

多|變|量|分|析 方法與操作

SPSS 之應用

溫福星 / 編著



Multivariate

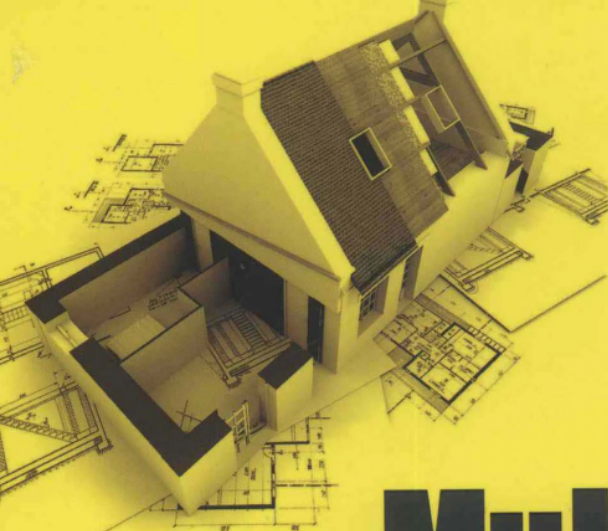
SPSS Operation and Application
Analysis



鼎茂圖書出版股份有限公司
<http://www.tingmao.com.tw>



內附光碟



Multivariate

SPSS Operation and Application
Analysis

- ◎ 詳細說明迴歸分析的原理與涵義，以及在調節與中介效果的應用。
- ◎ 利用迴歸分析解釋因素分析、主成分分析、典型相關分析、區別分析、邏輯斯迴歸的原理。
- ◎ 清楚說明因素分析與主成分分析、主軸與主成分萃取法的差異。
- ◎ 整合簡單相關、多元相關與典型相關的異同。
- ◎ 比較區別分析與邏輯斯迴歸分析、分類與分群的差異。
- ◎ 詳實解釋範例SPSS報表統計量運算過程與涵義。



鼎茂圖書出版股份有限公司

公司：台北市中正區開封街一段32號11樓

電話：02-2381-4314 傳真：02-2382-5963

台中辦事處：台中市區綠川東街20號2樓

電話：04-2221-1381 傳真：04-2221-1403

www.tingmao.com.tw



DB20680101

ISBN 986-226-786-0



9 789862 267868

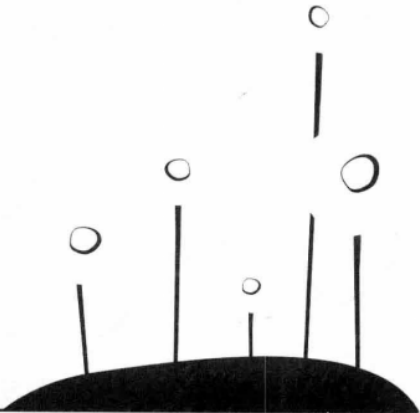
00390



多|變|量|分|析 方|法|與|操|作

SPSS 之應用

溫福星 / 編著



Multivariate

SPSS Operation and Application
Analysis



鼎茂圖書出版股份有限公司
<http://www.tingmao.com.tw>



內附光碟

國家圖書館出版品預行編目資料

多變量分析方法與操作-SPSS 之應用/ 溫福星 編

著 --一版--臺北市：鼎茂圖書，

民 101.09 面：公分

ISBN 978-986-226-786-8 (平裝)

1.多變量分析 2.統計分析 3.統計套裝軟體

511.7

101011291

多變量分析方法與操作-SPSS 之應用

作 者 溫福星 編著

企劃編輯 陳怡萱

專案行銷 張容甄

印務採購 賴銘銓、張育菁

發行通路

校園 張容甄

書局 高嘉聰

網路 黃志偉

發 行 所／鼎茂圖書出版股份有限公司

台北市中正區開封街 1 段 32 號 11 樓

電話：(02) 2381-4314 傳真：(02) 2382-5963

鼎茂網路書店 <http://www.tingmao.com.tw> 電子郵件信箱 service@tingmao.com.tw

郵政劃撥／18242879

鼎茂圖書出版股份有限公司

登 記 證／局版台業字第 5881 號

法律顧問／第一國際法律事務所 余淑杏律師

台北市中山區民生東路一段 43 號 3 樓

電話：(02)25215990

打字排版／恆新電腦排版打字行

校 稿／陳怡萱

本書編號／DB2068

I S B N／978-986-226-786-8《平裝》

出版日期／中華民國 101 年 9 月 一版

定 價／390 元

◎本著作物係著作人授權發行，若有重製、仿製或對內容之其它侵害，本公司將依法追究，絕不寬待！

◎書籍若有倒裝、缺頁、破損，請逕寄回本公司更換。

Answertree®, Clementine®, Decisiontime®, Lexiquist®, PASW®, PASW the Predictive Analytics Company®, Quancept®, Samplepower®, Showcase®, Smartscore®, SPSS®, and WHATIF?™ are trademarks or registered trademarks of SPSS, Inc. (or its affiliates), an IBM Company.

