07變異數分析

陳欣得

靜宜大學企管系

1

07變異數分析

變異數分析是一種特殊的假設檢定型式。不同於正規假設檢定處理母體參數與樣本統計量的關係，變異數分析處理樣本資料的變異數（其內容為均差平方和），將它細分為各類**可解釋變異**的加總，最後留下不可解釋的**隨機變異**。在變異數分析領域，若虛無假設成立，隱含某對應的可解釋變異會很小，因此該變異偏高則會導致拒絕假設；其中，可解釋變異的大小是與隨機變異比較的結果。

章節安排：

2

§7.1 變異數分析概論

假設檢定可執行的虛無假設型式如下

$$H_0: 母體參數 \begin{matrix} \geq \\ = \\ \leq \end{matrix} 常數，$$

因此，無法處理以下這種三個或以上母體平均數相等的虛無假設：

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_k \text{，} k \geq 3 \text{。}$$

就上一章學到的內容，我們可以檢定兩母體的 $H_0: \mu_1 = \mu_2$ 與 $H_0: \sigma_1^2 = \sigma_2^2$ 的假設，但無法處理如 $H_0: \mu_1 = \mu_2 = 0$ ， $H_0: \mu_1 = \mu_2 = \mu_3$ 這種假設。

3

變異數分析（***Analysis Of Variance***，***ANOVA***）是為了因應這種三組或三組以上母體平均數相等之虛無假設，新發展的一種檢定方法。ANOVA 的基本想法是將樣本的總均差平方和（樣本變異數的分子部分，接下來稱為**總變異**）分成兩部分的和：可解釋的變異加上不可解釋的部份。其中，後者又稱為**隨機變異**，而前者因為解釋理由的不同，可能又會細分成幾部分。所謂解釋理由指的就是虛無假設，若虛無假設成立（各母體的平均數相等），則以各樣本平均數計算出來的變異（均差平方和）應該等於零或很小。

若可解釋變異的數值太大，則會促成拒絕 H_0 的結論；這裡可解釋變異數值的大或小是與隨機變異比較的相對大小。我們有 F 分配可來檢定上述的兩變異相對比值是否偏高，其中，可解釋變異放分子，隨機變異放分母。所以，ANOVA 一律是右尾檢定。

4

§7.1.1 組間變異與組內變異

令觀察值為x_{ij}，$i=1,\ldots,n_j$，$j=1,\ldots,k$，其中，一共有k組觀察值，而n_j為第j組的觀察值數目。以下將所有觀察值的均差平方和分解成兩項均差平方和：

$$\begin{aligned}\sum_{ij}\left(x_{ij}-\mu\right)^2&=\sum_{ij}\left(x_{ij}-\mu_j+\mu_j-\mu\right)^2\\&=\sum_j\sum_i\left(\mu_j-\mu\right)^2+\sum_j\sum_i\left(x_{ij}-\mu_j\right)^2\\&=\sum_j n_j\left(\mu_j-\mu\right)^2+\sum_j\sum_i\left(x_{ij}-\mu_j\right)^2,\end{aligned}$$

其中，μ為全部k組的總平均數，μ_j為第j組的平均數。在虛無假設成立的狀況下，組間均差平方和應為零，$SSB=0$。

這個等式可以閱讀為

總均差平方和 ＝ 組間均差平方和 ＋ 組內均差平方和。

中英文名稱對照分別如，**總均差平方和**（***total sum of square***，***SST***），**組間均差平方和**（***sum of square between groups***，***SSB***），與**組內均差平方和**（***sum of square within groups***，***SSW***）。三種均差平方和的數學定義如

$$\text{總均差平方和} = SST=\sum_{ij}\left(x_{ij}-\mu\right)^2\;;$$

$$\text{組間均差平方和} = SSB=\sum_j n_j\left(\mu_j-\mu\right)^2\;;$$

$$\text{組內均差平方和} = SSW=\sum_j\sum_i\left(x_{ij}-\mu_j\right)^2\;。$$

相同等式的縮寫符號表示則如

$$SST = SSB + SSW \text{ 。}$$

由於均差平方和與變異數之間只差變異數需除樣本數(嚴格來說是自由度，$n-1$)，為方便起見，我們常稱之為**總變異**、**組間變異**、與**組內變異**，即

總變異 ＝ 組間變異 ＋ 組內變異。

另外，組間變異與虛無假設有關，又稱為**處置變異**（***sum of square of treatment***，***SSTR***）或**可解釋變異**；相對地，組內變異又稱為**誤差變異**（***sum of square of error***，***SSE***）、**隨機變異**、或**不可解釋變異**。

