

多變量分析 二版

管理上的應用

林師模・陳苑欽 著

0.5682247852135648

Multivariate Analysis
in Management

多變量分析 二版

管理上的應用

林師模・陳苑欽 著

0.56822478521356485

Multivariate Analysis
in Management

多變量分析 二版

管理上的應用

Multivariate Analysis in Management

林師模・陳苑欽 著

多變量分析：管理上的應用 / 林師模, 陳苑欽合著.

-- 二版. -- 臺北市：雙葉書廊, 2013.02

面；公分

ISBN 978-986-6018-43-5 (平裝)

1.多變量分析 2.統計套裝軟體

511.2

102001973

多變量分析：管理上的應用 二版

作 者 林師模、陳苑欽
發行人 張福隆
責任編輯 郭振揚
封面設計 陳慧欣
出版社 雙葉書廊有限公司
地 址 台北市羅斯福路三段 269 巷 12 號 1 樓
電 話 (02)2368-4198
傳 真 (02)2365-7990
網 頁 <http://www.yehyeh.com.tw>
讀者服務 pub@yehyeh.com.tw
登 記 證 局版北市業字第 239 號
出版日期 西元 2013 年 3 月 二版一刷
電腦排版 辰皓國際出版製作有限公司

I S B N : 978-986-6018-43-5

著作權所有◎侵害必究

本書如有缺頁、破損、裝訂錯誤，請寄回更換。

本書內文之翻譯文稿所有權為雙葉書廊有限公司所有，未經允許不得翻印、仿製或轉載。

二版序

本書自 2003 年底發行初版後，得到來自各大專院校教師、學生的迴響不斷。多數是肯定本書內容深淺適中、講解詳細，對初學的入門讀者幫助很大，使其不再畏懼學習像多變量分析這類看似艱深的學科。不過，也不乏是對本書有更高的期許，其中包括希望本書能夠增加更多的例題，或是能夠增添章節介紹其他方法等。有感於這些讀者的善意及厚愛，筆者無時不在思索如何盡快將本書改版，以更符合市場的需求。然而，教學、研究及行政工作依然佔據了大部分的工作時間，以致改版工作一拖再拖，遲遲無法完成。2012 年間，在各方不斷詢問下，終於下定決心、一鼓作氣，先將初版內容做全面的錯誤更正，也更新了大部分的軟體操作說明畫面，並增加了一個新的章節：多變量變異數分析。這樣的改版距離原先的理想顯然還有一段差距，惟筆者希望以此做個起頭，將全面改版當作不久的將來一個使命必達的目標，也希望讀者期待。

二版在軟體操作說明畫面的更新及內容的除錯方面，幸賴李秀娟博士及王淑卿小姐的鼎力幫助，得以順利完成，筆者特別在此致上謝意。學海無涯，尤其是在每次重新檢視及閱讀各種方法之理論及實務應用內容時，更有體會。謹以此期勉所有對多變量分析方法有興趣的讀者及同好。另外，本書改版時雖已盡量改正初版的錯字，然謬誤仍在所難免，還望讀者及先進指正。

林師模 謹識

於 桃園市中壢 中原大學

2013 年 1 月 20 日

目錄

第 1 章	多變量統計方法介紹	1
1-1	變數資料的類型	3
	以衡量尺度分類	3
	以變數的角色分類	5
1-2	多變量統計方法的分類	5
	依賴模式	5
	相依模式	9
	結構關係模式	12
1-3	本書架構及學習方法	13
第 2 章	多變量統計分析軟體介紹	17
2-1	SPSS 應用簡介	18
	啟動 SPSS 並輸入與儲存資料	18
	讀取 ASCII 資料	19
	定義變數	20
	選擇統計方法	21
2-2	SAS 應用簡介	23
	啟動 SAS For Windows 系統	23
	建立及讀取資料	25
	資料瀏覽及編輯	26
	統計分析方法	26
2-3	STATISTICA 應用簡介	27
第 3 章	相關與變異數分析	29
3-1	變數關係與相關分析	30
	變數的內涵	30
	變數之間的關係	31
	相關分析	31

