

變異數分析

—理論與應用—

吳宗正 編著

華泰書局

變異數分析

—理論與應用—

吳

江蘇工業學院圖書館
藏書章

華泰書局印行

版權所有 * 不准翻印

變異數分析—理論與應用

編著：吳宗正
發行者：吳茂根
發行所：華泰書局

台北市麗水街13巷2號

電話：三九三六六三三

總經銷：華泰圖書文物公司

台北市麗水街13巷2號

郵政劃撥一〇六五四六號

印刷者：瑞明彩色印刷有限公司

登記證：局版台業字第1201號

中華民國六十七年十一月十五日

定價：精裝 180元
平裝 160元

儲 序

統計在求“真”。

但什麼是真的？什麼是假的？

下面三則小故事，把它當真的亦可，當假的也無妨：

- 一、秦趙高指鹿爲馬，弄得當時的人們對鹿和馬分辨不清，不知那隻鹿是真的鹿，那隻是假的鹿，那匹馬是真的馬，那匹是假的馬。
- 二、莊國夢蝴蝶，也把張國自己弄糊塗了。畢竟自己是真的蝴蝶，在夢做莊國？還是真的張國，在夢做蝴蝶？
- 三、北平王麻子剪刀店，因剪刀好，生意賺錢；其子孫都衣鉢真傳開王麻子剪刀店；但競爭者也魚混珠的開王麻子剪刀店。於是真王麻子剪刀店看事情不妙，就在店招上加個“老王麻子剪刀店”字樣，但競爭者也來個老王麻子剪刀店；後來真王麻子剪刀店再在店招上加上“真老王麻子剪刀店”，但競爭者也同樣加上“真老王麻子剪刀店”字樣。結果都是真老王麻子剪刀店，究竟誰是真的，誰是假的，使消費者混淆不清。

下面再來四個小例子，觀察其是真？是假？也很有趣：

- 一、1899 年美國陸軍部長論駐菲律賓的美國士兵死亡率事。因當時外界對於美國駐菲律賓多數士兵的死亡稍多頗有責難，該陸軍部長乃出而置辯，大意謂駐菲士兵的死亡率不過萬分之一二七，與美國本土人口之死亡率相若，故駐菲士兵之死亡率不得謂爲過高也。

二、美國的約翰哈金斯大學初收女生時，致該校女生有百分之三十三又三分之一均與該校教師結成眷屬的事實。

三、美國某大學調查學生吸烟程度，以斷定吸烟爲學業不及格的原因，其結果如下表：

分 類	調查學生數	全年平均分數	不及格的百分數
完全不吸烟者	111	85.2	3.2
吸 烟 不 多 者	35	73.3	14.1
吸 烟 極 多 者	18	59.7	21.1

四、美國曾有反對種痘同盟職員投函紐約晚報（1914 年 5 月 4 日）引用英國的統計資料如下，意即就全國而論，種痘死者佔天花死者之半，就五歲以下而論，種痘死者尚多出天花死者之上幾有四倍之多，故強迫種痘殊可不必。

天花死亡總數（1905 ～ 1910 年） 199

種痘死亡總數（1905 ～ 1910 年） 99

五歲以下天花死亡總數（1905 年～1910 年） 26

五歲以下種痘死亡總數（1905 年～1910 年） 98

上述四個事例，是信以爲真？抑信以爲假？自從統計學術發展以來，應用統計方法，就很容易分析出各種事象真假的差異，從而判定其是真或是假；是真的，就假不了；是假的，也絕不可能混爲真的。吳講師宗正，最近編著這本變異數分析一書，就是分析並制定各種事衆真假的一塊明鏡。

儲全滋

序

變異數分析原本應用於農業實驗，醫藥研究、心理測驗等方面；近年來，由於工商業進步迅速，對於產品品質之控制、人事制度之稽核、工程變異之分析、以及產品設計之推廣方面，變異數分析亦擔負了重要的任務。尤其多年來學者專家們大力地推行及孜孜不倦地研究，使得變異數分析，在現代工商企業及學術領域中佔著重要的地位。

變異數分析與迴歸分析，實驗設計屬於線型統計學之領域。三者間之關係極為密切。本書之特色即將迴歸分析之理論及應用延伸擴展至變異數分析，使得兩學科間能相輔相成，尤其兩者之間的相互關係及轉換應用，除了以理論方式闡述外，另舉實例配合，以使讀者易於吸收及瞭解。

