

- 社会调查的方法
- SPSS系统中数据的编辑修改
- SPSS的基本语法和基本统计知识
- 数据变换
- 频率表与描述统计
- 应用Explore过程检测数据
- 交叉汇总和结合测量
- 两个子总体均值的比较
- 相关分析
- 多选项的统计技巧
- 非参数检验
- 线性回归分析
- 聚类法分析
- 结合分析的应用

高等学校文科类专业大学计算机规划教材

丛书主编 卢湘鸿

# 社会统计分析

## ——SPSS应用教程

周 爽 朱志洪 朱星萍 编著  
阮桂海 审定



清华大学出版社

高等学校文科类专业大学计算机规划教材

丛书主编 卢湘鸿

# 社会统计分析

## ——SPSS应用教程

周 爽 朱志洪 朱星萍 编著  
阮桂海 审定

清华大学出版社  
北 京

## 内 容 简 介

本书共分为14章。第1~第4章介绍SPSS的预备知识。第5~第7章介绍描述统计。第8章介绍两个子总体均值的比较方法。第9章介绍相关分析的方法。第10章介绍多选项的统计分析技巧,这是统计分析必须掌握的内容。第11章介绍非参数检验,可用于数据非正态分布时的检验。第12章介绍一元和多元线性回归分析。第13章介绍聚类法的分析。第14章介绍结合分析的应用。前12章涵盖了SPSS的所有基本统计分析方法,其中各种描述性图形经过特殊加工清晰易懂。第13章、第14章介绍聚类分析和结合分析等技术,是市场调研和品牌研究的重要工具。虽然SPSS的版本几乎每年更新一次,并增加一些模块及功能,但是常规的统计命令及其对话框基本上没有变化。因此,本书内容适用于各种版本的SPSS。

本书内容新颖、实例丰富。每章均附有丰富的习题,并在附录中,提供了习题解答。本书可作为高等学校文科类专业相关课程的教材,也可作为相关人员的自学教材或培训教材。

版权所有,翻印必究。举报电话:010-62782989 13501256678 13801310933

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。

本书防伪标签采用特殊防伪技术,用户可通过在图案表面涂抹清水,图案消失,水干后图案复现;或将表面膜揭下,放在白纸上用彩笔涂抹,图案在白纸上再现的方法识别真伪。

### 图书在版编目(CIP)数据

社会统计分析:SPSS应用教程/周爽,朱志洪,朱星萍编著. —北京:清华大学出版社,2006.3

(高等学校文科类专业大学计算机规划教材/卢湘鸿主编)

ISBN 7-302-12409-4

I. 社… II. ①周… ②朱… ③朱… 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(2006)第005029号

出 版 者:清华大学出版社

<http://www.tup.com.cn>

社 总 机:010-62770175

地 址:北京清华大学学研大厦

邮 编:100084

客户服务:010-62776969

组稿编辑:焦 虹

文稿编辑:徐跃进

印 刷 者:北京国马印刷厂

装 订 者:三河市新茂装订有限公司

发 行 者:新华书店总店北京发行所

开 本:185×260 印张:18.25 字数:449千字

版 次:2006年3月第1版 2006年3月第1次印刷

书 号:ISBN 7-302-12409-4/TP·7952

印 数:1~4000

定 价:24.00元

# 序

能够满足社会与专业本身需求的计算机应用能力已成为合格的大学毕业生必须具备的素质。

文科类专业与信息技术的相互结合、交叉、渗透,是现代科学技术发展趋势的重要方面,是不可忽视的新学科的一个生长点。加强文科类专业的计算机教育是培养能够满足信息化社会对文科人才要求的重要举措,是培养跨学科、综合型文科通才的重要环节。因此,使用一定层次、一定内容的计算机科学与技术知识来武装文科类专业(包括哲学、经济学、法学、教育学、文学、历史学等学科和管理学中的一些专业)的学生(包括研究生、本科生和高职高专生),开设具有文科专业特色的计算机课程是十分必要的。

为了指导文科类专业的计算机教学工作,教育部高等教育司组织制定了《高等学校文科类专业大学计算机教学基本要求(2003年版)》(以下简称《基本要求》)。

《基本要求》把文科类计算机教学的知识结构分为两大部分:一是大学计算机公共基础课;二是在开设计算机公共基础课之后,体现专业特色或与专业教学相结合的后续课程。

计算机公共基础课是为了满足文科类专业的共同需要,而后续课程是为了满足各文科专业的不同需要。

公共基础课由①计算机基础知识(软、硬件平台)、②微机操作系统及其使用、③多媒体知识和应用基础、④办公软件应用、⑤计算机网络基础、⑥Internet基本应用、⑦电子政务基础、⑧电子商务基础、⑨数据库系统基础和⑩程序设计基础等10个模块构筑。这些内容都是每个文科学生应该掌握的,可为他们在与专业紧密结合的信息技术应用方向上深入学习打下基础。根据这些模块可以编著3本书:一是由模块①~⑥组成计算机应用基础;二是由模块⑤~⑧组成网络应用;三是由模块⑨~⑩组成数据库与程序设计。

后续课程在深度上超过了公共基础部分的相应模块,或者是开拓了新的应用领域。它在更大程度上决定了学生应用计算机技术解决本专业实际问题的能力与水平。故建议至少设置16门后续课程。包括:微机组装与维护、计算机网络技术及应用、网页设计基础、动态网页设计——服务器端应用开发、电子政务应用、电子商务应用、法律事务信息处理基础(电子法务概论)、多媒体技术及应用、三维建模与动画设计、数据库基础及其应用、社会统计分析及应用(非财经类)、经济统计与分析软件应用基础(财经类)、信息处理基础、管理信息系统、财经信息化基础、程序设计及应用。

