

- 问卷设计必读
- SAS中最简捷的编程法
- SAS数据的定义、变换及汉化
- 统计方法的预备知识
- 单双变量的频数统计
- 绘制平面图形
- 用MEANS过程计算均值
- T检验与非参数检验
- 结合分析
- 线性相关分析
- 判别分析
- 聚类分析
- 调用GLM过程进行回归分析
- 多个随机实验组的协方差分析
- 典型相关过程CANCORR的应用
- 时序调整的自相关回归分析
- 时间序列分析
- 金融证券基本统计

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

SAS社会统计 实用教程

蔡建平 朱秀萍 等编著
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

SAS社会统计 实用教程

蔡建平 朱秀萍 等编著
阮桂海 审定

清华大学出版社
北京

内 容 简 介

本书分为两篇,共 18 章。第 1 篇是基础篇,为第 1 至第 10 章,基本上包括了统计分析系统(SAS)中的各种基本统计分析方法,面向 SAS 的初学者;第 2 篇是提高篇,为第 11 至第 18 章,包括常用的专业统计知识,面向具备 SAS 基础知识的读者,特别是其中的方差分析、结合分析及金融证券统计分析等知识,是进行市场调研分析及经济统计预测的指南。

本书可作为高等学校相关专业的教材,也可作为统计人员、科研人员、管理人员进行课题研究及定量分析的参考书。

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

SAS 社会统计实用教程/蔡建平 等编著. —北京:清华大学出版社,2006.10

(高等学校文科类专业大学计算机规划教材/卢湘鸿主编)

ISBN 7-302-13241-0

I. 社… II. 蔡… 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 068083 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:北京季蜂印刷有限公司

装 订 者:三河市化甲屯小学装订二厂

发 行 者:新华书店总店北京发行所

开 本:185×260 印张:17 字数:383 千字

版 次:2006 年 10 月第 1 版 2006 年 10 月第 1 次印刷

书 号:ISBN 7-302-13241-0/TP·8366

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、复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、社会统计学分析及应用(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。

• II •

前 言

SAS(statistical analysis system,统计分析系统)是 20 世纪 60 年代初美国 SAS 公司研发的统计分析软件。SAS 与 SPSS 及 BMDP 并称为国际上最有知名度的三大统计软件,如今,在我国社会科学乃至自然科学领域已得到了广泛的普及和应用。

笔者于 1987 年开始研究 SAS 和 SPSS 统计分析软件,一边研发一边教学,并于 1992 年出版了《SAS 与 SPSS/PC⁺ 4.0 高级教程》。本书根据新版本 SAS 进行编写,突出了其精华部分,简明实用。全书共 18 章,分为两篇(基础篇与提高篇)。第 1~10 章是基础篇,基本上包括了 SAS 中的各种基本统计分析方法,面向 SAS 的初学者。第 11~18 章是提高篇,包括常用的专业统计知识,面向具备 SAS 基础知识的读者,特别是其中的方差分析、结合分析及金融证券统计分析等知识,是进行市场调研分析及经济统计预测的指南。每一章均附有习题,并在附录 A 中提供了习题答案,以方便读者使用。部分程序和数据例子可在清华大学出版社网站上下载。

本书既介绍 SAS 的编程统计法;又讲解对话框统计法。编程统计法不受版本的约束,对话框统计法则对不了解 SAS 功能的读者十分适用。根据笔者的体会,使用对话框虽然省事,但是学习使用 SAS 和 SPSS 必须要学会编程,若不会编程则不能真正掌握 SAS 和 SPSS。限于本书篇幅 SAS 的有些内容未能入选,需要 SAS 或 SPSS 技术培训的读者请提前 1 个月发 Email(rghai@sohu.com 或 rgh114716@sina.com)与作者联系。

参加本书编写工作的有蔡建平、朱秀萍、于惠芳、朱志海、阮开林、蔡楠、谢力丹、刘萍、曲庆云、陈丹、吴宝科、李欣、严康敏、杨武栓、郭崇德、赵晓梅、秦小峰、蔡建瓴、蔡建琼、蔡淑娟、樊爱萍、崔博、顾林枝、周爽、阮开顺、孟峥、杨靖,全书由阮桂海审定。

由于时间仓促及笔者水平所限,书中难免有不妥之处,望读者批评指正。谢谢!

