心理與教育統計學講義

第 二 回

405590-2

社團法人 考友社 出版發行

405590-2

心理與教育統計學講義第二回　目錄

第三講　相關分析與概率分配 …… 1

壹、相關分析 …… 1

貳、概率分配 …… 27

✥精選試題✥ …… 60

第三講　相關分析與概率分配

◎命題重點◎

壹、相關分析

（壹）相關係數

一、相關

事物總是相互關聯的，它們之間的關係是多種多樣。分析起來，大概有以下幾種情況：一種是因果關係，即一種現象是另一種現象的因，而另一種現象則是果。例如學習的努力程度是學習成績好壞的因（至少是部分的因）；在一定刺激強度範圍內，刺激強度經常是反應強度的因等等。第二種是共變關係，即表面看來有關聯的兩種事物都與第三種現象有關，這時兩種事物之間的關係，便是共變關係。例如春天出生的嬰兒與春天栽種的小樹，就其高度而言，表面上看來都在增長，好像有關，其實，這二者都是受時間因素影響在發生變化，在它們本身之間並沒有直接的關係。第三種是相關關係，即兩類現象在發展變化的方向與大小方面存在一定的關係，但不能確定這兩類現象之間哪個是因，哪個是果，也有理由認爲這兩者並不同時受第三因素的影響，即不存在共變關係。具有相關關係的兩種現象之間，關係是較爲複雜的，甚至可能包含有暫時尚未認識的因果關係及其共變關係在內。例如，同一組學生的語文成績與數學成績的關係，就是屬於相關關係。

統計學中所講的相關是指具有相關關係的不同現象之間的關係程度。相關的情況有以下三種：一是兩列變量變動方向相同，

即一種變量變動時，另一種變量亦同時發生或大或小與前一種變量同方向的變動，這稱為正相關。如身高與體重的關係，一般講身長越長體重就越重。第二種相關情況是負相關，這時兩列變量中若有一列變量變動時，另一列變量呈或大或小但與前一列變量指向相反的變動。例如初學打字時練習次數越多，出現錯誤的量就越少等等。第三種相關是零相關，即兩列變量之間無關係。這種情況下，一列變量變動時，另一列變量作無規律的變動。如學習成績優劣與身高之間的關係，就屬零相關，即無相關關係，二者都是獨立的隨機變量。

二、相關係數

相關係數（correlation coefficient）是兩列變量間相關程度的數字表現形式，或者說是表示相關程度的指標。作為樣本間相互關係程度的統計特徵數，常用r表示，作為母體母數，一般用ρ表示，並且是指線性相關而言。

相關係數的數值介於－1.00～＋1.00之間，常用小數形式表示。它只是一個比率，不代表相關的百分數，更不是相關量的相等單位的度量。相關係數的正負號，表示相關方向，正值表示正相關，負值表示負相關。相關係數數值的大小表示相關的程度。相關係數為0時，稱零相關，即毫無相關，為1.00時，表示完全正相關，相關係數為－1.00時，為完全負相關。這二者都是完全相關。如果相關係數的絕對值在1.00與0之間不同時，則表示關係程度不同。接近1.00端一般為相關程度密切，接近0值端一般為關係不夠密切（注意：若是非線性相關關係，而用直線相關計算r值可能很小，但不能說二變量關係不密切）。關於這一點如何判定，尚須考慮計算相關係數時樣本數目的多少。如果樣本數目較少，受取樣偶然因素的影響較大，很有可能本來無關的兩類事物，卻計算出較大的相關係數來。

（貳）積差相關

一、概念及適用資料

積差相關（ product moment correlation ）是英國統計學家皮爾遜於20世紀初提出的一種計算相關的方法，因而也稱皮爾遜相關（ Pearson correlation ），是求直線相關的基本方法。

積差相關適用於哪種資料呢？首先，兩列數據都是測量的數據，而且兩列變量各自母體的分配都是常態的，即常態雙變量。爲了判斷計算相關的兩列變量其母體是否爲常態分配，一般要根據已有的研究資料查詢，若無資料可查，研究者應取較大樣本分別對兩變量作常態性檢驗 。

