

● 高等学校教材

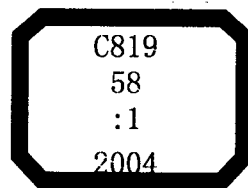
SPSS

统计分析 基础教程

主 编 张文彤 闫 洁



高等教育出版社



高等学校教材

SPSS 统计分析基础教程

主 编 张文彤
闫 洁

高等教育出版社

内容简介

SPSS 是最为优秀的统计软件之一,深受各行业用户的青睐。为满足广大读者学习统计学入门知识和统计软件入门操作的需求,本书改变了以往 SPSS 书籍对统计理论和软件操作“两条主线、各自表述”的编写方式,将两者完全融合起来。全书共分 15 章,以 SPSS 12.0 为准,针对统计初学者和 SPSS 初级用户的需求,以统计理论为主线,详细介绍了在 SPSS 中的界面操作、数据管理、统计图表制作、统计描述和常用单因素统计分析方法的原理与实际操作。其内容覆盖了目前国内大部分专业本科统计课程的教学范围,并结合 SPSS 的强大功能做了很好的扩展。各章后均附有参考文献和思考练习题,涉及统计理论的章节还提供了本章小结。全书内容深入浅出,风格简洁明快,是一本难得的统计理论与 SPSS 操作相结合的教材。

本书可用作各专业本科生和研究生的统计学教材,也可作为 SPSS 10~12 版的通用入门教材,可供各行业中非统计专业背景的人员以及希望从头学习 SPSS 软件的人员使用。

图书在版编目(CIP)数据

SPSS 统计分析基础教程 / 张文彤, 闫洁主编. —北京:
高等教育出版社, 2004.9(2006 重印)

ISBN 7-04-015855-8

I. S... II. ①张... ②闫... III. 统计分析 - 软件
包, SPSS - 高等学校 - 教材 IV. C819

中国版本图书馆 CIP 数据核字(2004)第 087691 号

策划编辑 耿 芳 责任编辑 欧阳舟 市场策划 韩 飞 封面设计 于文燕
版式设计 张 岚 责任校对 朱惠芳 责任印制 宋克学

出版发行 高等教育出版社
社 址 北京市西城区德外大街 4 号
邮政编码 100011
总 机 010-58581000

经 销 蓝色畅想图书发行有限公司
印 刷 高等教育出版社印刷厂

开 本 787 × 1092 1/16
印 张 24
字 数 580 000

购书热线 010-58581118
免费咨询 800-810-0598
网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
网上订购 <http://www.landaco.com>
<http://www.landaco.com.cn>
畅想教育 <http://www.widedu.com>

版 次 2004 年 9 月第 1 版
印 次 2006 年 12 月第 5 次印刷
定 价 32.00 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

版权所有 侵权必究

物料号 15855-00

SPSS 统计分析基础教程参编人员

主 编 张文彤(复旦大学)
 闫 洁(博塔(中国)有限公司)

编 者 (以姓氏笔画为序)
 王 莉(天津财经大学)
 邝春伟(华东师范大学)
 行智国(博塔(中国)有限公司)
 邹艳辉(博塔(中国)有限公司)
 赵 杨(南京医科大学)
 赵新平(复旦大学)
 高 峻(复旦大学)
 董 伟(博塔(中国)有限公司)

序 言

知识经济时代,数据成为宝贵的经济资源。在国外,电信、医疗、银行、证券、保险、制造、商业、科研、教育等行业已广泛采用先进的统计分析技术从数据中提取有价值的信息和知识。在国内,随着市场的蓬勃发展,了解成熟的统计分析产品,借鉴成功的统计分析行业应用经验,运用科学的统计分析方法,从数据中总结、归纳有用的知识,并将知识用于市场营销、运营决策和信用风险管理等领域,帮助企业、事业单位降低消耗、增加效益,从而提高整体运行效率,已成为各行业中有远见的人士普遍关注的问题。