显然,包括文科在内的大学非计算机专业的计算机教学有着广阔的前景。

清华大学出版社推出的高等学校文科类专业大学计算机规划教材,就是根据《基本要求》编写而成的。它可以满足文科类专业计算机公共基础课及其后续课程的教

学需要。

信息技术的发展日新月异,文科类专业的计算机教学还处在探索和不断完善的阶段,因此对教材中的不足或错误,需要同行和读者批评指正。

卢湘鸿

于北京中关村科技园

---

卢湘鸿 北京语言大学信息科学学院计算机科学与技术系教授、教育部高等学校文科计算机基础教学指导委员会副主任。负责教育部高等教育司组织制定的《普通高等学校文科类专业计算机基础课程教学大纲(2000年版)》和《高等学校文科类专业大学计算机教学基本要求(2003年版)》的统稿工作。

# 前 言

社会统计分析软件 SPSS 的全称为 Statistics Package for Social Science。本书是根据教育部高等教育司组织制定的《高等学校文科类专业大学计算机教学基本要求(2003 版)》中关于文科类专业计算机教学后续课程中的相关要求组织编写的。

本书共有 14 章。第 1~第 4 章介绍 SPSS 的预备知识,包括从问卷的正确设计到 SPSS 数据文件的建立以及数据的交换等内容。第 5~第 7 章介绍描述统计,包括 Custom Tables 过程、Frequencies 过程、Descriptives 过程、Crosstabs 过程、Explore 过程等的应用。第 8 章介绍两个子总体均值的比较方法,包括 Means、T-TEST 以及 ANOVA 等过程的应用。第 9 章介绍相关分析的方法,包括双变量相关(零阶相关)和一阶以上的偏相关。第 10 章介绍多选项的统计分析技巧,这是统计分析必须掌握的内容。第 11 章介绍非参数检验,可用于数据非正态分布时的检验。第 12 章介绍一元和多元线性回归分析。第 13 章介绍聚类法的分析。第 14 章介绍结合分析的应用。前 12 章涵盖了 SPSS 的所有基本统计分析方法,其中各种描述性图形经过特殊加工清晰易懂。第 13 章、第 14 章介绍聚类分析和结合分析等技术,是市场调研和品牌研究的重要工具。

本书内容新颖、实例丰富,是社会统计分析软件 SPSS 的入门教材。书中在介绍具体操作时以对话框为主,同时尽可能附上相应的命令进行对照,以便于教学。在对话框模式下统计数据时,可以不必输入命令,只使用对话框就可以输出结果。如果要快速地将数据输入到 SPSS 系统中,就应掌握第 3 章介绍的 DATA LIST 的语法知识和编辑数据文件的技巧,这是 SPSS 的精髓。只有通过 DATA LIST 命令输入原始数据,才能获得事半功倍的效果。遗憾的是,有些读者对 DATA LIST 命令的使用很不耐烦,不能很好地发挥其功能。除了要掌握 DATA LIST 命令的使用,还应掌握 COMPUTE 命令、RECODE 命令、IF 命令的使用,只有这样数据才能正确快捷地输入 SPSS 系统并进行必要的数据变换。在西方发达国家,SPSS 是大学学生的必修课和毕业后谋生的资本。在我国,SPSS 也已逐渐推广应用。

本书可作为高等学校文科类专业相关课程的教材,也可作为相关人员的自学教材或培训教材。为便于教学,书中每章均附有习题,并在附录 A 中提供了习题解答。书中所有程序可在清华大学出版社网站本书的网络资源中下载。虽然 SPSS 的版本几乎每年更新一次,并增加一些模块及功能,但是常规的统计命令及其对话框基本上没有变化。因此,本书内容适用于各种版本的 SPSS。

本书第 1~第 4 章由于惠芳、于小萍、朱志洪、朱志海、朱艳香、刘云、朱星萍、曲庆云执笔。第 5~第 7 章由李欣、阮桂海、阮潮海、吴宝科、孟峥、严康敏执笔。第 8 章由张才彬、陈丹、陈耀辉执笔。第 9 章由周爽、杨武栓、杨靖执笔。第 10 章由罗安娜、郭崇德执笔。第 11 章由赵晓梅、钟晓芳执笔。第 12 章由秦小峰、蔡文英执笔。第 13 章由蔡建平、蔡建瓴执笔。

第 14 章由周爽、蔡建琼、蔡素娟执笔。

北京大学社会学系杨善华教授提供了全套正版 SPSS 原著,在此表示由衷的感谢!

