

社会科学 智勝
研究方法丛书 BEST-WISE

多变量分析

· SPSS的操作与应用 ·

■ 林震岩 著



社会科学研究方法丛书

多变量分析

SPSS的操作与应用

■ 林震岩 著



北京大学出版社
PEKING UNIVERSITY PRESS

图书在版编目(CIP)数据

多变量分析:SPSS的操作与应用/林震岩著. —北京:北京大学出版社,2007.8
(社会科学研究方法丛书)

ISBN 978-7-301-12472-7

I. 多… II. 林… III. 多变量-统计分析-软件包,SPSS IV. C819

中国版本图书馆CIP数据核字(2007)第093104号

简体中文版由智胜文化事业有限公司(Taiwan)授权出版发行

《多变量分析:SPSS的操作与应用》,林震岩著,2007年1月再版,ISBN:957-729-615-3。

书 名:多变量分析——SPSS的操作与应用

著作责任者:林震岩 著

责任编辑:丁莉华

标准书号:ISBN 978-7-301-12472-7/C·0445

出版发行:北京大学出版社

地 址:北京市海淀区成府路205号 100871

网 址:<http://www.pup.cn>

电 话:邮购部 62752015 发行部 62750672 编辑部 62117788 出版部 62754962

电子邮箱:law@pup.pku.edu.cn

印 刷 者:北京大学印刷厂

经 销 者:新华书店

787毫米×1092毫米 16开本 44印张 834千字

2007年8月第1版 2007年8月第1次印刷

定 价:68.00元

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究

举报电话:010-62752024 电子邮箱:fd@pup.pku.edu.cn

再版序

Preface

多变量分析是社会科学研究者必备的工具,但因该分析牵涉到许多统计技术,学习相当不易。而 SPSS 包括初等统计与多变量分析的技术,通过统计软件来学习多变量分析技术,可加速学习速度。对教授多变量分析的老师而言,若只教授多变量分析的统计技术,学生不易理解,但若只教授统计分析软件,学生将因缺乏多变量分析理论背景无法理解统计分析的结果。故如何编撰或选择一本适合学生的多变量分析教科书,一直是统计学领域老师们的一个困扰。笔者因有感于目前的多变量分析教科书普遍存在以下问题,而引起撰写本书的动机:

- (1) 教科书内容过于理论化与深奥,不易理解。
- (2) 教科书内容不够生活化,无法引起读者的兴趣。
- (3) 教科书内容涵盖的多变量分析主题不够完整。
- (4) 教科书内容不适合非专业人士或不具备统计专业背景的同学阅读。
- (5) 教科书内容过时,跟不上多变量技术的发展。

SPSS for Windows 是 SPSS/PC 的 Windows 版本,因此具有 Windows 软件的共同特点。SPSS 在界面的友善性、易用性与分析结果的美观性上下了一番工夫,其操作界面大多是以 menu-driven 的方式呈现,因此很受学术界与实务界人士喜爱,并且也是目前国内大专院校最为风行的统计软件包。基于此,本书希望结合多变量分析与 SPSS 统计软件,以深入浅出的方式,通过 SPSS 带领学生认识多变量分析的奥妙。

现将本书特色分述如下:

(1) 以 SPSS 引出对多变量分析的学习兴趣。本书不只是 SPSS 也不只是多变量分析,而是将两者予以融合以利于学习,让读者不需再分别购置两本书。本书在每一章的第一节中,都会先针对该章所要探讨的多变量分析技术原理加以介绍,接着对 SPSS 的操作与输出作详细的说明与探讨。当老师在教学或学生在学习时,可先学习 SPSS 的操作,然后再回头学习原理,有时这样更容易了解原理;当然也可先学原理再学操作,通过这种多变量分析技术与 SPSS 软件的交互验证,可以收到事半功倍的效果。