本書共分八章，以一因子、二因子及三因子變異數分析為研究之對象，另外再闡因子效果之檢定及推定。樣本次數之規劃、資料與模式之配合方法，剩餘值檢定、以及迴歸研究等諸章節，以使全書連貫一氣，便於分析及研究，除了本文外，另於每章後加附習題以供練習。

本書之另一特色是於每章最後一節中，增闢問題解析，以使該章所提之理論及分析過程，在實例解析中予以詳細研討，以期收到理論與實務之最佳配合。

本書之編著承蒙成功大學統計系主任儲師全滋及工業管理研究所劉師漢容教授之多方鼓勵，以及系內諸位老師之多方指導，始能如期完成，在此表示衷心之謝意。又本書承李淑麗助教及、張敏詩、蔡淑敏、郭香蘭、黃女珍、黃子珍、蔡慧蘭、李燕孟諸位同學之抄錄及協

11/16/00/02

助，統此致謝，而本書之出版承蒙華泰圖書公司吳茂根先生之多方奔走及熱心協助，在此一併致謝。

著者學識淺短，且時間匆促，誤錯之處在所難免，尙祈斯學先進不吝賜教。

中華民國六十七年八月

吳宗正 謹識

目 錄

第一章 一因子變異數分析

1-1	意義	1
1-2	實驗之隨機化	2
1-3	變異數分析之觀念	4
1-4	迴歸與變異數分析之關係	8
1-5	變異數分析名詞之解釋	11
1-6	變異數模式分析之用途	13
1-7	固定效果模式	14
1-8	參數之估計	21
1-9	變異數分析表之製作	26
1-10	不偏變異數之期望值	36
1-11	因子水準平均數相等性之 F 檢定	44
1-12	t 之顯著差檢定	49
1-13	一般線型之 F 檢定	54
1-14	同一母群體多樣本之分析	56
1-15	多母群體抽樣之分析	63
1-16	F 檢定力	66
1-17	計算機輸入與輸出	71
1-18	問題之解析	74

第二章 因子效果之分析

2-1	意義.....	89
2-2	組間平方和之分解.....	90
2-3	因子效果之估計——平均數之推定.....	100
2-4	多重比較程序之意義.....	111
2-5	多重比較之 Tukey 方法——Q 值多重比較法.....	112
2-6	Tukey 考驗法之表列檢定法.....	121
2-7	多重比較之 Keuls 考驗法——序列考驗法.....	122
2-8	Keuls 考驗法之簡捷計算.....	123
2-9	多重比較之 Scheff'e 方法.....	130
2-10	多重比較之 Bonferroni 方法.....	134
2-11	Tukey, Scheff'e 及 Bonferroni 三種方法之比較.....	140
2-12	屬量因子之因子效果分析.....	140
2-13	迴歸分析之函數檢定及新觀察值之預測.....	147
2-14	問題之解析.....	151

第三章 變異數分析模式之應用

3-1	樣本大小之規劃——依檢定力方法.....	165
3-2	樣本大小之規劃——依估計方法.....	174
3-3	設定最佳處理以決定所需樣本之規劃.....	176
3-4	剩餘值之意義及性質.....	178
3-5	剩餘值之分析.....	179
3-6	轉換方法.....	186
3-7	變異數相等性之檢定.....	189
3-8	模式偏離之影響.....	198

第四章 變異數分析之探討——因子

4-1	t 化全距檢定法.....	205
-----	---------------	-----

4-2	Kruskal- Wallis	207
4-3	中位數檢定法.....	213
4-4	三種方法之評註.....	216
4-5	隨機效果之變異數分析模式.....	217
4-6	一因子變異數分析之迴歸研究.....	239
4-7	因子水準平均數之估計.....	249
4-8	缺失觀察值之估計.....	251
4-9	問題之解析.....	255

第五章 二因子變異數分析

5-1	二因子變異數分析之意義.....	267
5-2	實驗之隨機化.....	270
5-3	模式元素之涵意.....	273
5-4	二因子之固定效果模式.....	290
5-5	變異數分析及表之製作.....	293
5-6	不偏變異數之期望值.....	304
5-7	二因子變異數分析之 F 檢定法.....	312
5-8	t 之顯著差檢定.....	317
5-9	計算機之輸入與輸出.....	319
5-10	F 檢定力.....	322
5-11	問題之解析.....	323
	(i) 無交互作用之二因子變異數分析.....	324
	(ii) 有交互作用之二因子變異數分析.....	330

