心理與教育統計學講義

第　四　回

405590-4

社團法人 考友社 出版發行

405590-4

心理與教育統計學講義第四回　目錄

第六講　變異數及迴歸分析

◎命題重點◎

壹、變異數分析

變異數分析又稱作方差分析（analysis of variance，或縮寫ANOVA）是一種應用非常廣泛的變量分析方法。其主要功能在於分析實驗數據中不同來源的變異對總變異的貢獻大小，從而確定實驗中的自變量是否對因變量有重要影響。

（壹）變異數分析的原理及其基本過程

一、變異數分析的基本原理

變異數分析作為一種統計方法，是把實驗數據的總變異分解為若干個不同來源的分量。因而它所依據的基本原理是變異的可加性。不同來源的變異只有當它們可加時，才能保證總變異分解的可能。具體地講，它是將總平方和分解為幾個不同來源的平方和（這裡的平方和指實驗數據與平均數離差的平方和）。

表6-1是A.B.C三種實驗處理的數據，其中$k=3$表示有三種實驗處理；$n=5$表示每種實驗處理中有5個實驗數據；$\overline{X}_j$表示某一種實驗處理的平均數；$\overline{X}_t$表示總平均數。

為了直觀地顯示出全部數據在$\overline{X}_t$上下變異，以及各組數據在本組平均數上下變異的情況，將表6-1數據作圖如下：

表 6-1

	A	B	C	(K=3)
	10	15	10	
	14	20	12	
(n=5)	12	17	6	
	8	8	12	
	11	15	10	
$\overline{X}_j$	11	15	10	$\overline{X}_t=12$

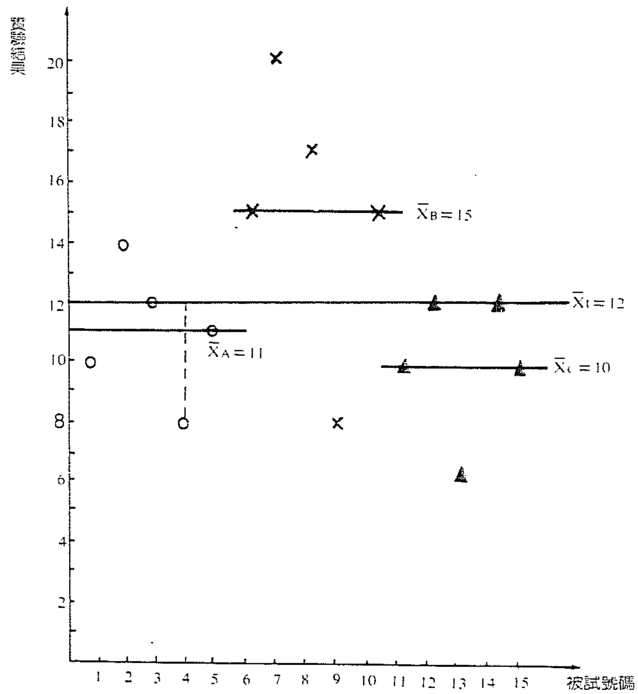

圖 6-1　數據變異示意圖

圖中1-5號被試的實驗結果（以0表示），即表6-1中A組的數據；6-10號實驗結果爲B組數據（以×表示）；11-15號實驗結果是C組數據（以▲表示）。三組各自的平均數$\overline{X}_A$ 、$\overline{X}_B$ 、$\overline{X}_C$

之間存在差異，每一組內的5個數據彼此也有差異，這兩部分的差異合起來即為實驗結果總的差異，或者說，每一個數據與$\overline{X}_t$的差異等於它與本組平均數之差加上小組平均數與$\overline{X}_t$的差。例如A組第4號被試（$X_4=8$）：$X_4-\overline{X}_t=-4$，而$X_4-\overline{X}_A=-3$，$\overline{X}_A-\overline{X}_t=-1$。

