多變量解析與其應用

曾國雄編著

華泰書局印行

多變量解析與其應用

曾國雄　編著

華泰書局印行

多變量解析與其應用

編　　著：曾　國　雄
發行者：吳　茂　根
發行所：華　泰　書　局
台北市金華街181～1號
電話：3936633・3416633
總經銷：華泰圖書文物公司
台北市金華街181～1號
電話：3936633・3416633
印刷者：瑞明彩色印刷有限公司
電話：3088928
登記證：局版台業字第1201號
中華民國七十一年九月一日
定價 230

序　　言

多變量解析是近年來新發展出的一門用於解析各種社會現象的數量方法，且目前已成爲現代市場、心理、社會、土木、建築、都市、交通、電機、醫學、地理、政治、農業、氣象、生物等所有學科資料解析的重要工具。但此學問在台灣目前尚無一本具備完整體系，可供國內研究者或實務家參考的中文書籍。因此，本書之目的在期待國內諸位研究者或實務家等讀者，能於讀後更廣泛，普遍地推廣應用於國內各行業界。

本書爲筆者於去年歸國任教於大同工學院事業經營研究所與國立交通大學運輸工程研究所時，分別講授「企業數量方法」與「都市數量方法」課程的講義之一編印而成。

筆者於留學日本國立大阪大學經濟學院博士課程時，蒙指導教授橫山保教授（經營統計學講座），與鈴木胖教授（系統工程講座），丘本正教授（數理講座）、毛利正光（都市交通講座），關田康慶助手，以及系統科學研究所的辻武夫先生，藤井英彥先生等對本方法之理論與實際應用上惠予啓示，至深感謝。

本講義編撰之初，得大同工學院事業經營研究所一年級研究生楊彰元君等之參與修正與協助校對，玆於本書付梓時，深致謝意。惟筆者才疏學淺，加以工作繁忙，時間匆促，本書疏陋錯誤之處，在所難免，尚祈諸先進賜予指正，推廣是幸。

國立交通大學運輸工程研究所

曾　國　雄　　謹啓

民國六十七年六月十五日於

台北市南京東路 4 段 75 巷 97 弄 50 號 2 樓

電話:（公）9626152-5　（宅）7521833

多變量解析與其應用 目 次

第一章　多變量解析之概況

1.1 多變量解析之意義

多變量解析爲由二種或二種以上的變量資料，利用多元空間之統計與線性代數等之方法，將複雜問題或現象抽象化、數量化後，對該問題或現象做合理而有系統之整理、分類、判斷、特性說明、評價，預測等爲目的之一種分析方法。

然問題之分析並不能僅考慮部分因素，而必須將所有有關之因素一併合之分析。如舉某個人之「體力」而言，與「體力」有關之因素有：性別、年齡、身高、體重、胸圍、跳躍力、腹筋力，…等之變量。而「體力」並非僅由性別可以決定的，也並非僅由體重可以決定的。必須同時考慮到性別、年齡、身高、體重等所有的因素，才能決定一個人的「體力」。這問題當「說明變量」不多時，處理上是沒有問題，但是像目前複雜的社會裡，情報（變量）經常有幾十個甚致幾百個以上，在此複雜的情況下，如僅用以前傳統的簡單統計方法，是無法解決的，必須藉用電腦及一些新的分析技術，才能解決此複雜的問題。此即爲本課程所探討的重心。

1.2 多變量解析之種類與其適用範圍

統計學上之多變量解析法，爲 1928 年由 J．Wishart 教授發現 Wishart 之分配以來，可以說是個劃時代的進步。其後，1958 年又由 T.W. Anderson 教授將多變量解析法成一有系統地整理，至今

還在發展當中，現在由於快速、精確，且可處理繁雜問題的電子計算機之使用後，本法已可以廣泛地被應用於市場分析、心理分析、社會分析、都市分析、土木、建築、交通、電機、醫學、地理、政治、農業氣象、生物等有關學科。因此，本書之目的有待讀者今後能更廣泛，普遍地推廣應用於各行各界。

多變量解析依其目的之不同，可如表 1.1 之分類，而各類分析方法之關係如圖1.1,分析體系如表1.2,又在經營學上常適用之範圍例可歸納如表 1.3，

表 1.1 分析方法之種類

型態	主要目的	分析方法	
		量的資料	質的資料
1 型	爲預測與建立關係式	多元回歸分析、正準相關分析	數量化Ⅰ類
2 型	各變量之綜合與分類	主成分分析、因子分析	數量化Ⅲ類
3 型	樣本之分類	判別分析	數量化Ⅱ類
4 型	檢定	變異數分析	