效果。

(2) 有关多变量分析原理的介绍既不过于专业也不过于简略。有关多变量分析原理的探讨如果过于强调分析方法的演算过程,会产生许多深奥的数学公式,特别是多变量分析牵涉到许多矩阵的处理,容易使一般数理基础不强的读者有非常深的挫折感,但如果不介绍这些分析方法的演算过程,又显得不够完整且无法掌握该多变量方法的精髓,对于想进一步理解的读者也不公平。事实上,对一般使用者而言,观念的理解远重于公式的推导。所以,本书除了理论与操作兼顾外,更强调多变量分析原理与方法的观念理解,并以浅显易懂的方式说明分析演算的过程,让读者能够易于理解且融会贯通。

(3) 涵盖完整的多变量分析功能。本书除了介绍主菜单上的 SPSS/Base 功能(如平均数检验、一般线性模型、因素分析、聚类分析、区别分析、回归分析)外,也对 Advanced 等模块功能作了说明,如一般书上少见的多维量表法、TREE、Logistic、时间序列分析等进阶的多变量分析技术也多有着墨。此外,不包括在主菜单上的技术,如规则相关分析与联合分析以及 SPSS 的外挂程序,如结构方程模型 AMOS、数据探勘——Clementine 等,在本书中也都有专门章节的介绍。

(4) 对使用经验加以深入的介绍与评论。本书除了对所有多变量分析方法与 SPSS 操作有完整介绍外,更加入笔者对这些方法的许多深入评论,如第一章有关变量命名的技巧;第九章如何对因素分析结果的因素命名;第十章如何对聚类分析结果的类别命名,或对各类别的数目的决定等。此外,也会举出使用者常犯的错误,以便提醒使用者避免,如第六章有关列联表的分析。

(5) 探讨如何制作与解释研究报告的分析表格。通常使用者会直接剪贴 SPSS 的结果输出窗口中的报表或图形来当做报告的一部分,虽然 SPSS 已对表格或图形进行了严谨的制作,但对学术研究者来说,仍有许多表格不符合论文格式的规定,故有必要重新制作。为此,本书特将各章中的多变量分析表格,另外搜集在“多变量分析参考表格”(请于智胜网站(http://www.bestwise.com.tw/user_book_data.asp?sel_serial_id=968)下载)中,使用者可以直接引用或修改这些表格范例,快速编制正确的表格。此外,本书也对如何解释这些表格进行了深入的说明。

(6) 每个范例都有完整的数据文件、程序文件与输出文件。本书的范例都会将其数据文件、程序文件与输出文件等储存在该章所属的范例文件中,其扩展名分别为 .sav、.sps、.spo 等,读者均可直接读取、执行与查阅,或进行数据与程序的修改,以尝试各种分析,从而使读者更能深入地了解多变量分析的奥妙,而不需要重新建立这些数据文件与程序文件。

(7) 每个范例都相当生活化,并且易于了解。本书的范例大部分是来自本土的调查研究,且范例也相当生活化。例如本书的每一章几乎都会以“1.1 学生”文件为例来说明。事实上,此文件只有 10 个字段,却可以进行所有多变量分析程序,让学生很快理解多变量分析如何应用。此外,像相当轰动的邇鸟侠、两岸三通、大陆投资环境评

估、大学评比、汽车排名、客户关系管理系统、3G 移动通信等都是生活中与管理中较普遍的应用,因此学生在阅读与试用这些范例时,就能够很容易地接受与理解多变量分析技术。

目前 SPSS 虽有中英文两种版本,但因中文版较不普及,并且许多中文用词翻译存在失真的问题,因此本书尽量对 SPSS 的用词保留英文并附上中文版翻译,以兹对照。本书编写的基本目标是简洁中不失详尽,深信本书精心挑选的内容,足以满足社会科学研究者 95% 的需求。

在 SPSS 的操作手册当中,有许多内容很重要但却是研究者很少用到的。举例来说,研究者在做因素分析时,经常会用到 7 种不同的萃取(extracting)法和 6 种不同的旋转(rotate)因子法的机会有多少?或者在单因素方差分析(One-Way)后,使用 18 种不同的事后比较法吗?一般在一个分析窗口中通常会提供数种不同的选择,但本书因限于篇幅而无法详细说明,所以仅对其中一到两种重要的方法加以介绍;其余建议有兴趣的读者可参考“SPSS 操作手册”。