在虛無假設成立的狀況下，組間均差平方和應為零，即 $SSB=0$ 。因此，若 SSB 越大，則意味著虛無假設不成立的機會越大。所以說，ANOVA 是右尾檢定。

§7.1.2 檢定統計量與變異數分析表

典型的變異數分析的檢定統計量為

$$F = \frac{\text{可解釋變異數}}{\text{不可解釋變異數}} = \frac{SSTR/(k-1)}{SSE/(N-k)} = \frac{MSTR}{MSE} \text{ ，}$$

其中，$N = n_1 + n_2 + \cdots + n_k$ ，***MSTR*** 稱為**處置均方和**（***mean sum of square of treatment***），***MSE*** 稱為**誤差均方和**（***mean sum of square of error***）。

變異數分析是特殊的假設檢定程序，其虛無假設一律為

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k \text{ ， } k \ge 3 \text{ ；}$$

其檢定統計量一律為

$$F_{df=(k-1,\, N-k)} = \frac{SSTR/(k-1)}{SSE/(N-k)} = \frac{MSTR}{MSE} \text{ ，右尾檢定。}$$

上面這個變異數分析的樣本檢定統計量值（$F^* = MSTR/MSE$）計算起來有些許複雜，因此，我們用**變異數分析表**（***ANOVA table***）來整合計算過程。**表 0.1** 是一個變異數分析表，其目的是一步步算出最右方的樣本檢定統計量值 F^*。請留意表中各數值間的關係。例如，(1)總變異之上的橫線表示加總的意思，即 $SSTR + SSE = SST$，$(k-1)+(N-k)=N-1$；(2)均方和為前兩欄（平方和、自由度）相除的結果；最後，(3)樣本檢定統計量值 F 是前一欄之兩均方和的比值。

表 0.1　變異數分析表

變異來源	平方和	自由度	均方和	F
組間變異	SSTR	k-1	MSTR = SSTR / k-1	F = MSTR / MSE
組內變異	SSE	N-k	MSE = SSE / N-k	
總變異	SST	N-1		

9

範例7.1　（變異數分析表，填表）

就以下變異數分析表，請推估並填滿表中其他數值：

變異來源	平方和	自由度	均方	F
處置	44.16	2		
組內				
總和	53.71	13		

【解】

$$SSE = SST - SSTR = 53.71 - 44.16 = 9.55$$
$$N - k = (N-1) - (k-1) = 13 - 2 = 11$$

10

範例7.1　（變異數分析表，填表）(續)

$$MSTR = \frac{SSTR}{k-1} = \frac{44.16}{2} = 22.08$$

$$MSE = \frac{SSE}{N-k} = \frac{9.55}{11} = 0.87$$

$$F = \frac{MSTR}{MSE} = \frac{22.08}{0.87} = 24.435$$

完整的變異數分析表如下

變異來源	平方和	自由度	均方	F
處置	44.16	2	22.08	25.43
組內	10	11	0.87	
總和	54	13		

§7.1.3 變異數分析的類別與假設

就可解釋變異有不同的定義，變異數分析分成三類：

(1)**單因子變異數分析**（***one-way ANOVA***）從總變異中分離出一個可解釋變異，留下來的是作為比較基準的隨機變異；

(2) **隨機區集設計**（***randomized block design***）除了可解釋變異的因子，實驗單位也依據某個外生變數分成幾個區集，如此，從總變異中可分離與因子有關以及與區集有關的兩個可解釋變異；

(3) **雙因子變異數分析**（***two-way ANOVA***）分析兩個因子所產生的變異，此兩因子也可能會有交互影響，因此，雙因子 ANOVA 有三個可解釋變異，分別來自兩個因子與因子間的交互影。

為了能夠有可用的檢定分配，所有變異數分析有以下三個假設：

(1)各組母體均為常態分配，

(2)各組母體的標準差相等，

(3)各組母體互相獨立。

以上假設來自常用的常態 $i.i.d.$，各組來自 $N(\mu_i,\sigma)$，$i=1,\ldots,k$，的一致且獨立分配，也假設各組間的樣本也是獨立的。