SPSS 软件是全球专业统计分析软件的领导者,一直致力于帮助企业事业单位提高科学运用统计分析方法的能力,20 世纪 80 年代就已经被许多学者引入中国市场。它包含了丰富的统计分析算法,而且在使用层面上更多地考虑了客户在整个统计分析过程中的应用感受,其简洁的界面、完善的数据准备功能和杰出的图表输出能力使得 SPSS 软件在全球有超过 25 万家的机构用户,并成为国内的主流统计分析软件。随着该产品在各行业应用的深入以及 SPSS 培训和认证的广泛开展,目前国内已涌现出一大批应用 SPSS 的专家。近两年,国内统计分析市场如火如荼,SPSS 在产品技术上也不断推陈出新,继 SPSS 英文版在国内成功应用之后,SPSS 公司在今年首次推出了简体中文版,该产品更加符合中国人的使用习惯,一经推出便受到各行业人士的喜爱。

为了让中国的 SPSS 软件爱好者更好地使用软件,我们在经过两年的筹备后向市场推出了 SPSS 统计分析大型丛书。该丛书是一套全面了解、认识和应用 SPSS 最新统计分析软件、掌握统计分析方法的专业书籍,以统计分析在国内的应用现状为切入点,本着学以致用原则,在介绍统计理论及 SPSS 软件功能模块的同时,更侧重于统计分析在各项工作中的实际应用,引导读者不仅掌握 SPSS 软件及技术原理,而且学会运用统计方法解决工作和学习中的实际问题。

该丛书由业内权威专家主笔编写,资料引用详实可靠,实例剖析切中肯綮,不仅融合了行业专家在统计应用领域多年的研究成果,而且还融入了很多 SPSS 软件新、老行业用户的实际应用经验。丛书总结了 SPSS 软件在各行业的实践应用状况,并结合 SPSS 最新行业应用方案,使各行业读者能通过学习提高 SPSS 软件的运用能力,解决工作中的实际问题。在丛书编写过程中,SPSS 公司和博塔(中国)有限公司(SPSS 中国地区分销商)的技术专家还及时提供了国际最新的行业发展信息和 SPSS 最新产品和技术信息,并结合 SPSS 的全球应用状况提出了宝贵意见。

丛书将分期分批出版相应的分册,其中首批面世的为通用教材《SPSS 统计分析基础教程》、《SPSS 统计分析高级教程》,均由复旦大学张文彤老师主编,全国多所高校的统计教师和统计专业人士参编。张文彤老师长期以来一直致力于积极推进统计分析工具在国内的普及应用,他在 2002 年编著的前作《SPSS 11 统计分析教程》基础篇和高级篇因内容翔实、风格独特,受到了广大读者的热烈欢迎,并被多所高校列为本科生或研究生教材,其中基础篇一书已通过教育部评审,成为 2003—2004 年度教育部研究生工作办公室推荐的“研究生教学用书”。他不仅拥有丰富的教学经验,而且熟谙统计分析产品,本次全新编写的这两册教材分别针对不同读者群,由浅入深、

II 序 言

结合实际应用全面介绍了 SPSS 产品和应用。全书实例引用突出,分析讲解透彻,读者可由本书管窥全套丛书“应用为本”的特色。

显然,上述两本书还不能完全覆盖 SPSS 的所有应用领域,因此,本套书从现在还有数本分册正在组织编写中,包括全新的行业应用分册。这里我们热忱邀请各行各业的 SPSS 资深用户,以及各高校的统计教师加入到本套丛书的编写工作中来,以共同推动我国各行业统计应用水平的迅速提高。

希望本套丛书能够让读者更清晰地了解统计分析,从而进一步促进统计分析在国内的普及。为便于读者交流和使用本套丛书,这里特公布相关网址如下:

SPSS 最新版本的全模块试用版下载:www.spssbj.com.cn

丛书相关案例数据下载:www.spssbj.com.cn、www.MedStatStar.com

读者答疑、经验交流:www.spssclub.com、www.StatStar.com

博塔(中国)有限公司 SPSS 丛书编委会

前 言

笔者前作《SPSS 11 统计分析教程》(基础篇)和《SPSS 11 统计分析教程》(高级篇)自 2002 年中面世以来,因其内容翔实、风格独特,受到了广大读者的热烈欢迎,这从读者用 E-mail 请教问题的数量即可看出,还有数位读者详细指出了书中的用字错误,可见其阅读的详细程度,在此笔者对读者深表谢意。同时,数所高校均将其作为本科生或者研究生教材,而基础篇一书已通过教育部评审,成为 2003—2004 年度教育部研究生工作办公室推荐“研究生教学用书”,这无疑都是对前作质量的充分肯定。

