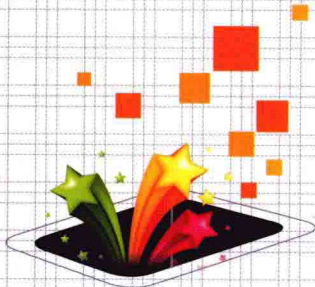




SPSS China



挖掘数据宝藏
助力决策支持

SPSS

统计分析标准教程

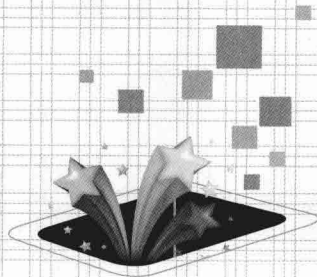
◎ 贾丽艳 杜强 编著
SPSS China 审



258分钟全程视频讲解
SPSS 17.0
正版试用软件



SPSS China



挖掘数据宝藏
助力决策支持

SPSS

统计分析标准教程

◎ 贾丽艳 杜强 编著
SPSS China 审

人民邮电出版社
北京

图书在版编目 (C I P) 数据

SPSS统计分析标准教程 / 贾丽艳, 杜强编著. -- 北京: 人民邮电出版社, 2010.5
ISBN 978-7-115-22281-7

I. ①S… II. ①贾… ②杜… III. ①统计分析—软件包, SPSS—教材 IV. ①C819

中国版本图书馆CIP数据核字(2010)第019384号

内 容 提 要

本书基于 SPSS 17.0 进行编写。首先介绍调查、概率、统计等方面的基本知识, 然后结合案例讲解如何在 SPSS 中进行分析。对几种常用的统计分析方法, 如探索性分析、假设检验、回归分析、方差分析、聚类分析等, 全书穿插有大量经典的案例分析, 读者在解决实际问题的过程中能得到锻炼和提高。每章最后还给出了例题, 读者可以通过解决它们来巩固和提高使用 SPSS 进行统计分析的能力。

本书将统计分析原理和软件操作完美结合起来, 既适合作为一般统计课程的参考用书, 也可供相关专业的本科生、研究生、教师以及其他统计分析人士自学与参考。

SPSS 统计分析标准教程

-
- ◆ 编 著 贾丽艳 杜 强
审 审 SPSS China
责任编辑 刘 浩
 - ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号
邮编 100061 电子函件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
北京铭成印刷有限公司印刷
 - ◆ 开本: 787×1092 1/16
印张: 20.25
字数: 501 千字
印数: 1—4 000 册
- 2010 年 5 月第 1 版
2010 年 5 月北京第 1 次印刷

ISBN 978-7-115-22281-7

定价: 48.00 元 (附光盘)

读者服务热线: (010)67132692 印装质量热线: (010)67129223

反盗版热线: (010)67171154

广告经营许可证: 京崇工商广字第 0021 号

SPSS 公司成立于 1968 年, 最早作为统计分析软件包的开发和发行商而著称。在过去的四十几年中, SPSS 公司以其丰富的预测分析产品获得了许多荣誉, 例如, 被评为纳斯达克上市公司中 22 个利润最高的公司之一, 被《财富》杂志评选为员工最愿意为之工作的公司之一, 其数据挖掘产品 Clementine (现名 Modeler) 被西班牙商业杂志评为最优秀的数据挖掘软件, 数据挖掘技术连续多年在 KD Nuggets 上被投票评选为使用率最高的数据挖掘软件。

SPSS 是世界上最早采用图形菜单驱动界面的统计分析和预测软件, 它最突出的特点就是操作界面极为友好, 输出结果美观漂亮。它将几乎所有的功能都以统一、规范的界面展现出来, 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项。用户只要掌握一定的 Windows 操作技能, 粗通统计分析原理, 就可以使用该软件为特定的数据处理和预测分析工作服务。

SPSS 软件功能强大、界面友好, 而且容易使用。目前在全球有超过 30 万家的客户, 分布在通讯传媒、科研教育、医疗卫生、市场研究、金融证券、保险、化工制造、政府管理、商业分析等多个领域和行业。

