第5章 變異分析
Analysis of Variance

1. 變異的分解（Decomposition of Variations）
2. 因子的重要性測試（Significance Tests）
3. 實驗誤差的統合（Pooling of Errors）
4. 信賴區間（Confidence Intervals）
5. Excel圖表製作（二）（Graphing/Tabling Using Excel）

第5.1節 變異的分解
Decomposition of Variations

5.1-1 實驗誤差（Experimental Errors）
5.1-2 變異的分解：1個控制因子（Decomposition of Variations）
5.1-3 變異的分解：多個控制因子（Multiple Control Factors）
5.1-4 實例：$L_4(2^3)$直交表實驗（Example: $L_4(2^3)$ Experiments）
5.1-5 實例：瓷磚製程實驗（Example: Tile Experiments）

表5.1-1 瓷磚製程實驗的數據（上表）及因子反應表（下表）

Exp.	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	y_1	y_2	y_3	y_4	y_5	y_6	y_7	$\overline{y}$	*S*
1	1	1	1	1	1	1	1	1	10.18	10.18	10.12	10.06	10.02	9.98	10.20	10.106	0.0870
2	1	1	2	2	2	2	2	2	10.03	10.01	9.98	9.96	9.91	9.89	10.12	9.986	0.0776
3	1	1	3	3	3	3	3	3	9.81	9.78	9.74	9.74	9.71	9.68	9.87	9.761	0.0641
4	1	2	1	1	2	2	3	3	10.09	10.08	10.07	9.99	9.92	9.88	10.14	10.024	0.0964
5	1	2	2	2	3	3	1	1	10.06	10.05	10.05	9.89	9.85	9.78	10.12	9.971	0.1293
6	1	2	3	3	1	1	2	2	10.20	10.19	10.18	10.17	10.14	10.13	10.22	10.176	0.0321
7	1	3	1	2	1	3	2	3	9.91	9.88	9.88	9.84	9.82	9.80	9.93	9.866	0.0476
8	1	3	2	3	2	1	3	1	10.32	10.28	10.25	10.20	10.18	10.18	10.36	10.253	0.0709
9	1	3	3	1	3	2	1	2	10.04	10.02	10.01	9.98	9.95	9.89	10.11	10.000	0.0698
10	2	1	1	3	3	2	2	1	10.00	9.98	9.93	9.80	9.77	9.70	10.15	9.904	0.1563
11	2	1	2	1	1	3	3	2	9.97	9.97	9.91	9.88	9.87	9.85	10.05	9.929	0.0713
12	2	1	3	2	2	1	1	3	10.06	9.94	9.90	9.88	9.80	9.72	10.12	9.917	0.1392
13	2	2	1	2	3	1	3	2	10.15	10.08	10.04	9.98	9.91	9.90	10.22	10.040	0.1199
14	2	2	2	3	1	2	1	3	9.91	9.87	9.86	9.87	9.85	9.80	10.02	9.883	0.0687
15	2	2	3	1	2	3	2	1	10.02	10.00	9.95	9.92	9.78	9.71	10.06	9.920	0.1295
16	2	3	1	3	2	3	1	2	10.08	10.00	9.99	9.95	9.92	9.85	10.14	9.990	0.0973
17	2	3	2	1	3	1	2	3	10.07	10.02	9.89	9.89	9.85	9.76	10.19	9.953	0.1473
18	2	3	3	2	1	2	3	1	10.10	10.08	10.05	9.99	9.97	9.95	10.12	10.037	0.0673

Ave = 9.984

Level	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
1	10.016	9.934	9.988	9.989	9.999	10.074	9.978	10.032
2	9.953	10.002	9.996	9.970	10.015	9.972	9.967	10.020
3		10.016	9.969	9.995	9.938	9.906	10.007	9.901

5.1-1 實驗誤差 (Experimental Error)

- 第一組實驗值： 10.18　10.18　10.12　10.06　10.02　9.98　10.20
- 在完全相同的製程條件下所量測到的不一樣的實驗值，我們將之全部歸因於「　　　　　　」，或簡稱「　　　」。
- 第一組實驗的變異向量：

$$\{\ 10.18\quad 10.18\quad 10.12\quad 10.06\quad 10.02\quad 9.98\quad 10.20\ \}$$