但是,再优秀的作品也有其生命周期,随着时间的推移,上述著作的不足之处也逐渐显现出来。对于基础篇而言,突出表现为以下两点:

1. 由于在 2002 年国内尚无系统、完整介绍 SPSS 统计功能及其操作界面的书籍,前作最终将风格定位在操作字典上,使读者能够全面了解 SPSS 各方面的分析功能及界面操作方法,应当说前作很好地完成了这一任务。但是现在随着 SPSS 中文版的正式面市,软件的界面操作已不是主要问题,再来编写这种新华字典式的教材已无必要。读者自然希望能够有一个更好的教材体系出现。

2. 对于统计软件教材而言,其本质应当是统计教材,软件仅仅是实现工具。前作顺应当时的需要,主要满足的是已学习过统计理论知识,但不了解如何使用统计软件来实现的读者群的需求,因此书中并未详细阐述统计基础知识。对于统计初学者而言,需要有一本统计教材与之配合。但是在几年的使用后,现在多所高校均希望直接采用基础篇进行本科生的统计教学,而不需要和其他教材配合使用。在软件工具已经越来越易用的时候,教材的最终归宿应当是以统计知识为主线,这无疑是我们重新投入编写工作的最大动力。

综上,在充分考虑了读者们的意见后,笔者毅然决定推翻原有的框架,完全从头编写基础教程。这本新的基础教程以 SPSS 12.0 为准,定位为统计软件和统计学入门书籍。他针对统计初学者和 SPSS 初级用户的需求,以统计理论为主线,严格按照本科生统计学教材方式编写,内容共分三大部分:第一部分的任务是 SPSS 操作入门,讲解了软件使用和数据管理的操作知识;第二部分以统计理论为主线,详细阐述了如何在 SPSS 中完成数据的统计描述和参数估计,以及如何使用统计图表来进行数据的完美呈现;第三部分则详细介绍了 t 检验、方差分析、秩和检验、 χ^2 检验、相关回归等常用的基本统计分析方法,内容覆盖了目前国内大部分专业统计课程的教学范围,并结合 SPSS 的强大功能作了很好的扩展。书后的附录根据初学者的特点加以编制,各章后均附有参考文献和思考练习题,第三部分的章节还专门提供了本章小结,更好地满足了本科生学习的需要。另外,书中大部分表为设计表格时自动生成的。因此,大部分表及表题为英文。

除作为教材外,本书还适用于各行业中非统计专业需要使用统计方法的人员,以及希望从头学习 SPSS 软件的人员。我们希望广大读者能一如既往地踊跃提出自己使用中的宝贵意见和建议,使得本书再版的时候能够更上一层楼,更完美地满足大家的学习和工作需求。