二、計算積差相關的基本公式

積差相關係數用r表示

$$r = \frac{\Sigma xy}{NS_xS_y} \qquad \langle 3\text{-}1 \rangle$$

式中：$x = X - \overline{X}$　　$y = Y - \overline{Y}$

N爲成對數據的數目

S_x 爲X變量的標準差

S_Y 爲Y變量的標準差

上式又可改寫成：

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}} \qquad \langle 3\text{-}2a \rangle$$

式中x 、y的含義同上，它是根據 $S_x = \sqrt{\frac{\Sigma x^2}{N}}$ 推導而來的。上述兩式都需要計算離均差。如果用原數目計算，可用下式：

$$r = \frac{\Sigma XY - \frac{\Sigma X \Sigma Y}{N}}{\sqrt{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}\sqrt{\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}}} \qquad \langle 3\text{-}2b \rangle$$

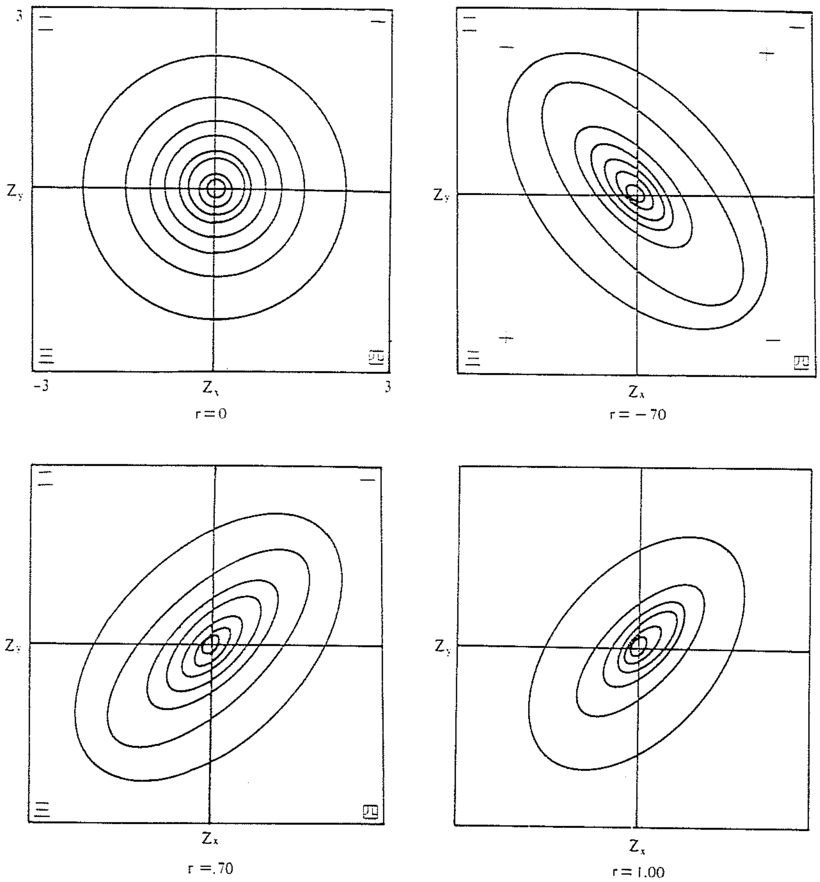

圖 3-1　以Z分數爲座標的不同相關程度散點分配

〔註〕：圖中不同圓表示散點密度不同，中心密，外周疏。

或寫作下式：

$$r=\frac{N\Sigma XY-\Sigma X\Sigma Y}{\sqrt{N\Sigma X^2-(\Sigma X)^2}\ \sqrt{N\Sigma Y^2-(\Sigma Y)^2}} \qquad \langle 3\text{-}2c \rangle$$

公式 < 3-2 *b* >、< 3-2 *c* > 是由公式 < 3-1 > 推演而來。

因爲 $\overline{X}=\frac{\Sigma X}{N}$，$\overline{Y}=\frac{\Sigma Y}{N}$，$x=X-\frac{\Sigma X}{N}$，$y=Y-\frac{\Sigma Y}{N}$，再將用原分數計算 S_X 及 S_Y 的公式代入公式 < 3-1 > 即得。