3-2	變異數分析原理.....	34
	變異數分析的定義.....	35
	因子——集區和處理.....	36
	變異數分析的數學模式.....	37
	固定效果與隨機效果.....	37
3-3	單因子變異數分析.....	38
3-4	二因子變異數分析.....	40
3-5	整合性例題.....	45
	例題一：國家競爭力.....	45
	例題二：企業財務資料.....	47

第4章 迴歸分析 57

4-1	迴歸分析原理.....	58
4-2	簡單線性迴歸分析.....	59
	簡單線性迴歸模型基本假設.....	60
	參數估計.....	61
	斜率 β 的抽樣分配、信賴區間及檢定.....	62
	截距 α 的抽樣分配、信賴區間及檢定.....	63
	判定係數.....	64
4-3	複迴歸模型與虛擬變數.....	66
	模型設定考量因素.....	70
	複迴歸模型.....	70
	虛擬變數迴歸模型.....	71
4-4	綜合整理.....	86
	迴歸分析的分析步驟.....	86
4-5	整合性例題.....	87
	例題一：國家競爭力.....	87

第5章 Logit 與 Probit 迴歸 93

5-1	線性機率迴歸模型原理.....	95
	線性機率模型.....	96
	隨機誤差項機率分配.....	96

線性機率模型的限制	97
5-2 Logit 模型	98
5-3 Probit 模型	99
Probit 模型與線性機率模型的比較	101
Probit 模型的參數估計	101
邊際效果	102
準確率	104
5-4 整合性例題	105

第 6 章 主成份分析 111

6-1 主成份分析應用簡介	112
6-2 主成份分析的理論架構	113
幾何架構	114
分析架構	119
6-3 觀念補充提示	125
6-4 整合性例題	126
例題一：國家競爭力	126
主成份分析結果	131
例題二：企業財務資料	139

第 7 章 因素分析 157

7-1 因素分析的架構	159
單一共同因子模型	159
多共同因子模型	161
因素分析求解的相關問題	164
因素分析的目的	167
7-2 因素分析的估計方法	167
主成份法	167
主因素軸法	176
其他估計方法	178
共同因子個數的選取	178
7-3 因素分析結果與資料適合度的判定	179



7-4	因素轉軸.....	180
	正交轉軸法.....	181
	斜交轉軸法.....	183
	因素分析—主成份法直交轉軸與斜交轉軸的比較.....	183
7-5	整合性例題.....	186
	例題一：國家競爭力.....	186
7-6	因素分析結果.....	191
	主成份法.....	191
	主因素法.....	202

第 8 章 區別分析 213

8-1	區別分析的原理.....	214
	選擇區別變數.....	216
	界定區別規則.....	218
	分類.....	220
8-2	區別規則.....	222
	基本假設.....	223
	區別規則的產生.....	223
	錯誤分類成本不同的情況.....	224
	多於兩群體的區別分析.....	225
8-3	變數選擇程序.....	225
	向前選擇程序.....	226
	向後消去程序.....	226
	逐步選擇程序.....	226
8-4	統計檢定.....	227
8-5	整合性例題.....	229
	金融危機預警系統.....	229

第 9 章 典型相關分析 237

9-1	典型相關分析的相關理論.....	238
9-2	典型相關分析架構.....	241
9-3	典型相關檢定與典型變量得點.....	248

典型變量組數的選取	248
重疊係數	249
9-4 整合性例題	250

第 10 章 集群分析 267

10-1 集群分析的基本概念	268
10-2 集群分析的步驟及作法	270
衡量相似性的方式	270
分群的原則與分群的方法	271
分群群數的決定	272
10-3 集群分析的分群方法	272
凝聚分層法	273
分離分層法	274
兩階段法	275
10-4 集群分析的相關檢定	275
10-5 整合性例題	277
例題一：國家競爭力	277
10-6 群集分析結果	282
完全聯結法	282
華德法	288
二階段法	292
二階段法集群分析結果	297

第 11 章 多變量變異數分析 305

11-1 單變量變異數分析	306
11-2 多變量變異數分析簡介	306
11-3 多變量變異數分析的檢定	307
兩組母體的檢定	307
多組母體的檢定	308
11-4 應用例題	310
單因子 MANOVA 分析	310
多因子 MANOVA 分析	312



