



非计算机专业计算机公共课系列教材

SPSS统计分析

主编 滕 冲 汪同庆



WUHAN UNIVERSITY PRESS

武汉大学出版社

014035223

C819-43
38

非计算机专业计算机公共课系列教材

SPSS统计分析

主编 滕冲 汪同庆

参编 李珊枝 肖慧 林迎迎 周宸



WUHAN UNIVERSITY PRESS

武汉大学出版社



北航

C1715422

C819-43
38

图书在版编目(CIP)数据

SPSS 统计分析/滕冲,汪同庆主编. —武汉:武汉大学出版社,2014.2
非计算机专业计算机公共课系列教材
ISBN 978-7-307-12819-4

I. S… II. ①滕… ②汪… III. 统计分析—软件包 IV. C819

中国版本图书馆 CIP 数据核字(2014)第 027014 号

责任编辑:林 莉 责任校对:汪欣怡 版式设计:马 佳

出版发行:武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件:cbs22@whu.edu.cn 网址:www.wdp.com.cn)

印刷:湖北恒泰印务有限公司

开本:787×1092 1/16 印张:13.75 字数:347千字 插页:1

版次:2014年2月第1版 2014年2月第1次印刷

ISBN 978-7-307-12819-4 定价:33.00元

版权所有,不得翻印;凡购买我社的图书,如有质量问题,请与当地图书销售部门联系调换。



内 容 提 要

本书讲解了统计分析软件 SPSS 的基本用法。主要内容包括 SPSS 系统概述、SPSS 数据文件的创建和管理、SPSS 数据的预处理、描述性统计分析 with 统计图表、参数检验、非参数检验、相关分析、回归分析和方差分析。

全书体系完整,结构清晰。用通俗简明的语言介绍基本理论,同时,配合了丰富典型的案例增强知识的应用性和可操作性。本书适用高等学校本、专科学生学习统计分析使用,也可为数据统计分析的学习者和研究者提供参考。



前 言



人类已步入大数据时代,凡是有数据的地方,就有数据分析的需求。如何高效地获取有价值的信息,从而作出科学的评估和决策?

SPSS 统计分析软件是在社会学研究、调查统计、市场研究、医学统计、政府和企业的数据分析应用中久享盛名的统计分析软件。在国际学术界的各个领域,其自动统计绘图、数据的深入分析、易学易用、功能齐全等方面得到了高度的评价和广泛的称赞。在国际学术界有条不紊的规定,即在国际学术交流中,凡是用 SPSS 软件完成的计算和统计分析,可以不必说明算法,由此可见其影响之大和信誉之高。

SPSS 软件版本更新很快,本书内容以 SPSS 20.0 中文版为操作平台,以统计分析为目的,讲解了创建数据文件、编辑和管理数据、为后续统计分析打好基础作数据的预处理、描述性统计分析、绘制统计图、假设检验、相关分析、回归分析、方差分析和非参数检验等一系列内容。

全书深入浅出,图文并茂,在简明阐述统计方法的理论与思想的同时,精心设计了大量的案例,给出了详细的操作指导及统计分析结果的解读,真正做到了理论与实践相结合,可操作性强。章节前的引例则是提纲挈领地展示了章节要点,启发式地提出问题,随后解决问题,举一反三。每一章最后附上本章小结,配以思考与练习题来强化重点和加深理解。如此,读者能切实地掌握一门统计分析方法,让自己的实验数据生动起来,为自己的专业领域添砖加瓦,令自己的专业技能如虎添翼。

本书共包括 8 章。其中,第 1 章由汪同庆编写,第 2 章由滕冲编写,第 3 章和第 4 章由肖慧编写,第 5 章和第 7 章由林迎迎编写,第 6 章和第 8 章由李珊枝和周晟编写。在本书编写过程中,陈波在语言表达、全文架构等方面给予了诸多指导和大力支持,在此深表感谢。同时,感谢武汉大学出版社对本书出版给予的支持。