张文彤

2004 年 7 月于复旦公共卫生学院

目 录

第一部分 数据管理与软件入门

第1章 数据分析概述与软件入门	3
1.1 SPSS 软件概述	3
1.1.1 软件的基本特点	3
1.1.2 SPSS 的 Client/Server 结构	4
1.1.3 SPSS 的模块式结构	5
1.1.4 SPSS 的安装	6
1.2 SPSS 操作入门	8
1.2.1 SPSS 软件的启动与退出	8
1.2.2 SPSS 的 5 个窗口	8
1.2.3 SPSS 的 4 种运行方式	10
1.2.4 SPSS 的 4 种结果输出	13
1.2.5 SPSS 的帮助系统	16
1.3 数据分析概述	20
1.3.1 数据分析方法论介绍	20
1.3.2 SPSS 系列产品对数据分析流程的支持	21
1.3.3 本书内容介绍	22
思考与练习	22
参考文献	22
第2章 数据录入与数据获取	23
2.1 数据格式概述	23
2.1.1 统计软件中数据的录入格式	23
2.1.2 变量属性介绍	23
2.2 数据的直接录入	27
2.2.1 操作界面说明	27
2.2.2 开放题和简单单选题的录入	28
2.2.3 多选题的录入	31
2.3 外部数据的获取	34
2.3.1 电子表格数据如何导入 SPSS 中	34
2.3.2 文本数据如何导入 SPSS 中	36
2.3.3 数据库格式数据如何导入 SPSS 中	37
2.4 数据的保存	39
2.4.1 存为 SPSS 格式	40
2.4.2 存为其他数据格式	40
思考与练习	41
参考文献	41
第3章 数据管理	42
3.1 变量级别的数据管理	42
3.1.1 计算新变量	42
3.1.2 对变量值进行分组合并	45
3.1.3 连续变量的可视化分段	47
3.1.4 将字符变量转换为数值变量	49
3.1.5 变量的编秩	50
3.1.6 Transform 菜单中的其他功能	51
3.2 文件级别的数据管理(一)	52
3.2.1 记录排序	52
3.2.2 记录拆分	53
3.2.3 记录筛选	54
3.2.4 记录加权	55
3.2.5 数据汇总	56
3.3 文件级别的数据管理(二)	57
3.3.1 数据字典的定义与应用	57
3.3.2 查找重复记录	60
3.3.3 数据文件的重新排列与转置	62
3.3.4 多个数据文件的合并	68
思考与练习	71
参考文献	71

第二部分 统计描述与统计图表

第4章 连续变量的统计描述与

参数估计 75

4.1 连续变量的统计描述概述	75
4.1.1 统计描述中可用的工具	75
4.1.2 连续变量的统计描述指标体系	76
4.1.3 SPSS 中的相应功能	77
4.2 集中趋势的描述指标	78
4.2.1 算术均数	78
4.2.2 中位数	80
4.2.3 其他集中趋势描述指标	80
4.3 离散趋势的描述指标	81
4.3.1 全距	82
4.3.2 方差和标准差	82
4.3.3 百分位数、四分位数与四分位数 间距	83
4.3.4 变异系数	84
4.4 连续变量统计描述实例	85
4.4.1 数据背景介绍	85
4.4.2 使用 Explorer 过程进行分析	85
4.4.3 使用其他过程进行分析	88
4.5 连续变量的参数估计	90
4.5.1 正态分布	90
4.5.2 参数的点估计	93
4.5.3 参数的区间估计	94
思考与练习	96
参考文献	96

第5章 分类变量的统计描述与参数

估计 97

5.1 分类变量的统计描述概述	97
5.1.1 分类变量的统计描述指标体系	97
5.1.2 分类变量的联合描述	99
5.1.3 SPSS 中的相应功能	100
5.2 分类变量统计描述实例	100
5.2.1 使用 Frequencies 过程输出 频数表	100
5.2.2 使用 Crosstabs 过程输出列联表	101

5.3 多选题的统计描述	103
5.3.1 多选题的描述指标体系	103
5.3.2 分析实例	104
5.4 分类变量的参数估计	107
5.4.1 二项分布的参数估计	107
5.4.2 其他分布类型简介	109
思考与练习	110
参考文献	111

第6章 数据的报表呈现(上) 112

6.1 SPSS 报表概述	112
6.1.1 SPSS 中的报表功能	112
6.1.2 报表的基本绘制步骤	113
6.2 表格入门	114
6.2.1 表格的基本框架	114
6.2.2 表头、数据区与汇总项	116
6.2.3 单元格的数据类型	116
6.2.4 几种基本表格类型	117
6.3 用 Original Tables 模块制表	119
6.3.1 功能简介	119
6.3.2 Basic Tables 过程	119
6.3.3 General Tables 过程	126
思考与练习	130
参考文献	131

第7章 数据的报表呈现(下) 132

7.1 用 Custom Table 模块自由制表	132
7.1.1 操作主界面	132
7.1.2 简单分析实例	133
7.1.3 其他选项卡功能	138
7.2 表格的编辑	140
7.2.1 基本编辑操作	140
7.2.2 主要编辑菜单功能介绍	143
7.2.3 表格属性的详细设置	146
7.3 表格高级应用技术	147
7.3.1 模板技术	147
7.3.2 在报告中直接使用 SPSS 表格	150
7.3.3 如何解决表格的中文兼容问题	151