第六章 二因子變異數分析之研究

6-1	二因子皆為屬質之變異數分析.....	341
-----	--------------------	-----

6-2	一因子或二因子爲屬量之變異數分析.....	353
6-3	二因子變異數分析模式之應用.....	361
6-4	問題之解析.....	364

第七章 變異數分析之探討—二因子

7-1	每個處理包含一個觀察值 ($n = 1$) 之分析.....	377
7-2	樣本次數不相等之情況.....	386
7-3	二因子變異數分析之隨機效果模式與混合效果模式.....	392
7-4	二因子變異數分析之迴歸研究.....	407
7-5	缺失觀察值之估計.....	423
7-6	問題之解析.....	429

第八章 多因子變異數分析

8-1	三因子變異數分析之固定效果模式.....	439
8-2	變異數分析及表之製作.....	454
8-3	因子效果之檢定.....	466
8-4	因子效果之分析.....	469
8-5	三因子變異數分析之實例解析.....	472
8-6	三因子變異數分析之隨機效果模式與混合效果模式.....	481

第一章 一因子變異數分析

1.1 意義

在統計檢定中，有關一個常態母全體平均數之檢定（ $H_0: \mu \cong \mu_0$ ， $H_1: \mu \not\cong \mu_0$ ）可因 σ 已知之小樣本或 σ 未知之大樣本時以常態分配檢定，而 σ 未知且為小樣本時以 t 分配檢定。而二個常態母全體之平均數差之檢定（ $H_0: \mu_1 \cong \mu_2$ ， $H_1: \mu_1 \not\cong \mu_2$ ），同樣可因 σ^2 已知而 $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 之小樣本或 σ^2 已知而 $\sigma_1^2 \neq \sigma_2^2$ 及大樣本時使用常態分配檢定，或 σ^2 未知，而 $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 小樣本時以大分配檢定，但是對於兩組以上常態母全體平均數間之檢定，則無法以上述方法求解，因此，在1920年間R. A. Fisher.（費雪）創造檢定多組因子水準平均數之方法，稱為變異數分析法。

由於單一樣本應用於統計實驗者常可見於一般統計書籍上，但是實驗不能只以單一樣本為主，在實際應用領域中，樣本數通常隨因子數之增加而需要擴展。一般在計數資料中由於 X^2 之相加性，樣本數可以任意增加，但是在計量資料中，二組樣本尚適合於 t 值檢定，而兩組以上之因子水準之平均數的處理，唯有賴“變異數分析”之方法以解決。

變異數分析之基本來源基於總平方和（或總變異）及自由度之分割，即資料之收集會產生各種變異，這些變異之總和即總變異，根據變異之加成性分解成組間（已知各原因之變異）和誤差變異（或稱組內變異），而將組間變異與誤差變異相除求得 F 之比值，再與 F 之臨界值比較檢定變異是否顯著。通常一因子變異數分析之總變異可以以

下式說明：

$$\begin{aligned} \text{即} \quad \text{總平方和 (總變異)} &= \text{組間變異 (或已知各原因之變異)} \\ &+ \text{誤差變異 (組內變異)} \end{aligned}$$

由於費雪教授之發現變異數分析法，目前已為實驗學者廣為採用。由於分析方法之繼續發展，遂使統計學術更為精進，舉凡醫藥檢定、教育實驗、農業推廣等應用皆以此方法做為研究之依據。此統計方法已為實驗統計 (Experimental Design) 之基礎，因此如何從事變異數之分析以及選擇一適當的統計方法來對資料做一合理之探討將為本文之重要課題。

1.2 實驗之隨機化

【例 1】 在飼料對豬仔之成長上，欲知使用 4 種飼料 B_1 、 B_2 、 B_3 、 B_4 ，試問此 4 種飼料對於豬仔之重量有無影響。

本例即為因子水準間平均數差異之檢定，屬於一因子變異數分析，只要對豬仔之平均重量加以檢定即可以瞭解飼料對豬仔之成長有無影響。

【例 2】 研究廣告製作是否對銷售量有影響，若某公司採用 3 種廣告設計 B_1 、 B_2 、 B_3 ，試問此 3 種設計對銷售量有無影響。

本例為一因子變異數分析，只要對因子水準間平均數差異加以檢定，即可瞭解廣告設計對銷售量有無影響。

通常做因子變異數分析時，須對實驗因子以外之條件加以固定或對實驗因子採隨機化實驗之，因使用實驗順序之隨機化，可以消除實驗順序與因子交絡。

【例 3】 以例 1 之豬仔成長實驗說明隨機化之使用，今 4 種飼料水準，每個水準採用 3 個重覆數實驗，亦即共有 12 個重覆數，其方法可以用(1)亂數表及(2)卡片法。

