

／ 大學用書 ／

變異數分析

呂金河博士 著

三民書局 印行

大學用書

變異數分析

呂金河 著

三民書局 印行

國家圖書館出版品預行編目資料

變異數分析 / 呂金河著. -- 初版四刷. -- 臺北市：
三民，2005
面；公分
ISBN 957-14-1903-6 (平裝)

1. 變異數分析

511.2

81006554

網路書店位址 <http://www.sanmin.com.tw>

◎ 變異數分析

著作人 呂金河
發行人 劉振強
發行所 三民書局股份有限公司
地址 / 臺北市復興北路386號
電話 / (02)25006600
郵撥 / 0009998-5
印刷所 三民書局股份有限公司
門市部 復北店 / 臺北市復興北路386號
重南店 / 臺北市重慶南路一段61號
初版一刷 1993年2月
初版四刷 2005年10月
編號 S 510070
基本定價 捌元貳角
行政院新聞局登記證局版臺業字第〇二〇〇號

有著作權 不准侵害

ISBN 957-14-1903-6 (平裝)

序

變異數分析是研究我們關心的事物是否與某些重要變因存在因果關係的統計方法。在各種變因中，來探求那個變因在什麼情況下對結果的影響最大，最顯著，是變異數分析的主題。本書的主要目的，正在介紹這個重要的資料分析方法，以推廣應用至各種領域。

本書的編寫約有以下幾個特點：

1. 爲了便於說明，書中第二章複習了一些基本統計推論的觀念，並視之爲較簡單的比較實驗。用此來引出，第三章以後較複雜的比較實驗。

2. 書中介紹單因子，雙因子及三因子變異數分析的方法及其基本觀念。文共八章，涵蓋內容應已足夠讓讀者對變異數分析有了整體的概念。

3. 本書假設主要讀者是專校學生，因此書中例題頗多，期使每一觀念及方法都有例題加以說明，以利瞭解。例題中分析計算的過程亦盡量詳細並規則化，以便讀者易於依循並應用。

本書的完成，要感謝三民書局劉發行人的熱心支持。還要感謝我家人不斷的鼓勵與策勉。著者才疏學淺，編校疏忽，誤謬之處，敬請學者先進不吝指正。

呂金河

民國八十一年十二月

變異數分析 目次

序

第一章 緒論

1-1 引言	1
1-2 基本名詞	3
習題一	5

第二章 簡單的比較實驗

2-1 緒言	7
2-2 基本統計概念	8
2-2-1 機率分配	8
2-2-2 均值、變異數及期望值	10
2-3 抽樣分配	12
2-3-1 隨機樣本、樣本均值及樣本變異數	12
2-3-2 估計式及其特性	13
2-3-3 抽樣分配	13

2-4 兩均值差的統計推論	18
2-4-1 信賴區間	19
2-4-2 假設檢定	23
2-5 配對資料均值差的統計推論	27
2-6 常態母體變異數的統計推論	29
習題二	31

第三章 單因子變異數分析

3-1 變異數分析模式	35
3-2 參數的估計	41
3-3 變異數分析表	44
習題三	57

第四章 因子水準效果的分析

4-1 圖示法	65
4-2 因子水準效果的估計	72
4-2-1 因子水準均值的估計	73
4-2-2 兩因子水準均值差的估計	76
4-2-3 因子水準均值對比的估計	78
4-3 聯合信賴區間	82
4-3-1 Bonferroni 法	82
4-3-2 Tukey 法	85
4-3-3 Scheffé 法	88
4-4 單一自由度的檢定	91

4-4-1 對比的檢定	91
4-4-2 最小顯著差法	94
4-5 多重比較	96
4-5-1 Duncan 法	97
4-5-2 Tukey 法	100
4-5-3 Scheffé 法	103
4-6 隨機效果模式	106
4-6-1 μ 的估計	112
4-6-2 變異數分量的估計	113
習題四	117