書籍內文之操作介面圖片皆由 IBM 公司授權提供

序

教了十年碩士班的多變量分析，一直覺得沒有適合的中文教材可以當教科書，而英文的多變量分析教科書卻不少，但不是太過於數理推導就是太過於文字表述。對於商學或是管理的碩士班學生而言，數理導向的多變量分析等於是天書，毫無實務應用價值，而以電腦操作為主的統計書，同學依樣畫葫蘆，跑出來的結果常常不知所以然，就如大陸所說的傻瓜書，不知如何解釋。

這本書的問世，是嘗試將多變量分析的數理原理，與研究方法做個結合，從研究方法的觀點，導出統計原理，配合統計軟體報表的結果，一一解釋這些統計量背後的意義與運算過程。這本書的主軸以迴歸分析為主，利用大家對它的瞭解，延伸去解釋多變量分析的其他統計方法，例如是因素分析、主成分分析、典型相關分析與邏吉斯迴歸等，讓研究生可以輕鬆上手，不僅懂得如何操作統計軟體，更能夠瞭解方法背後的原理，以及運算的結果。

這本書的前身是個人在碩士班多變量分析的授課教材，但一直存在我的腦袋中，從來沒有紙本的複製品，每年的課程根據同學的背景與程度，以及碩士論文方向的方法，作一些調整與變更。感謝鼎茂圖書的陳社長、CoCo 與 Ivy 等人的協助，可以將這個中文教材編纂而成，雖然還不是很完善，但對個人而言至少已經跨出第一步。這個講義書的呈現，是從 2011 年第一學期國貿所多變量分析修課同學的教材，由我的研究生辜盛峰、葉信佑與劉惠榕三位的繕打，再經 2012 年第二學期心理所研究生的使用，最後以書的形式呈現。

由於統計背景的關係，本書不免有線性代數的特徵值與特徵向量的計算，與多變量分析方法的矩陣運算，但本書盡可能將內容簡化，但過程詳細描述，除了可以讓同學瞭解這些統計方法原理以外，順便提升這些方法的數理基礎。如果對線性代數沒有興趣的同學也可以略過這部分，仍可以從方法與運算過程來學會多變量分析的內容。個人期望在未來很短時間內，可以將方法再加以擴充，且對內容可以更詳實的介紹。

溫福星

2012.09

Multivariate

SPSS Operation and Application
Analysis

目錄

第一章 多變量分析概說	3
第一節 本書統計分析方法架構	6
第二節 統計基本觀念	8
第三節 資料庫簡介	12
第二章 迴歸分析原理	15
第一節 簡單迴歸分析	15
第二節 相關分析與簡單迴歸分析的關係	19
第三節 兩獨立樣本平均數差異 t 檢定與簡單迴歸分析的關係	20
第四節 簡單迴歸分析的意義	21
第五節 多元迴歸分析的原理	24
第六節 多元迴歸分析的重要假設	32
第七節 結論	37
第三章 迴歸分析範例解讀	41
第一節 簡單迴歸分析	41
第二節 解釋變項中心化與標準化迴歸係數	60
第三節 迴歸分析殘差項假設的驗證	68
第四節 迴歸分析與兩獨立樣本 t 檢定：虛擬迴歸分析	75
第四章 迴歸分析的應用：調節效果的檢測	87
第一節 調節迴歸分析	87
第二節 虛擬迴歸分析	90
第三節 共變數分析與調節迴歸分析	94
第四節 中心化對多元共線性的影響	100
第五節 範例介紹	102
第五章 路徑分析與中介效果檢驗	117
第一節 路徑分析	117
第二節 中介效果的檢測	126
第三節 調節效果與中介效果的組合	138