SPSS China 成立于 2006 年, 总部位于上海。作为提供优秀统计产品和服务解决方案的国际公司, SPSS China 致力于将最优秀的统计分析、数据挖掘、收集发布类产品及优质服务提供给中国客户。

《SPSS 17 统计分析标准教程》把统计分析知识融合到 SPSS 软件的学习之中, 在学习 SPSS 统计分析的同时, 读者也可以回顾或者学习统计分析的理论和技巧。本书内容涵盖了 SPSS 统计软件的大部分功能, 从 PASW Statistics 软件 (即 SPSS 17.0) 的安装、数据文件的建立和操作, 到 T 检验、回归分析、到生存分析、对应分析、试验设计与分析。本书是熟悉和掌握 SPSS 各个模块统计分析功能最好的参考教材之一, 它具有以下特点:

- 采用最新的 PASW Statistics 17.0, 是目前市场上 SPSS 书籍中版本最新的。
- 每章增加了习题, 便于读者复习和测试。
- 图示清晰明了, 每个图示都增加了详细的注解。
- 每部分都辅以统计知识的讲解。给出了具体的数学公式, 方便读者查阅。
- 覆盖了从基本的描述性统计分析、推断统计分析, 到多元统计分析、生存分析、对应分析、信度分析。

更进一步, 作者给出了在实际生活中应用统计分析的六个经典案例, 它们分别是: 上市公司财务危机预警、影响汇率的因素分析、学生成绩综合评价、高校办学条件的聚类分析、信度分析、多因素试验设计与分析。所有这些都是极好的应用统计分析工具解决实际问题的参考。读者能从中知其所以然, 真正做到举一反三, 事半功倍。

SPSS 中国 董事长 郑镇燮 先生



当前,数据分析软件越来越被各种类型的企业所重视,市场需求迫切,特别是在最近的金融危机情况下,当某些软件销量出现下滑的时候,数据分析软件市场依旧保持增长。

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于 20 世纪 60 年代末开发,同时成立了 SPSS 公司,并于 1975 年在芝加哥组建了 SPSS 总部。1984 年 SPSS 总部首先推出了世界上第一个统计分析软件微机版本 SPSS/PC+,开创了 SPSS 微机系列产品的开发方向,极大地扩充了它的应用范围,并使其能很快地应用于自然科学、技术科学、社会科学的各个领域,世界上许多有影响的报刊杂志纷纷就 SPSS 的统计与绘图功能、深入的数据分析能力、使用方便、功能齐全等方面给予了高度的评价与称赞。

2009 年 4 月 9 日,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 PASW (Predictive Analytics Software),其中,统计分析软件工具 PASW Statistics 就是传统 SPSS 统计软件的延续,只不过重新命名而已。

本书基于统计分析软件 PASW Statistics 17.0 (相当于 SPSS 17.0)中文操作环境编写,一方面帮助读者了解 PASW 新产品的基本功能和操作方式,另外在数据整理、统计分析等方面给出许多实际的案例分析,更加突出其实用性。

本书的主要内容

- 第 1 章介绍了统计调查和概率统计的一些基本知识,这是进行统计分析的必要前提,了解这些原理,可以加深对分析过程和结果的理解。
- 第 2 章介绍如何在 PASW 中对数据、数据文件进行一些常用的操作,包括排序、拆分合并、编码、加权等;此外还重点介绍了如何在 PASW 中进行全面的缺失值分析,掌握这部分内容,有助于深入进行统计分析。
- 第 3~12 章综合介绍了各种常用的统计分析方法,如探索性分析、假设检验、回归分析、方差分析、聚类分析等;每一章中都有经典的案例分析,读者可以在解决问题的同时得到学习和进步;每一章的最后,还给出了习题,读者可以通过练习巩固和提高用 PASW 进行统计分析的能力。

本书的特点

书中图文搭配,在讲解实例时,首先把对话框分析的整个过程演示完毕,保持操作的连续性;然后对一些常用的设置参数进行解释;在最后的結果分析上,也是尽量多地加以说明和分析,增强实用性。