由于时间关系,加之编者水平有限,书中不妥之处谨望读者批评指正。

编者

2005 年 9 月

编者电子信箱: rgh114716@sina.com, rgh114716@mail.china.com

# 目 录

|                                      |    |
|--------------------------------------|----|
| <b>第 1 章 社会调查的方法</b> .....           | 1  |
| 1.1 调查研究的主要方法 .....                  | 1  |
| 1.2 问卷调查法 .....                      | 2  |
| 1.2.1 问卷的主要类型 .....                  | 2  |
| 1.2.2 问卷的基本结构 .....                  | 2  |
| 1.2.3 问卷设计的原则 .....                  | 8  |
| 1.2.4 问卷中问题的设计与<br>答案的构思 .....       | 9  |
| 1.2.5 问卷设计中常见的错误 .....               | 12 |
| 1.3 社会调查的量表法 .....                   | 14 |
| 1.3.1 量表法的作用 .....                   | 14 |
| 1.3.2 利克特量表的应用 .....                 | 15 |
| 习题 1 .....                           | 16 |
| <b>第 2 章 SPSS 中数据的编辑修改</b> .....     | 17 |
| 2.1 SPSS 的 Data Editor 窗口 .....      | 17 |
| 2.2 小规模数据输入法 .....                   | 19 |
| 2.3 大规模数据输入法 .....                   | 21 |
| 2.4 变量合并 .....                       | 21 |
| 2.4.1 变量横向合并的对话框 .....               | 22 |
| 2.4.2 存储变量合并后的结果 .....               | 24 |
| 2.5 个案合并 .....                       | 25 |
| 2.6 个案排序 .....                       | 27 |
| 2.7 数据的行列互换 .....                    | 28 |
| 2.8 限选个案进行区域性统计 .....                | 29 |
| 2.8.1 限选某个子总体 .....                  | 29 |
| 2.8.2 抽取随机样本 .....                   | 31 |
| 2.8.3 限选前 $n$ 个个案 .....              | 34 |
| 2.9 个案加权 .....                       | 35 |
| 2.10 Data 菜单中的其他<br>子菜单命令 .....      | 37 |
| 习题 2 .....                           | 38 |
| <b>第 3 章 SPSS 的基本语法和基本统计知识</b> ..... | 39 |
| 3.1 SPSS 的基本语句 .....                 | 39 |
| 3.2 数据的编码 .....                      | 40 |
| 3.3 SPSS 变量属性的定义 .....               | 42 |

|                                    |    |
|------------------------------------|----|
| 3.4 SPSS 数据的格式 .....               | 43 |
| 3.5 SPSS 数据格式实例 .....              | 43 |
| 3.6 建立 SPSS 的命令文件 .....            | 44 |
| 3.7 4 种 SPSS 命令文件的格式 .....         | 44 |
| 3.8 编辑 SPSS 新命令文件实例 .....          | 46 |
| 3.9 利用新命令文件进行<br>统计分析 .....        | 48 |
| 3.10 调用已有命令文件<br>进行统计分析 .....      | 50 |
| 3.11 统计分析的最佳方案 .....               | 50 |
| 3.12 总体与样本 .....                   | 51 |
| 3.13 参数与统计量 .....                  | 52 |
| 3.14 SPSS 描述性统计常识 .....            | 53 |
| 3.14.1 4 种测量水平 .....               | 53 |
| 3.14.2 3 种综合统计量 .....              | 53 |
| 3.15 正态分布的重要特征 .....               | 56 |
| 3.16 正态分布的经验规则 .....               | 56 |
| 3.17 正态性的假设与检验 .....               | 56 |
| 3.17.1 正态性检验 .....                 | 57 |
| 3.17.2 检验正态性的其他<br>附加方法 .....      | 58 |
| 3.18 显著性水平与临界域 $\alpha$ 值 .....    | 58 |
| 习题 3 .....                         | 59 |
| <b>第 4 章 数据变换</b> .....            | 62 |
| 4.1 创建新变量 .....                    | 62 |
| 4.1.1 使用对话框创建新变量 .....             | 62 |
| 4.1.2 使用 COMPUTE 命令<br>创建新变量 ..... | 63 |
| 4.1.3 数学函数 .....                   | 64 |
| 4.1.4 缺失值函数 .....                  | 65 |
| 4.2 采用 IF 命令进行条件变换<br>和逻辑校验 .....  | 65 |
| 4.2.1 IF 命令格式 .....                | 65 |
| 4.2.2 表达式中的关系符 .....               | 66 |
| 4.2.3 逻辑表达式 .....                  | 66 |
| 4.2.4 两种缺失值的逻辑函数 .....             | 66 |