11-5	使用 MANOVA 的必要性	312
11-6	軟體應用	314
	結果分析.....	318
第 12 章	問卷設計及調查	323
12-1	問卷設計原理.....	324
	問卷設計的流程.....	324
	問卷設計的常見錯誤.....	328
12-2	問卷調查方法及技巧.....	332
	抽樣調查之目的.....	332
	確定研究對象.....	333
	抽樣方法.....	333
	擬定訪談內容.....	336
	調查方式.....	336
	調查過程的考量因素.....	340
	調查的注意事項.....	341
第 13 章	測量的信度與效度分析	345
13-1	變數尺度及資料蒐集.....	346
	測量與信度、效度的關係.....	347
13-2	信度分析	348
	信度的標準.....	349
	信度的衡量模型.....	349
	信度的評估方法.....	350
13-3	效度分析	356
	內容效度與表面效度.....	356
	效標關聯效度.....	356
	建構效度.....	357
	效度的衡量模型.....	357
	效度的評估方法.....	360

第 1 章

多變量統計 方法介紹

學習內容

- 變數資料的類型
- 多變量統計方法的分類
- 本書架構及學習方法

一般企業在管理上，經常會面臨有大量的資料需要處理，以從中獲取一些有用資訊，來幫助定下各種管理的決策。所謂「資料會說話」通常指的是，在各種工作及經營情境下所產生的紀錄數據，可以適當地反映組織在不同情境下，各種生產、銷售、人力資源、財務、研發等種種的行為反應。例如，廠商生產某種產品的數量，在不同時間或情境會依據市場的需求、產品的價格、生產的技術及成本……等而決定；而廠商銷售某種產品的金額也會視各不同時間或情境下，產品的價格、行銷的方式、競爭對手的作法、消費者的偏好及消費能力……等而有所不同。透過分析這些紀錄數據，找出一些規律或是影響的因子，將有助於廠商制定有效的管理決策，提高其經營的績效，進而為股東創造最高的價值。

分析資料的方法很多，從簡單的資料列表、繪圖，到基本的**敘述統計分析**（descriptive statistical analysis），再到複雜的**多變量統計分析**（multivariate statistical analysis），其間可以應用的方法不勝枚舉。本書以一般管理上常用的**多變量分析方法**為主，介紹這些方法的基本理論及實務操作運用方式，給大專院校商學及管理相關科系的學生，以及一般從事企業經營的管理者，以幫助其了解這些方法應用上的範圍及奧秘，以及學習如何正確又有效的操作及運用這些方法。由於本書篇幅有限，因此介紹的方法僅涵蓋多變量統計方法中的一部份，其餘部份則將在另一本進階應用的書中，再詳加闡述。

本章主要是對多變量統計方法的面貌做概要的介紹，同時也對本書所涵蓋的範圍及學習的方式做一說明，希望能對讀者的閱讀及學習有所幫助。

1-1 變數資料的類型

企業或廠商各種行為的結果，或消費者的行為特性，有些是可以具體觀察並加以衡量的，而有些則是無法觀察及具體衡量的。例如，企業或廠商的**經營績效**可以用投資報酬率、資產報酬率等可具體計算的數據來衡量；而企業或廠商的**經營能力**則比較像是一個抽象的概念，雖然可以用一些數據資料（例如：報酬率、技術效率、生產力……等）共同來表現，但是通常並無法直接觀察。此外，消費者的特性一般可以其年齡、教育程度、所得、性別、品牌偏好等可以觀察及衡量的變數資料來代表，但消費者的消費傾向則可能無法直接觀察及衡量，必須藉由一些可以觀察及衡量的實際消費數據間接地衡量出來。

以衡量尺度分類

對於可以觀察到並加以衡量的行為變數，在衡量時會牽涉到要使用的衡量尺度應該為何的問題。通常衡量的尺度可以分為四類：**名目尺度**（nominal scale）、**序列尺度**（ordinal scale）、**區間尺度**（interval scale）、**比率尺度**（ratio scale）。以下我們舉例簡略說明這四種衡量的尺度。

▶ 名目尺度

名目尺度通常用來衡量分類或名目資料，其作用在於分類或命名，例如，性別、教育程度、職業、產業別等類別或等級資料。雖然我們通常將這些類別以數字表示，但這些數值並沒有大小次序，也不能做有意義的運算。例如，將職業分為「士」= 1、「農」= 2、「工」= 3、「商」= 4，並不代表商的地位或所得最高，而「士（1）」+「農（2）」也不會等於「工（3）」。

