



课堂实录

马晓玉 刘艳春 / 编著

SPSS 统计分析 课堂实录

- 构思独特，所有案例来自一线实战场景；
- 实用性强，将抽象的理论结合到实战案例上；
- 内容全面，结构清晰，体例丰富；
- 视频教学，专业教学视频帮助读者快速上手。

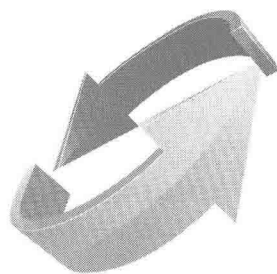
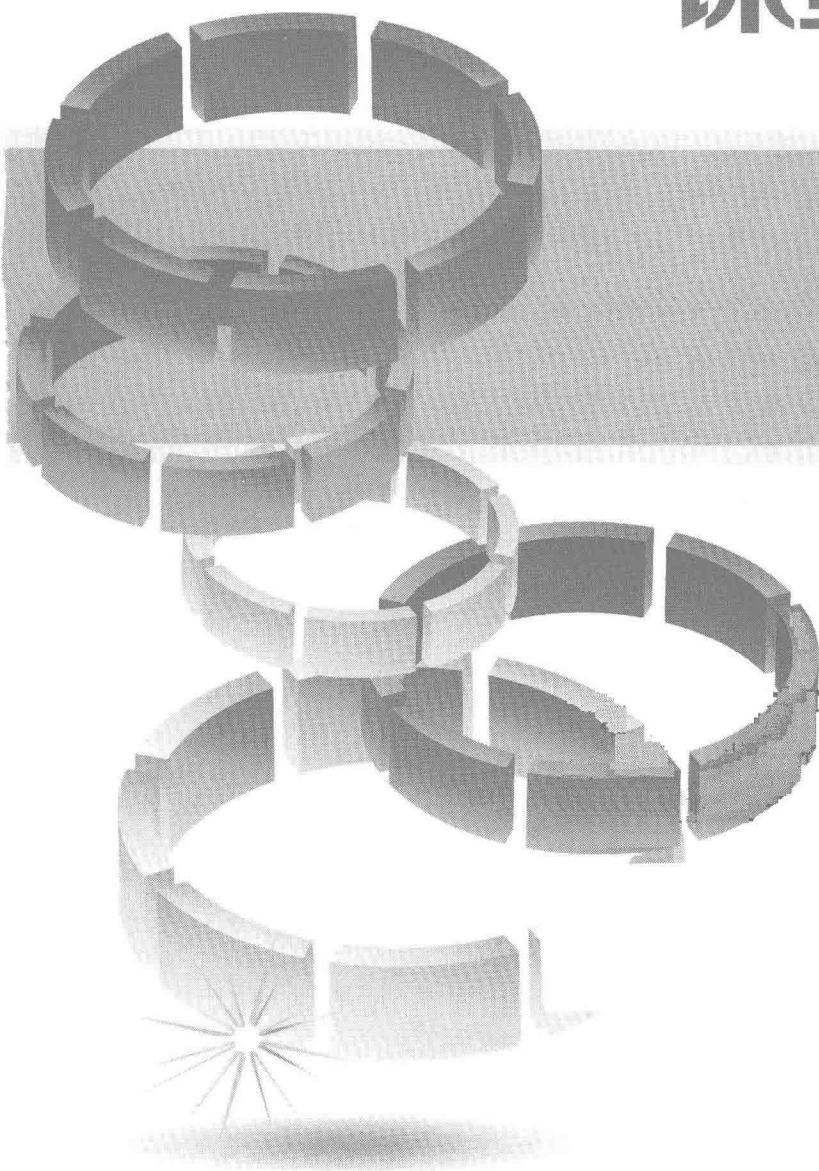


清华大学出版社

课堂实录

马晓玉 刘艳春 / 编著

SPSS 统计分析 课堂实录



清华大学出版社
北京

内 容 简 介

本书结合作者多年授课经验和高职高专学生的实际情况,针对非统计专业类学生或者刚入门学生的认知特点而编写的。本书详细阐述了基础理论知识,重点安排与实际应用相联系的内容。该书理论与实际相结合,在每章后面安排 2~3 个实验指导内容,可以强化学生的动手操作能力。全书共 10 章,详细介绍了统计学的一些基础知识、使用 SPSS 建立和管理数据文件、对数据文件进行预处理等内容。

全书结构编排合理、图文并茂、实例丰富,不仅可以作为大、中专院校教学和企业培训的教材,也可以作为在职人员的自学指导用书。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

SPSS 统计分析课堂实录/马晓玉,刘艳春编著. —北京:清华大学出版社,2016
(课堂实录)