(1) 應用隨機亂數表：首先從第一列開始每隔 2 行選取一組號碼，因為有 12 個重覆數，因此自變數中各別減去 12 之倍數使成為 12 以下之數，將其依次分配於 B_1 、 B_2 、 B_3 、 B_4 之上，則可以得隨機化實驗之順序，其求法如下：

$47 - 36 = 11$ A_1	$11 - 0 = 11$ 重覆	$7 - 0 = 7$ A_4
$73 - 72 = 1$ A_1	$10 - 0 = 10$ 重覆	$24 - 12 = 12$ 重覆
$36 - 24 = 12$ A_1	$97 - 96 = 1$ A_3	$79 - 72 = 7$ 重覆
$47 - 36 = 11$ 重覆	$24 - 12 = 12$ 重覆	$73 - 72 = 1$ 重覆
$61 - 60 = 1$ 重覆	$62 - 60 = 2$ 重覆	$76 - 72 = 4$ 重覆
$98 - 96 = 2$ A_2	$81 - 72 = 9$ 重覆	$27 - 24 = 3$ A_4
$71 - 60 = 11$ 重覆	$67 - 68 = 9$ 重覆	
$33 - 24 = 9$ A_2	$42 - 36 = 6$ A_3	
$16 - 12 = 4$ A_2	$32 - 24 = 8$ A_3	
$45 - 36 = 9$ 重覆	$32 - 24 = 8$ 重覆	

上列只剩下最後一號碼 5，即分配於 A_4 ，茲將實驗隨機化列成表如下：

表 1.1 隨機實驗之順序

B_1	B_2	B_3	B_4
11	2	10	7
1	9	6	3
12	4	8	5

(2) 應用卡片法：即將一副牌挑一色之 12 張，依照大小順序（即 A 、 2 、 3 ………、 10 、 J 、 Q ），然後事先規定 A 、 2 、 3 為 B_1 ， 4 、

5、6爲 B_2 ，7、8、9 B_3 ，10、J、Q 爲 B_4 ，然後將其混合，以隨機決定，依順序出現之先後而予以實驗。

1.3 變異數分析之觀念

通常在一定條件下對搜集而得之資料依分組而加以分析時，難免會發生差異，此差異顯示在一定條件下有分配存在，亦即有群體存在，變異數分析即在對此群體平均數做差異之檢定。一般變異數之模式分爲固定效果模式及隨機效果模式兩項，而前者以研究各因子水準間平均數差異爲主，故可以用 F 分配或 t 分配檢定之，後者之隨機效果模式以研究因子水準平均數間變異爲主，可以用 F 分配檢定。若欲在因子水準平均數間差異顯著下研究成對比較 ($u_i - u_j$) 或對比比較

$$\left(\frac{u_{.1} + u_{.2}}{2} - \frac{u_{.3} + u_{.4}}{2} \right) \text{ 爲主時，可以應用群組多重比較之方式}$$

予以推定其區間之可靠範圍。一般對多組平均數間檢定其差異性，可以編製變異數分析表以求出已知變異原因之均方和與誤差變異之均方和，然後以比值之方式得 F 統計量再與 F 分配的臨界值比較，以決定其是否差異顯著之方法即爲變異數分析。同時也可以對因子水準之單一平均數或平均數差異及多組平均數差異（對比）求推定。若因子 B 之各水準母平均數設爲 $u_{.1}$ 、 $u_{.2}$ 、 $u_{.4}$ …………… $u_{.r}$ ，根據變異數分析模式之假設其誤差是服從常態分配 $N(O, \sigma^2)$ 具有一致之變異數 σ^2 ，因此對於這批資料可以得到下列分析之程序：