-

=

自由度（Degrees of Freedom）

- 5.1-1式中的$\{y_i\}$含有_____個獨立量測的數據；$\{y_i\}$有_____個自由度。
- $\{\bar{y}\}$ 代表平均值，因為7個數值都是一樣的，事實上只是_____個獨立的數據；$\{\bar{y}\}$有_____個自由度。
- $\{y_i - \bar{y}\}$代表每一個數據與_________的距離，這7個距離完全獨立嗎？它們之間有什麼關係？

其中 r 是實驗重複次數。 $\{y_i - \bar{y}\}$ 有_____個自由度。

- 綜上所述， 5.1-1式隱藏著一個關係：______________________________，亦即

平方和（Sum of Squares, SS）

- 5.1-1式中也隱藏著另一個關係：____________________________________，

  亦即

  以本例而言，

- How to prove the relationship shown in equation (5.1-4)?

向量分解：畢氏定理（Pythagorean Theorem）

- 5.1-1式中的$\{\bar{y}\}$和$\{y_i - \bar{y}\}$是__________的向量，亦即，它們的______為 0，可以表示成

How to prove the above equation?

- 5.1-1式相當於將一個向量$\{y_i\}$分解成互相直交的兩個向量，______和__________。
- 依據畢氏定理，斜邊「長度」的_______會等於其它兩邊「長度」的__________，亦即5.1-4式中的關係。
- 畢氏定理可以直接延伸至n度空間的情況：若將一個向量分解成許多互相直交的向量（在n度空間的一個向量，可以分解成n個互相直交的向量），則原向量「長度」的_______會等於分解後所有向量「長度」的_________。

變異向量與實驗誤差

- 5.1-1式中，$\{y_i - \bar{y}\}$代表量測值與平均值的差距，或稱為相對於平均值的「　　　」，此向量稱為「　　　　　　」。
- 在此例中，「變異」乃__________造成的，所以我們可以用$\{y_i - \bar{y}\}$來代表__________。
- 但是$\{y_i - \bar{y}\}$是一個______，處理起來並不容易，統計學家喜歡以單一數值來代表實驗誤差。
- 對 r 個獨立的數據$\{y_i\}$而言，實驗誤差被定義為此「變異向量」內各數值的平方和除以它們的「自由度」後再取平方根，亦即：

以本例而言，

5.1-2 變異的分解：1個控制因子

- 將A因子從水準1調到水準2，其他因子都不變：

表5.1-2 一個控制因子的實驗及因子反應表

Exp.	A	y_1	y_2	y_3	y_4	y_5	y_6	y_7	$\bar{y}$	S
1	1	10.18	10.18	10.12	10.06	10.02	9.98	10.20	10.106	0.0870
2	2	10.06	10.18	10.05	9.88	9.96	9.93	10.04	10.014	0.0996

Ave = 10.060

	A
Level 1	10.106
Level 2	10.014

- 表5.1-2的上表中陰影部份代表14個實驗數據，它們會有如此的變異除了________的影響外，另外也受到____________的影響，我們要將這兩種影響因素區隔出來。
- 首先，我們將____________________________，其差異稱為**總變異向量**，再將此總變異向量分解成兩部份：(1)因為____________而產生的變異，及(2)因為___________而產生的變異。

- 我們將資料的「演化」分解成兩個階段：

$$\begin{bmatrix} 10.060 & 10.060 & 10.060 & 10.060 & 10.060 & 10.060 & 10.060 \\ 10.060 & 10.060 & 10.060 & 10.060 & 10.060 & 10.060 & 10.060 \end{bmatrix}$$

$$\to \begin{bmatrix} 10.106 & 10.106 & 10.106 & 10.106 & 10.106 & 10.106 & 10.106 \\ 10.014 & 10.014 & 10.014 & 10.014 & 10.014 & 10.014 & 10.014 \end{bmatrix}$$

$$\to \begin{bmatrix} 10.180 & 10.180 & 10.120 & 10.060 & 10.020 & 9.980 & 10.200 \\ 10.060 & 10.180 & 10.050 & 9.880 & 9.960 & 9.930 & 10.040 \end{bmatrix}$$