北京大学社会学系阮桂海

目 录

第 1 篇 基础篇

第 1 章 问卷设计必读	3	3.10 用 LABEL 汉化变量标签	26
1.1 问卷中问题与答案的设计技术	3	3.11 用 FORMAT 汉化数据	27
1.2 如何避免问卷设计中的错误	6	3.12 用 RETAIN 预置变量值	39
1.3 社会调查的量表法	8	3.13 创建运算的新变量	39
习题 1	9	3.14 数据的重新编码	42
第 2 章 SAS 中最简捷的编程法	11	3.15 语法小结	44
2.1 SAS 运行的基本环境	11	习题 3	45
2.2 SAS 编程实例	12	第 4 章 统计方法的预备知识	46
2.2.1 SAS 编程的一个简单实例	12	4.1 总体与样本的区别	46
2.2.2 SAS 的控制命令	15	4.2 参数与统计量的区别	46
2.2.3 SAS 的数据编码	16	4.3 正态分布的假设与检验	47
2.2.4 SAS 编程的另一个简单实例	18	4.4 经验规则的应用	48
习题 2	19	4.5 对正态分布的假设与检验	48
第 3 章 SAS 数据的定义、变换及汉化	20	4.5.1 检验正态性的重要统计量	48
3.1 SAS 数据的定义法	20	4.5.2 检验正态性的其他统计量	50
3.2 栏目位置不固定时的变量定义法	21	4.6 实际的显著性水平与 α 理论值	52
3.3 栏目位置固定时的变量定义法	21	习题 4	53
3.4 在同一行上输入几个不同观察值的数据	22	第 5 章 单双变量的频数统计	54
3.5 用 CARDS 语句嵌入数据的编程法	22	5.1 用 FREQ 命令统计频数	54
3.6 用 INFILE 语句调用外部数据的编程法	23	5.2 用 Analyst 对话框统计频数	59
3.7 用 IF 语句有条件地转换数据	24	5.3 用 FREQ 过程中的 TABLES 命令加权数据	61
3.8 用 DELETE 语句有条件地删除数据	25	5.4 用 UNIVARIATE 过程产生更详尽的频率表	64
3.9 用 GO TO 语句有条件地跳读数据	26	5.5 用 Analyst 对话框进行双变量的交叉汇总分析	69
		5.6 用编程法进行双变量的交叉汇总分析	74
		5.7 有附加控制变量时的交叉汇总及结合测量	76
		5.8 分析交叉汇总表中的单元内容	79
		5.9 应用实例	80