§7.2 單因子變異數分析

單因子變異數分析只從總變異中抽取出一組可解釋變異，過程如上一節的介紹，其分解結果如下：

$$\sum_{ij}\left(x_{ij}-\bar{x}\right)^2=\sum_j n_j\left(\bar{x}_j-\bar{x}\right)^2+\sum_j\sum_i\left(x_{ij}-\bar{x}_j\right)^2$$

其中，令

$$SST=\sum_{ij}\left(x_{ij}-\bar{x}\right)^2=\Sigma x_{ij}^2-n\bar{x}^2=\Sigma x_{ij}^2-\frac{\left(\Sigma x_{ij}\right)^2}{N}\text{，}$$

$$SSTR=\sum_j\sum_i\left(\bar{x}_j-\bar{x}\right)^2=\sum_j n_j\left(\bar{x}_j-\bar{x}\right)^2=n_1\bar{x}_1^2+n_2\bar{x}_2^2+\cdots+n_k\bar{x}_k^2-N\bar{x}^2\text{，}$$

$$SSE=\sum_j\sum_i\left(x_{ij}-\bar{x}_j\right)^2\text{。}$$

實務上（同學的課堂紙筆考試），我們可以用下列公式計算 SST、$SSTR$ 與 SSE；

$$SSTR=\left(\frac{(\Sigma A_1)^2}{n_1}+\frac{(\Sigma A_2)^2}{n_2}+\cdots+\frac{(\Sigma A_k)^2}{n_k}\right)-\frac{(\Sigma T)^2}{N}\ ,$$

$$SST=\Sigma x^2-\frac{(\Sigma T)^2}{N}\ ,$$

$$SSE=SST-SSTR\ ,$$

其中，

ΣA_j 為第 j 因子（行）觀察值之和，

ΣT 為全部觀察值之和，

Σx^2 為全部觀察值之平方和。

15

範例7.2　（從原始資料產生單因子ANOVA表）

請就以下資料計算並繪製單因子變異數分析表：

A	B	C
10	8	7
11	9	8
12	10	6
10	8	7
12		6

【解】

將資料作以下整理：

x	A	B	C	合計
	10	8	7	
	11	9	8	
	12	10	6	
	10	8	7	
	12		6	
ΣA	55	35	34	124
n1	5	4	5	14
(ΣA)²/n1	605	306.25	231.2	1142.45

x^2	A	B	C	Σx²
	100	64	49	
	121	81	64	
	144	100	36	
	100	64	49	
	144		36	
合計	609	309	234	1152

得到：$\Sigma T=124$，$N=14$，$\frac{(\Sigma A_1)^2}{n_1}+\frac{(\Sigma A_2)^2}{n_2}+\frac{(\Sigma A_3)^2}{n_3}=1,142.45$，$\Sigma x^2=1,152$。

16

範例7.2 （從原始資料產生單因子ANOVA表）(續)

由上面資料可以計算 $SSTR$、SST、與 SSE：

$$SSTR=\left[\frac{(\Sigma A_1)^2}{n_1}+\frac{(\Sigma A_2)^2}{n_2}+\frac{(\Sigma A_3)^2}{n_3}\right]-\frac{(\Sigma T)^2}{N}=1{,}142.45-\frac{(124)^2}{14}=44.16\text{；}$$

$$SST=\Sigma x^2-\frac{(\Sigma T)^2}{N}=1{,}152-\frac{(124)^2}{14}=53.71\text{；}$$

$$SSE=SST-SSTR=53.71-44.16=9.55\text{。}$$

最後，將之填入變異數分析表，並填滿其餘的部分，結果如下：

變異來源	平方和	自由度	均方	F
處置	44.16	2	22.08	25.435
隨機	9.55	11	0.87	
總和	53.71	13		

§7.2.1 變異數分析的假設檢定流程

變異數分析的假設檢定流程如下：

步驟一：虛無假設 $H_0:\mu_1=\mu_2=\cdots=\mu_k$；

步驟二：檢定統計量 $F_{df=(k-1,\,N-k)}=\dfrac{MSTR}{MSE}$，右尾檢定；

步驟三：拒絕區域 $R_0=\left\{F>F_{\alpha,\,df=(k-1,\,N-k)}\right\}$；

步驟四：計算樣本定統計量 F^*，若 $F^*\in R_0$，即 $F^*>F_{\alpha,\,df=(k-1,\,N-k)}$，則拒絕 H_0。

以上步驟一至步驟三每一個 ANOVA 除了組數 k 與顯著水準 α 有差異，其它部分完全一樣，因此常被忽略。因此常見的報告，只展示變異數分析表，然後直接將表中的 F^* 與臨界值 $F_{\alpha,\,df=(k-1,\,N-k)}$ 比較，並下拒絕 H_0 或無法拒絕 H_0 的結論。