對於一般情況而言，任意一個數據$\overline{X}_{ij}$（第j組的第i個數據）與總平均數$\overline{X}_t$的離差（$\overline{X}_{ij}-\overline{X}_j$），同樣等於$X_{ij}$與該組平均數的離差（$X_{ij}-\overline{X}_j$）加上該組平均數與總平均數的離差（$\overline{X}_j-\overline{X}_t$）。

即：$(X_{ij}-\overline{X}_t)=(X_{ij}-\overline{X}_j)+(\overline{X}_j-\overline{X}_t)$

把第j組的n個數據的平方和相加：

$$\sum_{i=1}^{n}(X_{ij}-\overline{X}_t)^2=\sum_{i=1}^{n}[(X_{ij}-\overline{X}_j)+(\overline{X}_j-\overline{X}_t)]^2$$

$$=\sum_{i=1}^{n}(X_{ij}-\overline{X}_j)^2+2(\overline{X}_j-\overline{X}_t)\sum_{i=1}^{n}(X_{ij}-\overline{X}_j)$$

$$+\sum_{i=1}^{n}(\overline{X}_j-\overline{X}_t)^2$$

$$\because\ \sum_{i=1}^{n}(X_{ij}-\overline{X}_j)=0$$

$$\therefore\ \sum_{i=1}^{n}(X_{ij}-\overline{X}_t)^2=\sum_{i=1}^{n}(X_{ij}-\overline{X}_j)^2+n(\overline{X}_j-\overline{X}_t)^2$$

然後將K組的這種關係全加起來：

$$\sum_{j=1}^{k}\sum_{i=1}^{n}(X_{ij}-\overline{X}_t)^2$$

$$=\sum_{j=1}^{k}\sum_{i=1}^{n}(X_{ij}-\overline{X}_j)^2+n\cdot\sum_{j=1}^{k}(\overline{X}_j-\overline{X}_t)^2 \quad \text{< 6-1 >}$$

式中：$\sum_{i=1}^{n}$ 表示各組的數據從1加到n的和

$\sum_{j=1}^{k}$ 表示從第1組加到第K組之和

令 $SS_t = \sum_{j=1}^{k} \sum_{i=1}^{n} (X_{ij} - \overline{X}_t)^2$ ＜ 6-2 ＞

$SS_b = n \cdot \sum_{j=1}^{k} (\overline{X}_j - \overline{X}_t)^2$ ＜ 6-3 ＞

$SS_w = \sum_{j=1}^{k} \sum_{i=1}^{n} (X_{ij} - \overline{X}_j)^2$ ＜ 6-4 ＞

則 $SS_t = SS_b + SS_w$ ＜ 6-5 ＞

其中SS表示平方和；SS_t 表示總平方和，t代表全部（total）的意思；SS_b 表示組間平方和，即由於不同的實驗處理而造成的變異，b代表組間（between groups）之意；SS_w 表示組內平方和，即由實驗誤差（包括個體差異）造成的變異，w代表組內（within group）之意。

對於表6-1所列數據，實際計算結果如下：

$$
\begin{aligned}
SS_t &= [(10-12)^2 + (14-12)^2 + \cdots\cdots + (11-12)^2] \cdots\cdots\cdots ① \\
&\quad + [(15-12)^2 + (20-12)^2 + \cdots\cdots + (15-12)^2] \cdots\cdots\cdots ② \\
&\quad + [(10-12)^2 + (12-12)^2 + \cdots\cdots + (10-12)^2] \cdots\cdots\cdots ③ \\
&= 192
\end{aligned}
$$

$$SS_b = 5[(11-12)^2 + (15-12)^2 + (10-12)^2] = 70$$

$$
\begin{aligned}
SS_w &= [(10-11)^2 + (14-11)^2 + \cdots\cdots + (11-11)^2] \cdots\cdots\cdots ① \\
&\quad + [(15-15)^2 + (20-15)^2 + \cdots\cdots + (15-15)^2] \cdots\cdots\cdots ② \\
&\quad + [(10-10)^2 + (12-10)^2 + \cdots\cdots + (10-10)^2] \cdots\cdots\cdots ③ \\
&= 122
\end{aligned}
$$