此外,为让读者清楚地了解本书所提到的各式输出,笔者还将其分为三大部分,分别置于智胜文化的网站上,供读者下载与应用。分别为:(1) 本书所有范例的数据文件、程序文件与输出文件;(2) 多变量参考表格;(3) SPSS 系统所提供的范例文件。笔者均已将这些范例文件的变量名称与变量值名称标记转为中文,读者使用时可直接贴上数据文件的内容,便可得到中文的数据文件。

本书得以完成,首先感谢政治大学企业管理研究所多变量分析课程的启蒙老师——黄俊英,因为他的教导使得笔者有机会从事多变量分析的教学以及 SAS/PC 教科书的撰写工作。此外,高孔廉老师的回归分析、许士军老师的研究方法、赖士葆与刘水深老师的数量方法、周文贤老师的 LISREL 技术等多变量分析方法的传授也让笔者受益匪浅。此外,还需感谢中原大学企业管理研究所诸承明所长购置 SPSS 与 AMOS 等软件,使本书的写作能够顺利进行;台湾政治大学企业管理学系同学郭宗霖与郭柏青协助软件的测试与数据的搜集与整理、研究助理林晓如与杨雅琪协助本书的图片与投影片制作,他们都对本书的完成有相当大的贡献。

本书的完成还需感谢中原大学教务处现任与前任教务长江彰吉与王晃三教授所推动的“教科书撰写奖励计划”,这是笔者撰写此书的原始动力。本书的出版获得读者相当多的回应,但因部分多变量技术的理论陈述仍稍显艰涩,故针对这些章节加以改写,以期让读者更易理解;希望本书的再版能够获得更多读者支持。最后,感谢智胜文化万总经理与所有同仁的协助,使本书得以丰富的内容来呈现。

因本人才疏学浅且多变量分析技术浩瀚,使得本书难免有疏漏,尚祈各界人士不吝指教。

中原大学企业管理学系教授

林震岩

2007 年 1 月

CONTENTS



目 录

第一部分 基本操作与数据处理

第一章 SPSS 系统总论与基本操作 3

第一节 SPSS 系统简介 5

第二节 系统的安装与启动 12

第三节 SPSS 数据文件的建立 19

第四节 数据分析、语句命令与结果输出 32

第二章 数据转换与处理 45

第一节 数据转换 47

第二节 数据处理 61

第三节 SPSS 的函数 76

第三章 SPSS 的统计图形制作 83

第一节 SPSS 绘图功能 85

第二节 交互式统计图形 111

第二部分 单变量与多变量总论

第四章 单变量描述性分析 125

第一节	描述性统计原理	127
第二节	报告	131
第三节	描述性统计	139
第四节	表格	152
第五节	复选题分析	158
第六节	描述性分析范例——CRM	165

第五章 多变量分析总论与信度分析 173

第一节	统计检验与 P 值	175
第二节	多变量分析程序总论	178
第三节	信度分析与效度分析的原理	183
第四节	SPSS 的信度分析	192
第五节	信度分析范例	200

第六章 两变量关联分析 205

第一节	两变量关联分析原理	207
第二节	SPSS 的 Correlate 分析	213
第三节	相关分析范例	224
第四节	SPSS 的 Crosstabs 分析	230
第五节	Crosstabs 的应用范例	238

第三部分 平均数与方差分析

第七章 平均数比较分析 249

第一节	平均数比较检验原理	251
第二节	One-Way ANOVA 检验原理	255
第三节	SPSS 的 Means 分析	260
第四节	SPSS 的单一样本 T 检验	267
第五节	SPSS 的独立样本 T 检验	269
第六节	SPSS 的成对样本 T 检验	275
第七节	SPSS 的 One-Way ANOVA 检验	277
第八节	平均数比较范例	286

