可做單相關分析。在做相關分析前我們應先考慮所用的數學模式。最簡單的模式為直線型，其式如下

$$\hat{Y} = a + bX$$

將兩種觀測值分為X與Y。把Y視為隨變數，把X視為自變數。其意為Y隨X而變。究竟將那一個觀測值視為隨變數或自變數，應由試驗者決定。$\hat{Y}$稱為Y的理論值，應該視為Y的最理想推測值。a及b各為常數。b有個特別稱呼叫迴歸係數。相關分析的第一步在求a及b。有了a及b之後，我們即可代入所得X而求$\hat{Y}$。

$$b = \frac{\Sigma XY - \dfrac{(\Sigma X)(\Sigma Y)}{N}}{\Sigma X^2 - \dfrac{(\Sigma X)^2}{N}} , \quad a = \frac{\Sigma Y - b\,\Sigma X}{N}$$

a及b 可由上述兩個公式求得。現舉例說明。設有十五個地 ，測得株數與產量，資料如下。

	株數（X）	產量（Y）	X^2	Y^2	XY
	12	15	144	225	180
	14	11	196	121	154
	14	12	196	144	168
	15	8	225	64	120
	17	6	289	36	102
	10	5	100	25	50
	11	6	121	36	66
	8	10	64	100	80
	8	9	64	81	72
	12	4	144	16	48
	7	4	49	16	28
	5	7	25	49	35
	5	6	25	36	30
	8	3	64	9	24
	7	4	49	16	28
合計	153	110	1755	974	1185

$$b=\frac{1185-\frac{(153)(110)}{15}}{1755-\frac{(153)^2}{15}}=0.3241 \qquad a=\frac{110-(0.3241)(153)}{15}=4.0275$$

$$Y=4.0275+0.3241(X)$$

上式雖然表示X與Y之關係，但這關係的密切程度應該由相關係數來表示。其公式如下。

$$r=\frac{\Sigma XY-\frac{(\Sigma X)(\Sigma Y)}{N}}{\sqrt{(\Sigma X^2-\frac{(\Sigma X)^2}{N})(\Sigma Y^2-\frac{(\Sigma Y)^2}{N})}}$$

以同資料為例

$$r=\frac{1185-\frac{(153)(110)}{15}}{\sqrt{(1755-\frac{(153)^2}{15})(974-\frac{(110)^2}{15})}}=0.352$$

相關係數的範圍在－1至＋1之間。此性質可做計算錯誤之判斷根據。當 r 之絕對值為1時，Y與$\hat{Y}$完全符合，可稱完全相關。相關的程度最好用 r 之平方表示，因 r^2 有如下意義。

$$r^2=\frac{\Sigma(\hat{Y}-\bar{Y})^2}{\Sigma(Y-\bar{Y})^2}$$

$\bar{Y}$為Y的平均值。分母顯然為總平方和，而分子為理論值的平方和。由此可見 r^2 表示相關所產生的平方和在總平方和所佔的比率。相關係數的顯著性測定可用F測驗，其公式如下。

$$F=\frac{r^2}{1-r^2}\times(N-2)$$

以同資料為例

$$F=\frac{(0.352)^2}{1-(0.352)^2}\times(15-2)=1.83$$

由F表查看機差自由度13，相關自由度為1時，5％的顯著標準為4.67。故該相關顯然不顯著。

以上所述最簡單的直線型單相關的分析方法。如果試驗者認為該數學模式並不理想，可另設模式。這些模式如果仍屬直線型計算並不難。例如試驗者認為相關直線必須經過坐標的原點，那麼模式應該採用 $\hat{Y}=bX$。如果試驗者發現Y與X之間的關係並非簡單的直線關係，而有凸或凹的曲線關係，可以將X或Y轉換成$\sqrt{X}$或$\sqrt{Y}$，或者取對數如 log X 或 log Y 等。這樣轉換的結果這個數學模式仍屬直線型，因為模式的直線型是針對迴歸係數講，並非對變數講。例如 $\hat{Y}=a+b\sqrt{X}$，$\hat{Y}=a+b\log X$，$\log\hat{Y}=a+bX$ 等均為直線型，所以可以用上述的方法做相關分析。有些數學模式雖只含兩個變數（這裡講的所謂兩種觀測值），但非屬直線型，故叫非直線型。例如 $\hat{Y}=A(1-e^{BX})$，雖然試驗者覺得這模式有用，但因為它是非直線型，所以相關分析較困難。遇到這種問題，通常需借用電腦，在此不便詳述。

三 複相關

當試驗者所得觀測值不只兩種而有兩種以上時，試驗者可以考慮每兩種觀測值互相間之關係。如果試驗者只考慮一種觀測值與其他兩種以上的觀測值的關係時，即所謂複相關。做複相關分析時同樣須先設 數學模式 。複相關與單相關一樣以直線型爲最簡單。

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_P X_P$$

雖然複相關分析，如果變數數目不多（四個以下 ）時可用筆算（當然用電算機來幫忙 ），通常都用電腦來幫忙（筆者在後面附電腦程式供參考 ）。因此詳細的計算方法從略。

複相關係數的意義與單相關相同，現以R來代表時，

$$R^2 = \frac{\Sigma(\hat{Y}-\bar{Y})^2}{\Sigma(Y-\bar{Y})^2} \qquad \text{其F值則} \quad F = \frac{R^2}{1-R^2} \times \frac{N-P-1}{P}$$

N爲觀測值的點數。P爲X變數數目，在此即爲複相關的自由度。故複相關的顯著性測定與單相關相同。

在複相關還可以對每一個X做顯著性測定。測定仍然可用F測定。但其F值計算較難，還是借用電腦吧。由這些F值可以看出每一個X對複相關的重要性。在這裡要特別注意的是，這樣看一個X對相關的重要程度並非絕對的。只能說在這樣X的組合裡它的重要性有多大而已。同一個X，在另一種組合裡它的重要性也會改變。所以複相關的解釋常須小心。