|                                             |     |                                             |     |
|---------------------------------------------|-----|---------------------------------------------|-----|
| 4.2.5 缺失值的处理 .....                          | 66  | 6.3 统计分析前对假设的检验 .....                       | 105 |
| 4.2.6 IF 命令的对话框 .....                       | 67  | 6.4 集中趋势分布的 3 种较佳<br>平稳测度 .....             | 107 |
| 4.3 数据重新编码 .....                            | 69  | 6.5 在 Windows 对话框中进行<br>Explore 的数据检测 ..... | 109 |
| 4.3.1 通过对话框重新编码<br>数据的实例 .....              | 69  | 6.5.1 利用 Explore 对话框<br>进行统计 .....          | 109 |
| 4.3.2 RECODE 命令的格式 .....                    | 73  | 6.5.2 利用图形探测数据 .....                        | 111 |
| 4.3.3 RECODE 命令在命令<br>文件中的位置 .....          | 75  | 6.5.3 缺失值处理 .....                           | 112 |
| 4.4 计算相同情况的次数 .....                         | 75  | 6.6 在 Syntax 窗口编程 .....                     | 112 |
| 4.4.1 应用 COUNT 命令<br>计数 .....               | 75  | 习题 6 .....                                  | 113 |
| 4.4.2 COUNT 对话框的用法 .....                    | 76  |                                             |     |
| 4.5 SPSS 的运算次序 .....                        | 78  | <b>第 7 章 交叉汇总和结合测量</b> .....                | 116 |
| 习题 4 .....                                  | 79  | 7.1 Crosstabs 过程的应用实例 .....                 | 116 |
| <b>第 5 章 频率表与描述统计</b> .....                 | 81  | 7.1.1 调查所涉及的主要变量 .....                      | 116 |
| 5.1 采用 Frequencies 过程进行<br>频次统计 .....       | 81  | 7.1.2 抽样调查的数据和程序 .....                      | 117 |
| 5.1.1 Frequencies 过程的<br>预备知识 .....         | 81  | 7.1.3 二维交叉汇总表 .....                         | 118 |
| 5.1.2 温差频次统计 .....                          | 86  | 7.1.4 输出结果及分析 .....                         | 118 |
| 5.1.3 一个完整的频率表 .....                        | 87  | 7.2 附加一个控制变量 .....                          | 121 |
| 5.1.4 Frequencies 过程的<br>命令表 .....          | 88  | 7.2.1 统计法 .....                             | 121 |
| 5.2 采用 Custom Tables 过程制表 .....             | 88  | 7.2.2 输出结果及分析 .....                         | 121 |
| 5.2.1 Tables 过程的对话框 .....                   | 89  | 7.3 独立性卡方检验 .....                           | 123 |
| 5.2.2 一个双变量的表格 .....                        | 90  | 7.4 结合测量 .....                              | 124 |
| 5.3 采用 Descriptives 过程描述数据 .....            | 91  | 7.4.1 “标称—标称”数据<br>的测量 .....                | 125 |
| 5.3.1 存储标准分 .....                           | 92  | 7.4.2 “次序—次序”数据<br>的测量 .....                | 127 |
| 5.3.2 在对话框中进行 Descriptives<br>过程的描述统计 ..... | 92  | 7.4.3 “标称—区间”(及以上)<br>数据的测量 .....           | 128 |
| 5.3.3 Descriptives 过程的子<br>命令及其关键词 .....    | 93  | 7.4.4 “比例—比例”数据<br>的测量 .....                | 128 |
| 习题 5 .....                                  | 95  | 7.5 相对风险率估计 .....                           | 128 |
| <b>第 6 章 应用 Explore 过程检测数据</b> .....        | 99  | 7.5.1 研究范例 .....                            | 128 |
| 6.1 Explore 过程可检测的数据错误 .....                | 99  | 7.5.2 相对风险率估计<br>的统计法 .....                 | 129 |
| 6.1.1 错误数据的来源 .....                         | 100 | 7.6 CROSSTABS 命令及交叉汇<br>总和结合测量 .....        | 130 |
| 6.1.2 数据检测的重要性 .....                        | 100 | 7.6.1 CROSSTABS 命令、子命令、<br>关键词 .....        | 130 |
| 6.2 图形检测 .....                              | 101 | 7.6.2 子命令的用法 .....                          | 130 |
| 6.2.1 图形检测之一：直方图 .....                      | 101 | 7.7 采用 Crosstabs 对话框进行<br>交叉汇总和结合测量 .....   | 132 |
| 6.2.2 图形检测之二：茎叶图 .....                      | 101 | 习题 7 .....                                  | 134 |
| 6.2.3 图形检测之三：框图 .....                       | 102 |                                             |     |

|                                       |     |
|---------------------------------------|-----|
| <b>第 8 章 两个子总体均值的比较</b> .....         | 137 |
| 8.1 描述子总体均值的差异 .....                  | 137 |
| 8.1.1 Means 过程应用实例 .....              | 137 |
| 8.1.2 MEANS 过程的命令、<br>子命令的功能及用法 ..... | 141 |
| 8.2 独立样本 T-TEST 的应用 .....             | 143 |
| 8.3 配对样本 T-TEST 的应用 .....             | 145 |
| 8.3.1 采用对话框法实现配对<br>样本 T-TEST .....   | 145 |
| 8.3.2 配对样本 T-TEST 的<br>结果分析 .....     | 146 |
| 8.3.3 采用命令统计法实现配<br>对样本 T-TEST .....  | 146 |
| 8.4 单样本 T-TEST .....                  | 147 |
| 8.5 单因素方差分析 .....                     | 149 |
| 8.6 多因素方差分析 .....                     | 151 |
| 8.6.1 ANOVA 过程的应用之一：<br>求描述性统计量 ..... | 151 |
| 8.6.2 ANOVA 过程的应用之二：<br>方差分析 .....    | 152 |
| 8.6.3 主效应的检验 .....                    | 153 |
| 8.6.4 ANOVA 过程的命令<br>一览表 .....        | 153 |
| 8.6.5 采用对话框进行多因素<br>方差分析 .....        | 154 |
| 习题 8 .....                            | 155 |
| <b>第 9 章 相关分析</b> .....               | 158 |
| 9.1 双变量相关分析 .....                     | 158 |
| 9.1.1 数据 .....                        | 158 |
| 9.1.2 利用相关矩阵进行分析 .....                | 159 |
| 9.1.3 缺失值的剔除法对相关<br>矩阵的影响 .....       | 160 |
| 9.1.4 Correlation 的对话框及<br>其解释 .....  | 162 |
| 9.1.5 运行 Correlation 过程<br>命令 .....   | 163 |
| 9.2 偏相关分析 .....                       | 164 |
| 9.2.1 计算偏相关系数的原理<br>与偏相关的“阶” .....    | 165 |
| 9.2.2 辨别变量之间的虚假<br>相关 .....           | 165 |
| 9.2.3 确定被控制的变量 .....                  | 166 |

|                              |     |
|------------------------------|-----|
| 9.2.4 通过对话框进行偏<br>相关分析 ..... | 167 |
| 9.2.5 偏相关主对话框<br>的应用说明 ..... | 167 |
| 9.2.6 偏相关命令表 .....           | 169 |
| 习题 9 .....                   | 169 |

|                               |     |
|-------------------------------|-----|
| <b>第 10 章 多选项的统计技巧</b> .....  | 171 |
| 10.1 多选项的问卷实例与<br>计算机编码 ..... | 171 |
| 10.2 多选项中分类法与二<br>分法的区别 ..... | 172 |
| 10.2.1 多选项分类法 .....           | 172 |
| 10.2.2 多选项二分法 .....           | 172 |
| 10.3 多选项的数据及程序 .....          | 173 |
| 10.4 多选项频次分布表及其分析 .....       | 175 |
| 10.5 多选项的交叉汇总表实例 .....        | 176 |
| 10.5.1 交叉汇总表的统计法 .....        | 176 |
| 10.5.2 交叉汇总表分析 .....          | 176 |
| 习题 10 .....                   | 177 |