光盘说明

光盘中包含 PASW Statistics 17.0 试用版软件, 应用软件自带的试用授权码, 试用期为一个月。自带授权码到期后, 可以继续使用如下 license (使用期限为 1 个月): db7b8722083290c38686、f2591f9e1860102b6594、d808416e1830d4e3254d、ff400a4c50e35643c454、ef28963249c0b32b8e86, 也可以发送邮件至 book_better@sina.com 再次进行申请。

光盘中还配有 258 分钟全程教学视频, 以便读者更轻松地学会、掌握软件的应用。

致谢

在此, 感谢林光华教授对我的教诲, 感谢所有编写人员的辛勤工作, 感谢金羽工作室张铮博士的帮助和指导。感谢 SPSS China 对本书的大力支持。

由于统计理论和软件的发展都非常迅速, 本书所涉及的内容难免出现一些疏漏和不足之处, 恳请广大读者提出批评与指正 (E-mail: book_better@sina.com)。

作者

2010 年 3 月

目 录

第1章 统计学基础知识1

1.1 统计学概述1

1.1.1 统计学的发展历程1

1.1.2 统计学观念2

1.1.3 统计学现状3

1.2 统计调查基础3

1.2.1 统计调查简述4

1.2.2 资料整理简述7

1.2.3 统计分组介绍8

1.2.4 数据的计量与类型9

1.2.5 统计指标简介11

1.3 常用概率分布12

1.3.1 事件与概率12

1.3.2 概率分布原理14

1.3.3 常用概率分布14

1.4 多元统计分析简介18

习题20

第2章 数据编辑和整理21

2.1 数据和文件概述21

2.1.1 数据编辑器概览21

2.1.2 数据编辑操作23

2.1.3 数据类型和格式24

2.1.4 操作符和表达式29

2.1.5 数据文件简述30

2.2 数据文件处理31

2.2.1 数据排序31

2.2.2 文件拆分32

2.2.3 文件合并34

2.2.4 文件转置38

2.2.5 变量求秩39

2.2.6 变量编码42

2.2.7 计算新变量46

2.2.8 记录加权47

2.3 分类汇总49

2.4 数据文件的结构重组52

2.4.1 数据格式介绍52

2.4.2 PASW 文件重组简介53

2.4.3 对横向格式的重组53

2.4.4 对纵向格式的重组57

2.5 缺失值分析60

2.5.1 缺失值分析简介60

2.5.2 案例应用——收入调查分析62

习题67

第3章 数据探索69

3.1 汇总报告分析69

3.1.1 个案汇总过程69

3.1.2 按行汇总72

3.1.3 按列汇总75

3.1.4 OLAP 汇总78

3.2 探索性分析81

3.2.1 频率分析81

3.2.2 描述性分析84

3.2.3 探索性分析85

3.2.4 交叉表分析89

3.3 作图分析94

3.3.1 作图基本操作95

3.3.2 作图实例97

习题101

第4章 假设检验103

4.1 均值检验105

4.1.1 均值比较简介105

4.1.2 案例应用——袋装食品的

重量检验	105	第 5 章 相关分析	137
4.2 T 检验	107	5.1 相关分析简述	137
4.2.1 单样本 T 检验	107	5.1.1 常用相关系数	138
4.2.2 两独立样本 T 检验	109	5.1.2 PASW 中的相关分析	139
4.2.3 配对样本 T 检验	112	5.2 简单相关分析	140
4.3 非参数检验简介	114	5.3 偏相关分析	142
4.3.1 非参数检验与参数检验的 比较	115	5.3.1 偏相关分析的基本原理	142
4.3.2 非参数检验的特点	115	5.3.2 案例应用——身高与肺活量的 相关分析 2	143
4.4 卡方检验	116	习题	145
4.4.1 卡方检验简述	116	第 6 章 线性回归分析	147
4.4.2 案例应用——测验得分 检验	116	6.1 回归分析简介	147
4.5 二项检验	118	6.2 一元线性回归的基本原理	148
4.5.1 原理简介	119	6.3 多元线性回归	149
4.5.2 案例应用——产品合格率 检验	119	6.3.1 多元线性回归原理简介	150