如果試驗者做複相關分析時，對Y及X都有預先安排的話，上述的分析應該大致可以滿足試驗者的要求。如果試驗者對Y雖有安排，但對X都沒有安排，目的只在求得最好相關的X組合的話，這問題並不簡單。還是到電腦中心去吧。

複相關的分析仍然與單相關一樣。如果發現所設模式不理想可另設模式，包括X或Y的轉換等（參閱單相關 ）。在複相關裡還可以考慮X之間的交感，而增設 $X_i X_j$ 等項。應採用那一樣的模式最好哪？這問題電腦是無法答覆的。電腦只能告訴您，當您採用那樣模式時，其結果如何而已。所以試驗者對試驗性質的了解，加上他在這方面的造詣是解決這問題不可缺的因素。

四 結　論

以上簡單介紹相關分析的概廓，筆者覺得還是說了一些廢話，因爲到了緊要關頭都建議試驗者到電腦中心去，不就等於說了廢話了嗎？最後筆者要强調，相關分析很重要，但千萬要小心以免導致錯誤的結論如下面的笑話：死亡與床在相關分析上應該很密切，因人大都死在床上。但不能說床是死亡的原因，人因病而躺在床上，也因病而死亡的緣故。

五 補　充

筆者在本文一再强調電腦在相關分析上的重要。現在以複相關爲例說明電腦如何處理這些繁雜的工作。複相關分析的計算方法以矩陣來表示最簡單。我們將X及Y的原始資料以如下矩陣來表示。

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & X_1 & X_2 \end{array} \\ \begin{bmatrix} 1 & 4 & 2 \\ 1 & 6 & 3 \\ 1 & 8 & 9 \\ 1 & 7 & 8 \\ 1 & 5 & 4 \end{bmatrix} \end{array} = X \qquad \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} Y \\ \begin{bmatrix} 9 \\ 6 \\ 7 \\ 4 \\ 3 \end{bmatrix} \end{array} = Y$$

爲了簡單起見我們假設X變數有2個。在X矩陣最左的一行全部爲1，其餘爲原資料，假定有5個點。這樣矩陣有5列及3行（橫爲列，直爲行的緣故 ）。最簡單的矩陣演算爲倒轉，即行與列掉換的意思。

$$\begin{array}{c} \\ 1 \\ X_1 \\ X_2 \end{array}\begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 4 & 6 & 8 & 7 & 5 \\ 2 & 3 & 9 & 8 & 4 \end{bmatrix} \end{array} = X'$$

此爲X矩陣的倒轉通常以X'代表。下面爲矩陣相乘。如A及B各代表矩陣，AB即爲兩個矩陣的相乘。現在假定A有3列5行，B有5列2行，相乘所得的矩陣爲3列2行的矩陣。在矩陣的相乘前面的矩陣的列的數目決定積的列的數目，而後面矩陣的行的數目決定行的數目。前面的矩陣的列數必與後面矩陣的行數相同。

$$\begin{bmatrix} 3 & 7 & 1 & 7 & 2 \\ 4 & 6 & 5 & 3 & 6 \\ 8 & 9 & 4 & 6 & 4 \end{bmatrix} \begin{bmatrix} 5 & 7 \\ 3 & 8 \\ 2 & 6 \\ 4 & 4 \\ 8 & 9 \end{bmatrix} = \begin{bmatrix} 72 & 129 \\ 108 & 172 \\ 131 & 212 \end{bmatrix} \qquad AB = C$$

如果矩陣是方矩陣（即列數與行數相同）　求反矩陣。反矩陣的結果還是方矩陣陣，行列數與原來矩陣相同。反矩陣通常以A^{-1}來代表A的反矩陣。

現在假定我們所用的數學模式爲

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2$$

目的在求b_0，b_1，b_2等。這三個數可寫成矩陣

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = B$$

那麼由矩陣數學可得

$$B = (X' X)^{-1} (X' Y)$$

由上式可知電腦只要處理矩陣的倒轉，相乘及反矩陣即可完成這任務。所以下面分三部份加以說明。以下所用電腦語言爲BASIC語言。

設有矩陣X（M，N）。括弧內M爲列數，N爲行數，倒轉結果爲T（N，M），注意M與N已掉換）。其程式如下。

```
1 FOR I = 1 TO M
2 FOR J = 1 TO N
3 T(I,J)= X(J,I)
4 NEXT J
5 NEXT I
```

設有矩陣A（M，L）及B（L，N）相乘積爲C（M，N）（注意行列數的關係）。其程式如下。

```
1 FOR I=1 TO M
2 FOR J=1 TO N
3 C(I,J)=0
4 FOR K=1 TO L
5 C(I,J)=C(I,J)+A(I,K)+B(K,J)
6 NEXT K
7 NEXT J
8 NEXT I
```

設有矩陣A（N，N），將它反矩陣之後仍在A矩陣的位置，其程式如下。

```
1 FOR K=1 TO N
2 S=A(K,K)
9 S=A(I,K)
10 A(I,K)=0
```

```
3  A( K , K )= 1
4  FOR J = 1 TO N
5  A( K , J )=A( K , J )/S
6  NEXT J
7  FOR I=1 TO N
8  IF I=K THEN 14
11 FOR J=1 TO N
12 A( I , J )=A( I , J )- S*A( K , J )
13 NEXT J
14 NEXT I
15 NEXT K
```

以上三種程式加以組合起來就可得所要的程式。

以上簡單說明電腦程式的寫法。目的在鼓勵大家對這門的興趣。