7.3.4 宏技术与 OMS 系统简介	152		
思考与练习	154		
参考文献	154		
第8章 数据的图形展示(上)	156	第9章 数据的图形展示(下)	195
8.1 统计图概述	156	9.1 线图	195
8.1.1 统计图的基本结构	156	9.1.1 简单线图	195
8.1.2 统计图的种类	158	9.1.2 多线图、垂线图与对数线图	196
8.1.3 SPSS 12 的常规统计图 功能简介	162	9.1.3 线图的编辑	198
8.1.4 交互式绘图简介	164	9.1.4 交互式点图、线图、条带图与 垂线图	200
8.2 直方图与茎叶图	166	9.2 散点图	201
8.2.1 常规图中的直方图	166	9.2.1 简单散点图	201
8.2.2 直方图的编辑	167	9.2.2 散点图矩阵与重叠散点图	202
8.2.3 用交互图绘制累积直方图与直方 图组	172	9.2.3 三维散点图	203
8.2.4 茎叶图	175	9.2.4 散点图的编辑	206
8.3 箱图	177	9.3 其他统计图	208
8.3.1 常规图中的箱图	177	9.3.1 P-P 图和 Q-Q 图	208
8.3.2 箱图的编辑	179	9.3.2 ROC 曲线	210
8.4 饼图	181	9.3.3 面积图	213
8.4.1 常规图中的简单饼图	181	9.3.4 Pareto 图	213
8.4.2 饼图的编辑	182	9.3.5 误差图	214
8.4.3 用交互图绘制复式饼图和散点 饼图	184	9.3.6 控制图	215
8.5 条图	187	9.3.7 高低图	217
8.5.1 简单条图	188	9.3.8 时间序列分析中使用的图形	218
8.5.2 复式条图、分段条图与百分条 图的绘制	189	9.4 交互式统计图的编辑	218
8.5.3 条图的编辑	190	9.4.1 编辑界面概述	218
8.5.4 用交互图绘制带误差线的条图	191	9.4.2 图形管理员	220
思考与练习	193	9.4.3 变量的重新分配	223
参考文献	194	9.4.4 Utility 工具栏的其他选项	223
		9.5 SPSS 绘图中的注意事项	224
		9.5.1 汉字兼容性问题的解决	224
		9.5.2 默认图形格式的更改	224
		9.5.3 图形模板的应用	225
		思考与练习	227
		参考文献	227

第三部分 常用假设检验方法

第10章 分布类型的检验	231	10.1.3 假设检验的两类错误	233
10.1 假设检验的基本思想	231	10.1.4 假设检验中的其他问题	235
10.1.1 问题的提出	231	10.2 正态分布检验	235
10.1.2 假设检验的基本思想	232	10.2.1 K-S 检验的原理	235

IV 目 录

10.2.2 分析实例	236	的解释	269
10.3 二项分布检验	238	12.2.5 分析实例	269
10.3.1 二项分布检验的原理	238	12.3 各组均数的精细比较	271
10.3.2 分析实例	238	12.3.1 方法原理	271
10.4 游程检验	239	12.3.2 分析实例	272
10.4.1 游程检验的原理	239	12.3.3 事先计划的比较	274
10.4.2 分析实例	240	12.4 组间均数变化的趋势检验	275
10.5 本章小结	243	12.5 本章小结	277
思考与练习	243	思考与练习	277
参考文献	243	参考文献	278
第 11 章 连续变量的统计推断(一) ——		第 13 章 有序分类变量的统计推断 ——	
<i>t</i> 检验	244	非参数检验	279
11.1 <i>t</i> 检验基础	244	13.1 非参数检验概述	279
11.2 样本均数与总体均数的比较 ..	246	13.1.1 非参数检验的意义	279
11.2.1 分析实例	246	13.1.2 非参数检验预备知识	280
11.2.2 单样本 <i>t</i> 检验中的其他问题	248	13.2 两个配对样本的非参数检验 ..	281
11.3 成组设计两样本均数的比较 ..	248	13.2.1 方法原理	281
11.3.1 方法原理	248	13.2.2 分析实例	283
11.3.2 分析实例	249	13.2.3 确切概率的计算	285
11.3.3 适用条件与方差齐性检验	251	13.3 两个独立样本的非参数检验 ..	286
11.4 配对设计样本均数的比较	253	13.3.1 Mann-Whitney <i>U</i> 检验	286
11.4.1 方法原理	253	13.3.2 分析实例	287
11.4.2 分析实例	253	13.3.3 其他两样本非参数检验方法	288
11.5 本章小结	255	13.4 多个独立样本的非参数检验 ..	289
思考与练习	256	13.4.1 方法原理	289
参考文献	256	13.4.2 分析实例	290
第 12 章 连续变量的统计推断(二) ——		13.4.3 多个样本的两两比较	291
单因素方差分析	257	13.5 多个相关样本的非参数检验 ..	292
12.1 方差分析入门	257	13.5.1 Friedman 检验	292
12.1.1 为什么要进行方差分析	257	13.5.2 分析实例	293
12.1.2 方法原理	258	13.5.3 Kendall 协和系数检验与 Cochran 检验	294
12.1.3 单因素方差分析的应用条件	261	13.6 秩变换分析方法	296
12.1.4 单因素方差分析的 SPSS 实现 ..	263	13.6.1 原理简介	296
12.2 均数间的多重比较	266	13.6.2 应用实例	296
12.2.1 直接校正检验水准	266	13.7 本章小结	299
12.2.2 专用的两两比较方法	267	思考与练习	299
12.2.3 两两比较方法的选择策略	268	参考文献	300
12.2.4 多重比较结果出现矛盾时			