第五章 單因子變異數分析的其他問題

5-1 模式診斷	121
5-1-1 常態假設的驗證	122
5-1-2 對時間的殘差圖	126
5-1-3 對估計值 $\hat{Y}_{ij}$ 的殘差圖	127
5-1-4 使變異數穩定的變數轉換	137
5-1-5 對其他變數的殘差圖	146
5-2 樣本大小的決定	146
5-2-1 作業特性曲線	147
5-2-2 誤差界限	154
5-3 反應曲線	156
5-4 無母數的變異數分析法	160
5-4-1 Kruskal-Wallis 檢定法	160

5-4-2 秩數轉換	164
5-4-3 多重成對比較	166
5-4-4 中位數檢定法	168
習題五	172

第六章 二因子變異數分析

6-1 簡單概念	175
6-2 二因子變異數分析模式	183
6-3 參數的估計與變異數分析	188
6-4 多重比較	220
6-4-1 交互作用不顯著	220
6-4-2 交互作用顯著	240
6-5 模式診斷	251
6-6 樣本大小的決定	260
習題六	264

第七章 二因子變異數分析的其他問題

7-1 沒有交互作用的二因子模式	273
7-2 沒有反覆的二因子模式	281
7-3 隨機效果模式	296
7-4 混合效果模式	304
7-5 樣本大小的決定	312
7-6 另一種混合效果模式	314
7-7 反應曲面	317

7-8 樣本大小不等的問題	320
7-8-1 比例資料	320
7-8-2 近似方法	326
7-8-3 迴歸模式法	337
習題七	348

第八章 多因子變異數分析

8-1 三因子固定效果模式的分析	357
8-2 因子效果的比較	371
8-3 三因子隨機效果模式和混合效果模式	383
習題八	393

附 錄

表 I. 標準常態分配的累加機率表	400
表 II. t 分配的分位數表	401
表 III. 卡方分配的分位數表	402
表 IV. F 分配的分位數表	403
表 V. 固定效果模式的作業特性曲線	409
表 VI. 隨機效果模式的作業特性曲線	413
表 VII. Duncan 多重全距檢定的顯著全距表	417
表 VIII. 學生全距統計量的分位數表	419
表 IX. 亂數表	421

第一章

緒論

1-1 引言

變異數分析是一個很重要的統計方法。它和迴歸分析類似，都是用來研究**自變數**(Independent variable)與**因變數**(Dependent variable)間的關係。兩者的因變數常取為**屬量變數**(Quantitative variable)。而迴歸分析和變異數分析的主要不同在，前者的自變數常取為屬量變數，而且需事先假設自變數與因變數的關係為直線或曲線等函數。變異數分析則無上二項條件。因此，相對而言，其應用範圍更廣更大，而成為資料分析時，不可或缺的工具。

例 1 某機構想研究年所得(X)與家庭衣著消費額(Y)間的關係。若取 $X = \$50, \$60, \$70$ 萬元，三種不同年所得的家庭各幾家，詢問其家庭年衣著消費額。資料圖示如下。若假設 X, Y 的關係呈一曲線，如圖(a)，則可利用迴歸分析，估計下圖(a)曲線的方程式。