习题 5	80	8.5.2 用 Univariate 过程进行配 对差值的 T 检验	121
第 6 章 绘制平面图形	82	8.5.3 分析 Univariate 配对差值 T 检验的结果	122
6.1 曲线图的画法	82	8.6 配对差值的 Wilcoxon 非参数 检验	122
6.1.1 用 ASSIST 中的对话框 绘制曲线图	82	8.7 对两个总体均值比较的总结	124
6.1.2 用 PLOT 命令编程绘出 散点图	83	习题 8	125
6.2 用编程法绘制条形图	84	第 9 章 结合分析	127
6.3 用编程法绘制区域图	90	9.1 结合分析的基本原理	127
6.4 用 CHART 过程的 PIE 语句 绘制空心圆形图	92	9.2 采用正交设计建立卡片	127
6.5 用 GRAPHICS 中的对话框 绘制实心圆形图	94	9.3 问卷调查简例	128
6.6 分别用编程法与对话框法 绘制等高线图	97	9.4 如何进行结合分析	129
6.6.1 用编程法绘制等高 线图	97	9.5 对结合分析的进一步研究	131
6.6.2 用对话框法绘制等高 线图	99	9.6 用结合分析进行市场预测	132
习题 6	101	习题 9	133
第 7 章 用 MEANS 过程计算均值	102	第 10 章 线性相关分析	134
7.1 用 Analyst 中的对话框计算 均值	102	10.1 4 种类型的变量	134
7.2 用 PROC MEANS 编程法计算 均值	105	10.2 “比例-比例”型变量的相关 测度	135
习题 7	107	10.3 “二分-区间以上”型变量的相关 测度	138
第 8 章 T 检验与非参数检验	109	10.4 “次序-次序”型变量的相关 测度	138
8.1 两组独立样本的数据描述	109	10.5 “次序-比率”型变量的相关 测度	142
8.2 两组独立样本 T 检验的 假设与检验	111	10.6 “次序-次序”型变量的斯皮尔曼 等级相关系数	145
8.2.1 用对话框进行独立组 样本 T 检验	112	10.7 “标称-标称”型变量的相关 测度	147
8.2.2 用 PROC TTEST 过程进 行独立组样本 T 检验	115	10.8 Cronbach 的 Alpha 与 Spearman 相关系数	148
8.3 两组独立样本的 Wilcoxon 非 参数检验	117	10.9 利用 PROC CORR 编程进行 各种相关分析	148
8.4 描述配对组样本中的数据	119	习题 10	153
8.5 配对差值的 T 检验	120		
8.5.1 配对差值 T 检验的假设与 检验	121		

第 2 篇 提 高 篇

第 11 章 判别分析	157
11.1 判别分析的规则	157
11.2 三种主要的判别分析法	157

11.3	一般判别分析法	157	14.2.1	设计多因变量协方差 分析的程序	195
11.4	逐步判别分析法	163	14.2.2	多因变量协方差的 结果分析	197
11.5	典型判别分析法	164	习题 14		198
习题 11		165			
第 12 章	聚类分析	166	第 15 章	典型相关过程 CANCERR 的 应用	200
12.1	4 个聚类过程	166	15.1	计算典型相关的 SAS 数据	200
12.2	CLUSTER 凝聚法	166	15.2	用 CANCERR 过程命令计算 典型相关	201
12.3	大样本的 FASTCLUS 聚类法	170	15.3	用 CANCERR 对话框计算 典型相关	201
12.4	应用过程 VARCLUS 对变量 进行聚类	172	15.4	输出结果与分析	203
12.5	应用 PROC TRRE 绘制 树状结构图	176	习题 15		207
习题 12		177			
第 13 章	调用 GLM 过程进行回归分析	178	第 16 章	时序调整的自相关回归分析	209
13.1	GLM 过程的一元线性回归	178	16.1	时序自相关的数据	209
13.1.1	数据与程序	179	16.2	数据文件的编辑	209
13.1.2	结果分析	179	16.3	用 ASSIST 对话框进行时序 自相关回归	210
13.1.3	编写回归方程	183	16.4	结果分析	213
13.2	GLM 过程的多元线性回归	184	16.5	最小二乘法回归与调整的自 相关回归相对误差比较	216
13.2.1	GLM 的源程序	184	习题 16		217
13.2.2	多元线性回归的输出 结果	184			
13.2.3	多元线性回归的结果 分析	185	第 17 章	时间序列分析	219
13.3	GLM 过程的多项式回归	186	17.1	季节性调整	219
13.3.1	多项式回归的一般 模型	186	17.1.1	用 ASSIST 对话框进行 统计分析	220
13.3.2	多项式回归的实例	186	17.1.2	结果与分析	222
13.4	虚拟变量的用法	188	17.2	从回归分析中建立回归 预测模型	226
习题 13		190	17.2.1	回归分析的程序	226
			17.2.2	回归分析的输出与 预测分析	227
第 14 章	多个随机实验组的协方差分析	191	习题 17		227
14.1	两组随机数据的单因变量 协方差分析	191			
14.1.1	编程进行单因变量协 方差分析	191	第 18 章	金融证券基本统计	229
14.1.2	单因变量协方差的 结果分析	193	18.1	收益率计算	229
14.2	2 组随机数据(多因变量) 协方差分析	195	18.1.1	内生收益率计算	229
			18.1.2	债券到期的收益率 计算	232