範例7.2 (續) （ANOVA的假設檢定）

請檢定三母體的平均數是否相等（ $\alpha = 0.05$ ）。

【解】

變異數分析表如：

變異來源	平方和	自由度	均方	F
處置	44.16	2	22.08	25.435
隨機	9.55	11	0.87	
總和	53.71	13		

因 $F^* = 25.435 > F_{\alpha=0.05,\, df=(2,11)} = 3.9823$ ，拒絕 H_0 ，三母體的平均數不會完全相等。

範例7.2 (續) （ANOVA的假設檢定）(續)

完整的檢定流程：

(1) $H_0: \mu_A = \mu_B = \mu_C$ （右尾檢定）；

(2) $F_{df=(2,11)} = {}^{MSTR}\!/_{MSE}$ ；

(3) $R_0 = \left\{ F > F_{\alpha=0.05,\, df=(2,11)} = 3.9823 \right\}$ ；

(4) 完成變異數分析表，得到 $F^* = 25.435 \in R_0$ ，拒絕 H_0 ；

(5) 資料充分證明三組母體的平均數不完全相等。

§7.2.2 變異數分析的事後分析

在 ANOVA 中，若拒絕虛無假設，得到各組母體平均數不完全相等的結論；需要瞭解，這個結論不否認可能有部分母體的平均數相等。此時，我們有興趣知道產生不完全相等的原因何在。例如，若 H_0: $\mu_1=\mu_2=\mu_3$ 不成立，那麼問題是 $\mu_1\neq\mu_2$， $\mu_1\neq\mu_3$，或者是 $\mu_2\neq\mu_3$？

我們有兩種事後分析方法，(1)兩兩檢定母體間平均數是否相等，(2)以信賴區間法一次檢定所有母體的平均數。

變異數分析（ANOVA）的三個假設有常態分配、獨立、以及各組母體標準差皆相等，也就是， $\sigma_1=\sigma_2=\cdots=\sigma_k=\sigma$。這個共同標準差的不偏估計量為 $\sqrt{MSE}$ ，即

$$\hat{\sigma}=\sqrt{MSE}\ 。$$

其中， $\hat{\sigma}$ 為 σ 的估計量。

假設要檢定第 i 組與第 j 組間的平均數是否相等，即 $H_0:\mu_i=\mu_j$ ，檢定統計量為

$$t_{df=N-k}=\frac{\bar{x}_i-\bar{x}_j}{\sqrt{MSE\left(1/n_i+1/n_j\right)}}\ 。$$

在母體組數 k 很多時，兩兩組和的數量很大，此時改用信賴區間法來檢定比較有效率。先分別寫出各組平均數 μ_i ， $i=1,\ldots,k$ ，的信賴區間如下

$$CI_{\mu_i}=\{\bar{x}_i\pm\varepsilon\}=\{\bar{x}_i-\varepsilon\leq\mu_i\leq\bar{x}_i+\varepsilon\}\ ；$$

其中，$\varepsilon=t_{\frac{\alpha}{2},df=N-k}\times\sqrt{MSE}\big/\sqrt{n_i}$。然後檢測信賴區間是否有交集，例如若 $CI_{\mu_\ell}\cap CI_{\mu_m}=\varnothing$，則拒絕虛無假設 $H_0:\mu_\ell=\mu_m$ ，其中， $\ell,m\in\{1,2,\ldots,k\}$ ， $\ell\neq m$ 。

範例7.2 (續2) （ANOVA的假設檢定）

請兩兩檢測兩組平均數是否相等（ $\alpha=0.05$ ）。

【解】

檢定 μ_A 、 μ_B 是否相等：

(1) $H_0:\mu_A-\mu_B=0$ （雙尾檢定）；

(2) $t_{df=11}=\dfrac{\bar{x}_A-\bar{x}_B}{\sqrt{MSE\left(1/n_A+1/n_B\right)}}$ ；

(3) $R_0=\left\{|t|>t_{\alpha/2=0.025,\,df=11}=2.201\right\}$ ；

(4) $t^*=\dfrac{55/5-35/4}{\sqrt{0.87\left(1/5+1/4\right)}}=3.5960\in R_0$ ，拒絕 H_0 。