第六章 因素分析	141
第一節 因素分析原理	141
第二節 因素分析方法的應用	148
第三節 萃取技術與觀念再整理	156
第四節 信度原理	165
第五節 因素分析 SPSS 範例說明	166
第七章 主成分分析	197
第一節 主成分分析原理	197
第二節 主成分分析示範	200
第三節 主成分分析與因素分析的差異	206
第四節 特徵值與特徵向量的求算	212
第八章 典型相關分析	225
第一節 典型相關分析原理	225
第二節 典型相關分析示範說明	230
第三節 典型相關分析數理模型	237
第九章 區別分析	243
第一節 區別分析原理	243
第二節 區別分析範例說明	248
第十章 邏輯斯迴歸分析	259
第一節 邏輯斯迴歸原理	259
第十一章 集群分析	281
第一節 集群分析原理	281
第二節 集群分析 SPSS 範例說明	286
第十二章 結語	299

參考文獻

第一章 多變量分析概說



Multivariate

SPSS Operation and Application
Analysis



多變量分析(multivariate analysis)顧名思義就是研究好多個變項之間的關係，在大學部初等統計學研究過單一樣本平均數的檢定、兩獨立樣本平均數差異檢定、以及變異數分析，這些都是在探討一個類別自變項(independent variable)對一個連續量尺依變項(dependent variable)的影響。除此之外，亦會教授相關分析與迴歸分析，迴歸分析包含簡單迴歸分析與多元迴歸分析，簡單迴歸分析是研究兩個連續量尺變項之間的關係，一個是自變項一個是依變項。而多元迴歸分析則是好多個連續量尺的自變項對一個連續量尺依變項的影響，如果依狹義的多變量分析定義就是屬於多變量分析的範疇。而相關分析是研究兩個變項之間的關聯(association)，它包含連續對連續的簡單相關分析，也包含類別、次序、連續變項之間的關聯。最特別是兩個類別變項之間的關聯，例如 ψ 相關，以及牽涉到一個次序變項的等級相關等。在兩個類別之間的關聯以外，亦會計算卡方檢定統計量，可能進行獨立性、同質性或是適合度（單一類別變項）的檢定。如果再有時間，會簡單介紹無母數統計以及初等的時間數列分析。我們簡單將上述的關係以表 1 來做個整理，簡單將變項分成自變項與依變項，再將自變項與依變項分成連續量尺（等距與比率）與類別尺度（名義與次序）整理初等統計與多變量分析的內容。基本上，本書是以橫斷面分析為主的統計研究方法的介紹。

表 1-1 多變量分析與初等統計方法的分類

統計方法		自變項（一個或多個）	
		類別尺度	連續尺度
依變項 （一個）	類別 尺度	列聯相關 卡分檢定 邏吉斯迴歸分析 對數線性模式	點二系列或系列相關 等級相關 區別或判別分析 邏吉斯迴歸分析
	連續 尺度	t 檢定 變異數分析 迴歸分析	迴歸分析
依變項 （多個）	連續 尺度	多變項變異數分析(MANOVA)	多變項多元迴歸分析(multivariate multiple regression)

註：框線部分屬於多變量分析範疇

從表 1-1 可以發現初等統計的內容教的不少，基本上可以應付一般的資料分析。但至於要從事學術研究或是更深入的資料分析方法，上表中的區別分析、邏輯迴歸與對數線性模式都是很好的建模工具，因為存在一個依變項，也就是研究者關心的變項，研究到底哪些因素或是變項可以幫助我們對這個依變項有更進一步的了解。有關於對數線性模式，事實上它是沒有依變項的變數，而是許多類別變項的列聯表所組成，其模式的依變項為計數資料，此模式相當於變異數分析，探討各個類別變項對於計數資料的影響。此外，表 1-1 中右下角的連續自變項對連續依變項的影響只有一個迴歸分析，從這裡可以看出迴歸分析在統計學中的重要角色，基本上高階統計或是多變量分析很多的原理是建立在迴歸分析的基礎上。