第八章 一般线性模型 295

第一节	GLM 分析的原理	297
第二节	SPSS 的 GLM 单变量分析	309
第三节	SPSS 的 GLM 多变量分析	320

第四节	SPSS 的 GLM 重复测量分析	327
第五节	GLM 一般范例	331
第六节	GLM 综合范例——3G 移动通信	337

第四部分 数据简化分析

第九章	主成分分析与因素分析	345
第一节	主成分分析原理	347
第二节	因素分析原理	354
第三节	SPSS 的因素分析	363
第四节	因素分析范例	374
第十章	聚类分析	387
第一节	聚类分析原理	389
第二节	SPSS 的层次聚类分析法	400
第三节	SPSS 的 K-Means 聚类分析法	409
第四节	SPSS 的 Two Step 聚类分析法	415
第五节	聚类分析范例	423
第十一章	多维量表分析法	431
第一节	多维量表分析原理	433
第二节	SPSS 的多维量表分析	438
第三节	多维量表分析范例	448
第四节	个体与属性的联合空间图	456

第五部分 预测分析技术

第十二章	区别分析、TREE、Logistic 与 Probit 回归	469
第一节	区别分析原理	471
第二节	SPSS 的区别分析程序	487
第三节	区别分析的范例	497
第四节	SPSS 的 TREE 程序	504
第五节	Logistic 回归分析程序	512
第六节	Probit 程序	523
第十三章	回归分析与典型相关分析	529
第一节	回归分析原理	531

第二节	SPSS 的线性回归分析	542
第三节	线性回归分析范例	555
第四节	SPSS 的曲线估计程序	559
第五节	典型相关分析原理	566

第六部分 因果关系分析与其他技术

第十四章	结构方程模型与路径分析	579
第一节	结构方程模型的原理	581
第二节	路径分析原理	591
第三节	SPSS 的 AMOS 系统	597
第四节	路径分析与结构方程模型范例	612
第十五章	其他统计检验程序	627
第一节	非参数检验	629
第二节	时间序列分析	644
第三节	联合分析	672
第四节	数据探勘——Clementine	684
参考文献		691

第一部分

基本操作与数据处理

PART



Multivariate Analysis

第一章

SPSS 系统总论与基本操作

学习目标

1. 认识 SPSS 有哪些模块及各模块的功能。
2. 了解 SPSS 的功能与特点。
3. 理解 SPSS 统计分析的步骤。
4. 认识 SPSS 的安装与启动。
5. 了解 SPSS 菜单中的分析 (Analyze) 功能。
6. 了解 SPSS 菜单中的辅助说明 (Help) 功能。
7. 探讨 SPSS 数据文件建立的编辑。
8. 数据文件的建立与调用,特别是从 Excel 与 Text 档调用。
9. 认识变量的定义与数据的定义,包括缺失值的处理。
10. 介绍如何进行分析 (Analyze) 的操作。
11. 介绍如何进行语句编辑窗口 (Syntax Editor) 的操作。
12. 介绍如何进行结果输出窗口 (Viewer) 的操作。

本章简介

本章对 SPSS 的功能与基本操作作一整体介绍,可视为是 SPSS 总论。第一节对 SPSS 系统加以简介,包括 SPSS 的发展历程、完整的 SPSS 系统模块,以作为选购模块的参考。此外,本节还介绍 SPSS 统计软件的功能、特点与 SPSS 统计分析的步骤。第二节探讨 SPSS 系统的安装与启动,除介绍 SPSS 12.0 版如何安装与启动外,还探讨了系统的目录规划、SPSS 系统的菜单以及菜单中分析(Analyze)功能与辅助说明(Help)功能。第三节探讨 SPSS 数据文件的建立,介绍数据文件的建立与调用,强调 SPSS 可直接调用其他格式的文件;探讨数据编辑窗口中的 Data View 和 Variable View 功能;并探讨变量名称、变量类型、缺失值、变量定义等,以及与变量和数据定义有关的重要功能。在第三节建立了数据文件后,紧接着即可进行数据的分析(Analyze),并可将 SPSS 分析所用的语句粘贴(Paste)在语句编辑窗口(Syntax Editor)中,且分析的结果将输出在结果输出窗口(Viewer)中。第四节则将介绍三者间的关联与运作。