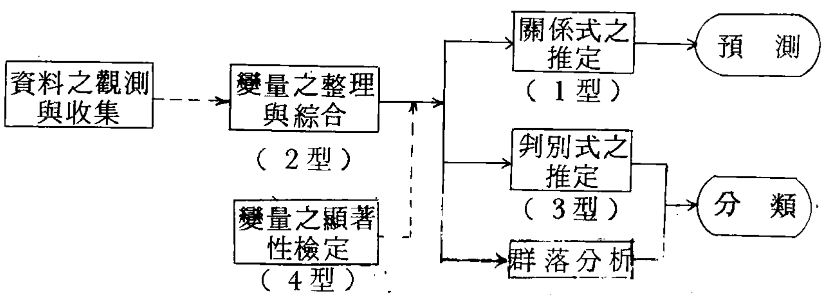

圖 1.1　資料分析之流程圖

表 1·2　多變量解析法之分析體系

外部基準	問　　題	要　　因	重　要　方　法
有	量的推定	量的 質的 and ／ or 量的	多元回歸分析 多元相關分析 數量化理論（第Ⅰ類）
	質的分類	量的 質的 and ／ or 量的	判別函數 數量化理論（第Ⅱ類）
無	質的分類	量的 質的 and ／ or 量的	因子分析，主成分分析 數量化理論（第Ⅲ類） 數量化理論（第Ⅳ類）

表 1.3　多變量解析之適用範圍例

分析方法＼適用範圍	一般：過去之結構分析	一般：經濟預測	一般：政策效果分析	一般：景氣動向指數	預測：需要預測	預測：銷售預測	預測：利潤預測	預測：事務量預測	預測：人員預測	管理：生產管理	管理：品質管理	管理：材料管理	管理：設備管理	管理：成本管理	管理：銷售管理	管理：資金管理	管理：預算管理	管理：人事管理	管理：情報管理	計劃：生產計劃	計劃：銷售計劃	計劃：資金計劃	計劃：利潤計劃	計劃：人事計劃	計劃：經營計劃
多元回歸分析、數量化Ⅰ類	○	○	○		○	○	○	○	○	○		○	○		○	○	○			○	○	○	○	○	○
主成分分析、因子分析		○		○	○	○					○						○	○	○		○			○	○
判別分析、數量化Ⅱ類				○		○						○			○	○					○	○			
正準相關分析														○	○								○		

○………表示特別重要的適用範圍

1.3 適用質料之形態

多變量解析適用資料之形態可分成「量的資料」與「質的資料」兩類（如表 1.1 ）。實際之例即如，

量的資料……如「所得」，「人口」，「道路輔裝率」、「交通量」，…，等的數值資料。

質的資料……如「性別」（男、女）、「職業」（管理職、事務職…）、「滿意度」（非常滿意、滿意、馬馬虎虎、不滿意，非常不滿意），…，等的數值。

量的資料

樣本	年所得（萬元）
1	12
2	15
3	18
⋮	⋮

質的資料

樣本	職業			
	管理職	事務職	技工職	其他
1	0	1	0	0
2	0	0	1	0
3	1	0	0	0
⋮				

1 …是

0 …非

第二章 多元回歸分析

(*Multiple Regression Analysis*)

2.1 多元回歸分析之內含

2.1.1 分析目的

本章分析之目的，爲建立某一變量（目的變量），與影響此一變量之數個變量（說明變量）間之（線性）函數關係，由此，可以依據此一函數關係以測出各說明變量對「目的變量」之影響強度，又若已知各「說明變量」之數值時，即可預測所需「目的變量」之大小。

2.1.2 適用例

若有某一清涼飲料公司經檢討之結果，該公司產品之銷售總額 Y（目的變量），與廣告費（X_2）以及零售商店數（X_3）有關。根據過去 6 年來的實績，欲建立各變量間之（函數）關係。若預定下個年

表 2.1

變量 / 樣本	銷售總額（Y）	廣告費（X_2）	零售商店數（X_3）
1（年次）	8 百萬元	0.5 萬元	6 家
2	9	0.5	8
3	12	0.7	10
4	11	0.5	12
5	13	0.8	12
6	17	1.2	12
7（下年度）	?	(1.7)	(14)

度之零售商店數爲 14 家，廣告費用爲 1.7 萬元時，如何預測下年度可能之銷售總額呢？

2.2 模 型

今若有某一目的變量 Y，受（P－1）種說明變量 X_2，X_3，…，X_P 之影響。此時，設下列之線性關係式成立，

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_P X_P + e \qquad (2.1)$$

此 e 爲測定 Y 之「誤差部分」，又稱之爲「誤差項」或「擾亂項」或「殘差項」，β_2，β_3，…，β_P 爲「偏回歸係數」。

如表 2.1 之 Y 與 X_2，X_3，…，X_P 爲樣本數 n（≒6）之樣本觀測值，以下表 2.1 之資料表示如下。

$$(Y_\upsilon, X_{2\upsilon}, X_{3\upsilon}, \cdots, X_{P\upsilon})\ ;\ \upsilon = 1, 2, \cdots, n \qquad (2.2)$$