$$192 = 70 + 122$$

由上可見，總變異可以分解為組間變異和組內變異。

我們知道，平方和除以自由度所得的樣本變異數即可作為其母體變異數的無偏估計。那麼，變異數分析中各組間的變異數就等於 $\frac{SS_b}{df_b}$ ，df 為組間自由度，組間變異數一般稱作組間均方，以MS_b 表示。

$$MS_b = \frac{SS_b}{df_b} \qquad < 6\text{-}6 >$$

同樣，組內變異數又稱組內均方。

$$MS_w = \frac{SS_w}{df_w} \qquad < 6\text{-}7 >$$

在變異數分析中，比較組間變異與組內變異，要用各自的均方來比較，而不能直接比較各自的平方和。這是因為在求平方和時，是若干項的平方和，其大小與項數（即k 或n ）有關，應該將項數的影響去掉，求其均方才能比較。因此必須除以各自的自由度，求均方。

組間自由度　$df_b = k - 1$

組內自由度　$df_w = k(n - 1)$

總自由度　$df_t = nk - 1$

$$df_t = df_b + df_w \qquad < 6\text{-}8 >$$

檢驗兩個變異數之間的差異要用F 檢驗，因此比較MS_b 與MS_w 也要用到F 檢驗即

$$F = \frac{MS_b}{MS_w} \qquad < 6\text{-}9 >$$

F 為組間變異與組內變異相比較而得出的一個比值F，如果F ＜1，說明數據的總變異中由分組不同所造成的變異只占很小的比例，大部分由實驗誤差和個體差異所致，也就是說不同的實驗

處理之間差異不大，或者說實驗處理基本上無效。如果F＝1，同樣說明實驗處理之間的差異不夠大。當F＞1而且落入F分配的臨界區，則表明數據的總變異基本上由不同的實驗處理所造成，或者說不同的實驗處理之間存在著顯著差異。

二、變異數分析的基本過程

在實際應用變異數分析時，爲了方便，一般直接從原始數據求平方和，很容易証明，這時平方和的公式爲：

$$SS_t = \Sigma\Sigma X^2 - \frac{(\Sigma\Sigma X)^2}{nk} \qquad \langle 6\text{-}10 \rangle$$

$$SS_b = \Sigma \frac{(\Sigma X)^2}{n} - \frac{(\Sigma\Sigma X)^2}{nk} \qquad \langle 6\text{-}11 \rangle$$

$$SS_w = SS_t - SS_b$$

$$= \Sigma\Sigma X^2 - \Sigma \frac{(\Sigma X)^2}{n} \qquad \langle 6\text{-}12 \rangle$$

現以表6-1所列數據，介紹變異數分析的基本過程。目的是討論A、B、C三種實驗處理之間是否有顯著差異。從表上看，三種實驗處理的平均數分別爲11、15和10，這似乎是由於實驗處理不同所造成，但是在同一種實驗處理中的5個數據並不相等，造成這個現象的原因在於實驗的某些偶然誤差及各組被試彼此間的個別差異，可以統稱之爲組內變異，也就是說，這15個數據彼此的差異，有實驗處理的原因，也有「組內變異」的原因，只有實驗處理的作用顯著地大於組內變異的作用時，才能確認實驗處理的有效作用，即A、B、C三種處理之間差異顯著。所以，必須通過變異數分析，看看組間變異數是否在統計上顯著地大於組內變異數。具體步驟如下：