第一节 SPSS 系统简介

一、SPSS 的发展历程

在学术研究中,常常需要对大量的资料进行统计处理,这是一项细致而烦琐的工作,如果完全依靠手工来进行,工作量较大,且难以保证准确性,也得不到高精度。为了减轻整理和计算大量数据的负担、提高工作效率,必须充分利用现代化的统计工具。随着电脑软件技术的发展,电脑在统计分析数据方面发挥了相当大的作用,它功能多、速度快、计算精确、较易利用,并且电脑统计软件可以完成更为精确的资料分析与统计计算。

目前常见的统计软件有 SPSS 与 SAS 统计分析系统,这些软件都有大型电脑和个人电脑的版本,可方便统计数据的工作。善用这些软件包,一则可加速处理复杂又庞大的信息,二则可以使统计分析者能更进一步地计算各种统计量与检验结果。这些软件包在统计功能上并没有太大的差别,但在使用的方法上,SPSS 具有亲切的使用者界面,对于一般使用者较易学习,且价格也较为便宜,故目前较为风行。

SPSS(Statistical Program for Social Sciences,社会科学统计软件)是一套统计分析系统程序包,由 Norman H. Nie、C. Hadlai (Tex) Hull 和 Dale H. Bent 等为进行数据分析所发展出来的软件语言。SPSS 是社会科学研究人员首选的统计软件,也是目前世界上最流行的统计软件之一。因而被广泛应用于社会科学和自然科学的各个领域中。它是一个组合式软件包,兼有数据管理、统计分析、统计绘图和统计报表功能,接口友善、使用简单,广泛用于教育、心理、医学、管理、人口、体育、农业、林业、商业、金融、保险等研究领域。

SPSS 公司陆续购并了 SYSTAT、BMDP、Quantime 等公司,使 SPSS 的产品线更加完整。近年来,SPSS 也往数据仓储与数据探勘领域整合,扩大了 SPSS 产品线的深度与广度,更将它的英文名称更改为 Statistical Product and Service Solution,意为“统计产品与服务解决方案”。

SPSS 公司近年来已将许多公司及其主打产品纳入旗下,从而使原来的单一统计产品开发与销售,转向对企业、教育科研以及政府机构提供全面信息统计决策支持服务,成为走在最新流行的“数据仓储”(Data Warehouse, DW)和“数据探勘”(Data Mining, DM)领域前端的一家综合统计软件公司。

自从 1975 年开始,SPSS 系统主要应用于企业事业单位,但自 1984 年 SPSS 公司首先推出了世界上第一个 PC 版本的统计分析软件——SPSS/PC V1.0 以来,其版本不断更新,软件功能日臻完善,操作越来越简便,与其他软件的整合接口也越来越多,从而确立了其在个人电脑用户市场中第一的地位。至 1993 年 6 月,SPSS

公司正式推出 SPSS for Windows 6.0 版本,该版本不仅修正了以前版本的错误,改写了一些模块使运行速度大大提高,而且根据统计理论与技术的发展,更增加了许多新的统计分析方法,使其功能日臻完善。

SPSS 统计软件采用电子表格的方式输入与管理数据,能方便地从其他数据库中读取数据(如 dBase、Excel、Text File 等)。它的统计过程包括描述性统计、平均值比较、相关分析、回归分析、聚类分析、资料简化、存活分析、多重响应等几大类,每类中又包含同类多种统计过程,比如回归分析中又分线性回归分析、非线性回归分析、曲线估计等多个统计过程,而且在每个过程中都允许用户选择不同的方法及参数进行统计分析,因此,除了可以实现常规的各种统计外,还可用来作一些不常用的分析处理。