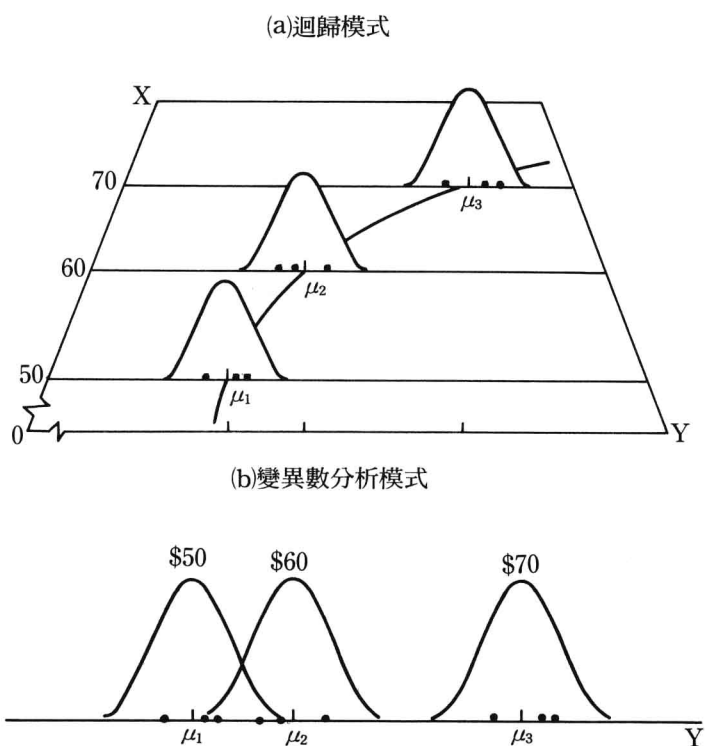


圖 1.1 迴歸與變異數分析圖

若只想知道所得不同是否確實消費額不同，則可用變異數分析，如上圖(b)，探討三種不同所得的各個家庭衣著消費平均消費額 μ_1 , μ_2 , μ_3 是否差異顯著。

在實務上，若自變數是屬量變數，而分析者需說明自變數與因變數的關係型態時，選用迴歸分析。若只需證實自變數是否真的影響因變數及什麼情況下影響最大，則採用變異數分析。或者，先用變異數分析證

實自變數確實影響因變數後，再用迴歸分析探討變數間的關係函數。兩種統計方法若能相輔相成，自然最為完美。

1-2 基本名詞

變異數分析中，一些常用的術語分別說明於下：

1. **反應值 (Response) 或因變數**：研究目的所要測量的目標值。其量測值多屬連續性資料。
2. **因子 (Factors) 或自變數**：會影響反應值的變數。這些因子是否真的對反應值有影響效果 (Effect) 是變異數分析的主要目的。因子可能是屬量變數，可能是屬性變數，但都以分類資料的形式表示。
3. **水準 (Levels)**：因子以分類資料表示時的各種不同值，稱為該因子的水準。

例如上述年所得與家庭衣著消費額的問題。消費額的數量是反應值，為連續量。年所得是因子，為屬量變數。但只取\$50, \$60, \$70 萬元三個水準，以這三種分類來研究問題。又如研究四個電視節目的收視率是否大小不同。收視率為反應值，電視節目為因子，其為屬性變數，取四個節目共四個水準來研究問題。

4. **處理 (Treatments)**：各種因子水準的組合稱為處理。

例如上述年所得與衣著消費額的問題。若只考慮年所得為因子，則\$50, \$60, \$70 萬元三個水準的組合仍只三種，即對應三個處理。因此，單一因子時，處理和水準意義相同。但影響家庭衣著消費額的因子，也許不只年所得一項。個人的消費習慣可以是因子二，若分成奢侈、中庸、節儉三個水準。家庭中是否有小孩可以是因子三，若分成有、無小孩二個水準，則總共三個因子，水準分別為三、三、二的水準組合，共有 $3 \times 3 \times 2 = 18$ 種處理。

5. 實驗單位 (Experimental units)：能接受各種不同處理的標的物，在其上我們要量測反應值。實驗單位要力求其條件，背景因素一致(Uniform)。

例如上述衣著消費額的問題，若考慮三個因子，18 種處理，則資料搜集時，至少要找 18 個家庭以對應 18 種處理的情況。家庭即為實驗單位。這 18 個家庭除了對應三種不同因子的 18 種不同處理外，其他條件和背景因素要力求一致。比如，各家庭人數應相同，18 個家庭的社區物價水準應相同，18 個家庭地區氣候條件相似等。在其他條件因素一致的情況下，才能真正凸顯因子不同所產生的影響效果。