18.2 清算日处于两个到期日之间的 收益率计算	233	附录 A 习题答案	243
18.3 确定债券的价格	234	附录 B 如何快速入门学好 SAS 和 SPSS	258
18.4 债券价格与必要收益率的 关系	235	参考文献	259
习题 18	238		



第1篇

基础篇

第 1 章 问卷设计必读

笔者在所编著的其他书中已介绍过社会调查的方法以及问卷设计的技巧,本书在此着重强调问卷设计中常见的错误,希望能引起市场调研人员的重视。

1.1 问卷中问题与答案的设计技术

问卷中问题的“答案”是用以了解被访者对一系列现象、需求等的意见和看法。如何构思问题的答案,这关系到问卷设计的成败。

这里的所谓“答案”,是封闭式问卷中“问答题”的答案。但在开放式问卷中,因为问答题是让被访者自由回答的,因此没有统一固定的答案。

1. 问题的类型

归纳起来讲,问卷中的问题有以下 5 类。

1) 回答题

如表 1.1 所示,在问题的后面划出一道横线,答案要填在横线上。

表 1.1 回答题

您的年龄: _____ 周岁
您的月工资: _____ 元
全家的人均月工资: _____ 元
.....

这一类问题应该非常简明扼要。

2) 是非题

如表 1.2 所示,在问题后面列出两种判断(是与否、好与坏、对与错),让被访者任选其中一个答案。

表 1.2 是非题

您的户口?	1-居民户	2-农户
您的性别?	1-男	2-女
您家有××电视机吗?	1-有	2-无
.....		

是非题也称“是否”题,是问卷设计中占比例最多的一种问题,其优缺点如下所述。

① 优点: 简单易填,数据错误率较少。

② 缺点: 容易走两个极端,使得信息反映不充分。

例如:

“您家有××电视机吗? 1-有 2-无”

这个问题如果改为表 1.3,那么,信息反馈将会更为充分。

表 1.3 对表 1.2 的改进

您的户口?	1-城市户口	2-农村户口			
您的性别?	1-男	2-女			
您家的电视机是下列哪一种品牌?	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	

3) 多项选一

在一个问题的后面列举出两个以上的答案,让被访者选择最符合自己实际情况的答案。表 1.4 是多项选一的简例。

表 1.4 多项选一的简例

您的婚姻状况是(请在您的答案前画“○”):					
1-未婚	2-已婚	3-离婚	4-再婚	5-丧偶	<input type="checkbox"/>

4) 矩阵式

即把问题排成行,而各种答案排成列,如表 1.5 所示。

表 1.5 矩阵式的答案

您觉得下列电视机的质量如何?					
(请在每一行选择一个答案,然后画“○”):					
	差	较差	一般	较好	好
长虹牌	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
牡丹牌	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
北京牌	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
上海牌	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
福日牌	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

5) 表格式

这是表 1.5 矩阵式的特殊形式,它将表 1.5 用横线和竖线围成一个表格,见表 1.6 所示。

表 1.6 表格式的答案

您觉得下列电视机的质量如何?					
(请在每一行选择一个答案,然后画“○”):					
	差	较差	一般	较好	好
长虹牌	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
牡丹牌	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
北京牌	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
上海牌	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
福日牌	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