二、SPSS 系统模块

以 SPSS for Windows V12.0 为例,完整的 SPSS 系统包括底下十一个模块,这些模块同时具有英文版与中文版。SPSS 的英文版网址为 www.spss.com,在我国台湾地区的中文版由宏德国际软件咨询顾问公司来负责,其网址为 www.sinter.com.tw/SPSS/,有关 SPSS 的产品信息可进一步探询。

SPSS 系统中以 Base 最为重要,为核心且必须安装的模块,而其他子系统则可依使用者的需求来购买与附加。为了协助使用者快速学习 SPSS 系统,以便快速发挥 SPSS 的强大功能,本书将以 Base 这一核心系统为介绍重点,但仍包括其他常用的模块。

(一) SPSS/Base 12.0

使用 Client/Server 架构,其中 ODBC 接口可与各种数据库联结和读取各种格式的数据文件,以及多样化输出(将报表汇出至 Html、XML、text 文件)与多种文件数据的计算、转换与管理。包括统计分析中最受欢迎的统计方法,如描述性统计、多维度报表(OLAP Cube)、交叉分析(Crosstabs)、复回归(Multiple Regression)、独立与成对样本的平均数比较、非参数检验、方差分析(ANOVA)、多变量分析(区别分析、因素分析、聚类分析)、实时互动的高品质 3D 图形、各式统计图表、时间数列图表、信度分析(Reliability)与多维量表分析(Multidimensional Scaling, MDS)等。汇总如下表:

· 简单因子 ANOVA 模块	· 比较平均数	· 次数分析	· 多维量表法
· K 平均数聚类分析	· 区别分析	· 线性回归	· 复选题分析
· 模型建立统计量	· 相关分析	· 曲线估计	· 偏相关分析
· 单因子 ANOVA	· 因素分析	· 综合报表	· 非参数检验
· 描述性统计量	· 交叉分析	· 聚类分析	· Proximities
· 拟合优度分析	· 信度分析	· T 检验	· 其他等

(二) SPSS/Complex Samples 12.0

是一种帮助制定抽样计划、执行抽样(几率抽样),并以样本估计总体参数的系统。可以交互式交谈窗口建立抽样计划,该计划包含抽样方法与样本大小的信息,其中,抽样方法包括简单随机抽样、分层与聚类抽样以及多阶段抽样等。此外,还可用交互式交谈窗口帮您从已经抽样的样本与权数信息中建立抽样计划或再次进行抽样。

(三) SPSS/Regression Models 12.0

在 Base 版中只有 Multiple Regression(包含 Curve Estimation)功能,但在此模块中,则包括下面各种进阶的回归模型,如下所示:

- (1) 二项式与多项式(Binary & Multinomial)、逻辑回归(Logistic Regression)。
- (2) Probit Analysis。
- (3) 非线性回归(Nonlinear Regression)。

(四) SPSS/Advanced Models 12.0

在 Base 版中只有 ANOVA 功能,但在此模块中,尚包括以下各种进阶的分析:

- (1) 一般化线性模型(GLM-General Linear Models, ANOVA、ANCOVA、MANOVA、Repeated Measures)。
- (2) 方差成分估计(Variance Components)。
- (3) Model Selection Loglinear Analysis(Hierarchical,层次性的对数线性模型)。
- (4) 对数线性分析(General Loglinear Analysis)和顺序尺度回归(Ordinal Regression)。
- (5) 生存分析(Survival Analysis)、Kaplan-Meier Survival Analysis、Cox Regression。

(五) SPSS/Tables 12.0

以可视化式拖曳来自定多个变量的表格呈现方式,在一个表格中展示各种信息,制作高品质报表。

- (1) 可产生组合型的复杂表格。
- (2) 制作问卷调查常用的复选题分析表。
- (3) 多个变量在同一个表格展现的次数分布表。

(六) SPSS/Trends 12.0

是一种专为执行时间数列分析的模块,包含:ARIMA、Autoregression、Exponential Smoothing、Seasonal Decomposition、Spectral Analysis。