由于编者水平和时间有限,书中难免有错漏之处,敬请广大读者提出宝贵意见,欢迎致邮:tengchong@whu.edu.cn。

编 者

2014 年 2 月

于珞珈山



目 录



第 1 章 SPSS 统计分析软件概述	1
1.1 初识 SPSS	1
1.1.1 SPSS 的发展简史	1
1.1.2 SPSS 的主要特点	2
1.1.3 SPSS 的结构	3
1.2 SPSS 操作入门	6
1.2.1 SPSS 软件的安装和授权	6
1.2.2 SPSS 的启动和退出	6
1.2.3 SPSS 的基本操作方式	7
1.3 SPSS 的界面	8
1.3.1 SPSS 的工作窗口	8
1.3.2 SPSS 的菜单	10
1.4 SPSS 系统选项及常用设置	12
1.5 利用 SPSS 作统计分析的基本步骤	15
本章小结	16
思考与练习	17
第 2 章 SPSS 数据文件的建立和管理	18
2.1 引例	18
2.2 SPSS 数据文件	19
2.2.1 SPSS 数据结构	19
2.2.2 变量属性	19
2.3 数据文件的新建、保存和打开	24
2.3.1 数据文件的新建	24
2.3.2 数据文件的保存	25
2.3.3 数据文件的打开	26
2.4 数据文件的编辑	26
2.4.1 编辑数据结构	26
2.4.2 编辑数据内容	27
2.4.3 编辑格式	29
2.5 数据文件的合并与拆分	30
2.5.1 添加个案	30
2.5.2 添加变量	33



2.5.3	拆分数据文件	37
2.6	外部数据的获取	38
2.6.1	读取 Excel 数据文件	39
2.6.2	读取文本文件	40
2.6.3	读取数据库文件	40
	本章小结	42
	思考与练习	43
第3章	SPSS 数据的预处理	45
3.1	数据预处理的常用操作	45
3.1.1	引例	45
3.1.2	数据排序	46
3.1.3	个案选择	46
3.1.4	变量计算	50
3.1.5	分类汇总	53
3.2	数据校验	54
3.2.1	引例	55
3.2.2	数据验证	55
3.2.3	标志重复个案	60
3.2.4	标志异常个案	62
3.3	可视离散化	65
3.4	数据的重编码	68
3.4.1	重新编码为相同变量	68
3.4.2	重新编码为不同变量	70
3.4.3	自动重新编码	71
3.5	缺失值	73
	本章小结	75
	思考与练习	76
第4章	描述性统计分析与统计图表	78
4.1	引例	78
4.2	频率分析	79
4.3	连续变量的描述性统计分析	82
4.3.1	连续变量的描述性统计量	82
4.3.2	连续变量的描述性统计分析	85
4.4	分类变量的描述性统计分析	94
4.4.1	分类变量的描述性统计量	94
4.4.2	分类变量的描述性统计分析	94
4.5	应用统计图表作描述性统计分析	96
4.5.1	连续变量的图形描述	96



4.5.2 分类变量的图形描述	102
本章小结	104
思考与练习	105
第5章 参数检验	106
5.1 假设检验概述	106
5.1.1 假设检验的基本思想	106
5.1.2 假设检验的相关概念	107
5.1.3 假设检验的一般步骤	108
5.2 均值过程	109
5.3 单样本 T 检验	112
5.3.1 引例	112
5.3.2 数据准备	113
5.3.3 单样本 T 检验的实现	114
5.4 独立样本 T 检验	116
5.4.1 引例	117
5.4.2 数据准备	118
5.4.3 独立样本 T 检验的实现	121
5.5 配对样本 T 检验	123
5.5.1 引例	123
5.5.2 配对样本 T 检验的实现	124
本章小结	126
思考与练习	126
第6章 非参数检验	128
6.1 非参数检验概述	128
6.2 单样本非参数检验	129
6.2.1 卡方检验	129
6.2.2 二项式检验	137
6.2.3 K-S 检验	138
6.2.4 游程检验	141
6.3 独立样本非参数检验	145
6.4 相关样本非参数检验	152
本章小结	158
思考与练习	158
第7章 相关分析与回归分析	160
7.1 相关分析概述	160
7.2 简单相关分析	161
7.3 偏相关分析	166