ISBN 978-7-302-37548-7

I. ①S… II. ①马… ②刘… III. ①统计分析-软件包 IV. ①C819

中国版本图书馆 CIP 数据核字(2014)第 174641 号

责任编辑:夏兆彦

封面设计:张 阳

责任校对:胡伟民

责任印制:刘海龙

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印装者:北京鑫丰华彩印有限公司

经 销:全国新华书店

开 本:190mm×260mm 印 张:16.75

字 数:485 千字

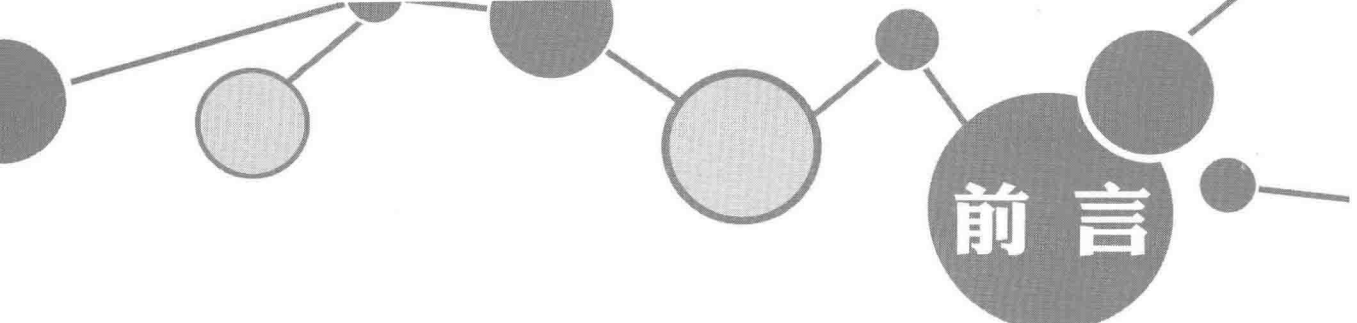
版 次:2016 年 2 月第 1 版

印 次:2016 年 2 月第 1 次印刷

印 数:1~3000

定 价:39.00 元

产品编号:056612-01



前言

SPSS 统计分析软件是世界上最早的统计分析软件,迄今为止 SPSS Statistics 软件已有 40 余年的成长历史。它功能强大,易学、易用,且操作界面友好,输出结果美观,是世界上公认的标准统计分析软件之一。在国际学术界有条不成文的规定,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,可以不需要说明算法,由此可见其影响之大和信誉之高。目前 SPSS 已经广泛应用于通信、医疗、银行、证券、保险、制造、商业、市场研究、科研教育等众多领域和行业,成为世界上应用最广泛的专业统计软件。

本书使用 SPSS 20,内容安排由易到难,逐层推进,便于初学者掌握,有一定经验的人员也能从本书中找到可以学习的知识。本书每章都配有大量实例,每个实例都经过精挑细选,使之更具有针对性,更加方便读者练习和学习,提高学习效果。通过学习本书的内容,不但可以使读者全面掌握 SPSS 的操作和使用,而且还可以提高分析和解决实际问题的能力。

1. 本书主要内容

本书介绍了统计学的一些基础知识、使用 SPSS 建立和管理数据文件、对数据文件进行预处理、使用 SPSS 进行基本的数据分析、常用的分析方法介绍、绘制统计图形等内容。全书共分为 10 章,各章内容概括如下。

第 1 章讲解 SPSS 基础知识,主要介绍 SPSS 的产生与发展、SPSS 的特点及功能、SPSS 在各个领域的应用、SPSS 软件界面的介绍等内容。

第 2 章讲解建立和管理数据文件,主要介绍数据的结构、数据的定义方法、数据文件的读取、数据库文件的编辑等内容。

第 3 章讲解数据库文件的预处理,主要介绍数据的排序、变量计算、数据选取、计数、分类汇总、数据分组等内容。

第 4 章讲解 SPSS 基本统计分析方法,主要介绍频数分析、描述性分析、探索分析、交叉表分析、多选项分析、比率分析等内容。

第 5 章讲解报表与 Bootstrap,主要介绍简单报表的制作、高级报表的制作、Bootstrap 的设置等内容。

第 6 章讲解均值比较与检验,主要介绍参数检验、单一样本 t 检验、两独立样本 t 检验、两配对样本 t 检验等内容。

第 7 章讲解方差分析,主要介绍单因素方差分析、多因素方差分析、协方差分析等内容。

第 8 章讲解相关分析,主要介绍散点图、相关系数、偏相关分析、距离相关分析等内容。

第 9 章讲解回归分析,主要介绍线性回归分析、曲线回归、二项 Logistic 回归等内容。

第 10 章讲解统计图的绘制,主要介绍条形图、线形图、面积图、高低图、箱图等内容。



2. 本书主要特色

本书结合非统计专业学生和刚入门学生的需求，详细介绍了数据统计的基础知识和实际应用，具有以下特色。

- **丰富实例** 每章以实例形式演示计算机网络的操作应用知识，便于读者学习与模仿操作，同时方便教师组织授课。
- **彩色插图** 本书提供了大量精美的实例，在彩色插图中读者可以感受逼真的实际效果，从而迅速掌握、了解计算机网络设备的内容。
- **思考与练习** 扩展练习测试读者对本章所介绍内容的掌握程度；上机练习理论结合实际，引导学生提高上机操作能力。
- **视频教学** 本书完整地提供了书中实例效果和大量全程配音视频文件，便于读者学习使用。所有视频教学文件均已上传到 www.ztydata.com.cn，读者可自行下载。

3. 本书适用对象

全书结构编排合理、图文并茂、实例丰富，适用于初、中级用户学习使用。本书不仅可以作为大、中专院校教学和企业培训的教材，也可以作为在职人员的自学指导用书。

除了封面署名人员之外，参与本书编写的人员还有孙岩、马海军、张仕禹、夏小军、赵振江、李振山、李文采、吴越胜、李海庆、何永国、李海峰、陶丽、吴俊海、安征、张巍屹、崔群法、王咏梅、康显丽、辛爱军、牛小平、贾栓稳、王立新、苏静、赵元庆、郭磊、徐铭、李大庆、王蕾、张勇、郝安林等。在编写过程中难免会有漏洞，欢迎读者通过清华大学出版社网站 www.tup.tsinghua.edu.cn 与我们联系，帮助我们改正提高。