類似的，

μ_A 、 μ_C 間： $t^*=\dfrac{55/5-34/5}{\sqrt{0.87\left(1/5+1/5\right)}}=7.1197\in R_0$ ，拒絕 H_0 ；

μ_B 、 μ_C 間： $t^*=\dfrac{35/4-34/5}{\sqrt{0.87\left(1/4+1/5\right)}}=3.1165\in R_0$ ，拒絕 H_0 。

範例7.2 (續3) （ANOVA的假設檢定）

請以信賴區間法作事後檢定（ $\alpha=0.05$ ）。

【解】

三母體平均數的信賴區間分別如下：

$t_{\alpha/2=0.025,\,df=11}=2.201$ ， $MSE=0.87$ ；

$CI_{\mu_A}=\left\{\dfrac{55}{5}\pm 2.201\times\sqrt{0.87/5}\right\}=\left\{10.083\le\mu_A\le 11.917\right\}$ ；

$CI_{\mu_B}=\left\{\dfrac{35}{4}\pm 2.201\times\sqrt{0.87/4}\right\}=\left\{7.725\le\mu_B\le 9.775\right\}$ ；

$CI_{\mu_C}=\left\{\dfrac{34}{5}\pm 2.201\times\sqrt{0.87/5}\right\}=\left\{5.883\le\mu_C\le 7.717\right\}$ 。

因 $CI_{\mu_A}\cap CI_{\mu_B}=\varnothing$ ，故可推論在 $\alpha=0.05$ 下會拒絕 $H_0:\mu_A=\mu_B$ 。同理，虛無假設 $H_0:\mu_A=\mu_C$ 與 $H_0:\mu_B=\mu_C$ 也都不會成立。

範例7.2 (續3) （ANOVA的假設檢定）(續)

以下是電腦報表：

變數	平均	信賴區間 (1-α = 95%)		
A	11.000	10.083	11.917	--------------------------------(-----------)----
B	8.750	7.725	9.775	-----------------(------------)-----------------
C	6.800	5.883	7.717	-------(----------)-------------------------------

§7.2.3 變異數分析與兩母體平均數差的檢定

兩母體平均數差的檢定：

$$H_0 : \mu_1 = \mu_2 \text{，}$$

若 $X_1 \sim N(\mu_1, \sigma_1^2)$ 、 $X_2 \sim N(\mu_2, \sigma_2^2)$，母體變異數未知但相等，即 $\sigma_1^2 = \sigma_2^2 = \sigma^2$，則我們也可以用 ANOVA 來檢定。事實上，兩者完全一樣。

$$F_{df=(k-1,N-k)} = F_{df=(1,n_1+n_2-2)} = \frac{MSTR}{MSE} = \left(\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{MSE(1/n_1 + 1/n_2)}} \right)^2$$

$$= \left(\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2 (1/n_1 + 1/n_2)}} \right)^2 = t^2_{df=n_1+n_2-2}$$

§7.3 隨機區集設計

先看一下接下來範例的數據。某作物有三種施肥方式（稱為**處置**，***treatment***），這些作物種在五種不同的區域（稱為**區集**，***block***），兩者交叉一共有 15 筆收穫數據如下：

	T1	T2	T3
B1	7	9	12
B2	11	12	14
B3	13	11	8
B4	8	9	7
B5	9	10	13

我們想知道處置間以及區集間對作物的收穫是否有影響。也就是說，這個例子的變異數分析有兩種分組方式：縱向的處置分組，與橫向的集區分組。在實驗設計領域，這種作法稱為**隨機區集設計**（***Randomized block design***），將實驗單位分成幾個集區，然後將不同處置的實驗隨機分配給各集區中的實驗單位。

若以 ANOVA 來分析隨機區集設計的資料，會面臨兩種可解釋的變異：**處置變異**（***Treatments Variation***，***SSTR***），與**區集變異**（***Block Variables***，***SSBK***）。假設有 k 種處置方式（縱向分組）、b 個集區（橫向分組），總變異會有如下分解：