7.4	回归分析概述	169
7.5	简单线性回归分析	170
7.6	多元线性回归分析	175
	本章小结	178
	思考与练习	178
第8章	方差分析	180
8.1	引例	180
8.2	方差分析概述	181
8.3	单因素方差分析	182
8.3.1	单因素方差分析的基本思想	182
8.3.2	单因素方差分析的实现	184
8.3.3	单因素方差分析的进一步分析	186
8.4	多因素方差分析	195
8.4.1	多因素方差分析基本思想	195
8.4.2	多因素方差分析的实现	195
8.4.3	多因素方差分析的进一步分析	201
	本章小结	208
	思考与练习	209
	参考文献	211



第1章 SPSS 统计分析软件概述



SPSS 软件是世界上应用最广泛的统计分析软件之一。目前,在全球约有 28 万用户,它们分布于电子商务、银行金融、保险证券、市场调研、政府税务、教育科研、医疗卫生、化工行业、制造业、零售业、通信等多个领域和行业,全球 500 强中约有 80% 的公司使用 SPSS 用于数据统计分析。在市场研究和市场调查领域,SPSS 软件有超过 80% 的市场占有率,深受广大用户的青睐。

1.1 初识 SPSS

1968 年,美国斯坦福大学的三位研究生研制开发了最早的统计分析软件包,用于解决社会学研究中的统计分析问题,将其称为“Statistical Package for the Social Sciences”即“社会科学统计软件包”,简称 SPSS。1975 年,SPSS 公司在美国芝加哥成立。

2000 年,随着 SPSS 产品服务领域的扩大和服务深度的增加,SPSS 公司正式将英文全称更改为“Statistical Product and Service Solutions”,意为“统计产品与服务解决方案”,没有改变英文缩写形式——SPSS。

2009 年 4 月,随着 SPSS 公司成为预测分析领域的领导者,SPSS 公司将旗下主要产品的名称前面统一冠以 PASW 字样,全称是“Predictive Analysis Software”,意思是“预测分析软件”。

2009 年 10 月,IBM 宣布成功收购 SPSS 公司,随后的产品不再以 PASW 为名,重新命名为 IBM SPSS Statistics。

虽然,软件历经名称的演变,但是功能定位在统计分析这一点上始终未变,用户已习惯称其为 SPSS。

1.1.1 SPSS 的发展简史

最初的 SPSS 软件主要是面向中小型计算机和企业级用户,统称为 SPSSx 版。1984 年,推出了运行在 DOS 操作系统上的 SPSS 微机第 1 版,与随后推出的第 2 版和第 3 版统称为 SPSS/PC+,开创了 SPSS 微机系列产品的开发方向,从而确立了个人用户市场第一的地位。

20 世纪 90 年代,随着图形界面 Windows 操作系统的出现和普及,称为 SPSS for Windows 的第 5 版、第 6 版相继出现。90 年代中后期,又推出了以交互式对话为主要特征的第 7 版,后续版本纷纷诞生,其功能不断增加和完善。加之,随后的 SPSS 公司大刀阔斧地并购 SYSTAT 公司、BMDP 公司、ISL 公司、Showcase 等一系列同行公司,并将这些公司的主打产品收纳到 SPSS 旗下,从而使 SPSS 公司由原来单一统计产品开发与销售转向为企业、教育科研和政府机构等提供全面信息统计决策支持服务。

新的版本纷纷登场,SPSS 公司大约用一年时间就会推出一款新的版本。11 版软件界面

有了较大的改动、提供多语言输出、结果能直接导出为 XLS 电子表格文件；经典的 13 版加入了树模型等智能统计分析方法、复杂抽样模块、输出接口加入了 OMS 系统；17 版提供了包含简体中文的多语言界面；18 版作了一系列的更新，醒目的 PASW 标记也是昙花一现；19 版是并购后的第一个版本，展现出 IBM 风貌，提供了广义线性模型、增强语法编辑器等；20 版改进结果输出、服务器功能，广义线性模型上增加有序分类因变量的建模功能；21 版在 Base 模块的“分析”菜单下增加“模拟”菜单项，在“数据”菜单下增加“比较数据集”菜单项等。