编者



目录

第 1 章 了解 SPSS

1.1 SPSS 介绍	2	1.3 使用 SPSS 软件	12
1.1.1 SPSS 的产生与发展	2	1.3.1 安装 SPSS 软件	12
1.1.2 SPSS 的优点	2	1.3.2 SPSS 的启动与退出	15
1.1.3 SPSS 的功能	4	1.3.3 SPSS 编辑窗口	16
1.2 SPSS 的应用	8	1.3.4 SPSS 输出窗口	19
1.2.1 SPSS 在科研、教育中的应用	8	1.3.5 语法编辑器	20
1.2.2 SPSS 在医疗、心理研究中的应用	9	1.4 获取 SPSS 帮助	20
1.2.3 SPSS 在金融、保险中的应用	10	1.4.1 在窗口中获取帮助信息	21
1.2.4 SPSS 在制造业中的应用	10	1.4.2 获取 SPSS 教程信息	22
1.2.5 SPSS 在市场调研中的应用	10	1.4.3 在对话框中获取帮助信息	23
1.2.6 SPSS 在政府机构中的应用	11	1.5 思考与练习	24
1.2.7 SPSS 在通信行业中的应用	11		

第 2 章 建立和管理数据文件

2.1 SPSS 数据的结构和定义方法	26	2.3.3 从数据库中读取数据	36
2.1.1 定义变量名称	26	2.3.4 从文本文件读取数据	39
2.1.2 定义变量的类型	27	2.4 编辑数据	41
2.1.3 变量标签和值标签	28	2.4.1 剪切、复制和粘贴数据	41
2.1.4 定义变量缺失值	29	2.4.2 查找与替换数据及属性	44
2.1.5 定义列宽和对齐方式	30	2.4.3 操作个案数据	45
2.1.6 定义变量度量	30	2.4.4 选择个案	46
2.1.7 定义变量角色	30	2.5 数据文件的存储	47
2.2 录入数据	31	2.5.1 保存已修改的数据文件	47
2.2.1 数据录入规则	31	2.5.2 以外部格式保存数据文件	47
2.2.2 输入字符串	31	2.5.3 以 Excel 格式保存数据文件	48
2.2.3 使用值标签输入数据	32	2.5.4 以 SAS 格式保存数据文件	49
2.2.4 通过编辑栏输入数据	32	2.5.5 保存变量子集	50
2.2.5 输入日期数据	32	2.5.6 导出到数据库	50
2.2.6 输入货币型数据	33	2.6 实验指导：输入数据	52
2.2.7 通过【语法编辑器】输入内容	33	2.7 实验指导：导入数据文件	54
2.3 读取数据文件	34	2.8 实验指导：个人消费记录合并	56
2.3.1 读取 SPSS 数据文件	34	2.9 思考与练习	57
2.3.2 从电子表格中读取数据	35		

第3章 预处理 SPSS 数据

3.1 排序和拆分数据	60	3.4.1 转换值	71
3.1.1 排序操作	60	3.4.2 重新编码	73
3.1.2 数据拆分处理	61	3.5 个案排序和加权个案	77
3.2 标识与互换操作	61	3.5.1 个案排序	78
3.2.1 数据的行列互换	61	3.5.2 加权个案	79
3.2.2 标识重复的个案	62	3.6 分类汇总数据	80
3.2.3 标识异常个案	63	3.7 实验指导：应用数据权重探索吸烟、 喝酒与患病的关系	82
3.3 变量计算和计数	67	3.8 实验指导：分析工人的生产水平	83
3.3.1 变量计算	67	3.9 思考与练习	85
3.3.2 计数操作	69		
3.4 转换值与重新编码	71		

第4章 基本统计分析

4.1 描述统计简介	88	4.4.2 探索性统计分析的操作过程	95
4.1.1 常用描述统计量	88	4.5 交叉列联表分析	99
4.1.2 数据的统计指标	89	4.5.1 交叉列联表分析的主要功能	99
4.1.3 统计分析的方法	89	4.5.2 交叉表分析的操作步骤	99
4.1.4 统计分析方法的选择	89	4.6 比率分析	103
4.2 频率统计	90	4.6.1 比率分析的主要功能	103
4.2.1 频数统计的主要功能	91	4.6.2 比率分析的操作步骤	103
4.2.2 频率统计的操作步骤	91	4.7 实验指导：分析图书销量与定价的 关系	104
4.3 描述统计	93	4.8 实验指导：分析工人的生产水平	105
4.3.1 描述统计的主要功能	93	4.9 实验指导：某高校学生自尊水平 评测	107
4.3.2 描述统计的操作步骤	93	4.10 思考与练习	110
4.4 探索性统计分析	95		
4.4.1 探索性统计分析的主要功能	95		