（2.1）式之關係，即

$$\begin{aligned}
Y_1 &= \beta_1 + \beta_2 X_{21} + \cdots + \beta_P X_{P1} + e_1 = \beta_1 + 0.5\beta_2 + 6\beta_3 + e_1 \\
Y_2 &= \beta_1 + \beta_2 X_{22} + \cdots + \beta_P X_{P2} + e_2 = \beta_1 + 0.5\beta_2 + 8\beta_3 + e_2 \\
&\cdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (2.3) \\
Y_n &= \beta_1 + \beta_2 X_{2n} + \cdots + \beta_P X_{Pn} + e_n = \beta_1 + 1.2\beta_2 + 12\beta_3 + e_6
\end{aligned}$$

$$(n = 6,\ p = 3)$$

若以矩陣符號表示時，即

$$\underset{\sim}{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 8 \\ 9 \\ \vdots \\ 17 \end{pmatrix}, X = \begin{pmatrix} 1 & X_{21} \cdots X_{P1} \\ 1 & X_{22} \cdots X_{p2} \\ \vdots & \vdots \qquad \vdots \\ 1 & X_{2n} \cdots X_{pn} \end{pmatrix} = \begin{pmatrix} 1 & 0.5 & 6 \\ 1 & 0.5 & 8 \\ \vdots & \vdots & \vdots \\ 1 & 1.2 & 12 \end{pmatrix}, \qquad (2.4)$$

$$\underset{\sim}{\beta}' = (\beta_1\ \ \beta_2 \cdots\cdots \beta_P),\ \ \underset{\sim}{e}' = (e_1\ \ e_2 \cdots\cdots e_n)$$

$$\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{e} = \begin{pmatrix} 1 & 0.5 & 6 \\ 1 & 0.5 & 8 \\ 1 & 0.7 & 10 \\ 1 & 0.5 & 12 \\ 1 & 0.8 & 12 \\ 1 & 1.2 & 12 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \end{pmatrix} \qquad (2.5)$$

此 e 之各元素（隨機變數）e_υ **，滿足（2.3）式之樣本觀測值（（2.2）式）應以何種形式表示呢？在多回歸分析裡，擾亂項之向量 $\underset{\sim}{e}$ 與說明變量之矩陣 X，有如下之基本假定作爲前提，即

1°	$E(\underset{\sim}{e}) = \underset{\sim}{O}$	（各擾亂項 e_υ 之期望值爲 0）	(2.6)
2°	$E(\underset{\sim}{e}\underset{\sim}{e}') = \varepsilon I_n$	（各擾亂項 e_υ 之變異數爲常數 ε，互變異數爲 0）	(2.7)
3°	X 爲定常數之集合	（非隨機變數之集合）	(2.8)
4°	X 的階數 $= p < n$		(2.9)

$E(\underset{\sim}{e}\underset{\sim}{e}') = \varepsilon I_n$ 之假定是很重要的（I_n 爲 n 階之單位矩陣）換言之，即

---

* $\underset{\sim}{Y}$ 與 $\underset{\sim}{e}$ 皆爲隨機（機率）變數，說明變量 $X_2, X_3, \cdots, X_P$ 不爲隨機變數。“～”表向量

** e_υ 爲第 υ 個樣本之擾亂項（隨機變數）。

$$E\left\{\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}(e_1, e_2, \cdots, e_n)\right\} = E\begin{pmatrix} e_1e_1 & e_1e_2 & \cdots & e_1e_n \\ e_2e_1 & e_2e_2 & \cdots & e_2e_n \\ \vdots & \vdots & & \vdots \\ e_ne_1 & e_ne_2 & \cdots & e_ne_n \end{pmatrix}$$

$$= \varepsilon \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \cdots & & 1 \end{pmatrix} = \begin{pmatrix} \varepsilon & & & O \\ & \ddots & \varepsilon & \\ O & & \ddots & \\ & & & \varepsilon \end{pmatrix}$$

變異數 $E(e_\upsilon^2) = \varepsilon$（常數），互變異數 $E(e_\upsilon e_{\upsilon'}) = 0$（$\upsilon \neq \upsilon'$）。各樣本擾亂項 e_υ（$\upsilon = 1, 2, \cdots, n$）之相互間爲成獨立性（線性獨立，無相關）。

根據上述之假定，吾人欲推定滿足（2.1）式之係數 $\underset{\sim}{\beta}$。若樣本數 n（sample size）相當大時，所得出之 $\underset{\sim}{\beta}$ 值將較理想，但本例爲說明簡單起見，僅取樣本 $n = 6$，作爲 $\underset{\sim}{\beta}$ 之推定。今 $\underset{\sim}{\beta}$ 推定值（估計值）之向量若以 $\wedge$ 符号表示時，即

$$\hat{\underset{\sim}{\beta}}' = (\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_p), \ p = 3$$

且（2.5）式之擾亂項（殘差項）之向量爲，

$$\hat{\underset{\sim}{e}}' = (\hat{e}_1, \hat{e}_2, \cdots, \hat{e}_n), \ n = 6 \quad (\hat{\underset{\sim}{e}} \text{ 爲擾亂項之行向量})$$

此時（2.5）式即成如下之形式，

$$\boxed{\underset{\sim}{Y} = X\hat{\underset{\sim}{\beta}} + \hat{\underset{\sim}{e}} \qquad (2.10)}$$

在此之（2.10）式適用於「最小平方法之原理」，以「殘差之平