至 2013 年底，SPSS 22 版已经在国内正式发行，其新增了语法文件加密，增强了编程能力，使得创建 Python 程序更简单或者使用已有的 Python 扩展更方便等功能。

1.1.2 SPSS 的主要特点

SPSS 统计软件的易学易用以及专业化的统计分析结果是 SPSS 软件强大的竞争优势之一，也是广大数据分析者对其特别偏爱的主要原因。大量成熟的统计分析方法、完善的数据定义界面、开放的数据接口以及灵活的统计表格和统计图形，为 SPSS 的长盛不衰提供了有力保障。其主要特点如下：

1. 操作简单

除数据输入工作需要使用键盘外，大多数操作通过菜单、图标按钮、对话框来完成。用户不需要精通统计分析的各种算法，只需要了解统计分析原理，在对话框中告之系统要做什么，即可得到统计分析结果。

2. 良好的数据管理功能

SPSS 具有良好的数据管理功能，包括：定义变量属性，利用值标签来快捷地录入数据，插入变量，数据重组，合并数据文件，汇总数据，重复记录和异常数据的发现等，这些数据管理功能为后续的统计分析提供了很好的基础。

3. 强大的统计功能

用户可根据不同的需要，选择不同的统计模块完成问题的基本统计分析和专业统计分析任务。SPSS 为各种研究提供了全方位的统计学算法，如方差分析、回归分析、多元统计分析方法、生存分析方法等，方法体系覆盖全面。

4. 兼容多种数据格式

SPSS 软件可以与很多其他软件文件格式进行数据传输。可以将日常工作中常用的 Excel 表格数据、文本格式数据、数据库文件等导入 SPSS 中进行分析，不仅节省大量时间，而且可以避免因复制数据引起的错误。SPSS 分析的结果也可直接导出并保存为 Word、文本、网页、Excel 等文件。

5. 丰富的图表功能

用户可根据需要选择不同风格、不同功能的表和图展示数据。SPSS 提供了条图、线图、饼图、直方图、散点图等丰富的图形和多种表格形式供用户选择。

6. 强大的编程扩展功能

在 SPSS 命令语法语言的基础上提供与其他编程语言的结合，例如，用 Python 和 R 语言编写的程序代码，可以管理使用 SPSS 语法所编写的任务流。这种扩展编程功能和特性，令 SPSS 软件如虎添翼，成为了更强大的统计开发平台。



1.1.3 SPSS 的结构

1. SPSS 的客户机/服务器结构

从 SPSS 10 版开始,已发展为 Client/Server 结构的体系。对于大规模数据的分析,用户可以购买 SPSS Server,以利用 Server 的计算能力来解决速度慢、网络不畅等由于大规模数据而引起的问题;对于小规模数据,用户选择 SPSS Client 就可以了。

2. SPSS 的模块结构

无论是 SPSS 客户机还是 SPSS 服务器,都是模块式结构,SPSS 把所有的功能分散为多个模块。用户可以根据分析中可能用到的数据处理和统计分析方法,选择适当的模块进行购买,而不必花更多的钱购买所有模块。

SPSS 的模块数量随版本的不同有所变化,18 版以前的版本,SPSS Statistic Base 是必需的。软件的整个框架、基本的数据获取、数据准备和获取等基本功能都集中在 SPSS Statistic Base 这个模块上,其他模块在它搭建的平台上才能工作。从 18 版起,其余模块也可以脱离 Base 单独存在并运行。但对于普通用户而言,仍然是以 Base 和其余模块的并用更为常见。

下面以 IBM SPSS Statistics 20 为例,了解其包含的 16 个功能模块:

- SPSS Statistics Base

基础模块,管理整个软件平台,管理数据访问、数据处理和输出,并能进行很多种常见基本统计分析。

- SPSS Advanced Statistics

广义线性模型(GZLMS)、广义估计方程(GEES)、混合模型、一般线性模型(GLM)、方差成分估计、MANOVA、Kaplan-Meire 估计、Cox 回归、多因子系统模式的对数线性模型、对数线性模型、生存分析。