|                                                  |     |
|--------------------------------------------------|-----|
| <b>第 11 章 非参数检验</b> .....                        | 179 |
| 11.1 非参数检验过程的菜单<br>和数据要求 .....                   | 179 |
| 11.2 单样本卡方检验 .....                               | 180 |
| 11.3 二项分布检验 .....                                | 181 |
| 11.4 链检验 .....                                   | 182 |
| 11.5 单样本 K-S 检验 .....                            | 183 |
| 11.6 两个样本 K-S 检验 .....                           | 184 |
| 11.7 $k$ 个独立样本的 Kruskal-Wallis<br>检验 .....       | 185 |
| 11.8 两个相关样本的检验 .....                             | 186 |
| 11.8.1 两个相关样本的 Wilcoxon<br>Signed Ranks 检验 ..... | 186 |
| 11.8.2 两个相关样本的 Sign<br>检验 .....                  | 187 |
| 11.8.3 两个相关样本的<br>McNemar 检验 .....               | 187 |
| 11.9 $k$ 个相关样本无差异的检验 .....                       | 188 |
| 11.9.1 $k$ 个相关样本无差异的<br>Friedman 检验 .....        | 188 |
| 11.9.2 $k$ 个相关样本无差异的<br>Kendall 检验 .....         | 189 |

|                                                |     |                                  |     |
|------------------------------------------------|-----|----------------------------------|-----|
| 11.9.3 $k$ 个相关样本无差异的<br>Cochran 的 $Q$ 检验 ..... | 190 | 13.1.4 二阶聚类的结果分析 .....           | 221 |
| 习题 11 .....                                    | 191 | 13.1.5 二阶聚类小结 .....              | 223 |
| <b>第 12 章 线性回归分析</b> .....                     | 193 | 13.2 谱系聚类 .....                  | 224 |
| 12.1 线性回归分析简介 .....                            | 193 | 13.3 个案 $Q$ 聚类 .....             | 224 |
| 12.2 一元回归模型的拟合优度 .....                         | 193 | 13.3.1 $Q$ 聚类法 .....             | 226 |
| 12.3 预测值及其标准误差 .....                           | 196 | 13.3.2 从冰柱图看聚类结果 .....           | 226 |
| 12.4 数据进行回归分析的基本条件 .....                       | 196 | 13.3.3 平均连接法的图表 .....            | 227 |
| 12.5 设定异常值和影响点 .....                           | 200 | 13.3.4 从树状图分析聚类<br>成员 .....      | 227 |
| 12.5.1 只有 1 个预测变量时的<br>异常值测量法 .....            | 200 | 13.3.5 图形显示的改正 .....             | 228 |
| 12.5.2 有多个预测变量时的<br>异常值测量法 .....               | 201 | 13.4 变量 $R$ 聚类 .....             | 228 |
| 12.5.3 通过删除残差及 Cook 距离<br>法检测影响点 .....         | 202 | 13.5 利用 SPSS 对话框进行<br>聚类分析 ..... | 229 |
| 12.6 向前选择变量法 .....                             | 204 | 13.5.1 对话框的命令按钮 .....            | 230 |
| 12.7 自后淘汰变量法 .....                             | 206 | 13.5.2 选择统计量 .....               | 230 |
| 12.8 逐步回归法 .....                               | 207 | 13.5.3 聚类分析图 .....               | 231 |
| 12.9 多元回归中影响点的检测 .....                         | 208 | 习题 13 .....                      | 231 |
| 12.10 多元共线性诊断 .....                            | 209 | <b>第 14 章 结合分析的应用</b> .....      | 233 |
| 12.11 解释回归模型 .....                             | 211 | 14.1 结合分析的基本构思 .....             | 233 |
| 12.12 在对话框中进行线性回归 .....                        | 212 | 14.2 通过正交设计建立卡片 .....            | 233 |
| 习题 12 .....                                    | 214 | 14.3 问卷调查实例 .....                | 234 |
| <b>第 13 章 聚类法分析</b> .....                      | 217 | 14.4 如何做好结合分析 .....              | 235 |
| 13.1 二阶聚类 .....                                | 217 | 14.5 品牌预测与决策 .....               | 236 |
| 13.1.1 二阶聚类的特点 .....                           | 217 | 习题 14 .....                      | 237 |
| 13.1.2 二阶聚类的数据实例 .....                         | 218 | <b>附录 A 习题答案</b> .....           | 238 |
| 13.1.3 二阶聚类的对话框<br>实例 .....                    | 219 | <b>参考文献</b> .....                | 277 |

# 第 1 章 社会调查的方法

在市场经济大环境中,每一个人都要学会通过社会调查了解市场动态,把握一切机遇,这样才能在激烈的竞争中稳操胜券。

所谓了解社会、了解市场,重要的一点是要学会调查,“没有调查就没有发言权”。从长远看,学会调查研究,可以掌握第一手的数据信息,从而获得科学的指导,这对于培养优秀人才、加快发展国民经济,具有十分重要的现实意义和深远的战略意义。从当前看,调查研究是一门学问,也是学习统计技术的先决条件。

## 1.1 调查研究的主要方法

调查研究的方法有多种,其中最主要的为以下几种方法。

### 1. 问卷调查

问卷调查可以对数据信息进行定量统计分析。这是一种被广泛采用的、先进的调查方法。其特点是由研究人员根据课题研究的项目内容,采用“问卷”的形式设计出问答式的“题目”(详见下面的问卷例子),然后由调查员(最好是亲自)将问卷交给被访者笔答。