4.6 游程检验	120	6.3.2 关于模型假设的其他检验	151
4.6.1 原理简介	120	6.3.3 案例应用——环境对植物 生长量的影响分析	151
4.6.2 案例应用——投掷硬币的 随机性检验	121	习题	160
4.7 单样本 K-S 检验	122	第 7 章 逻辑回归	161
4.7.1 原理简介	122	7.1 逻辑回归简介	161
4.7.2 案例应用——身高和体重的 分布检验 1	123	7.2 二元逻辑回归	162
4.8 两独立样本检验	124	7.2.1 二元逻辑回归原理	162
4.8.1 原理简介	124	7.2.2 案例应用——信用评估分析	163
4.8.2 案例应用——收入水平的 一致性检验	125	7.3 多元逻辑回归分析	174
4.9 k 个独立样本检验	127	7.3.1 多元逻辑回归原理	174
4.9.1 原理简介	127	7.3.2 案例应用——旅游偏好分析	175
4.9.2 案例应用——身高和体重的 分布检验 2	128	习题	184
4.10 两相关样本检验	129	第 8 章 非线性回归	186
4.10.1 原理简介	130	8.1 非线性回归	186
4.10.2 案例应用——减肥药的 疗效检验	131	8.1.1 PASW 中的非线性回归	186
4.11 k 个相关样本检验	132	8.1.2 案例应用——投入与产出的 关系研究	187
4.11.1 原理简介	132	8.2 加权回归分析	194
4.11.2 案例应用——商场促销的效果 检验	134	8.2.1 PASW 中的加权回归分析	194
习题	135	8.2.2 案例应用——新产品的销售 分析	195
		8.3 有序回归	198
		8.3.1 案例应用——药品的疗效	

分析	198	10.1.1 聚类方法简述	254
8.4 概率单位回归分析	204	10.1.2 聚类分析原理	254
8.4.1 概率单位回归分析简介	205	10.2 快速聚类	257
8.4.2 案例应用——杀虫剂的效果 分析	205	10.2.1 快速聚类简介	257
8.5 最佳尺度回归	210	10.2.2 案例应用——鸢尾花的归类 分析 1	257
8.5.1 最佳尺度回归简介	210	10.3 系统聚类	262
8.5.2 案例应用——不同饮料的 偏好分析	210	10.3.1 系统聚类简述	262
习题	218	10.3.2 案例应用——汽车车型的 归类分析	262
第 9 章 方差分析	220	10.4 两步聚类	270
9.1 方差分析简介	220	10.4.1 两步聚类简述	270
9.2 单因素方差分析	222	10.4.2 案例应用——鸢尾花的归类 分析 2	271
9.2.1 基本原理	222	习题	278
9.2.2 案例应用——雨量对产量的 影响分析	224	第 11 章 判别分析	280
9.3 多因素方差分析	230	11.1 判别分析简介	280
9.3.1 原理与方法	230	11.2 一般判别分析	282
9.3.2 案例应用——影响销售量的 因素分析	232	11.3 逐步判别分析	291
9.4 协方差分析	241	11.3.1 案例应用——家庭收支的 判别分析	291
9.4.1 协方差分析介绍	241	习题	297
9.4.2 案例应用——肥料的增产 效果分析	242	第 12 章 因子分析	299
9.5 重复测量设计的方差分析	245	12.1 因子分析简介	299
9.5.1 原理与方法	245	12.1.1 因子分析原理	299
9.5.2 案例应用——环境对人反应 速度的影响分析	246	12.1.2 因子分析和主成分分析	300
习题	251	12.1.3 因子分析的基本步骤	300
第 10 章 聚类分类	253	12.2 案例应用——社会经济状况 分析	301
10.1 聚类分析简介	253	习题	310
		附录 SPSS China 培训认证中心简介	312

第 1 章 统计学基础知识

现在的统计学可以认为是应用数学的一个分支,它以概率论为基础建立各种数学模型,对所收集的关于某个问题的数据,进行量化分析、总结,进而进行推断和预测,从而为相关决策提供依据和参考。统计分析广泛应用于各门学科,从物理科学、社会科学到人文科学,还被用于工商业及政府的情报决策之中。