- SPSS Bootstrapping

SPSS Statistics Bootstrapping 模块可以更有效地使用小样本量的数据,通过数据自身重采用的功能,让用户可以模拟大样本情况下的采样结果,从而对数据结构特征和偏差有更直接的认识。当某些参数估计或者假设检验值得怀疑时,也可以直接使用该功能,从而以一种更加直观的方式来执行结果的检查。

- SPSS Categories

Categories 是优秀的对应分析程序,用启发性的二维图和感知图可以清晰地看到数据中的关系,从而更完整和方便地分析数据。PROXSCAL 把观测数据分配到“概念空间”的具体位置,可以在低维度空间描述相似性和不相似性,以便对变量间的关系有一个“空间位置”的理解。

- SPSS Complex Samples

如果使用了特别复杂的抽样方案,那么该模块可以计算复杂样本的统计数据,得到更精确的结果。它拥有专门的规划工具和统计方法,提供各种向导来制定取样方案或详细定义样本,并提供专门的技术来解决样本设计,能够减少得出错误或误导性推论的风险。Complex Sample 将抽样设计融入调查分析之中,使复杂抽样数据的总体得到更加有效的统计推论,对于调查、市场、民意研究人员或者社会科学家来说是必不可少的统计工具。

- SPSS Conjoint

SPSS Conjoint 是包含三个相互关联过程的一个系统,用于进行全特征联合分析。联合分

析使研究人员了解消费者的偏好，或在一定产品属性及其水平条件下的产品评定。联合分析考虑研究时应包括的产品属性、考虑属性水平、产品卡片的数量，用正交设计生成一个包含适量产品卡片的正交主效果片段因子设计。Conjoint 帮助市场研究人员和新产品开发部门了解在消费者心目中什么产品属性是重要的，了解最偏爱的属性水平是什么，进行定价研究，进行品牌价格研究。在产品投入大批量生产之前进行这些研究，以避免可能的失误。

- SPSS Custom Tables

Tables 用更深入的分析，轻松地处理复选题与缺失值，用包括所有统计量、易于理解的表格来展现分析结果。SPSS Tables 广泛应用于调查研究和市场研究等领域，交互式创建各种表格(堆积表、嵌套表、分层表等)。

- SPSS Data Preparation

利用 SPSS Statistics Data Preparation，用户可以获得多个简化数据准备过程的程序。这个附加模块能够在预处理数据时轻易地识别虚假的和无效观测、变量和数据值；确认可疑的或者残缺的案例；查看数据缺失模式；描述变量分布以备分析；更准确地应用针对于分类变量的算法；还可以用为分类变量而设计的运算法则来做更多精确的工作。使用 Data Preparation，可以迅速找到多元的极端值，执行数据检验，为建模预处理数据。

- SPSS Decision Trees

Decision Trees 模块基于数据挖掘中发展起来的树结构模型对分类变量或连续变量进行预测，可以方便、快速地对样本进行细分。建立决策树来确认分组并预测结果，利用直觉式的树形图，颜色分类图，和表格协助研究人员轻松确认和评估区隔。它提供四种强大的决策树算法(CHAID、Exhaustive CHAID、CART 和 QUEST)，提供训练数据及测试数据的模型评估比较，提供 Gain、Response、Index、Profit、ROI 图，帮助评估风险及效益。

- SPSS Direct Marketing

Direct Marketing 主要用来处理市场直销中的一些分析需求。目前提供 RFM 客户评分，客户分群，目标客户轮廓概括，客户响应评分，不同营销行为响应测量等模型。

RFM 是客户价值分析的一种常用方法，此方法使用客户的购买时期，购买频率和消费金额数据对客户进行一些区隔和客户价值分析。客户分群根据客户的不同特征，将客户划分成不同的特征群体。可用来进行市场细分和客户分群。目标客户轮廓概括，用来概括某种不同目标划分(是否购买某产品)的客户群的特征，帮助快速掌握不同客户的特征。客户响应评分，可以根据不同的特征群来判断客户某种行为的倾向，并给予倾向性的评分。