问卷调查至少有 3 种形式:

(1) 普查法。此法反馈的信息比较全面,但信息重复的可能性也比较大,而且调查的时间长、耗资大,因此它仅仅适合于小范围内的调查。不过,我国几年一度的全国人口普查、工农业普查等例外。

(2) 抽样调查(即抽查)法。此法是“随机”抽取一个省(或一个县)中的几家企业或部门进行调查研究,即通常所说的“以点带面”。它所获得的信息具有代表性和可靠性,而且时间短、费用省,因而被广泛应用于全国性的抽样调查。比如,抽查全国各地居民购买商品房的状况,抽查国内几个大中城市对某些股市的评价等。

(3) 典型调查法。此法抓住对全局具有指导意义的某事件、某产品进行调查分析,然后总结出经验,进而推广到全局。比如对某几部电视剧开播效果的评估,对国企改革样本的调研与总结,甚至对某种抗癌药物的临床试验与效果检测等。此法可起到典范的或榜样的作用,其优点在于缩小目标,节省人力、物力和财力。但最大的缺点是很难选准、选一个好的典型,因此可靠性比较欠缺。

### 2. 文献追踪

文献是宝贵的信息资源,它记载着各企业、各部门的生产史和营销史。文献包括各种信息报、市场报以及各类广告,特别是政府部门的统计数据及新闻报刊的信息报道等。对各种有价值的文史档案进行跟踪调查,从中分门别类地进行分析对比、去粗存精,就不难预测近期市场营销的发展趋势,进而制定出相应的竞争对策。文献追踪法对社会调查与市场营销具有普遍的实用价值。

### 3. 内部资料调研

知己知彼,方能百战不殆。知己知彼首先要重新认识自己、了解自己。只有从本部门的生产发展史出发,通过对历年积累的生产与供销、成本与利润、人力与资源的统计资料进行综合调研,然后站在全国高度乃至全球高度,与同行业比产值、比利润、比效益,从中预测出产销对路的前景,才能为今后的市场营销打开局面。

### 4. 访谈调查

所谓“访谈”,可以召集一部分代表对营销的状况进行座谈,直接掌握第一手的材料;也可以参加各种经济信息的发布会、商品交易会以及商品订货会等,通过访问与参加,亲自收集可靠的营销信息,了解市场行情。因此,访谈调查法近来逐步为世人所采纳。

### 5. 亲自实践

这是一种值得倡导和实践的调查研究方法。一位伟人曾比喻说:你要知道梨子的滋味,就得亲口尝一尝。同样,作为一名研究人员和决策人员,若要知道本企业产品的供销趋势,就应该注意质量跟踪,甚至亲自站柜台,在与顾客接触的同时,也能真正听到买方对产品质量与价格的评论。

下面着重介绍最常用的问卷调查法。

## 1.2 问卷调查法

问卷调查法是将被调查的问题有条理、有逻辑地罗列在调查表中,让被访者如实回答。此法的优点是清晰、直观、不易遗漏。因此,问卷调查法是社会调查和经济预测的关键手段。

本节将着重介绍问卷的类型、结构及设计方法。

### 1.2.1 问卷的主要类型

由于调查的内容和目的不同,问卷的类型也可以不同。它一般分为“自填”和“他填”两种。

#### 1. 自填类型的问卷

自填类型的问卷,是问卷设计者或研究人员按照数据的时间紧迫性及资料的复杂性,亲自登门调查并将调查的答案填在问卷上,以减少错填率,避免因返工而贻误时间。这类问卷有很强的时间概念,目的在于抢占市场。

#### 2. 他填类型的问卷

此类问卷的数量多,时间性不强,只要求对问题填写清楚即可。因此,研究人员可将这类问卷按调查地点分发到被访者手中,然后派人将填好的问卷收回进行整理与编码;也可以将问卷封好后,寄往外省市或外单位的被访者手中,被访者填好之后邮寄回来进行处理。

### 1.2.2 问卷的基本结构

一般情况下,问卷的基本结构有以下两种形式,每种形式各有3部分结构形体。

第一种形式:①封面致敬信;②正文问答题;③封底指导语。

第二种形式:①封面致敬信;②扉页指导语;③正文问答题。

下面进一步说明问卷中的这3部分结构的形体和内容。

### 1. 封面致敬信

假设生产电视机的某厂家,为了向顾客了解对电视机系列产品的质量与售后服务的意见,设计了“××电视机质量追踪调查”的问卷。为了赢得被访者对此次调查的充分理解和紧密配合,在问卷的封面是如表 1-1 所示的一篇致敬信。

表 1-1 问卷的封面内容

|                                                                                                        |                                                                                                                              |
|--------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|
| ××电视机质量追踪调查                                                                                            |                                                                                                                              |
| 问卷号:                                                                                                   | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
| 被访者所在省份编号:                                                                                             | <input type="checkbox"/> <input type="checkbox"/>                                                                            |
| 所在市(县)编号:                                                                                              | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>                          |
| 致 敬 信                                                                                                  |                                                                                                                              |
| 亲爱的朋友:                                                                                                 |                                                                                                                              |
| 您好!                                                                                                    |                                                                                                                              |
| 产品的质量和售后服务关系到您和千家万户的切身利益。为了提高产品的质量,进一步做好售后服务,尽可能地降低售价,我们特地组织了此次产品质量跟踪调查,以便联络和沟通供需信息。恳切地希望您在百忙中予以支持与帮助。 |                                                                                                                              |
| 本调查表是供企业内部改进产品质量时参考,您不必填写真实单位和姓名,只须根据自己的使用情况或所了解的信息来回答问题。选答时,您觉得哪一项内容符合实际情况,就在该项前面的序号上画一个圆圈○(请勿打√或×)。  |                                                                                                                              |
| 感谢您的合作!并衷心地祝愿您及家人安康欢乐!笑口常开!                                                                            |                                                                                                                              |
| ××省××市××电视机厂销售科                                                                                        |                                                                                                                              |
| 联系人:                                                                                                   | ××× 电话:××××××××                                                                                                              |
| 2005 年 6 月 2 日                                                                                         |                                                                                                                              |