總變異＝處置變異＋區集變異＋隨機變異 ，

$$(SST = SSTR + SSBK + SSE)$$

這些變異的計算公式如

$$SST=\Sigma x^2-\frac{(\Sigma T)^2}{N}\ ;$$

$$SSTR=\left(\frac{(\Sigma A_1)^2}{b}+\frac{(\Sigma A_2)^2}{b}+\cdots+\frac{(\Sigma A_k)^2}{b}\right)-\frac{(\Sigma T)^2}{N}\ ;$$

$$SSBK=\left(\frac{(\Sigma B_1)^2}{k}+\frac{(\Sigma B_2)^2}{k}+\cdots+\frac{(\Sigma B_b)^2}{k}\right)-\frac{(\Sigma T)^2}{N}\ ;$$

$$SSE=SST-SSTR-SSBK\ 。$$

其中，

ΣA_j 為第 j 行觀察值之和，

ΣB_i 為第 i 列觀察值之和，

ΣT 為全部觀察值之和，

Σx^2 為全部觀察值之平方和。

隨機區集設計有兩組檢定，(1)檢定處置分組的平均數是否完全相等，(2)檢定區集分組的平均數是否完全相等。兩檢定的檢定統計量分別為

$$F_{TR}=\frac{MSTR}{MSE}=\frac{SSTR/(k-1)}{SSE/(k-1)(b-1)}\text{，自由度為 } df=(k-1,(k-1)(b-1))\ ;$$

$$F_{BK}=\frac{MSBK}{MSE}=\frac{SSBK/(b-1)}{SSE/(k-1)(b-1)}\text{，自由度為 } df=(b-1,(k-1)(b-1))\ 。$$

其變異數分析表如**表 0.1** 所示。請留意，不可解釋的變異（這裡稱為組內變異）永遠放在變異來源的最後一組；這是兩組可解釋變異是否偏高的比較基準。

表 0.1 隨機區集設計變異數分析表

變異來源	平方和	自由度	均方和	*F*
處置變異	SSTR	k-1	MSTR = SSTR / k-1	F_{TR} = MSTR / MSE
區集變異	SSBK	b-1	MSBK = SSBK / b-1	F_{BK} = MSBK / MSE
組內變異	SSE	(k-1)(b-1)	MSE = SSE / (k-1)(b-1)	
總變異	SST	N-1		

範例7.3　（隨機區集設計）

請就下列資料分別檢定處置間（T1~T3）與集區間（B1~B5）的平均數是否有差異。

x	T1	T2	T3
B1	7	9	12
B2	11	12	14
B3	13	11	8
B4	8	9	7
B5	9	10	13

範例7.3　（隨機區集設計）(續)

【解】

將資料作以下計算、整理：

x	T1	T2	T3	ΣB	n2	$(\Sigma B)^2/n2$
B1	7	9	12	28	3	261.33
B2	11	12	14	37	3	456.33
B3	13	11	8	32	3	341.33
B4	8	9	7	24	3	192.00
B5	9	10	13	32	3	341.33
ΣA	48	51	54	153	15	1592.33
n1	5	5	5	15		
$(\Sigma A)^2/n1$	460.8	520.2	583.2	1564.2		

x^2	T1	T2	T3	Σx^2
B1	49	81	144	
B2	121	144	196	
B3	169	121	64	
B4	64	81	49	
B5	81	100	169	
合計	484	527	622	1633

範例7.3 （隨機區集設計）(續2)

得到

$$\Sigma T = 153 \text{，} N = 15 \text{，} \frac{(\Sigma A_1)^2}{b} + \frac{(\Sigma A_2)^2}{b} + \cdots + \frac{(\Sigma A_k)^2}{b} = 1,564.20 \text{，}$$

$$\frac{(\Sigma B_1)^2}{k} + \frac{(\Sigma B_2)^2}{k} + \cdots + \frac{(\Sigma B_b)^2}{k} = 1,592.33 \text{，} \Sigma x^2 = 1,633 \text{。}$$

整理成變異數分析表如下：

變異來源	平方和	自由度	均方	F	p	F*
處置	3.60	2	1.80	0.388	0.690	4.459
區集	31.73	4	7.93	1.712	0.240	3.838
隨機	37.07	8	4.63			
總和	72.40	14			α =	0.05

其中

$$SSTR = 1,564.20 - \frac{153^2}{15} = 3.60 \text{，} SSBK = 1,592.33 - \frac{153^2}{15} = 31.73 \text{，}$$

$$SST = 1,633 - \frac{153^2}{15} = 31.73 \text{，} SSE = SST - SSTR - SSBK = 37.07 \text{。}$$

33

範例7.3 （隨機區集設計）(續3)