- SPSS Exact Tests

为了确定现有变量之间的关系，研究人员经常首先查看交叉表和非参数检验中的 p-值。如果数据符合假设条件，那么用传统的计算方法是可行的。但是，如果数据属于小样本或零星的数据，又将用户的资料细分到多个类别，或数据变量中有超过 80% 的观测值集中在某一类别，传统方法算出的答案便会不正确。在传统情况下，要得到更好更正确的结论，每一单元需要有五个以上数据，SPSS STATISTICS Exact Tests 为用户解除了这种限制。

超过 30 个精确检验涵盖了小型或大型数据集所有的非参数和分类数据问题。包括独立或相关样本的单样本、两样本和 K-样本检验，拟合度检验，RxC 列联表独立性检验和联合测度检验等。无论用户的数据结构为何，SPSS STATISTICS Exact Tests 都能给用户正确的 p-值，为用户提供可信赖的结果。如果需要做调查研究、医学研究、生物统计、社会科学研究、市场调查或各类的实验，那么用户需要使用 SPSS STATISTICS Exact Tests。



• SPSS Forecasting

Forecasting 是目前功能最强的时间序列分析工具，是分析历史资料、建立模型与预测未来事件的强有力的工具，能帮助研究人员做更好的预测。Forecasting 利用完备的时间序列提高预测能力，包括多重曲线拟合、平滑以及自回归方程估计。利用专家建模器，可自动从 ARIMA 和指数平滑模型中选择最佳拟合用户的时间序列和因变量的模型，避免反复选择模型的工作。Forecasting 使用这些高级建模技术，用深入的建模支持、丰富的诊断方法、图形界面使分析变得简单容易。预测模块提供了一流的评估过程，用简单、有效的平滑技术进行高质量的预测，使用 SPSS STATISTICS Forecasting 可完成多种任务，包括生产管理、数据处理、预算管理、公共政策研究等。

• SPSS Missing Values

缺失数据会带来偏差或错误的分析结果，简单代入法或者简单的回归法都不能正确地填补缺失值，SPSS STATISTICS Missing Values Analysis 帮助研究人员在分析过程中排除数据中隐含的偏差，得出更精确的结论。

Missing Values Analysis 用六种灵活的诊断报告来评估缺失值是否会影响分析结论，更好地了解它们的特性。它通过快捷地诊断缺失值，得到更精确的摘要统计量，方便地用估计值替换缺失值，得到精确的结论。

• SPSS Neural Networks

神经网络是一种模拟人类大脑处理信息的方式的简化算法模型。通过多个神经元层的输入输出运算给出一个判断结果。和传统的模型不同，每个神经元节点内部运算是一个黑箱，我们能清晰获得的是结果而不能精确了解内部的过程。

通常在神经网络中有三个部分：一个输入层；一个或多个隐藏层；一个输出层。这些单元通过可变的连接强度(或权重)连接。输入数据显示在第一层，其值从每个神经元传播到下一层的每个神经元。最终从输出层中输出结果。

神经网络可检查输入的记录，并为每个记录生成预测，通过预测结果正确与否进行判断和调整，在达到某个标准之前，神经网络可以进行自学习来不断改进模型，此过程会不断重复，持续提高其预测准确度。

神经网络执行预测分析的能力不逊于其他传统技术，并且只需很少的统计或数学知识即可进行应用。神经网络对字段类型没有限制，可以处理多种类型的格式。能够广泛应用于各个领域进行预测分析，如银行保险的欺诈判断，客户的反馈判断等。

• SPSS Programmability Extension

SPSS Programmability Extension 模块可以让用户使用外部语言来执行 SPSS 一连串的分析动作以达到自动化的目的。用户可藉由撰写内嵌在 SPSS 18.0 里的 Python 程序来控制 SPSS 的各式语法工作执行，如执行设定变量属性、观察程序输出、错误码或条件状态等。运用外部程序结合新的后端 API 处理，可扩大语法执行工作的弹性。透过条件执行式(If/Then/Else)、循环控制式(For/While)，这些几乎都会出现在外部语言的叙述，套用在新增的 SPSS Programmability Extension 模块上，使得控制语法工作的程序更有效率、更方便。

• SPSS Regression

大量的非线性建模工具、多维尺度分析帮助研究人员进行非线性回归分析。它将数据从数据约束中解放出来，方便地把数据分成两组，建立可控制的模型及表达式进行非线性模型的参数估计，能够建立比简单线性回归模型更好的预测模型。