第一個階段代表：因為________________而產生的影響，其值是

第二個階段代表：因為_____________而產生的影響，其值是

- 所以，此14個實驗數據可以作如下的分解：

- 5.1-6式可視為一個向量（14個自由度）分解成三個向量，其中第一個向量為____________向量（___個自由度），第二個向量為____________向量（___個自由度），第三個向量為____________向量（_____個自由度）。

自由度與平方和

- 5.1-6式中的各向量有下列關係：

- 仔細檢驗這三個向量，你會發現它們互相垂直。如何驗證？
- 因此， _______定理應該會成立， 以本例而言：

5.1-3 變異的分解：多個控制因子

- 以上結論可以直接推廣到多個控制因子的情況：一個代表實驗數據的向量$\{y_{ij}\}$可以分解為互相垂直的幾個向量，包括____________$\{\bar{y}\}$、____________（$\{E_A\}, \{E_B\}, \{E_C\}, \ldots$等）、及____________$\{e\}$，亦即

- 自由度的關係：

- 平方和的關係：

自由度與平方和的計算

- 當實驗數據或控制因子______時，我們可以將向量分解，再計算各項的平方和及自由度。
- 當實驗數據_______時，我們不太可能進行如5.1-6式的向量分解的工作。
- 事實上，變異分析的計算工作並不需知道這些向量中的每一個數值，只需知道每一個向量的__________及___________ 。

總變異向量的自由度與平方和

- 假設$\{y_{ij}\}$是一個$n \times r$的向量，它是由L_n直交表實驗中，量測到的$n \times r$個數據所組成。
- 這個直交表實驗包含了A, B, C, ...等控制因子，其中A因子有L_A個水準，B因子有L_B個水準，C因子有L_C個水準，...等。
- 這些數據是獨立量測的，所以向量$\{y_{ij}\}$的自由度是______，而平方和是________。
- 總平均值向量只包含____個值$(\bar{y})$，所以其自由度是____，其平方和是__________。
- 所以總變異向量（$\{y_{ij}\} - \{\bar{y}\}$）自由度（DOF_T）與平方和（SS_T）分別是：

$$DOF_T =$$

$$SS_T =$$

因子效應向量：1個控制因子

- 5.1-6式中，

$$\begin{bmatrix} 0.046 & 0.046 & 0.046 & 0.046 & 0.046 & 0.046 & 0.046 \\ -0.046 & -0.046 & -0.046 & -0.046 & -0.046 & -0.046 & -0.046 \end{bmatrix}$$

代表「因為A因子的變動而產生的變異」，或稱「　　　　　　　　」。

- A因子在水準1時其反應比總平均值__________：

- A 因子在水準2時其反應比總平均值_________：

因子效應向量的自由度與平方和

- 一般而言，P因子的因子效應向量是一個$n \times r$陣列，其中每一行都相同，而第 i 列的數值是_____________ ，其中 k 為直交表中，P因子所相對的直行中第 i 列的水準值。以表5.1-1的實驗為例（**只列出其中一行**）：

$$[E_A] = \begin{bmatrix} \bar{y}_{A1}-\bar{y} \\ \bar{y}_{A1}-\bar{y} \\ \bar{y}_{A1}-\bar{y} \\ \bar{y}_{A1}-\bar{y} \\ \bar{y}_{A1}-\bar{y} \\ \bar{y}_{A1}-\bar{y} \\ \bar{y}_{A1}-\bar{y} \\ \bar{y}_{A1}-\bar{y} \\ \bar{y}_{A1}-\bar{y} \\ \bar{y}_{A2}-\bar{y} \\ \bar{y}_{A2}-\bar{y} \\ \bar{y}_{A2}-\bar{y} \\ \bar{y}_{A2}-\bar{y} \\ \bar{y}_{A2}-\bar{y} \\ \bar{y}_{A2}-\bar{y} \\ \bar{y}_{A2}-\bar{y} \\ \bar{y}_{A2}-\bar{y} \\ \bar{y}_{A2}-\bar{y} \end{bmatrix} = \begin{bmatrix} 10.016-9.984 \\ 10.016-9.984 \\ 10.016-9.984 \\ 10.016-9.984 \\ 10.016-9.984 \\ 10.016-9.984 \\ 10.016-9.984 \\ 10.016-9.984 \\ 10.016-9.984 \\ 9.953-9.984 \\ 9.953-9.984 \\ 9.953-9.984 \\ 9.953-9.984 \\ 9.953-9.984 \\ 9.953-9.984 \\ 9.953-9.984 \\ 9.953-9.984 \\ 9.953-9.984 \end{bmatrix} = \begin{bmatrix} 0.032 \\ 0.032 \\ 0.032 \\ 0.032 \\ 0.032 \\ 0.032 \\ 0.032 \\ 0.032 \\ 0.032 \\ -0.032 \\ -0.032 \\ -0.032 \\ -0.032 \\ -0.032 \\ -0.032 \\ -0.032 \\ -0.032 \\ -0.032 \end{bmatrix}$$