以上五個專有名詞非常基本。欲了解一個變異數分析的問題，都應該先將上述五個名詞在問題中的意義指明，才能正確的分析題目。

6. 實驗計劃 (Experimental design) 與 觀察研究 (Observational study)：變異數分析的資料如何搜集？主要有二種方法，一為實驗計劃，一為觀察研究所得。研究者控制或設定各種因子處理，並依隨機化等方法，將各種處理指派到各實驗單位。透過這種有計劃的控制來進行實驗，如此所得資料，可以消除特定因子以外的其他因素對反應值的影響，是較理想的統計資料。若影響反應值的因子不加以控制，或無隨機化的搜集資料，則稱為觀察研究。這類資料的反應值，可能受其他外在因素的影響，而使變異數分析的統計結果可信度降低。

例如隨機選 20 家商店，隨機指派 2 種不同商品陳列方式，以研究陳列方式對商品銷售量的影響。100 個顧客，隨機讓他們試吃三種不同廠牌的冷凍水餃，以研究顧客對三種冷凍水餃的喜愛程度是否不同。這是屬於實驗計劃。而實際上，大部分在商業、經濟、社會科學裏，我們所看到的統計資料都屬觀察研究。例如調查 100 個有肺病的病人，詢問他們是否吸煙，用此資料判斷吸煙是否和患肺病有關係。這種觀察研究的資料較易取得。但是吸煙的人可能工作壓力較大，心情不開朗或生活不規

律。因此患肺病可能不是因為吸烟，而是因為生活不規律所致。如果改以實驗計劃，找一些身體健康狀況相似的 20 人，隨機分成二組，一組強迫吸烟，一組禁止吸烟，過了 20 年後，看看他們患肺病的人數。如此斷定吸烟是否和肺病有關係，所得結論應該比上述觀察研究所下結論來得更明確且有說服力。因此，如有可能，應多以實驗計劃來搜集資料。但事實上，實驗計劃有時不易進行，如上述吸烟計劃，有人願意被強迫吸烟嗎？所以大部分統計資料，如工廠製程產生的數據、公司管理產生的會計、產量、成本等數據、消費者行為、社會調查意見、經濟指標、儲蓄消費等數據都是觀察研究的資料。

習 題

- 1-1 試分別定義反應值、因子、水準、處理、實驗單位。
- 1-2 某品質工程師想研究纖維內棉含量的百分比是否對纖維強度產生顯著影響效果，於是在公司各種產品中，隨機抽取棉含量 15%，20%，25%，30% 的纖維各 10 段分別測度其張力強度，問此實驗所顯示的反應值、因子、水準、處理、實驗單位各指什麼？
- 1-3 某農場主人想比較不同肥料對玉米收穫量的影響。他取鉀、鈉、氨三種肥料各 1 公斤，分別在每平方公尺大的試驗田 12 塊，各分別施以三種肥料各 4 塊。收穫後，測量各試驗田所收玉米重量以做變異數分析，問此實驗所指的反應值、因子、水準、處理、實驗單位各是什麼？
- 1-4 某冰淇淋公司推出一新產品，分別以紅、藍、黃三色包裝，市場部門記錄了三種包裝下，各週的冰品銷售量，三個月後，以此資料做變異數分析以比較三種包裝法是否對銷售量產生顯著影響。問此問題所指的反應值、因子、水準、處理各是什麼？

第二章

簡單的比較實驗

本章將討論比較二種不同處理的簡單實驗。例如，為比較二種生產方法是否有相同產能，而進行某種觀察或實驗。如此搜集所得的資料如何進行統計分析？此類說明，順使用以複習一些諸如機率分配、期望值、變異數、隨機樣本、估計與檢定等基本統計觀念。

2-1 緒言

某便利商店想知道 A, B 兩牌的香烟銷售量是否相同。於是老板隨機選擇五週做記錄，而得下列資料。

週別	A 牌	B 牌
1	38	42
2	37	44
3	35	40
4	40	45
5	39	44