多项式逻辑斯蒂回归(MLR)帮助用户建模分析哪些因素可以用于预测客户会购买产品A、产品B还是产品C,使用户不再局限于是/否的二分类问题。二元逻辑斯蒂回归(BLR)帮助用户找到最好的模型来预测二元的结果,如“选”或“不选”。它在每一步骤中提供多种方式来选择能最好地预测响应变量的主效应和交互效应。Probit和Logit响应模型用于分析对某一刺激(如药物剂量、价格和激励机制)的响应力度。Probit过程通过Logit或Probit估算导致一定响应比例所需的刺激强度。

1.2 SPSS 操作入门

本书基于IBM SPSS Statistics 20版编写,如果没有特别指明,文中SPSS均指IBM SPSS Statistics 20,以下不再说明。

1.2.1 SPSS 软件的安装和授权

SPSS安装文件分为32位和64位两种,其在Windows系统下的安装类似其他软件的安装,启动安装程序,按照界面引导完成安装操作。在较老的版本中,需要选择希望安装的模块,而在新版本中,模块都是默认全部安装的。

SPSS软件是一个授权许可软件,如果未进行软件授权确认,那么只能获得一定的试用期,试用期满后,软件会被禁用。用户需要在“开始”菜单中找到IBM SPSS Statistics组,运行其中的“IBM SPSS Statistics 20许可证授权向导”,在联网状态下输入授权码激活软件后,购买的模块就可以正常使用了。

1.2.2 SPSS 的启动和退出

SPSS软件安装完毕后,系统自动在“开始”菜单中创建“IBM SPSS Statistics”组,选择执行其中的启动项“IBM SPSS Statistics 20”命令,即可启动SPSS软件,如图1-1所示为SPSS初始画面:数据编辑窗口和使用向导。



图 1-1 SPSS 初始画面

用户可在向导对话框下选择所需的操作,如果在以后的启动中,不希望该向导对话框出



现, 可以选中左下角的“以后不再显示此对话框”复选框, 并单击“确定”按钮即可。

如果要退出 SPSS 软件, 那么可以选择菜单“文件/退出”, 或单击窗口右上角的关闭按钮, 均可退出 SPSS。

1.2.3 SPSS 的基本操作方式

SPSS 为用户提供了四种基本操作方式, 即菜单对话框的操作方式、程序运行方式、混合运行方式和生产作业方式。

1. 菜单对话框的操作方式

用户在使用 SPSS 的过程中, 所有的数据录入和分析操作过程大多数通过选择菜单项、对话框引导的方式完成。

菜单对话框的操作方式是一种最常见和普遍的管理方式, 其最大的优点是图形用户界面友好、操作简单、形象直观。用户不需要任何计算编程的概念, 只要熟悉 Windows 的基本操作, 如复选框、单选框、下拉框和对话框等并懂得相应的统计知识, 就可以非常方便地完成统计分析工作。

该方式的缺点是, 对于大量的重复统计分析工作, 需要将相应菜单和所需参数操作一遍, 因此, 处理速度较慢且容易产生操作和输入错误。然而, 瑕不掩瑜, 菜单对话框的操作方式仍然是一般的统计分析人员和 SPSS 初学者选择的主要操作方式。后面的讲解主要以该方式为主。

2. 程序运行方式

程序运行方式是指, 用户在使用 SPSS 的过程中, 统计分析人员首先根据自己的分析需要, 在语法编辑窗口中输入程序, 提交给 SPSS 系统执行, 无需人工干预, SPSS 会自动按照程序命令语句的前后顺序, 自动逐句执行相应的命令并给出结果。例如: 对“数据集 1”, 按“教育水平”为类别作频数分析, 则可以编辑如下程序:

```
DATASET ACTIVATE 数据集 1.  
FREQUENCIES VARIABLES = 教育水平  
/ORDER = ANALYSIS.
```

在语法编辑窗口中, 如果要调用其他语法文件时, 那么也可以使用 include 命令。例如: 要调用 sample. sps 文件时, 只要在语法编辑窗口输入如下命令:

```
Include 'd: \ SPSS \ MYFILE \ sample. sps'
```