第5章 报表与 BootStrap

5.1 代码本	112	5.3.2 个案汇总的操作过程	119
5.1.1 报表分析概述	112	5.4 按行汇总	121
5.1.2 代码本的应用	112	5.5 BootStrap 简介	125
5.1.3 代码本的操作步骤	112	5.5.1 BootStrap 计算原理	125
5.2 在线分析处理	115	5.5.2 BootStrap 的使用方法	125
5.2.1 了解 OLAP 分析	115	5.5.3 BootStrap 的操作步骤	125
5.2.2 在线分析的应用	116	5.6 实验指导：制作摘要报表	127
5.2.3 OLAP 立方分析的操作步骤	116	5.7 实验指导：应用重复抽样精确分析致病 因素	128
5.3 个案汇总	119	5.8 思考与练习	130
5.3.1 个案汇总的应用	119		

第6章 均值比较与参数检验

6.1 假设检验	134	6.3.2 单样本 T 检验的操作步骤	141
6.1.1 问题的提出	134	6.4 两独立样本 t 检验	142
6.1.2 假设检验的原理	134	6.4.1 两独立样本 T 检验的原理	142
6.1.3 假设检验中的小概率事件	135	6.4.2 两独立样本检验的操作步骤	143
6.1.4 假设检验的两类错误	135	6.5 两配对样本 t 检验	144
6.1.5 单侧检验和双侧检验	136	6.5.1 两独立样本 T 检验的原理	144
6.1.6 假设检验的步骤	137	6.5.2 两独立样本 T 检验的操作步骤	145
6.1.7 假设检验的种类	137	6.6 实验指导: 分析影响上市公司出口的因素	146
6.2 均值比较	137	6.7 实验指导: 不同地区电信用户的消费习惯调查	148
6.2.1 均值比较的原理	137	6.8 实验指导: 学生素质评测	151
6.2.2 均值比较的操作步骤	138	6.9 思考与练习	155
6.3 单一样本 t 检验	140		
6.3.1 单样本 T 检验的原理	140		

第7章 方差分析

7.1 方差分析	158	7.4 多元方差分析	167
7.1.1 了解方差分析	158	7.4.1 多元方差分析的基本概念	167
7.1.2 方差分析假定条件	158	7.4.2 多元方差分析的操作步骤	168
7.1.3 方差分析的思想	159	7.5 协方差分析	170
7.1.4 方差分析的分类	159	7.5.1 协方差分析的基本概念	170
7.1.5 方差分析的原理	160	7.5.2 协方差分析的操作步骤	171
7.2 单因素方差分析	161	7.6 实验指导: 探索不同教学方式对学生成绩的影响	173
7.2.1 单因素方差分析的基本概念	161	7.7 实验指导: 分析不同药物的药效	175
7.2.2 单因素方差的操作步骤	162	7.8 实验指导: 工龄与肺活量的关系研究	178
7.3 两因素方差分析	164	7.9 思考与练习	180
7.3.1 两因素方差分析的基本概念	164		
7.3.2 两因素方差分析的适用条件	165		
7.3.3 两因素方差分析的操作步骤	165		

第8章 相关分析

8.1 相关分析的基本原理	184	8.2.3 散点图的操作步骤	187
8.1.1 相关分析的概念	184	8.3 双变量相关分析	188
8.1.2 相关分析的分类	184	8.3.1 双变量相关分析的基本原理	188
8.1.3 相关分析的结果输出	185	8.3.2 双变量相关分析的操作步骤	190
8.1.4 相关分析的性质	186	8.4 偏相关分析	191
8.2 散点图	186	8.4.1 偏相关分析的基本原理	191
8.2.1 散点图的基本概念	186	8.4.2 偏相关分析的适用条件	191
8.2.2 使用散点图的注意事项	187	8.4.3 偏相关分析的操作步骤	191

8.5	距离相关分析	192
8.5.1	距离相关分析的基本原理	193
8.5.2	距离相关分析的分类	193
8.5.3	距离相关分析的操作步骤	193
8.6	实验指导: 各产业发展对国内生产总值的影响	194

8.7	实验指导: 分析影响保险人员业绩的因素	197
8.8	实验指导: 探寻影响员工工资的主要因素	201
8.9	思考与练习	203

第9章 回归分析

9.1	回归分析概述	206
9.1.1	回归分析的目的	206
9.1.2	回归分析与相关分析的关系	206
9.1.3	回归分析的基本假定	207
9.1.4	回归分析的应用	207
9.1.5	回归分析的基本步骤	208
9.2	一元线性回归分析	208
9.2.1	一元线性回归的基本概念	208
9.2.2	一元线性回归模型的基本假定	209
9.2.3	参数的最小二乘估计	209
9.2.4	回归方程的检验	209
9.2.5	一元线性回归的结果预测	210
9.2.6	一元线性回归的操作步骤	210

9.3	多元线性回归分析	214
9.3.1	多元线性回归概念	214
9.3.2	选择多元线性回归分析方法	215
9.3.3	多重共线性相关问题	215
9.3.4	多元线性回归的操作步骤	215
9.4	实验指导: 某地区河流污染情况分析	217
9.5	实验指导: 预测正常体表温度下的代谢量	220
9.6	实验指导: 估算汽车油耗和加油金额	223
9.7	思考与练习	225

第10章 统计图形

10.1	统计图形概述	228
10.1.1	统计图绘制与原则	228
10.1.2	了解图表操作	228
10.2	条形图	228
10.2.1	简单条形图	229
10.2.2	复式条形图	232
10.2.3	堆积条形图	233
10.2.4	3-D 条形图	234
10.3	线图	236
10.3.1	简单线图	236
10.3.2	多线线图	237
10.3.3	垂直线图	238
10.4	面积图和饼图	240
10.4.1	简单面积图	240
10.4.2	堆积面积图	241