【说明】

矩阵式和表格式的答案,其优点是整齐美观、节省版面。其缺点是问题的答案只能简单地重复出现,因此,它仅用于简单的情况调查。

2. 答案的构思

问卷中的问题有两大类:封闭式的问题和开放式的问题。封闭式问题的答案可以由调查单位规定具体的答案。开放式的问题却有五花八门的答案,一般难于硬性规定其答案(收入、身高、体重等定距以上变量除外)。

这里主要是根据封闭式的问题讲述答案的构思。

1) 封闭式问题的答案

设计时必须注意到“互斥性”、“穷尽性”。

(1) 互斥性:是指同一个问题中,答案与答案之间相互不同,而且只能有一个答案是符合实情的,如性别,非男则女,不应是不男不女。因此,设计性别变量时,应该按照表 1.7 设计。比如性别,只有男女两个答案可供选择,不可能有第 3 种性别。

表 1.7 设计性别变量

您的性别?	1-男	2-女	<input type="checkbox"/>
您的户口?	1-居民户	2-农户	<input type="checkbox"/>

(2) 变“非穷尽性”的答案为“穷尽性”的答案。

① 非穷尽性:在封闭式问题中,诸如职业、年龄、爱好、个人专长、电视机品牌、洗衣粉品牌、电冰箱品牌等,其答案有多种多样,即具有“非穷尽性”。如果指定的答案为有限个,被访者则往往觉得答案不符合自身的情况而难于回答。比如表 1.8 中询问顾客喜欢购买哪一种电视机的答案就具有“非穷尽性”。

表 1.8 答案具有“非穷尽性”的问题

您喜欢购买下列哪一种电视机?(请画“○”)	<input type="checkbox"/>			
1-长虹牌	2-牡丹牌	3-红星牌	4-上海牌	5-福日牌
6-东芝牌	7-三洋牌			

对于上述几种品牌的电视机,顾客如果都感到不如意,而是喜欢北京牌的电视机那就无法在问卷中回答该问题,就有可能造成不答或拒答。若拒答的变量多了,问卷就失去了意义。因此,类似表 1.8 的答案属于非穷尽性的答案,应该加以避免。

② 变非穷尽性为穷尽性:其办法很简单,如表 1.9 所示,只要在答案的最后,增加一个“其他电视机”即可。

表 1.9 将表 1.8 的答案改为穷尽性的答案

您喜欢购买下列哪一种电视机?(请画“○”)	<input type="checkbox"/>			
1-长虹牌	2-牡丹牌	3-红星牌	4-上海牌	5-福日牌
6-东芝牌	7-三洋牌	8-其他电视机(自填一种)		

表 1.9 的“8-其他电视机(自填一种)”,使得原有“非穷尽性”的答案,一下子变成了“穷尽性”的答案。

此处用“8-其他电视机(自填一种)”这句话,比用“8-其他电视机”显得“一语双关”,既

可让被访者有答案可选,也便于研究者了解“其他电视机”里面,究竟还有什么品牌的电视机受人青睐。这就是设计的技巧。

2) 区间(定距)以上的变量宜用开放性的答案

所谓区间(定距)变量,英文称之为 Interval variables,是可以对变量值测量出一个区间的距离,例如,个人的月收入、家庭的人均月收入、年龄、身高、体重、分数、发表论文篇数等。

如表 1.10 所示,区间(定距)以上的变量,如果将其数据划分为几组进行回答,固然可以,但是不便于计算平均收入、平均年龄等,无疑会失去数据的真正意义。

表 1.10 铁板一块的错划范围

您的平均月收入处于下列什么范围?(请画“○”)			<input type="checkbox"/>
(1) 600 元以下	(2) 600~1000 元	(3) 1000~2000 元	
(4) 2000~3000 元	(5) 3000~4000 元	(6) 4000 元以上	

表 1.10 中的分组貌似有理,但是统计分析时就会犯难了。因为不能真正统计出本批问卷中人均月收入为多少。错误之处是很明显的:收入被分组后,原来的区间(定距)数据却成了次序(定序)型的数据,而次序型的数据只能计算它们的中位数,若计算其平均收入则失去了原有的意义。因此,应该将表 1.10 改为表 1.11 所示。