第 14 章 无序分类变量的统计推断——	
χ^2 检验	302
14.1 χ^2 检验基础	302
14.1.1 χ^2 检验原理	302
14.1.2 χ^2 值的计算与意义	303
14.1.3 χ^2 分布	303
14.2 拟合问题——样本率与已知	
总体率的比较	304
14.2.1 分析实例	304
14.2.2 检验方法的 SPSS 实现	306
14.2.3 单样本 χ^2 检验的其他话题	307
14.3 相关问题——两(多)个率或	
构成比的比较	308
14.3.1 分析实例	309
14.3.2 检验方法的 SPSS 实现	311
14.3.3 多样本 χ^2 检验的其他话题	312
14.4 两分类变量间关联程度的	
度量	314
14.4.1 相对危险度与优势比	314
14.4.2 分析实例	315
14.5 一致性检验与配对 χ^2 检验	317
14.5.1 Kappa 一致性检验	317
14.5.2 配对 χ^2 检验	318
14.6 分层 χ^2 检验	319
14.7 本章小结	322
思考与练习	323
附录 1 SPSS 13 版新增功能介绍	348
附录 2 SPSS 函数一览表	350
附录 3 各种情形下最常用统计检验方法索引	359
附录 4 统计术语英汉名词对照表	361
SPSS 产品简介	367
参考文献	324
第 15 章 相关分析与回归分析	325
15.1 相关分析简介	325
15.1.1 相关分析的指标体系	325
15.1.2 一些基本概念	328
15.1.3 SPSS 中的相应功能	328
15.2 简单相关分析	329
15.2.1 方法原理	329
15.2.2 分析实例	332
15.2.3 秩相关系数	334
15.2.4 Kendall's 等级相关系数	335
15.3 偏相关分析	335
15.3.1 方法原理	335
15.3.2 分析实例	336
15.4 Distances 过程	338
15.4.1 距离测量与相似性测量的指标	
体系	338
15.4.2 分析实例	340
15.5 简单回归分析	341
15.5.1 方法原理	341
15.5.2 分析实例	344
15.5.3 相关与回归分析的联系和	
区别	346
15.6 本章小结	346
思考与练习	346
参考文献	347

第一部分

数据管理与软件入门

第 1 章 数据分析概述与软件入门

1.1 SPSS 软件概述

SPSS 公司总部位于美国芝加哥,创立于 1975 年,一直以经营统计软件产品开发为主业。1994—1998 年间,SPSS 公司得到了很大的发展,陆续购并了 SYSTAT 公司、BMDP 软件公司、Quantum 公司、ISL 公司等,并将各公司的主打产品收纳 SPSS 旗下,从而使 SPSS 公司由原来的单一统计产品开发与销售转向为企业、教育科研及政府机构提供全面信息统计决策支持服务,成为最新的“数据仓库”和“数据挖掘”领域前沿的一家综合统计软件公司。