檢定結果，在顯著水準 $\alpha = 0.05$ 下，兩分組的平均數皆沒有差異。檢定的數值如下：

$F_{TR}^* = 0.39 < F_{\alpha=0.05,\, df=(2,8)} = 4.459$ ，無法拒絕 H_0 ；

$F_{BK}^* = 1.712 < F_{\alpha=0.05,\, df=(4,8)} = 3.838$ ，無法拒絕 H_0 。

34

§7.4 雙因子變異數分析

隨機區集設計中因子分組與區集分組是獨立的，而雙因子 ANOVA 中類似區集分組不是來自外生變數，而是另一個因子。兩個因子（內生變數）間就難避免會有交互影響的問題；這個交互作用就成為雙因子 ANOVA 的第三個可解釋變異的來源。

雙因子變異數分析中，總變異呈現以下的分解：

$$總變異 = 行間變異 + 列間變異 + 交互影響變異 + 隨機變異，$$
$$(SST = SSA + SSB + SSAB + SSE)$$

其中，行間變異（SSA）與列間變異（SSB）分別就是雙因子 ANOVA 的處置變異（SSTR）與區集變異（SSBK），這裡特意取一個與區集設計無關的名詞。

這些變異的計算公式如下：

$$SSA = \left(\frac{(\Sigma A_1)^2}{n_{A_1}} + \frac{(\Sigma A_2)^2}{n_{A_2}} + \cdots + \frac{(\Sigma A_a)^2}{n_{A_a}} \right) - \frac{(\Sigma T)^2}{N} ;$$

$$SSB = \left(\frac{(\Sigma B_1)^2}{n_{B_1}} + \frac{(\Sigma B_2)^2}{n_{B_2}} + \cdots + \frac{(\Sigma B_b)^2}{n_{B_b}} \right) - \frac{(\Sigma T)^2}{N} ;$$

$$SSE = \sum_{i=1}^{b} \sum_{j=1}^{a} (\text{第 } ij \text{ 方格內之變異}) = \sum_{i=1}^{b} \sum_{j=1}^{a} \left(x_{ij,1}^2 + x_{ij,2}^2 + \cdots + x_{ij,n_{ij}}^2 - \frac{(\Sigma x_{ij})^2}{n_{ij}} \right) ;$$

$$SSAB = SST - SSA - SSB - SSE 。$$

其中，

a 為行數，b 為列數，N 為總觀察樣本數量，
ΣA_j 為第 j 行觀察值之和，
ΣB_i 為第 i 列觀察值之和，
ΣT 為全部觀察值之和，
Σx^2 為全部觀察值之平方和。

在雙因子 ANOVA，因子 A 的各分組與因子 B 的各分組交叉組合成實驗方格，每一個實驗方格都需要有一個以上的觀察值；相對於隨機區集設計，每一方格只有一個觀察值。另外，在雙因子 ANOVA 的各變異中，最不好計算的是交互影響變異（SSAB），它最後由總變異減去前兩項變異與隨機變異（方格內變異）而來。

雙因子 ANOVA，除了行間平均數相等與列間平均數相等之外，另有第三個虛無假設，H_0：兩因子間沒有交互影響。三者的檢定統計量分別如下：

$$F_A = \frac{MSA}{MSE} = \frac{SSA/(a-1)}{SSE/(N-a\times b)}$$，自由度為 $df = (a-1, N-a\times b)$；

$$F_B = \frac{MSB}{MSE} = \frac{SSB/(b-1)}{SSE/(N-a\times b)}$$，自由度為 $df = (b-1, N-a\times b)$；

$$F_{AB} = \frac{MSAB}{MSE} = \frac{SSAB/(a-1)(b-1)}{SSE/(N-a\times b)}$$，自由度為 $df = ((a-1)(b-1), N-a\times b)$。

變異數分析表如**表 0.1** 所示。

表 0.1 雙因子變異數分析表

變異來源	平方和	自由度	均方和	F
行因子變異	SSA	a-1	MSA = SSA / a-1	F_A = MSA / MSE
列因子變異	SSB	b-1	MSB = SSB / b-1	F_B = MSB / MSE
交互影響變異	SSAB	(a-1)(b-1)	MSAB = SSAB / (a-1)(b-1)	F_{AB} = MSAB / MSE
隨機變異	SSE	N-ab	MSE = SSE / (N-ab)	
總變異	SST	N-1		

範例7.4 （雙因子ANOVA）

台大農場以 3 種不同品種（A、B、C）及 3 種不同施肥方式（甲、乙、丙）同時試驗，並測得玫瑰花的產量如下表：

品種	施肥方式	各區產量		
A	甲	64	71	67
	乙	72	73	71
	丙	76	84	90
B	甲	75	76	74
	乙	71	71	71
	丙	65	64	66
C	甲	57	67	62
	乙	59	71	65
	丙	61	77	69