10.4.3	饼图	243
10.5	高低图	244
10.5.1	简单高低关闭图	244
10.5.2	聚类高低关闭图	245
10.5.3	简单范围栏图	247
10.5.4	聚类范围栏图	248
10.5.5	差别面积图	249
10.6	箱图	250
10.6.1	简单箱图	251
10.6.2	复式箱图	252
10.7	误差条形图	253
10.7.1	简单误差条形图	254
10.7.2	复式误差条形图	255
10.8	实验指导: 人口金字塔	256
10.9	思考与练习	257

第 1 章

了解 SPSS

SPSS 采用类似 EXCEL 表格的方式输入与管理数据，数据接口较为通用，能方便的从其他数据库中读入数据。而 SPSS 软件的统计过程，包括了常用的、较为成熟的功能，完全可以满足非统计专业人士的工作需要。

通过 SPSS 统计或分析后的输出结果也十分美观。存储结果时可采用专用的 SPO 格式，可以转存为 HTML 格式和文本格式。

对于熟悉编程运行方式的用户，SPSS 还特别设计了语法生成窗口，用户只需在菜单中选好各个选项，然后按“粘贴”按钮就可以自动生成标准的 SPSS 程序，极大地方便了中、高级用户。

本章学习要点：

- 了解 SPSS
- SPSS 应用范围
- 使用 SPSS 软件
- 获取 SPSS 帮助信息

1.1

SPSS 介绍

SPSS 是世界上最早的统计分析软件,由美国斯坦福大学的三位研究生 Norman H. Nie、C. Hadlai (Tex) Hull 和 Dale H. Bent 于 1968 年研究开发成功,同时成立了 SPSS 公司,并于 1975 年成立法人组织,在芝加哥组建了 SPSS 总部。

1.1.1 SPSS 的产生与发展

SPSS 原意是“社会科学统计软件包”,是 Solution Statistical Package for the Social Science 的英文名称首字母的缩写,是世界上最早的统计分析软件。

SPSS 软件最初诞生时,是用于大型机的统计软件。后来随着微型计算机的问世与发展,SPSS 总部于 1984 年首先推出了世界上第一个微机版统计分析软件 SPSS/PC+,并很快地应用于自然科学、社会科学、技术科学的各个领域。

1992 年,SPSS 公司推出 Windows 版本统计软件,同时也开始了全球化的发展。1994 年至 1998 年,SPSS 公司先后并购了 SYSTAT、BMDP、Quantime、ISL 等公司,进而使 SPSS 公司从原来单一统计产品的开发与销售,转向为企业、政府机构及教育科研提供全面信息统计决策支持与服务。

到 2000 年 SPSS 11.0 起,SPSS 英文全称改为 Statistical Product and Service Solutions,即“统计产品和服务解决方案”。

到 2009 年,SPSS 公司宣布重新包装旗下的 SPSS 产品线,定位为预测统计分析软件 (Predictive Analytics Software)PASW,包括四

部分:

- ❑ PASW Statistics (formerly SPSS Statistics, 统计分析);
- ❑ PASW Modeler (formerly Clementine, 数据挖掘);
- ❑ Data Collection family (formerly Dimensions, 数据收集);
- ❑ PASW Collaboration and Deployment Services (formerly Predictive Enterprise Services, 企业应用服务)。

而到 2010 年,随着 SPSS 公司被 IBM 公司并购,各个子产品家族名称的前面不再以 PASW 为名,并统一修改为 IBM SPSS 字样。

SPSS 作为一种使用方便的集成化计算机数据处理软件,以其强大的统计功能、便捷的操作方式、灵活的分析报告和精美的图形展示,赢得全球约 26 万家产品用户,广泛应用于通信、医疗、银行、证券、保险、制造、商业、市场研究、科研和教育等众多领域和行业。

目前,SPSS 已成为世界最流行、应用最广泛的三大专业数据分析软件之一(它们是 SAS、SPSS 和 SYSTAT)。

1.1.2 SPSS 的优点

SPSS 软件除了具有强大的分析功能之外,还包含有以下优点。这些优点都是受众多领域不同用户欢迎的原因。

1. Windows 风格的界面极为友好

SPSS Statistics for Windows 最突出的特点就是操作界面极为友好,输出结果美观漂亮。SPSS 是第一个采用人机交互界面的统计软件,非常容易学习和使用。自从 1995 年 SPSS 公司与微软公司合作开发 SPSS 界面后,SPSS 界面变得越来越友好,操作也越来越简单,使熟悉微

软公司产品的用户学习 SPSS Statistics 操作时,很容易学习。

SPSS 界面非常类似于 Windows 界面风格,数据视图也类似 Excel 布局。SPSS 具有第四代语言的特点,告诉系统要做什么,无需告诉怎样做。只要了解统计分析的原理,无需通晓统计方法的各种算法,即可得到需要的统计分析结果。