SPSS 软件是 SPSS 公司赖以起家的产品,目前也仍然是该公司的主打产品之一,目前的最新版本为 12.0 本书也均以 12.0 版本为准进行讲解。SPSS 的英文名称原为 Statistical Package for Social Sciences,意为社会科学统计软件包。后来随着 SPSS 产品服务领域的扩大和服务深度的增加,SPSS 公司已于 2002 年将英文全称更改为 Statistical Product and Service Solutions,意为统计产品与服务解决方案。在近 30 年的发展中,虽然竞争对手不断出现,但 SPSS 却始终以其鲜明的特色鼎立于统计学软件之中,现在和 SAS(另一种统计分析软件)被并称为当今最权威的两大统计软件。

1.1.1 软件的基本特点

SPSS 得到用户广泛欢迎的原因在于 SPSS 强大的统计分析 with 数据准备功能,方便的图表展示功能,以及广阔的兼容性、界面的友好性满足了广大用户的需求,深受广大应用统计分析人员的喜爱。

1. 功能强大

(1) 囊括了各种成熟的统计方法与模型,为统计分析用户提供了全方位的统计学算法,为各种研究提供了相应的统计学方法。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。

(2) 提供了各种数据准备与数据整理技术。如利用值标签来快捷地录入数据,从而为数据审核与分析提供了便利条件。生成新的变量,对连续性变量进行离散性转换;将几个小类别合并为一个类别等。利用 SPSS 强大的数据整理技术,可使数据结构、内容更易于分析。

(3) 包括自由灵活的表格功能。特别是在 SPSS 11.5 版本中新增的自定义表格模块(Custom Table),使得制表变得更加简单和直接。

(4) 提供了各种常用的统计学图形,如条图、线图、饼图、直方图、散点图等多种图形,并且可

将表格图形直接拷贝到 Word 文档、幻灯片中,直接进行结果的展现。

2. 兼容性好

(1) 在数据方面,不仅可在 SPSS 中作数据录入工作,还可将日常工作中常用的 Excel 表格数据、文本格式数据导入 SPSS 中进行分析,不仅节省了相当大的工作量,并且避免了因拷贝粘贴可能引起的错误。

(2) 在结果方面,SPSS 的表格、图形结果可直接导出为 Word、文本、网页、Excel 格式等,也可以将表格、交互式图形作为对象选择性粘贴到 Word、PowerPoint 等中,并在其中再利用 SPSS 对它们进行编辑。

3. 易用性强

SPSS 之所以有广大的用户群,不仅因为它是一种权威的统计学工具,提供了强大的统计功能,也因为它是一种非常简单易用的软件。人机界面的友好、操作的简单,使得各位统计分析人员对它青睐不已。另外,SPSS 也向一些高级用户提供了编程功能,使分析工作变得更加节省时间和精力。

1.1.2 SPSS 的 Client/Server 结构

SPSS 软件自 10.0 版本以来,已发展为 Client/Server 的结构体系。用户可以选择只购买单机版,也可以选择购买服务器和单机版。对于大数据量客户,可以利用 SPSS Server 来解决速度慢、网络阻塞等由于数据量大而引起的问题。

在分析中使用 SPSS Server 的好处在于:

(1) 更快的分析速度。由于服务器端往往与数据仓库的物理距离更近,而 SPSS Server 也对计算进行了优化,加之应用服务器的硬件配置也远高于单机端(客户端),因此对于进行大数据量分析的客户,SPSS Server 可以使速度提高很多。

(2) 缓解网络阻塞。由于数据不需要全部传送到单机端,所以网络上的数据传输量大大减少,从而缓解了网络阻塞问题。

在使用时调用 Server 的具体做法是:在应用服务器端安装 SPSS Server,在单机端安装相同版本的 SPSS Client(参见图 1.1)。在单机端打开 SPSS for Windows 时,选择菜单 File→Switch Server,在如图 1.2(a)所示的对话框中指定要连接的 SPSS Server 所在服务器地址,如果是第一次使用,则单击“Add”按钮,出现如图 1.2(b)所示对话框,输入服务器名或 IP 地址、端口号,单击“OK”按钮,在服务器列表中出现相应的 Server 地址,然后输入用户名、密码、域名,单击“OK”按钮,即可登录到 SPSS Server。此时,在 SPSS for Windows 下方的状态栏中,就会显示“SPSS Processor on‘服务器名’:‘端口号’ is ready”,表示连接已经建立。