除了表 1-1 的分類以外，有些是無法以自變項與依變項來區分，資料可能是一團雜亂無章的數據，或是已經整理待分析的資料，我們可能關心的是這群資料是否是研究變項間的關係、還是研究受試者間的關係這兩大類。根據變項研究受試者間的異質性就是分群的意思，是屬於**集群分析(cluster analysis)**，或是研究受試者(subjects or objects)在變項間的相似性或相異性的**知覺定位圖(perceptual map)**是**多元尺度分析(multidimensional scaling)**。而研究變項間的關係可以分為三類，第一類是將研究變項事前區分為兩大類，每一大類都有許多變項，想要研究這兩類研究變項之間的相關為**典型相關分析(canonical correlation analysis)**；第二類是研究這一組研究變項間的相關是否有共同的影響原因，我們稱為**潛在變項(latent variables)**或**因素分數(factor scores)**，可以來解釋這一群研究變項之間為何存在相關的解釋變項，換言之這一群所蒐集到的研究變項，是看不到的潛在變項或因素分數的化身或是代言人，我們稱為**反應性(reflective)變數**，這種統計方法為**因素分析(factor analysis)**。第三類是將這群研究變項形成新的變項，建立新指標或新指數，我們稱為**主成分(principal components)**，用較少的新變數來代表原先這群較多的研究變項，新的變數是由這群舊變數所形成，所以這群舊變數稱為**形成性(formative)變數**，這個方法為**主成分分析(principal component analysis)**。

表 1-2 連續量尺多變量分析的統計方法

統計方法	研究對象	
	研究受試者間關係	研究變項間關係
多個連續量尺變項	根據研究變項研究受試者間的異質性→ 集群分析	兩組研究變項間的關係→ 典型相關分析
	根據研究變項研究受試者或研究對象、屬性間相似性或相異性的知覺定位圖→ 多元尺度分析	研究多個變項的共同因→ 因素分析 研究多個變項的線性組合→ 主成分分析

一般來說，多變量分析的許多統計方法的結果，可能是我們最終資料分析過程的一部分，例如因素分析除了驗證研究變項的因素效度以外，它誠如 SPSS 軟體中的『維度縮減』，和主成分分析、集群分析等都是研究的中間過程，因為社會科學的研究主要是在研究表 1-1 的依變項和其影響因素自變項之間的關係。鑒於迴歸分析的重要性，本書從迴歸著手，主要介紹因素分析、主成分分析、典型相關分析、區別分析、邏輯迴歸分析與集群分析等，做這樣的安排主要是迴歸分析除了本身是研究假說的分析方法外，它主要是跟研究方法有關，例如調節效果分析與中介效果分析，這兩個分析是社會科學研究方法很重要的概念，除了現在的結構方程模式 (structural equation modeling) 以外，迴歸分析都可以達到這樣研究方法的實踐。因此，本書前五章主要是詳細介紹這些方法的原理與實徵分析過程，以及統計軟體 SPSS 的執行與分析結果報表的解讀。其次，介紹『維度縮減』的兩個方法，分別是因素分析與主成分分析，當在迴歸分析碰到解釋變項多元共線性問題無法解決時，這兩個分析方法可以部份解決這樣的問題。除此之外，因素分析在驗證性因素分析 (confirmatory factor analysis) 未興起之前，它是有關研究構念效度的驗證方法，在因素分析這章除了介紹原理之外，我們將信度放入，示範有關社會科學以問卷進行研究構念的檢視程序。而主成分分析是將解釋變項做一線性組合，使得新產生的變項間可以沒有相關，適合進一步迴歸分析或建構新指標之用。這兩章的原理都利用到線性代數的特徵值與特徵向量，因此依序介紹這兩章。

之後，利用主成分分析的特徵值與特徵向量的概念，介紹典型相關分析，這個

方法與區別分析都利用到矩陣特徵值的求算，基本原理在於主成分分析。在典型相關分析之後介紹類別依變項的分析方法，以二分變項為例，當解釋變項為連續變項或是類別變項時，如何建構之間的迴歸方程式。區別分析會利用到特徵值，而邏吉斯迴歸利用到迴歸分析原理，這兩個分析方法原理不同但目的相同，不過之間有些差異與限制。最後是集群分析的介紹，研究受試者的異質性，根據研究變項來進行分群，然後可以利用變異數分析進行進一步的資料分析或假設檢定的驗證。