1.1 统计学概述

统计学可以分为描述统计学和推断统计学两个分支。给定一组数据,利用统计学的知识对其进行摘要描述,则称作描述统计学;观察者从数据出发,建立一个用以解释其随机性和不确定性的数学模型,并以此来推论研究步骤及总体的性质,这种用法就被称做推断统计学。这两种用法都可以被称作应用统计学,此外还有一个被称为数理统计学的学科,专门用来讨论应用统计背后的理论基础。

1.1.1 统计学的发展历程

统计学的英文 statistics,最早是源于现代拉丁文的 statisticum collegium (国会)以及意大利文的 statista (国民或政治家);德文中的 Statistik,最早是由 Gottfried Achenwall (1749) 使用,代表对国家资料进行分析的学问,也就是“研究国家的科学”。在 19 世纪,统计学在广泛的数据以及资料中探究事物的规律,并且由 John Sinclair 引进到英语世界。

统计学是一门很古老的科学,一般认为其学理研究始于古希腊的亚里斯多德时代,迄今已有 2300 多年的历史。它起源于研究社会经济问题,在 2000 多年的发展过程中,统计学至少经历了“城邦政情”、“政治算术”和“统计分析科学”3 个发展阶段。所谓“数理统计”并非指独立于统计学的新学科,确切地说它是统计学在第三个发展阶段形成的所有收集和数据分析数据的新方法的一个综合性名词。概率论是数理统计方法的理论基础,但是它不属于统计学的范畴,而属于数学的范畴。

统计学的发展过程经历了 3 个阶段。

(1) 第一阶段称之为“城邦政情”(Matters of state) 阶段。

“城邦政情”阶段始于古希腊亚里斯多德撰写的“城邦政情”或“城邦纪要”。他一共撰写了 150 余种纪要,其内容包括各城邦的历史、行政、科学、艺术、人口、资源和财富等社会和经济情况的比较分析,具有社会科学的特点。“城邦政情”式的统计研究延续了 1000 多年,直至十七世纪中叶才逐渐被“政治算术”这个名词所替代,并且很快演化为“统计学”

(Statistics), 但是统计学依然保留了城邦 (state) 这个词根。

(2) 第二阶段称之为“政治算术”(Political arithmetic) 阶段。

此阶段与“城邦政情”阶段没有很明显的分界点, 本质的差别也不大。“政治算术”的特点是统计方法与数学计算、推理方法开始结合, 分析社会经济问题时更加注重运用定量分析方法。

1690 年, 英国威廉·配弟出版《政治算术》一书作为这个阶段的起始标志, 他用数字、重量和尺度将社会经济现象数量化的方法是近代统计学的重要特征。因此, 这本书被后来的学者评价为近代统计学的来源, 威廉·配弟本人也被称为“近代统计学之父”。

(3) 第三阶段称之为“统计分析科学”(Science of statistical analysis) 阶段。

“政治算术”阶段出现了统计与数学相结合的趋势, 逐渐发展便形成了“统计分析科学”。

十九世纪末, 欧洲大学开设的“国情纪要”或“政治算术”等课程名称逐渐消失, 代之而起的是“统计分析科学”课程。当时“统计分析科学”的课程内容仍以分析研究社会经济问题为主。“统计分析科学”课程的出现标志着现代统计发展阶段的开端。

1908 年, “学生”(William Slessy Gosset 的笔名 Student) 发表了关于 t 分布的论文, 这是一篇在统计学发展史上具有划时代意义的文章, 它创立了小样本代替大样本的方法, 开创了统计学的新纪元。

现代统计学的代表人物首推比利时统计学家奎特莱 (Adolphe Quelet), 他将统计分析科学广泛应用于社会科学、自然科学和工程技术科学领域, 因为他深信统计学是可以用于研究任何科学的一般研究方法。

现代统计学的理论基础——概率论, 始于研究赌博的机遇问题, 大约开始于 1477 年。数学家为了解释支配机遇的一般法则进行了长期的研究, 逐渐形成了概率论理论框架。在概率论进一步发展的基础上, 到十九世纪初, 数学家们逐渐建立了观察误差理论、正态分布理论和最小平方法则等。于是, 现代统计方法便有了比较坚实的理论基础。