①求平方和：

$$SS_t = \Sigma\Sigma X^2 - \frac{(\Sigma\Sigma X)^2}{nk} = 192$$

✥精選試題✥

【題一】　為研究不同科目的教師當班主任，對學生某一學科的學習是否有影響。把40名學生隨機分派到5名教不同科目的班主任負責的班級中，經過一段時間以後對40名學生進行數學考試，結果見表 6-1。用變異數分析的方法檢驗5組不同班主任的學生數學成績是否有顯著差異。（其中A表示班主任教數學，B表示班主任教語文，C表示班主任教生物，D表示班主任教地理，E表示班主任教物理。）

答：設成績分配為常態

①平方和：

$$SS_t = \Sigma\Sigma X^2 - \frac{(\Sigma\Sigma X)^2}{nk} = 204508 - \frac{(2852)^2}{8\times 5}$$

$$= 204508 - 203347.6 = 1160.4$$

$$SS_b = \sum\frac{(\Sigma X)^2}{n} - \frac{(\Sigma\Sigma X)^2}{n\cdot k} = 203662 - 203347.6$$

$$= 314.4$$

$$SS_w = SS_t - SS_b = 846$$

②自由度：

$$df_t = nk - 1 = 39$$

$$df_b = k - 1 = 5 - 1 = 4$$

$$df_w = k(n-1) = 35$$

③均方：

$$MS_b = \frac{SS_b}{df_b} = \frac{314.4}{4} = 78.6$$

$$MS_w = \frac{SS_w}{df_w} = \frac{846}{35} = 24.17$$

表 6-1

	A	B	C	D	E	
	76	76	62	65	67	
	78	67	70	68	71	
	65	70	69	68	72	
	72	64	73	71	69	
	71	67	71	61	74	
	72	83	69	69	79	
	83	72	73	65	76	
	79	73	69	69	84	
ΣX	596	572	556	536	592	$\Sigma\Sigma X = 2852$
$\frac{(\Sigma X)^2}{n}$	44402	40898	38642	35912	43808	$\Sigma \frac{(\Sigma X)^2}{n} = 203662$
ΣX^2	44624	41152	38726	35982	44024	$\Sigma\Sigma X^2 = 204508$

④F檢驗：

$$F = \frac{MS_b}{MS_w} = \frac{78.6}{24.17} = 3.252$$

查附表4　　$F_{0.05(4,35)} = 2.64$

$F > F_{0.05}$　　$P < 0.05$

在不同班主的班級中數學成績有顯著不同。

⑤變異數分析表：

變異來源	平方和	自由度	均　方	F	$F_{0.05}$
組　間	314.4	4	78.60	3.253*	2.64
組　內	846.0	35	24.17		
總	1160.4	39			

〔註〕：*表示F值在0.05水準顯著，若在0.01水準顯著，則以**表示

【題二】　用不同強度的光做視覺反應時的實驗（光強分爲Ⅰ、Ⅱ、Ⅲ三個等級），把被試隨機分成三組，分別做某一種光強的反應時實驗，由於某些原因，各組人數沒能相同。試問，從下面的結果能否得出不同強度的反應時有顯著不同？

答：

光強等級	Ⅰ	Ⅱ	Ⅲ	
視反應時間（毫秒）	150	190	200	
	220	230	240	
	190	170	260	
	170	260	180	
	240	250	190	
	200	170	280	
	180	280		
		190		
		220		
ΣX	1350	1960	1350	$\Sigma\Sigma X=4660$
n_j	7	9	6	
$\frac{(\Sigma X)^2}{n_j}$	260357.14	426844.4	303750	$\Sigma\frac{(\Sigma X)^2}{n_j}=990951.5$
ΣX^2	265900	439800	312100	$\Sigma\Sigma X^2=1017800$

$$SS_b=\sum\frac{(\Sigma X)^2}{n_j}-\frac{(\Sigma\Sigma X)^2}{N}$$
$$=990951.5-987072.7$$
$$=3878.8$$

$$SS_w=\Sigma\Sigma X^2-\sum\frac{(\Sigma X)^2}{n_j}$$
$$=1017800-990951.5$$
$$=26848.5$$