当然对于数据量不大的客户,只用 SPSS Client 就可以了。现在国内绝大多数用户所说的 SPSS,实际上就是指的单机版。

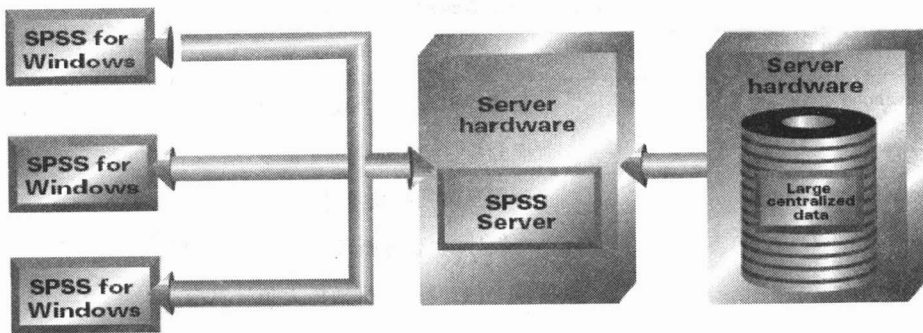
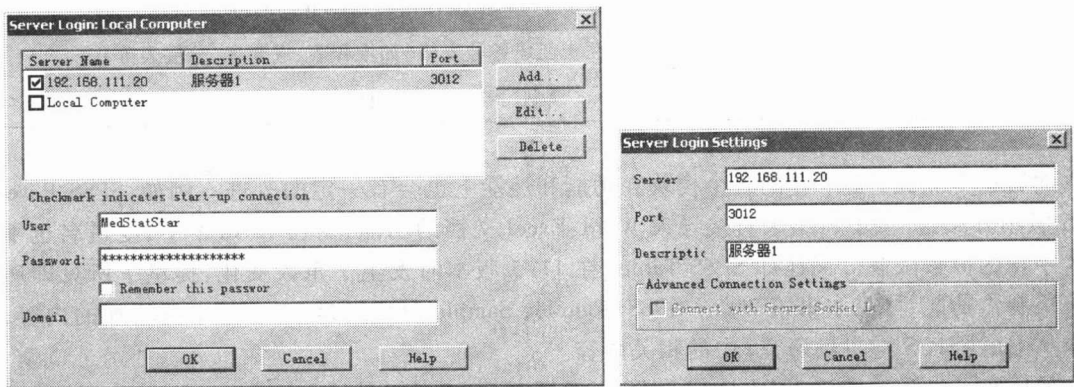


图 1.1 SPSS 软件的 Client/Server 体系



(a)

(b)

图 1.2 SPSS Client 调用 SPSS Server 的对话框设置

1.1.3 SPSS 的模块式结构

无论是 SPSS Client 还是 SPSS Server,均是模块式结构,即它把自己的所有功能分放在多个模块上。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,自己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 12 共由 11 个模块构成,它们分别是:SPSS Base、SPSS Advanced、SPSS Categories、SPSS Complex Sample、SPSS Conjoint、SPSS Exact Test、SPSS Maps、SPSS Missing Value Analysis、SPSS Regression、SPSS Tables 和 SPSS Trends。其中 SPSS Base 是必需的,因为 SPSS 软件的整个框架、基本的数据获取、数据准备等基本功能都被集中在这个模块上,其他模块必须在 SPSS Base 搭建的平台上才能工作。其他模块的功能分别如表 1.1 所示。

SPSS 软件通过其 License 来控制模块是否安装。一个模块安装上之后,在 SPSS for Windows 的菜单中就会出现相应的菜单项,所以不同客户的 SPSS for Windows 的菜单可能有所不同。如果没有购买 SPSS Trends 模块,软件中就不会有这样一个菜单:Analyze→Trends;如果没有购买 SPSS Maps 模块,软件中就不会有菜单:Graph→Maps。