表 1.11 修改后的问答题

您的平均月收入_____元?(请填写真实数字)	<input type="text"/>
-------------------------	----------------------

【说明】

统计出平均月收入之后,如果又要统计出表 1.10 中的收入范围,只要利用 SAS 或 SPSS 中的数据编码的命令,将数据重新分组即可。

1.2 如何避免问卷设计中的错误

问卷设计是人们最常用的一种调查方法。但是,由于询问问题的广泛性和复杂性,无论是初用者还是经验丰富的研究者,都难免产生意想不到的差错。

下面列举市场调查中最为常见的问卷设计的错误并予以分析。

1. 问卷设计超出计算机实际处理的能力(见表 1.12)

表 1.12 错误之一

您喜欢下列哪几种产品?(任选 3 种并排序)				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(1) 上海牙膏	(2) 氟化钠牙膏	(3) 洁银牙膏	(4) 上海特效牙膏			
(5) 洗必太牙膏	(6) 其他牙膏					

从问题的提出和答案的要求来看,被访者由于爱好不同,有可能作出以下五花八门的选答:

(123)、(214)、(512)、(215)、(312)等

类似这样的答案,输入给计算机进行统计分析时,其结果既看不出哪一种牙膏(或商品)最受欢迎,也看不出“123”或“214”是什么意思。若让计算机对它们排序,则要经历成百上千种的排列组合,而且排序出来的“123”或“124”等结果一点儿也没有实用价值!

对于此类错误,如果不懂得 SAS 中多项选择的统计法,不妨将表 1.12 中的 6 种牙膏作为 6 个变量(见表 1.13),然后统计出哪个变量的答案被选答得最多、次多、第 3 多……这样修改不仅简单易算,而且效果显著。

表 1.13 纠错之一

请对下列问题选择您的答案(请画“○”选择)			
(1) 您喜欢上海牙膏吗?	1-是	2-否	<input type="checkbox"/>
(2) 您喜欢氟化钠牙膏吗?	1-是	2-否	<input type="checkbox"/>
(3) 您喜欢洁银牙膏吗?	1-是	2-否	<input type="checkbox"/>
(4) 您喜欢上海特效牙膏吗?	1-是	2-否	<input type="checkbox"/>
(5) 您喜欢洗必太牙膏吗?	1-是	2-否	<input type="checkbox"/>
(6) 您喜欢其他什么牙膏?			

2. 区间(定距)以上的变量错划了范围

如上所述,身高、体重、年龄、月利润、月收入、月产量等变量属于区间(定距)以上的变量,但有人自觉或不自觉的就将它们错划为表 1.14 所示的范围。

表 1.14 错误之二

(1) 您的月收入: ① 2000 元以下	② 2000 元以上	<input type="checkbox"/>
(2) 您的月支出: ① 1000 元以下	② 1000 元以上	<input type="checkbox"/>

乍看有理,实际上则行不通!其一因为无法计算出每个人的月平均收入和月平均支出,其二无法计算每个人的月结余为多少。改错的方案如表 1.15 所示。

表 1.15 纠错之二

(1) 您的月收入_____元	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
(2) 您的月支出_____元	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

3. 问题的答案犯了嵌套的错误(见表 1.16)

表 1.16 错误之三

请在符合您的答案上画“○”:			
您最喜欢光顾的商场是: <input type="checkbox"/>			
1-城里大商场	2-附近大商场	3-名牌商场	4-价廉物美的商场

表 1.16 中的错误在于答案“3”有可能包含了其他三个答案。因为名牌商场有可能位于附近,也可能是在城里。所以说,答案“3”有可能嵌套了其他答案。

上面错误只是范例,实际设计中会碰到更多的形形色色的嵌套错误。