$$[E_B] = \begin{bmatrix} \bar{y}_{B1}-\bar{y} \\ \bar{y}_{B1}-\bar{y} \\ \bar{y}_{B1}-\bar{y} \\ \bar{y}_{B2}-\bar{y} \\ \bar{y}_{B2}-\bar{y} \\ \bar{y}_{B2}-\bar{y} \\ \bar{y}_{B3}-\bar{y} \\ \bar{y}_{B3}-\bar{y} \\ \bar{y}_{B3}-\bar{y} \\ \bar{y}_{B1}-\bar{y} \\ \bar{y}_{B1}-\bar{y} \\ \bar{y}_{B1}-\bar{y} \\ \bar{y}_{B2}-\bar{y} \\ \bar{y}_{B2}-\bar{y} \\ \bar{y}_{B2}-\bar{y} \\ \bar{y}_{B3}-\bar{y} \\ \bar{y}_{B3}-\bar{y} \\ \bar{y}_{B3}-\bar{y} \end{bmatrix} = \begin{bmatrix} 9.934-9.984 \\ 9.934-9.984 \\ 9.934-9.984 \\ 10.002-9.984 \\ 10.002-9.984 \\ 10.002-9.984 \\ 10.016-9.984 \\ 10.016-9.984 \\ 10.016-9.984 \\ 9.934-9.984 \\ 9.934-9.984 \\ 9.934-9.984 \\ 10.002-9.984 \\ 10.002-9.984 \\ 10.002-9.984 \\ 10.016-9.984 \\ 10.016-9.984 \\ 10.016-9.984 \end{bmatrix} = \begin{bmatrix} -0.050 \\ -0.050 \\ -0.050 \\ 0.018 \\ 0.018 \\ 0.018 \\ 0.032 \\ 0.032 \\ 0.032 \\ -0.050 \\ -0.050 \\ -0.050 \\ 0.018 \\ 0.018 \\ 0.018 \\ 0.032 \\ 0.032 \\ 0.032 \end{bmatrix}$$

因子效應向量的自由度與平方和

- A因子效應向量的自由度是_____。
- B因子效應向量的自由度是_____。
- 一般而言，若P因子有L_P個水準，則此向量有___________個自由度。亦即

$$DOF_P =$$

平方和可以表示為：

$$SS_P =$$

誤差向量的自由度與平方和

- 5.1-6式中

$$\begin{bmatrix} 0.074 & 0.074 & 0.014 & -0.046 & -0.086 & -0.126 & 0.094 \\ 0.046 & 0.166 & 0.036 & -0.134 & -0.054 & -0.084 & 0.026 \end{bmatrix}$$