除了数据录入及部分命令程序等少数输入工作需要键盘键入外,大多数操作可通过鼠标拖



曳、单击【菜单】、【按钮】和【对话框】来完成。

SPSS 功能强大界面友好,易学易用。SPSS 界面完全是菜单式,使用下拉菜单来选择所需要执行的复杂的统计命令,使用 Windows 的窗口方式展示各种管理和分析数据方法的功能,使用对话框展示出各种功能选择项,只要掌握一定的 Windows 操作技能,精通统计分析原理,就可以使用该软件为特定的科研工作服务。开放式的命令语句窗口,可以通过复制和粘贴的方法学习和使用其“统计程序”语言,同时也适合数据分析专家和研究员使用。

2. 易学易用的统计分析软件

SPSS 易于操作、易于入门,结果易于阅读。对于统计软件的学习不会影响统计的主题,这样研究人员就可以将精力集中在社会研究方法、市场研究方法、营销的业务问题上,而不是忙于编程和统计。一般稍有统计基础的人员经过几天的培训即可用 SPSS 做简单的数据分析,包括绘制图表、简单回归、相关分析等。

当然,真正应用好 SPSS 软件的关键在于如何进行科学地研究方案设计、严谨地收集数据、严密深入地对数据进行统计分析及解释,适度保守地研究结论和进行决策。这一方面要求研究者掌握数理统计的基本知识,另一方面也要求研究者经常进行实践,在实践中了解各种统计结果的实际意义。

从某种意义上讲,SPSS 软件还可以帮助数学功底不够的使用者学习运用现代统计技术。使用者仅需要关心某个问题应该采用何种统计方法,并初步掌握对计算结果的解释,而不需要了解其具体运算过程,就可以在使用手册的帮助下完成对数据定量分析。现在很多使用者只需要适当的练习,就能够掌握简单的操作分析,因此 SPSS 特别受非统计专业数据分析人员的青睐。

SPSS 采用类似 EXCEL 表格的方式输入与管理数据,数据接口通用,能方便地从其他数据库中读入数据。其统计方法包括了常用的、较为成熟的统计方法,完全可以满足非统计专业人士的工作需要,是非专业统计人员的首选统计软件。

3. 功能全面的统计分析软件

SPSS 针对初学者、熟练者及精通者都比较

适用。它提供了数据获取、数据管理与准备、数据分析、结果报告这样一个数据分析的完整过程,因此非常全面地涵盖了数据分析的整个流程,特别适合设计调查方案、对数据进行统计分析以及制作研究报告中的相关图表。

此外,SPSS 具有完整的数据输入、编辑、统计分析、报表、图形制作等功能。仅 SPSS Base 模块就提供了从简单的统计描述到复杂的多因素统计分析方法。如数据的探索性分析、统计描述、列联表分析、二维相关、秩相关、偏相关、一元方差分析、非参数检验、多元回归、生存分析、协方差分析、判别分析、因子分析、聚类分析等常见的分析方法。

4. 强大的编程能力,支持二次开发

对于常见的统计方法,SPSS 的命令语句、子命令及选择项的选择绝大部分由“对话框”操作完成。因此,用户无需花大量时间记忆大量的命令、过程、选择项。

由于 SPSS 20.0 具备强大的 Syntax 编程功能,以及包括了 SPSS Programmability Extension™ 模块的编程扩展功能,那些熟练或精通者也较喜欢 SPSS,因为他们可以通过编程,在 SPSS 命令语法语言的基础上提供与其他编程语言的结合功能,来实现更强大的功能。例如,用其他语言编写的程序代码,如 Python 和 R,可以管理使用 SPSS 语法所编写的任务流。使用 SPSS 20.0 提供的扩展编程功能和特性,让 SPSS for Windows 成为了最强大的统计开发平台之一。

5. 支持丰富的数据源,具备强大数据访问和管理能力

SPSS 可以同时打开多个数据集,方便研究时对不同数据库进行比较分析和进行数据库转换处理。软件提供了更强大的数据管理功能帮助用户通过 SPSS 使用其他的应用程序和数据库,能够读取及输出多种格式的文件。

比如,由 dBase、FoxBASE、FoxPro 产生的*.dbf 文件,文本编辑器软件生成的 ASC II 数据文件,Excel 的*.xls 文件等均可转换成可供分析的 SPSS 数据文件。它支持 Excel、文本、Dbase、Access、SAS 等格式的数据文件,能够把 SPSS 的图形转换为七种图形文件,结果可

保存为*.txt、word、PPT 及 html 格式的文件。

通过使用 ODBC(Open Database Capture) 的数据接口, 可以直接访问以结构化查询语言 (SQL) 为数据访问标准的数据库管理系统, 通过数据库导出向导功能可以方便地将数据写入到数据库中等。

从 SPSS Statistics 19.0 开始, 增加了 Salesforce.com 数据库驱动程序, 允许分析人员访问 Salesforce.com 中的数据, 就像访问 SQL 数据库中的数据一样。

6. 灵活的配置方案

对于商业用户而言, SPSS Statistics 是一种按照模块进行配置的软件产品, 主要包括

SPSS Statistics Base 模块和其他一系列扩充功能模块。SPSS Statistics Base 是基础软件平台, 具备强大的数据管理能力和输入输出界面管理能力, 并具备完整的常见统计分析功能, 而其他每个独立扩充功能模块均在 SPSS Statistics Base 的基础上, 为其增加某方面的分析功能家族。

7. 支持多种操作系统

SPSS 客户端支持 Windows XP (32 位)、Windows Vista/7 (32 位和 64 位)、Linux 和 Mac OS。服务器端支持 Windows Server 2003 (32 位和 64 位)、Windows Server 2008 (32 位和 64 位)、AIX、HP-UX、Solaris。