表 1-1 是一种问卷的封面设计,它包含两方面的内容:其一是问卷的题目和编号,其二是写在封面上的致敬信。如果觉得应该突出问卷的题目,则应采用大号字型并占整版版面,并将致敬信排在扉页上。

下面分别介绍题目和编号,以及致敬信的内容。

#### 1) 题目和编号

(1) 题目是问卷调查主题内容的高度概括,要对被访者具有很强的吸引力。正如书名一样,对读者要有吸引力。但是如果光有吸引力而内容污秽不堪,也会适得其反。

(2) 编号是对某问卷的流水记忆,它没有主次之分,而仅是用阿拉伯数字标识前一份问卷与后一份问卷的区别而已。当发现数据输入有误时,可通过问卷的编号查出该问卷,并进行数据的重新输入或改正。

① 问卷号:给全部问卷从第一份到最后一份编号排序。如问卷号由 5 个方格 ☐☐☐☐☐ 表示,则编号范围为 00 001~99 999。

② 省份编号:对全国各省进行抽样调查时,必须将省份作为 1 个变量并加以编号,以便在统计分析时既可统计出全国的总结果,又可统计出各个省份的分结果,还可利用 SPSS 程序抽查出某些省份的统计数字。

③ 县(市)编号:这一级编号也是抽样统计时必不可少的;而且即使不是为了抽样统计,也不能略去县(市)一级的变量及其编号。比如计算机本身不能认识大兴区或仙游县,只

有通过选择大兴区或仙游县的代码(编号),计算机才能从庞大的数据库中提取出北京市“大兴区”或福建省“仙游县”的数据进行分析处理。假定某数据库中北京市的编码为 01,“大兴区”的数据编码为 08,那么,利用 SPSS 统计处理大兴区的数据时,只需输入以下一条命令即可调出所需的数据:

```
SELECT IF (ID1 = 01 AND ID2 = 08)
```

同理,若福建省的编码为 35,“仙游县”的数据编码为 12;利用 SPSS 统计处理仙游县的数据时,只须输入以下一条命令即可调出所需的数据:

```
SELECT IF (ID1 = 35 AND ID2 = 12)
```

余者类推。

说明:

如果还需要统计乡镇企业的状况,就必须在市县变量的后面增加一个“乡镇”变量。

总之,从表面上看,在问卷中对省、市县、乡镇考虑得越周全,数据输入到计算机时的工作量也越大,但实质上,统计处理时却常常可以事半功倍。这样可避免设计问卷时因为省市、县等变量考虑不周到(漏设)而无法分别按省市县进行对比分析,也避免了“明知数据有误却查不到具体的问卷号码”的问题。

2) 致敬信的内容

表 1-1 中致敬信的内容是给被访者看的,其内容亲切诚恳与否,直接影响到被访者填表的态度认真与否,因此务必语气亲切,让人一看便能责无旁贷地、认真作答。因此致敬信应是一封简短亲切的求助信。

(1) 致敬信第一行称呼的写法需要根据被访者的身份而定。可以像表 1-1 开头称呼“朋友们”,也可以称呼“亲爱的顾客朋友们”等。总之,见什么对象就相应地称呼什么。

(2) 致敬信的最后落款应该写明问卷设计者的真实单位或课题名称,如表 1-1 倒数第 2 行的“××省××市××电视机厂销售科”,其中省份一定要写明,否则全国同名同类的销售科多如牛毛,被访者会对来历不明的调查心生疑虑而拒绝配合。

其次,除了写明调查者的真实单位外,最好还要注明调查者的详细地址、联系人的姓名、邮政编码、联系电话等。这样,既便于问卷的回收,又可消除被访者的疑虑,提高相互信赖的程度。示例见表 1-2 中的最后 3 行(调查者的落款)。

表 1-2 调查者的落款

|                                |
|--------------------------------|
| 亲爱的朋友们:                        |
| (正文略)                          |
| ××大学社会学系市场营销预测中心调查组            |
| 联系人:××× 电话:××××××××(邮编:××××××) |
| 2005 年 6 月 2 日                 |

(3) 致敬信的正文。

① 正文的内容要中心突出、简明扼要,以便被访者开门见山地了解此次调查的目的。而且,正文内容还要有一些让对方具有同感并产生共鸣的话语。比如,产品的质量和售后服务关系到千家万户的切身利益,为了提高产品的质量,进一步做好售后服务、降低售价等。

② 正文内容的保密性。调查内容往往涉及到个人对某个敏感问题的看法,比如对国企改革制的看法或月收入多少,甚至女人的年龄。对于这类问题一定要注意让被访者打消顾虑,或给予保密。

为此,在问卷中一定要表明意向,如表 1-1 中所强调的“本调查表是供企业内部改进产品质量时参考,请您不必填写真实单位和姓名”。或者干脆向被访者承诺“问卷中所填的数据绝对予以保密,若有泄露,您可追究我们的责任”等。

③ 填写正文时的注意事项。如何填写问卷中的问题,是关系到本次调查数据的可信度问题,也是节约开支、快速出成果的关键,因此,一定要向被访者指明填写的方法。关于这一点,可参阅 3.——指导语。