第一節

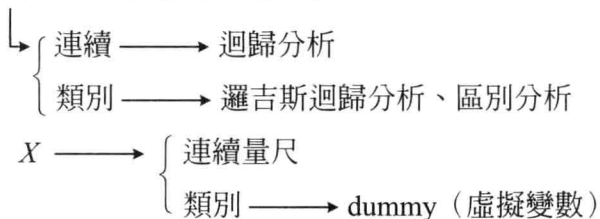
本書統計分析方法架構

基本原理一：簡單迴歸分析

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

基本原理二：多元迴歸分析

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$



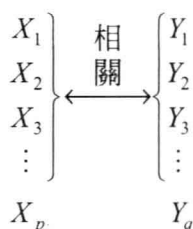
當解釋變項有多元共線性問題時

$\longrightarrow$ 因素分析
 主成分分析

當 Y 有好多個但不考慮解釋變項時

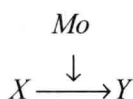
{ 重複觀測 (相同變項) $\longrightarrow$ MANOVA
 { 多依變項 (不同變項) $\longrightarrow$ 因素分析

兩組變項之間的相關 $\longrightarrow$ 典型相關分析



迴歸分析方法的應用

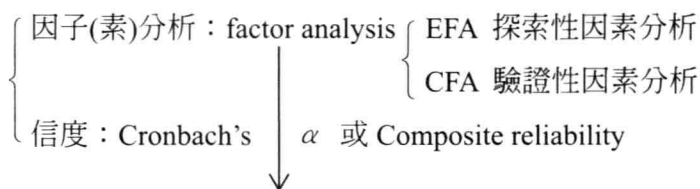
$\left\{ \begin{array}{l} \text{中介效果 mediation} \\ \text{調節效果 moderation} \end{array} \right.$



迴歸分析的目的

1. 描述 X 與 Y ， X 與 Y 的關係
2. 解釋 $X \rightarrow Y$ ， X 用來解釋 Y 的變異
3. 預測 $X \rightarrow ?$ ，有了 X 則 Y 會是多少
4. 控制 $? \rightarrow Y$ ，爲了得到 Y 要有多少 X

問卷的信效度：構念(constructs)



$$SEM = CFA + Path$$

(路徑分析 path analysis：多個迴歸分析組成)

實驗設計：ANOVA、MANOVA

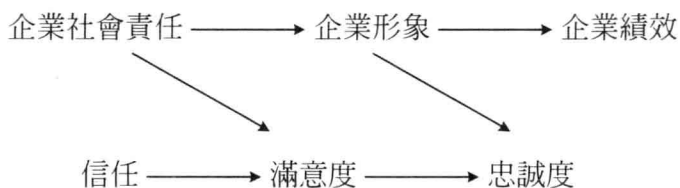
$\updownarrow$ 原理相同但方向相反

區別分析 $\longrightarrow$ 如果存在類別解釋變項以邏吉斯迴歸取代

變項和變項間的關係

id	X ₁	X ₂	X ₃	X ₄	X ₅	Y ₁	Y ₂	Y ₃	gender	edu	...
1											
2											
3											
4											
5											
6	人與人之間的關係					cluster analysis					
7						latent class analysis					
8											
...											

橫斷面變項間關係的研究架構，例：



1. 多個迴歸 —————→ path analysis
2. SEM（一次估完整個架構的迴歸係數）

第二節

統計基本觀念

統計學的世界裡，尤以連續尺度的參數統計學，主要探討兩個很重要的統計概念，一個是平均數、一個是變異數或是共變數，變異數則是自己和自己的共變數。而共變數的表達方式有三：第一是變異(variation)或是交乘項平方和、第二是變異數或共變數、第三是相關係數。

表 1-3 樣本變異數與共變數概念

公式	變異數	共變數
Variation (平方和)	$\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})$	$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$
Variance (變異數)	$\frac{\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})}{n-1}$	$\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1}$
Standard Deviation (標準差)	$\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})}{n-1}}$	
Correlation coefficient (相關係數)		$\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$

初等統計學的平均數是屬於集中趨勢，而集中趨勢包含下面幾個統計量

Central tendency (集中趨勢)：mean、median、mode

而變異數屬於離散程度的部份，其他離散程度尚包含：

Dispersion (離散程度)、range、(Q3-Q1)/2、variance、S.D.、absolute deviation

統計學可以說是探討與計算平均數與變異數(含共變數)的問題，假設母體有 5 筆資料，如下所示：

Data：1、2、3、4、5

$$\mu = 3 = \frac{1+2+3+4+5}{5}$$

平均數相當於重心的位置

而變異數相當於面積的概念

$$\text{var} = \frac{\sum (x_i - \mu)^2}{5} = \frac{4+1+0+1+4}{5} = 2$$

1. 面積概念

2. 假設資料獨立，故才能相加