1.1.3 SPSS 的功能

2011 年, 业界领先的预测分析软件 SPSS 推出其旗舰统计分析软件 SPSS Statistics 的最新版 SPSS Statistics 20.0。

新版本在原软件的基础上, 融合了多项业界领先的统计分析技术, 增加了新的功能模块, 大大提高了简便性和个性化, 进一步贴近用户需求。

SPSS 统计分析软件是一款按照功能模块进行配置的软件产品, 主要包括 SPSS Statistics Base 模块和其他一系列扩充功能模块, 每个独立扩充功能模块均可在 SPSS Statistics Base 模块基础上, 为其增加某方面的分析功能。SPSS Statistics 20.0 全模块软件的 16 个功能模块的功能描述如下。

1. SPSS Statistics Base

必需的基础模块, 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。

基本统计分析功能包括描述统计和行列计算, 还包括在基本分析中最受欢迎的常见统计功能, 如汇总、计数、交叉分析、分类比较、描述性统计、因子分析、回归分析及聚类分析等。具体如下:

- 数据访问、数据准备、数据管理与输出管理;
- 描述统计和探索分析 频数、描述、集

中趋势和离散趋势分析、分布分析与查看、正态性检验与正态转换、均值的置信区间估计;

- 交叉表 计数; 行、列和总计百分比; 独立性检验; 定类变量和定序变量的相关性测度;
- 二元统计 均值比较、T 检验、单因素方差分析;
- 相关分析 双变量相关分析、偏相关分析、距离分析;
- 线性回归分析 自动线性建模、线性回归、Ordinal 回归—PLUM、曲线估计;
- 非参数检验 单一样本检验、双重相关样本检验、K 重相关样本检验、双重独立样本检验、K 重独立样本检验;
- 多重响应分析 交叉表、频数表;
- 预测数值结果和区分群体 K-means 聚类分析、分级聚类分析、两步聚类分析、快速聚类分析、因子分析、主成分分析、最近邻元素分析、判别分析、尺度分析;
- 报告 各种报告、记录摘要、图表功能 (分类图表、条型图、线型图、面积图、高低图、箱线图、散点图、质量控制图、诊断和探测图等);
- 数据管理、数据转换与文件管理。

2. SPSS Advanced Statistics

在分析数据时,除了基本的数据分析外,如果还想建立分析过程数据,就需要使用 Advanced Models,为顺序结果建立更灵活、更成熟的模型,在处理嵌套数据时得到更精确的预测模型,可以分析事件历史和持续时间数据。具体功能包括:

- ☐ 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。

其中,一般线性模型(General Linear Model—GLM)为描述一个结果变量和一组影响变量之间的关系提供了更多的灵活性,模型包括线性回归、ANOVA、ANCOVA、MANOVA 和 MANCOVA。GLM 也包含重复测量、混合模型、Post Hoc 检验、重复测量的 Post Hoc 检验、四种类型的平方和、边缘期望值配对比较、处理复杂的缺失单元及储存设计矩阵和结果数据。

3. SPSS Bootstrapping

SPSS Statistics Bootstrapping 模块可以让用户更有效地使用小样本量的数据,通过数据自身重复采用的功能,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,也可以使用该功能进行直接采用,从而以一种更加直观的方式来执行结果的检查。

4. 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让用户清晰地看到数据中的关系,使用户可以更完整和方便地分析数据。Categories 提供非线性主成分分析来描述数据,并用图标清楚地展示数据中的关系,展示并分享动态、交互的分析结果,让用户从分类数据

中得到更丰富的信息。

使用 Categories,用户可以从大量变量或二维及多维表格中了解重要信息。通过类似传统的回归分析、主成份分析及典型相关分析,帮用户处理和了解顺序及名义数据可视化地探索多变量分类数据。

对应分析用于分析二维列联表或可以表示为二维表格的数据,比如课程偏好和习惯性选择数据间的关系。使用对应分析和分类回归分析过程,用户可以把分析结果用表格、图形和独特的“枢轴表”来展示。

最优尺度回归用一个优化的尺度量化分类变量,得到一个优化的转换变量线性回归方程。用户还可以用这个回归的结果方程来预测这三个变量的任意组合所影响的工作满意度。

用户还可以应用齐性分析(homogeneity analysis),又称多重对应分析,来分析分类多变量数据矩阵。齐性分析类似于对应分析,但不限于两个变量。齐性分析用于记录和分类变量赋予数值的方法来量化名义数据,在低维度空间中表述两个或多个名义变量之间的关系。例如,通过齐性分析来图形化地显示工作种类、少数民族种类和性别之间的关系。

5. SPSS Complex Samples

如果使用了特别复杂的抽样方案,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以及相伴标准误差,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,对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。

6. 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。

7. SPSS Custom Tables

Tables 提供 35 种单元和摘要统计量,能够更方便地显示多重序列数据,它能串接所有的维度,以在同一表格中显示包含不同统计量的各种变量。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,通过完整的表格控制权,研究人员还可以自制表格,创造优美外观。SPSS Tables 广泛应用于调查研究和市场研究等领域,可以完成很多“中国式报表”。

SPSS 直观的图形化界面使用户在制表的时候不需凭猜测进行操作,使用鼠标拖动的方式和预览的功能,用户能够在生成之前,对于将制成的表格结构一目了然。使用交互界面制表非常简单容易。