2. 问答题

表 1-3 是一份“问卷正文”的简单形式。

表 1-3    ××电视机质量跟踪调查

|                                                  |                                                                                                     |
|--------------------------------------------------|-----------------------------------------------------------------------------------------------------|
| 1. 基本情况: 数字请您填在横线上,选答题请您画○。                      |                                                                                                     |
| (1) 您的性别: 1—男      2—女                           | <input type="checkbox"/>                                                                            |
| (2) 您的年龄____周岁;                                  | <input type="checkbox"/> <input type="checkbox"/>                                                   |
| (3) 您的文化程度:                                      | <input type="checkbox"/>                                                                            |
| 1—文盲   2—初小   3—高小   4—初中   5—高中(含中专、职高等)        |                                                                                                     |
| 6—大专   7—大本   8—研究生(及以上)                         |                                                                                                     |
| (4) 您工作单位的性质:                                    | <input type="checkbox"/>                                                                            |
| 1—国有企业   2—机关事业单位   3—学校   4—乡镇企业   5—合资企业       |                                                                                                     |
| 6—独资企业   7—个体户      8—其他                         |                                                                                                     |
| (5) 月工资____元;                                    | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
| (6) 全家的人均月工资____元。                               | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
| 2. 对本品牌电视机的使用及评价(填法同题 1)。                        |                                                                                                     |
| (7) 您家电视机的品牌:                                    | <input type="checkbox"/>                                                                            |
| 1—长虹牌   2—牡丹牌   3—红星牌   4—上海牌   5—福日牌   6—国内其他品牌 |                                                                                                     |
| 7—东芝牌   8—三洋牌   9—国外其他品牌                         |                                                                                                     |
| (8) 如果是长虹牌,其规格是:                                 | <input type="checkbox"/>                                                                            |
| 1—12 英寸   2—18 英寸   3—24 英寸   4—30 英寸以上          |                                                                                                     |
| (9) 每台买价____元;                                   | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
| (10) 购买至今已____年;                                 | <input type="checkbox"/> <input type="checkbox"/>                                                   |
| (11) 修理过____次;                                   | <input type="checkbox"/> <input type="checkbox"/>                                                   |
| (12) 收视方式: 1—旋钮选台   2—遥控                         | <input type="checkbox"/>                                                                            |
| (13) 清晰度: 1—不清楚,伴有雪花   2—偶尔有雪花   3—清楚   4—很清楚    | <input type="checkbox"/>                                                                            |
| (14) 收视率: 1—天天收看   2—隔天收看   3—每周收看两次   4—每周收看一次  | <input type="checkbox"/>                                                                            |
| (15) 每次约看____小时;                                 | <input type="checkbox"/> <input type="checkbox"/>                                                   |
| (16) 您家收看最多的频道:                                  | <input type="checkbox"/>                                                                            |
| 1—中央一台   2—中央二台   3—中央三台   4—中央四台                |                                                                                                     |
| 5—北京一台   6—北京二台   7—北京三台   8—凤凰卫视                |                                                                                                     |
| 9—其他台                                            |                                                                                                     |

- (17) 收视时更换频道的次数: 1—频繁 2—比较频繁 3—不频繁 ☐
- (18) 电视机故障的主要部位: 1—声音 2—图像 3—声音图像 4—其他 ☐
- (19) 您家电视机故障时维修点的位置: ☐  
 1—在住宅区院内 2—约两站地远 3—约四站地远 4—四站地以外
- (20) 您家附近有无长虹电视机维修点: 1—有 2—无 ☐
- (21) 对长虹电视机的价格与性能比,您认为: 1—不合理 2—较合理 3—合理 ☐
- (22) 对长虹电视机的“价格与性能比”有何要求?

### 1) 问卷的编码格 ☐

表 1-3 是问卷的样例,从中可见:每一个问题的末尾一般都留有 ☐ 格。对这些空格,被访者不必理会,因为那是研究人员专用的。其作用有两方面:一方面可对反馈的问卷进行数据编排;另一方面可对被访者未填或漏填的问题,统一规定出一个值,这个值俗称“用户缺失值”(如表 1-4 的 0)。

表 1-4 数据编码

1. 基本情况: 数字请您填在横线上,选答题请您画○。

- (1) 您的性别: ①—男 2—女 1
- (2) 您的年龄 26 周岁 026
- (3) 您的文化程度: 0  
 1—文盲 2—初小 3—高小 4—初中 5—高中(含中专、职高等)  
 6—大专 7—大本 8—研究生(及以上)  
 .....

### 说明:

表 1-4 右端的数据 1、26、0 等,分别表示填写该问卷的被访者为“男性、26 周岁、文化程度不详(即用户缺失值为 0)”。

### 2) 编码须知

- (1) 一个编码格只能填入一个数字(尽量不用字母表示),之后每一份问卷都照此编码。
- (2) 数值类型要保持一致,不能混编。如果表 1-4 作为第一份问卷中的数字 1 表示男性,2 表示女性时,那么,下一份问卷则必须照此办理,而不能改用 M 表示男性,F 表示女性。
- (3) 数据超范围的处理。如对年龄等变量,设计的是两个编码格 ☐☐,但偶然遇到 100 岁以上的被访者,则属于数据超范围,此时,应该统一增加一个编码格。即 26 岁应编码为 026。

(4) 编码格 ☐ 是为了清晰可见,对于熟练的研究人员,可以省略不用编码格 ☐,而像表 1-4 所示直接在问题的右端编码。

### 3) 数据的栏目位置

通常编辑软件允许每一个物理行最多只能输入 78 个字符:即每一行数据只能输入 78 个数字(或字母),因此,每一行的栏目位置最多有 78 列。

如果一份问卷(one case)的数据超出 78 个字符,就要考虑“一个个案多行记录”的问题。