这样使得已有的代码段以便捷的方式得到重复利用。

3. 混合运行方式

混合运行方式是指在使用菜单对话框的操作方式的同时, 还可以编辑 SPSS 程序, 是以上两种方式的结合。

用户首先按照菜单对话框的操作方式, 选择统计分析的菜单和选项, 选择完毕后, 在对话框中单击“粘贴”按钮, SPSS 系统把相应的操作转换成 SPSS 的命令程序, 并粘贴到语法编辑器窗口, 用户选择菜单“运行/全部”命令, 完成程序的执行工作。如图 1-2 所示为混合运行方式。

可见, 混合运行方式弥补了菜单对话框的操作方式中每步分析操作都要人工干预的不足, 同时摆脱了程序运行方式中必须熟记 SPSS 命令和参数的制约, 因此是一种较为灵活且实用的操作方式。另外, 对于熟练的 SPSS 程序员, 可以借助该方式在程序中书写菜单和选

项中没有提供的参数。

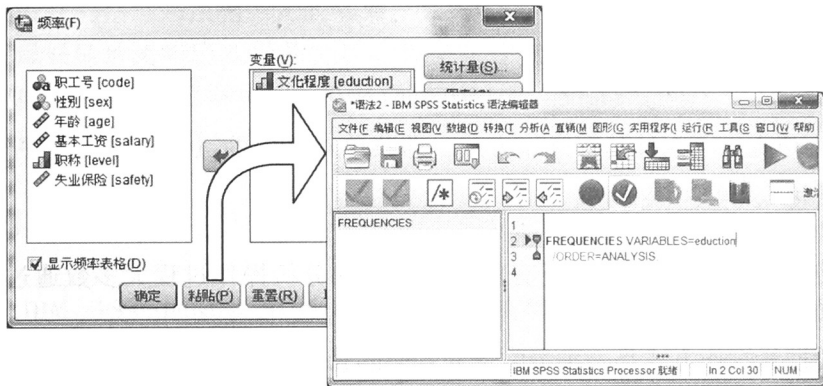


图 1-2 混合运行方式

4. 生产作业方式

SPSS 提供了程序全自动运行的生产作业 (Production Facility) 方式。在生产作业方式下, 程序在无人看管的情况下运行, 并在执行最后一条命令后终止。在 SPSS 菜单中, 选择“实用程序/生产设施”命令, 在弹出的对话框中设置: 需自动执行的一个或多个程序文件, 执行方式和结果输出方式等参数。

对于常常运行相同的一组耗时较长的分析, 例如: 周报告, 例行月度统计或预测, 采用生产作业方式使任务在无人看管且预定的时间自动运行。运行结束后, 系统弹出对话框提示任务完成, 并有设置的结果文件生成。

这四种运行方式各有千秋, 实际中应根据应用分析的需要和对 SPSS 的掌握程度进行合理的选择。

1.3 SPSS 的界面

SPSS 的界面主要指 SPSS 的工作窗口与 SPSS 的菜单。

1.3.1 SPSS 的工作窗口

SPSS 的工作窗口有数据编辑窗口、语法编辑窗口、结果输出窗口和脚本编辑窗口四种。其中数据编辑窗口和结果输出窗口是最常用到的两个窗口。

1. 数据编辑窗口 (Data Editor)

数据编辑窗口, 也称数据编辑器。启动 SPSS 后, 系统会自动打开数据编辑窗口。SPSS 处理数据的主要工作全在此窗口中进行, 该窗口的左下角显示两个视图: 数据视图 (Data View) 和变量视图 (Variable View), 数据视图用于显示数据, 变量视图则显示变量的信息: 变量名称、变量的类型、变量的格式等, 两个视区可以通过鼠标单击进行切换。

当同时打开多个数据文件时, 每个数据文件独占一个不同的数据窗口, 系统会对这些数据窗口自动按照“数据集 0”、“数据集 1”……这样的工作名称来以示区分。如图 1-3 所示, 数据编辑窗口, 可以用来创建、显示和编辑数据文件。