首先,用户可以预览,并进行修改报表。其次,用户能够分辨分类变量和连续变量,并立刻得到关于数据结构的信息。甚至只需轻轻单击鼠标即可更改变量类型。

制表时,只需将需要的变量拖动到表格预览窗口(Table preview builder),不需要写复杂的语法,也不再需要与难用的对话框打交道。并且,用户还可以轻松地将变量从行拖到列,以实现变量的精确定位。只要用户做出改动,表格的结构立刻发生变化,能够立刻看到改动的效果。

用户也可以直接在表格预览窗口对变量进行添加、交换或嵌套的操作,也可以隐藏统计量标签。在看到所有变量的前提下,将结构复杂的表格变得更简练。

8. SPSS Data Preparation

利用 SPSS Statistics Data Preparation 模块,可以获得多个简化数据准备过程的程序。这个附加模块使用户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。然后,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以及用户还可以为分类变量

而设计的运算法则,来做更多精确的工作。

使用 Data Preparation 模块,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。

9. SPSS Decision Trees

在 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。

Decision Trees 模块可直接在 SPSS Statistics 内做分类区分,用 Syntax 撰写或用 XML 来储存设定。使用 Classification Trees 还可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。

Decision Trees 模块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。

10. SPSS Direct Marketing

Direct Marketing 模块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。

SPSS Statistics 将直销中常用的分析定制为不同的模块,市场研究人员可藉由 Direct Marketing 模块,以简单的方式进行简易直觉的分析,锁定高价值顾客来进行各种营销分析。

- ❑ 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。



11. 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,用传统的计算方法是可以的。

但是,如用户的数据属于小样本或零星的数据,又将资料细分到多个类别或数据变量中有超过百分之八十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。

传统情况下,要得到正确的结论,每一个单元需要有五个以上数据,SPSS Statistics Exact Tests 会为用户解除这种限制。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, $R \times C$ 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,提供可信赖的结果。如果用户需要做调查研究、医学研究、生物统计、社会科学、市场调查或各类的实验,可使用 Exact Tests 模块。

12. 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。

13. 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。

14. SPSS Neural Networks

神经网络是一种模拟人类大脑处理信息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。能够广泛应用于各种领域进行预测分析,如银行保险的欺诈判断,客户的反馈判断等。

神经网络是一种综合了多种技术的新型算法技术。用户可以利用 SPSS Statistics Neural Networks 模块,以全新的方式探索数据。

15. 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。用户由撰写内嵌在 SPSS 20.0 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。

16. 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。

多项式逻辑斯谛回归 (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。二元逻辑斯谛回归 (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 的应用

正因为 SPSS 统计分析软件以其易学易用等特点,所以已经成为众多领域进行统计分析不可缺少的工具。因此,它广泛应用于科研、教育、医疗、心理研究、金融、保险、政府数据分析、企业和市场研究、制造等领域。

1.2.1 SPSS 在科研、教育中的应用

在科学研究、日常教学活动中要得出定量的结论,也需要经过大量而周密的数学运算。

现代数学分析方法已广泛应用于各门学科的科学研究的之中,不仅在自然科学,而且在社会科学、思维科学都已普遍使用数学,处理计量化的过程,产生了计量社会学、计量历史学、计量地理学、教育统计学、教育测量学等新的计量科学。

教育技术研究资料统计处理的对象,主要是来自试验或实验中的数据。当数据表现为分数或频数这样的数量形式时,研究者通常会用适当的统计方法进行数据分析。统计分析一般和实验、准实验、调查研究的数据结果处理相联系。

如果完全依靠手工来进行,工作量较大,且难以保证准确性,也得不到很高的精度。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技术手段。

例如,在为研究高校人文社会科学研究中立项课题数受哪些因素的影响? 研究者收集到某年全国各省市自治区高校关于社会科学研究的有关数据,以投入科研事业经费数、投入高级职称的人年数、投入人年数、论文数、专著数、获奖数等为解释变量,以立项课题为被解释变量。

该研究可以采用 SPSS 中线性回归分析的方法进行统计分析,最终找出影响立项课题的因素。

另外,在实际教学过程中,无论是阶段性评价,还是终结性评价,通常是以考试的方式来测试学生对知识的掌握情况。如考试后,可以对考生分析每一题的得分率高低的分析方法进行质量分析,但从这些分析中往往无法检测出教师的课堂教学情况,而 SPSS 分析软件则可帮助教师个体诊断并改进自己的教学方法。

为了检测教师在课堂教学中的不足,以执教的某班级在区 (县) 级两次统考中的数据为例,运用 SPSS 技术进行具体分析。如区 (县) 级第一次统考中,质量调研试卷满分 100 分,将考题中的 100 个得分点,按照物理知识体系划分为:基本概念理论 (32 分)、运动知识 (23 分)、物理计算 (8 分)、物理实验 (31 分) 和综合分析 (6 分) 共 5 部分。而区 (县) 组第二次统考中,调研满分 60 分,同样将 60 个得分点按知识点分成 5 部分。然后,使用单一样本 T 检验,通过检测班级样本均值与区 (县) 级的统考样本均值进行比较,看是否存在显著差异。单一样本 T 检验可适用于小样本的检验 (样本数 $N < 30$), 同时被检验的数据要服从正态分布,在描述性分