▶ 序列尺度

序列尺度通常用來衡量可以排列先後、等級、強弱的類別資料，一般問卷設計常用的**李克特量表**（Likert's scale）或**偏好排序**，就是一種序列尺度。例



如，某消費者對於國產汽車品牌的偏好，依序可能是：TOYOTA（1）、FORD（2）、NISSAN（3）、HONDA（4）；或是對於興建核四的意見，以非常同意=1、同意=2、無意見=3、不同意=4、非常不同意=5進行問卷調查。序列尺度的數值僅能代表順序，數值之間沒有等距性，而且其運算也沒有意義。例如，某人對於TOYOTA和FORD喜好的差距，可能遠大於FORD和NISSAN之間的差距。

▶ 區間尺度

區間尺度是衡量有任意原點的數量資料，其數值有大小順序，數值間的距離有意義，所以可做加減的運算，但由於其不具絕對原點，在乘除和比率的運算上則無意義。以溫度為例， 10°C 和 20°C 的差距大於 5°C 和 10°C 的差距，但不代表 10°C 比 5°C 熱兩倍。

▶ 比率尺度

比率尺度是用來衡量具有絕對原點的數量資料，例如，體重、GDP、營業額等，其數值間有大小順序，也可以做有意義的四則運算。例如，甲國外匯存底為800億美元，乙國為400億美元，則甲國外匯存底比乙國多400億美元，且為乙國的兩倍。

補充說明 1.1

李克特量表為一種量表設計方法，其利用陳述性語句，配合衡量受測者態度或意見的選項及分數，來衡量屬質變數。其步驟如下：

1. 針對研究命題蒐集大量的論點（例如：開放三通將有利於台灣經濟發展）。
2. 針對每一個論點設定其反映的類別以及分數（由同意至不同意可以

設定幾點尺度，尺度的數目依研究目的及變數性質而定）。

3. 將論點隨機排列。
4. 由研究對象中選取小樣本，建立基本的資料組，將樣本依總分數由大至小排列、分組，計算各組、各論點的平均值。
5. 選擇組間平均值差異大的論點為問卷題目。

以變數的角色分類

一般而言，我們可以依據變數在研究中的角色區分為**應變數**（dependent variables）和**自變數**（independent variables）。應變數通常是研究者所關心或希望了解、預測的變數，又稱為被解釋變數、依變數；自變數則是用來解釋、區別、分析或預測應變數的變數。在統計分析中，我們會視自變數和應變數的個數、變數的性質（屬質或屬量），以及研究的目的，來決定所採用的統計方法。在本書中將會介紹其中部份的方法。

1-2 多變量統計方法的分類

多變量統計方法（multivariate statistical methods）可分為三大類，即**依賴**（dependence）或**因果**（cause-and-effect）**模式**、**相依**（interdependence）**模式**、**結構關係**（structural）**模式**。依賴模式通常應用在當一群變數中，可以大致區分出反應變數（response or dependent variables）與解釋變數（explanatory or independent variables）的時候。依賴模式並不必然用以說明變數間的因果關係，惟當有相關的理論佐證時，則可以進一步建立及驗證這兩群變數間的因果關係。相依模式通常是用在當一群變數或資料中，沒有辦法清楚地區隔出反應變數及解釋變數的時候。利用相依模式，可以分析出變數或觀測值間到底是如何（how）及為什麼（why）相關，但是卻無法看出哪些變數是如何受到哪些變數的影響。結構關係模式通常用於描述一群可觀測或不可觀測變數間的結構及系統關係，為第二代的多變量統計方法。以下略為介紹上述三種方法，同時也對其作進一步的分類。

依賴模式

依賴模式可以用來建立反應變數與解釋變數間的函數關係，其另可依：

- (1) 反應變數的個數；
- (2) 解釋變數的個數；
- (3) 反應變數的衡量尺度（屬質或屬



量)；(4)解釋變數的衡量尺度的不同，而進一步細分為幾種不同的模式。表 1-1 列出以這些區分基準定義出來的一些依賴模式。

表 1-1 依賴模式的分類

			反應變數			
			一個		多個	
			屬量	屬質	屬量	屬質
解釋變數	一個	屬量	迴歸	區別分析 Logit、Probit	典型相關	多元區別分析
		屬質	t 檢定	間斷型區別分析	多變量變異數分析	間斷型多元區別度分析
	多個	屬量	多元迴歸	區別分析 Logit、Probit	典型相關	多元區別分析
		屬質	變異數分析	間斷型區別分析 聯合分析	多變量變異數分析	間斷型多元區別度分析