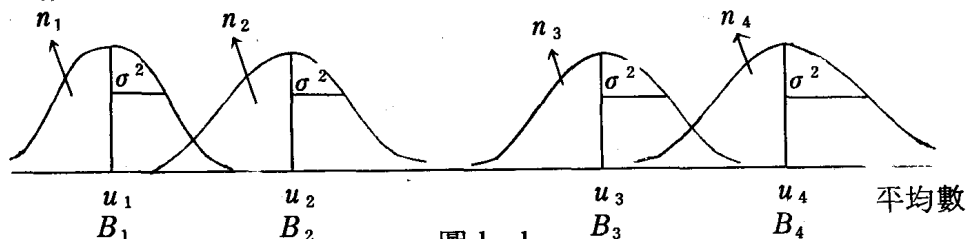


圖 1.1

(1)各因子水準 B_1 、 B_2 、 B_3 、 B_4 之母平均數是否有顯著差異

$$\text{即} \quad \begin{cases} H_0 : u_{.1} = u_{.2} = u_{.3} = u_{.4} \\ H_1 : u_{.i} \text{ 不相等} \end{cases}$$

可依變異數分析之 F 分配檢定之，或類似的 t 化全鉅檢定法，Kruskal-wallis 等級檢定法及中位數檢定法。

(2)若依 F 分配檢定結果具有顯著性，可進一步地分析兩母子水準平均數間是否有顯著差異。

$$\text{即} \quad \begin{cases} H_0 : u_{.j} = u_{.j'} \\ H_1 : u_{.j} \neq u_{.j'} \end{cases}$$

可應用 t 顯著檢定法檢定之（固定效果模式）。

(3)若依 F 分配檢定結果具有顯著性，可進一步應用群組多重估計係數檢定多組間平均數互差之考驗即可以應用 Q 值考驗法，或哈特列式考驗法及 Tukey，Scheffe' 和 Bonferroni 區間檢定法檢定之（固定效果模式）。

(4)若因子為屬量時，可以做對應函數（迴歸係數）之估計檢定，若迴歸函數配合資料設為二次式，

$$\text{即} \quad \begin{cases} H_0 = E(Y) = B_0 + B_1 X + B_{11} X^2 \\ H_1 = E(Y) \neq B_0 + B_1 X + B_{11} X^2 \end{cases}$$

可以明瞭資料與模式之配合是否為二次式，或以連結分配之 $U_{.1}$ 、 $U_{.2}$ 、……、 $U_{.r}$ 以判定是否為直線或曲線（固定效果模式）。

(5)若因子為屬量時，可以依據設定之迴歸模式以從獨立變數（ X ）估計相依變數（ Y ）。（固定效果模式）

(6)各因子水準單一母平均數之推定；即（固定效果模式）

$$\bar{Y}_{.j} - T(1 - \frac{\alpha}{2}; n_T - r) S(\bar{Y}_{.j}) \leq u_{.j} \leq \bar{Y}_{.j} + T(1$$

$$-\frac{\alpha}{2}; n_T - r) S(\bar{Y}_{.j})$$

(7)兩因子水準母平均數間差異之推定，即（固定效果模式）

$$(\bar{Y}_{.j} - \bar{Y}_{.j'}) - t(1 - \frac{\alpha}{2}, n_T - r) S(\bar{Y}_{.j} - \bar{Y}_{.j'}) \leq (u_j - u_{j'}) \leq (\bar{Y}_{.j} - \bar{Y}_{.j'}) + t(1 - \frac{\alpha}{2}; n_T - r) S(\bar{Y}_{.j} - \bar{Y}_{.j'})$$

(8)因子水準平均時間對比之估計

即：若定義

$$L = \sum_{j=1}^r C_j u_{.j}$$

$$\hat{L} = \sum_{j=1}^r C_j \bar{Y}_{.j}$$

其限制式為 $\sum_{j=1}^r C_j = 0$

則 $S^2(L) = MSE \sum_{j=1}^r \frac{C_j^2}{n_j}$

其對此區估計為：（如 $L = u_{.1} - u_{.2}$ ， $L = \frac{u_1 + u_2}{2} - \frac{u_3 + u_4}{2}$

或 $L = \frac{u_1 + u_3}{2} - \frac{u_2 + u_4}{2}$ ）

$$\hat{L} - t(1 - \frac{\alpha}{2}, n_T - r) S(\hat{L}) \leq L \leq \hat{L} + t(1 - \frac{\alpha}{2}, n_T - r) S(\hat{L})$$

(9)在群組係數 $(1 - \alpha)$ 下，求成對比較 $(u_j - u_{j'})$ 及對比