代表「因為實驗誤差而產生的變異」，或稱為「　　　　　」。

- 此向量的自由度是 ______ 。
- 「實驗誤差」定義為「　　　　　」各分量的平方和除以其自由度後再開根號：

$$S =$$

誤差向量的自由度與平方和

- 實驗誤差亦可利用表5.1-2中兩組個別實驗誤差（0.0870及0.0996）計算出來：

$S =$

亦即將兩個實驗誤差______(pooling)為一個實驗誤差。

- 一般而言，若有n組實驗，而每組實驗有r個實驗數據，則此$n \times r$個數據的實驗誤差為：

式中的S_i表示第 i 組實驗數據的標準偏差。

誤差向量的自由度與平方和

- 5.1-16式中，根號內分母及分子則是誤差向量的________及________，亦即

$$DOF_e =$$

$$SS_e =$$

- 5.1-18式中的標準偏差是以2.1-4式（s）計算的，若以2.1-6式（s_n）計算則應改寫成：

5.1-4 實例：$L_4(2^3)$直交表實驗

表5.1-3 $L_4(2^3)$直交表實驗數據及因子反應表

Exp.	A	B	C	y_1	y_2	y_3	y_4	y_5	$\bar{y}$	S
1	1	1	1	69.7	72.4	55.0	65.3	68.7	66.2	6.77
2	1	2	2	62.1	77.2	72.5	69.8	71.9	70.7	5.52
3	2	1	2	100.2	89.6	95.2	102.5	98.8	97.3	5.03
4	2	2	1	99.1	104.4	100.7	105.8	95.6	101.1	4.10

Ave. = 83.825

	A	B	C
Level 1	68.46	81.74	83.67
Level 2	99.19	85.91	83.98

向量分解

$$\begin{bmatrix} 69.7 & 72.4 & 55.0 & 65.3 & 68.7 \\ 62.1 & 77.2 & 72.5 & 69.8 & 71.9 \\ 100.2 & 89.6 & 95.2 & 102.5 & 98.8 \\ 99.1 & 104.4 & 100.7 & 105.8 & 95.6 \end{bmatrix} - \begin{bmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} \end{bmatrix} + \begin{bmatrix} \end{bmatrix}$$

$$+ \begin{bmatrix} \end{bmatrix} + \begin{bmatrix} \end{bmatrix}$$

DOF：

SS：

利用公式計算

$DOF_T =$

SS_T

$DOF_A =$

SS_A

$DOF_B =$

SS_B

$DOF_C =$

SS_C

$DOF_e =$

SS_e

$S =$

變異分析表

表5.1-4 變異分析表

Factor	*SS*	*DOF*	*Var*
A	4721.67	1	
B	86.94	1	
C	0.48	1	
Error	473.67	16	29.60
Total	5282.76	19	

Var：變異數；誤差平方和除以自由度，亦即實驗誤差的________。

5.1-5 實例：瓷磚製程實驗（Tile Experiments）

表5.1-5 瓷磚製程實例的變異分析表

Factor	*SS*	*DOF*	Contribution	*Var*
A	0.1264	1	4.79%	
B	0.1642	2	6.23%	
C	0.0165	2	0.63%	
D	0.0143	2	0.54%	
E	0.1378	2	5.23%	
F	0.6005	2	22.78%	
G	0.0361	2	1.37%	
H	0.4421	2	16.77%	
Others	0.0343	2	1.30%	
Error	1.0640	108	40.36%	0.0099
Total	2.6363	125	100.00%	0.0211

此例中，以下兩式不成立，因為此例是＿＿＿＿＿＿＿＿實驗。

$$DOF_T = DOF_A + DOF_B + DOF_C + \cdots + DOF_e \qquad (5.1\text{-}10式)$$

$$SS_T = SS_A + SS_B + SS_C + \cdots + SS_e \qquad (5.1\text{-}11式)$$

那麼，該如何計算Others那一列的數值?

貢獻度（Contribution）

- 貢獻度是指每個因子的平方和（SS）所佔的比例。
- 貢獻度代表該因子（或實驗誤差）的變異造成總品質損失的比例。
- 品質損失包含兩部份：(1) 總平均值偏離目標值（9.984 – 10.000）及 (2) 每個品質特性量測值（y_i）偏離總平均值的程度如下：

- 假設存在調整因子，使得總平均值可以調整至與目標值一致，則品質損失可以表示成5.1-22式乘以____________________。
- 5.1-22式的計算結果事實上就是表 5.1-5右下角的數字（0.0211），稱為___________（total variance）。
- 因此，總變異數（乘以k後）代表可能的______________，貢獻度代表該因子的變異佔總品質損失的________。

貢獻度可以作為重要性指標

- 「貢獻度」可以作為一種簡單的指標來代表一個因子的變動所帶來的影響力，因此可以作為因子「重要性」的指標。
- 然而，就作為重要性指標而言，雖然「貢獻度」很簡單，但卻缺乏進一步的統計意義。
- 第5.2節會介紹更具統計意義的重要性測試方法，稱為 F測試法（F-test），本書大部份的時候使用 F測試法來測試因子的重要性，有時則依經驗或直覺判斷。