資料來源：Sharma (1996)，“Applied Multivariate Techniques”。

- **迴歸 (regression)**：當反應變數與解釋變數都各只有一個，且兩者都是屬量的變數時，則兩者之間所建立的線性函數關係即為一（簡單）迴歸。例如，當探討廠商的廣告支出與其營收間的關係時，反應變數為營收，而解釋變數為廣告支出。在一種特殊的情況下，若解釋變數只有一個且為屬質的變數時，則此一迴歸基本上就是一個變異數分析。但是以迴歸的方式估計及計算的結果，與以變異數分析計算的結果之間，在部份結果的呈現方面仍有一些差異。有關這方面的探討，我們將在第三及第四章中做較詳盡的說明。
- **多元迴歸 (multiple regression)**：當反應變數只有一個，而解釋變數有兩個以上，且兩者（反應及解釋變數）都是屬量的變數時，則兩者之間所建立的線性函數關係即為一多元迴歸或複迴歸。在某些情況下，解釋變數中可能同時包含屬質及屬量的變數，此時屬質變數的內涵通常是轉換以 0 或 1 來代表（此變數通常稱為**虛擬變數 (dummy variable)**），而迴歸則稱為**虛擬變數迴歸**。例如，我們可能想了解在某一企業中，員工的性別及工作年資與其薪資間是否有顯著關係，則此時所建立的迴歸分析模式中，薪資為



反應變數，性別及工作年資為解釋變數，而其中薪資與工作年資均為屬量變數，但是性別卻為屬質變數。在這種情況下，我們以 0 及 1 分別代表女性及男性員工，然後與其他變數一起進行迴歸分析。

● **變異數分析 (analysis of variance, ANOVA)**：當反應變數只有一個且為屬量變數，而解釋變數也只有一個但為屬質的變數時，則探討解釋變數對反應變數是否有影響關係的分析架構，即為一因子單變量變異數分析。如果解釋變數有兩個以上，且均為屬質的變數，則稱為多因子單變量變異數分析。在二因子單變量變異數分析的情況，又可分為有重複實驗及無重複實驗兩種，而有重複實驗的情況又可分為重複實驗次數均等及不等的兩種情況。一因子單變量變異數分析的例子如：某企業人事部門要探討該企業中，畢業自不同學校的 MBA 員工，其薪資是否有顯著差異。此時，薪資為屬量的反應變數，畢業學校為屬質的解釋變數，而其內涵可能有兩種以上（即員工畢業自兩種以上的學校）。如果員工的畢業學校只有兩種，則一般可以 T 檢定進行分析，探討該兩群員工薪資的平均數是否相等；但是當員工的畢業學校有三種以上時，則以成對的 T 檢定進行分析將增加誤差的機率，而較理想的是以變異數分析來取代。

● **區別分析 (discriminant analysis)**：當反應變數只有一個且為屬質的變數，解釋變數有一個以上且為屬量的變數時，則兩者之間所建立的線性函數關係之一即為區別分析。在區別分析中，如果反應變數的內涵只有兩種情況，而解釋變數有兩個以上，則此時的分析架構是多元迴歸的一個特例情況。區別分析的例子如：某大學某科系申請入學的審查作業中，是否准予入學考量的是申請人大學學測成績、高中在校平均成績、口試成績、資料審查成績等，而審查的結果只有兩種情況：准予入學或拒絕入學。在這個例子當中，反應變數是入學審查結果，解釋變數是大學學測成績、高中在校平均成績、口試成績、資料審查成績。區別分析將先依據該科系過去（前一、二年）作業所得結果，建立一個線性區別函數，然後將申請人資料代入此函數中，以判斷是否應准予申請人入學。

● **Logit 或 Probit 迴歸**：當反應變數只有一個且為屬質的變數，解釋變數有一個以上，且為屬量或屬量與屬質混合（即其中一些變數為屬量、一些變數為屬質）的變數時，則兩者之間所建立的線性函數關係之一即為 Logit 或