公式 < 3-1 > 中的 $\frac{\Sigma xy}{N}$ 稱爲協方差（covariance，或共變

數）。兩列變量離均差的乘積xy的大小，能夠反映兩列變量的一致性，即當x大y也大時，xy也大，當x小y小時，xy也小。Σxy的絕對值大小隨X、Y二列變量的一致性程度而變。表3-1的結果直觀地顯示了這個關係。

從表3-1結果可看出，Σxy絕對值大，相關係數的絕對值也

表 3-1　Σxy與X、Y的一致性關係及r的計算

被試	X	Y_1	x	y_1	xy_1	X	Y_2	x	y_2	xy_2
1	7	7	2	2	4	7	3	2	−2	−4
2	6	6	1	1	1	6	4	1	−1	−1
3	5	5	0	0	0	5	5	0	0	0
4	4	4	−1	−1	1	4	6	−1	1	−1
5	3	3	−2	−2	4	3	7	−2	2	−4
N=5	$\overline{X}=5$ $\overline{Y}=5$		$S_x=\sqrt{2}$ $S_{y1}=\sqrt{2}$		$\Sigma xy_1=10$ $r=1.00$		$S_x=\sqrt{2}$ $S_{y2}=\sqrt{2}$		$\Sigma xy_2=-10$ $r=-1.00$	

被試	X	Y_3	x	y_3	xy_3	X	Y_4	x	y_4	xy_4
1	7	4	2	−1	−2	7	5	2	0	0
2	6	3	1	−2	−2	6	6	1	1	1
3	5	7	0	2	0	5	3	0	−2	0
4	4	6	−1	1	−1	4	7	−2	2	−2
5	3	5	−2	0	0	3	4	−2	−1	2
	$S_x=\sqrt{2}$ $S_{y3}=\sqrt{2}$		$\Sigma xy_3=-5$ $r=-0.50$				$S_x=\sqrt{2}$ $S_{Y4}=\sqrt{2}$		$\Sigma xy=1$ $r=0.10$	

大，Σxy絕對值小相關係數也小。但是不能直接用 $\frac{\Sigma xy}{N}$ 表示一致性,因爲它是具有不同測量單位，而且其值大小隨應用測量單位的大小而變化的不穩定的量。爲了克服這一缺點，將各變量的離均差x、y分別用各自的標準差S_x與S_y除，使其成爲沒有實際測量單位的標準分數，然後求其共變數。這樣，不同測量，不同測量單位表示的兩列變量的一致性便可測量了。這就是求相關

係數的公式中所以用比率的由來。

將公式＜3-1＞稍加分析：

$$r=\frac{\Sigma xy}{NS_xS_y}=1/N\cdot\Sigma\left(\frac{x}{S_x}\right)\cdot\left(\frac{y}{S_y}\right)=(1/N)\,\Sigma Z_x Z_y \qquad ＜3\text{-}3＞$$

前文已講到一組變量Z分數的均數$\overline{Z}=0$，$S_Z=1$

$$\because \Sigma(Z-\overline{Z})^2=\Sigma(Z-0)^2=\Sigma Z^2$$
$$\Sigma Z^2=N\cdot S_Z^2=N$$

如果兩列變量中X變量增加Y變量也增加（相同或按比例增加），其各成對數據的標準分數$Z_X=Z_Y$，代入公式＜3-3＞，得：

$$r=(1/N)\,\Sigma Z_X\cdot Z_Y \qquad \because Z_X=Z_Y$$
$$r=1/N\cdot\Sigma Z^2$$
$$r=1/N\cdot N$$
$$r=1.00$$

這是完全正相關，r值最大不能超過1.00。同理可知，若$Z_x=-Z_Y$，則r值為-1.00；若成對數據的Z分數絕對值不相等，則r的絕對值就介於1.00～-1.00之間。