假設次資料適合 ANOVA 分析，試以 $\alpha = 0.05$ 檢定：

(1) 品種不同會不會影響產量？

(2) 施肥方式不同會不會影響產量？

(3) 品種與施肥方式之間是否有交互作用？

範例7.4 （雙因子ANOVA）(續)

【解】

將原始資料整理成以下標準形式：

x_{ij}	A			B			C		
甲	64	71	67	75	76	74	57	67	62
乙	72	73	71	71	71	71	59	71	65
丙	76	84	90	65	64	66	61	77	69

x_{ij}^2	A			B			C		
甲	4096	5041	4489	5625	5776	5476	3249	4489	3844
乙	5184	5329	5041	5041	5041	5041	3481	5041	4225
丙	5776	7056	8100	4225	4096	4356	3721	5929	4761

範例7.4　（雙因子ANOVA）(續2)

將每方格中的數值加總，並整理成雙因子 ANOVA 的工作表：

x	A	B	C	ΣB	n2	$(\Sigma B)^2$/n2
甲	202	225	186	613	9	41752.11
乙	216	213	195	624	9	43264.00
丙	250	195	207	652	9	47233.78
ΣA	668	633	588	1889	27	132249.89
n1	9	9	9	27		
$(\Sigma A)^2$/n1	49580.44	44521	38416	132517.4		

x^2	A	B	C	Σx^2
甲	13626	16877	11582	
乙	15554	15123	12747	
丙	20932	12677	14411	
合計	50112	44677	38740	133529

其中，請留意各行、列的樣本數量 $n_A = n_B = \cdots = n_{丙} = 9$，需要點算各行與各列的實際樣本數量。

範例7.4　（雙因子ANOVA）(續3)

以及計算方格內變異所需的數值：

$x^2-(\Sigma x)^2/n$	A	B	C	SSE
甲	24.67	2	50	
乙	2	0	72	
丙	98.67	2	128	
合計	125.33	4	250	379.33

方格內變異計算，如（甲，A）方格的變異(均差平方和)為

$x^2-(\Sigma x)^2/n = 13,626 - 202^2/3 = 24.67$ 。

範例7.4 （雙因子ANOVA）(續4)

然後將以上工作表中的數值整理成以下變異數分析表：

變異來源	平方和	自由度	均方	F	p	F*
A因子	357.41	2	178.70	8.480	0.003	3.555
B因子	89.85	2	44.93	2.132	0.148	3.555
交互影響	542.37	4	135.59	6.434	0.002	2.928
隨機	379.33	18	21.07			
總和	1368.96	26			α =	0.05

其中，

$$SSA = 132,517.4 - \frac{(1,889)^2}{27} = 375.4 \text{ ；}$$

$$SSB = 132,249.89 - \frac{(1,889)^2}{27} = 89.89 \text{ ；}$$

$$SSE = 379.33 \text{ ；}$$

$$SST = 133,529 - \frac{(1,889)^2}{27} = 1,369 \text{ ；}$$

$$SSAB = SST - SSA - SSB - SSE = 542.37 \text{ 。}$$

43

範例7.4 （雙因子ANOVA）(續5)

檢定結果，在顯著水準 $\alpha = 0.05$ 下，A 因子（品種）會影響產量，B 因子（施肥方式）不會影響產量，兩因子間（品種與施肥方式）有交互影響。檢定的數值如下：

$F_A^* = 8.480 > F_{\alpha=0.05,\, df=(2,18)} = 3.555$ ，拒絕 H_0 ；

$F_B^* = 2.132 < F_{\alpha=0.05,\, df=(2,18)} = 3.555$ ，無法拒絕 H_0 ；

$F_{AB}^* = 6.434 > F_{\alpha=0.05,\, df=(4,18)} = 2.928$ ，拒絕 H_0 。

A 因子（品種間）的事後分析如下表：

變數	平均	信賴區間 (1-α = 95%)		
A	74.22	71.01	77.44	--------------------------------(---------)------
B	70.33	67.12	73.55	--------------------------(---------)------------
C	65.33	62.12	68.55	-------------------(---------)--------------------

顯示， $\mu_A \neq \mu_C$ 。

44

The End

45

©陳欣得 2025/8/18 《113-4 統計學(一)》07變異數分析

46