三、計算積差相關的其它公式

㈠差法公式：

這是利用離均差x、y相加或相減的方法，求積差相關係數的兩種方法。公式如下：

1.減差法

$$r=\frac{S_x^2+S_y^2-S_{x-y}^2}{2S_xS_y} \qquad ＜3\text{-}4a＞$$

2.因F為兩個變異數的比率，故F總為正值。

3.F分配表是根據F分配函數計算得來（F分配函數在一般的數理統計書中皆有）。因F分配是一族分配，故不能像常態表那樣列出不同F值以下或以上的概率。只能根據需要，列出最常用的0.95、0.99（指某F值左側，F分布曲線下的概率）或α為0.05、0.01（即某F值右側F分配曲線的概率，分別為1～0.95，1～0.99）。

4.當分子的自由度為1，分母的自由度為任意值時，F值與分母自由度相同概率的t值（雙側概率）的平方相等。例如分子自由度為1時，分母自由度為20時 $F_{0.05}=4.35$ $F_{0.01}=8.10$，查t值表df＝20時，$t_{0.05}=2.086$ $t^2_{0.05}=4.35$ $t_{0.01}=2.845$ $t^2_{0.01}=8.10$。這一點可以說明當組間自由度為1時（即分子的自由度為1)F檢驗與t檢驗的結論相同 。

✥精選試題✥

【題一】 今有10名學生的數學和語文考試成績（見表），問數學與語文成績是否相關？

学 生	語 文	数 学	R_x	R_y	D	D^2
1	59	47	4.5	6	−1.5	2.25
2	35	40	10	10	0	0
3	59	42	4.5	8	−3.5	12.25
4	57	55	6	3.5	2.5	6.25
5	50	49	7	5	2	4
6	71	63	1	1	0	
7	62	55	3	3.5	−0.5	0.25
8	47	42	8	8	0	0
9	43	42	9	8	1	1
10	68	57	2	2	0	0

答：

$$N=0 \qquad \Sigma D^2=26$$

$$\Sigma C_x=\frac{2(2^2-1)}{12}=0.5$$

$$\Sigma C_y=\frac{2(2^2-1)}{12}+\frac{3(3^2-1)}{12}=0.5+2=2.5$$

$$\Sigma x^2=\frac{N^3-N}{12}-\Sigma C_x=\frac{10^3-10}{12}-0.5=82$$

$$\Sigma y^2=\frac{N^3-N}{12}-\Sigma C_y=\frac{10^3-10}{12}-2.5=80$$

$$r_{RC}=\frac{83+80-26}{2\times\sqrt{82\times 80}}=\frac{136}{161.99}=0.84$$

因為一般的考試成績很難保證其分配為常態，因此成績分數雖屬等距變量，但不能用積差相關計算，而應該用等級相關計算相關係數。

故數學與語文成績有相關。

【題二】 今有9名學生的兩門功課成績評定分數（見表），問該兩門功課成績是否具有一致性？

學生	成績評定		等級		D	D^2
	課程A	課程B	R_x	R_y		
1	良	良	7	7.5	−0.5	0.25
2	優	優	2.5	3	−0.5	0.25
3	優	良	2.5	7.5	−5	25.00
4	良	優	7	3	4	16.00
5	優	優	2.5	3	−0.5	0.25
6	良	良	7	7.5	−0.5	0.25
7	中	中	11	11	0	0
8	良	優	7	3	4	16.00
9	良	中	7	11	−4	16.00
10	中	良	11	7.5	3.5	12.25
11	優	優	2.5	3	−0.5	0.25
12	中	中	11	11	0	0

答：

$N=12$ $\qquad \Sigma D^2=86.5$

$$\Sigma x^2=\frac{12^3-12}{12}-(\frac{3^3-3}{12}+\frac{4^3-4}{12}+\frac{5^3-5}{12})=126$$

$$\Sigma y^2=\frac{12^3-12}{12}-(\frac{3^3-3}{12}+\frac{4^3-4}{12}+\frac{5^3-5}{12})=126$$

$$r_{RC}=\frac{126+126-86.5}{2\times\sqrt{126\times126}}=\frac{165.5}{252}=0.6567$$

在心理與教育方面的研究，經常採取評價量表的方法，對成績或某些心理量進行評定，評價量表的分級越少，重複的等級數目就越多。這時若用等級相關法計算相關，則應該用公式計算。

因r_{RC}較大，故可以說兩門課程的成績具有一致性。