1.1.2 统计学观念

为了将统计学应用到科学、工业以及解决社会问题上, 人们往往由研究总体开始。所谓总体, 可能是一个国家的人民, 或是石头中的水晶, 或者是某家特定工厂生产的商品; 甚至可能由许多次同样的观察所组成, 时间序列就是由这种数据所组成的总体。

根据实际需要, 人们经常选择研究总体的一个子集来代替研究整个总体, 这个子集称作样本。资料和数据是统计分析的对象, 并且有做两种相关的用途: 描述和推论。

描述统计学用于处理有关叙述的问题: 资料是否可以被有效地摘要, 不论是以数字或是图片、表格, 来代表总体的性质; 基础的数学描述包括平均数和标准差等; 图像的摘要则包含了许多种表和图。

推断统计学被用来将资料中的数据模型化, 计算它的几率并且做出对于总体的推论。这个推论可以是如下一些内容: 某个假设是否成立, 对于数字特征量的估计, 对于未来观察的预测, 关联性的预测, 或是将关系模型化 (如回归), 等等。

如果样本足以代表总体, 那么由样本所做的推论和结论可以被引申到整个总体之上。因此最大的问题在于决定样本是否足以代表整个总体, 为此, 统计学提供了许多方法来估计和修正样本和资料收集过程中的随机性 (误差)。数理统计 (通常又称为统计理论) 是应用数学的分支, 是使用概率论来分析并且验证统计的理论基础。

任何统计方法都是有效的,只要数据资料或是所讨论的总体满足该方法的基本假设。误用统计学可能会导致描述面或是推论面严重错误。即使统计学被正确应用,其分析结果对于不是专家的人来说可能也会难以陈述。举例来说,统计资料中显著的改变可能是由样本的随机变量所导致,但是这个显著性可能与大众的直觉相悖,人们需要一些统计的技巧(或怀疑)以面对日常生活中通过统计数据所获得的信息。

常用的统计分析方法包括描述性统计、方差分析、回归分析、聚类分析和判别分析等。

1.1.3 统计学现状

在科学技术飞速发展的今天,统计学广泛吸收和融合相关学科的新理论,不断开发、应用新技术和新方法,深化和丰富了统计学传统领域的理论与方法,并拓展了新的领域。今天的统计学已展现出强有力的生命力。随着我国社会主义市场经济体制的逐步建立,实践发展的需要,对统计学提出了更多新的、更高的要求。

第一,对系统性及系统复杂性的认识为统计学的未来发展增加了新的思路。由于社会实践广度、深度地迅速拓展,以及科学技术的高度发展,人们对客观世界的系统性及系统复杂性的认识也更加全面和深入。随着科学融合趋势的兴起,统计学的研究触角已经向新的领域延伸,新兴起了对探索性数据统计方法的研究,研究的领域也向复杂客观现象扩展。21世纪统计学研究的重点将由确定性现象和随机现象转移到对复杂现象的研究上,如模糊现象、突变现象及混沌现象等新的领域。可以说,复杂现象的研究给统计学开辟了新的研究领域。

第二,定性与定量相结合的综合集成法将为统计分析方法的发展提供新的思想。定性与定量相结合的综合集成方法是钱学森教授于1990年提出的,这一方法的实质就是将科学理论、经验知识和专家判断相结合,提出经验性的假设,再用经验数据和资料以及模型对它的确实性进行检测,经过定量计算及反复对比,最后形成结论。它是研究复杂系统的有效手段,而且在问题的研究过程中处处渗透着统计思想,为统计分析方法的发展提供了新的思维方式。

第三,统计科学与其他科学渗透将为统计学的应用开辟新的领域。现代科学发展已经出现了整体化趋势,各门学科不断融合,已经形成一个相互联系的整体。由于事物之间具有的相互联系性,各学科之间研究方法的渗透和转移已成为现代科学发展的一大趋势。许多学科取得的新进展为其他学科发展提供了全新的发展机遇。模糊论、突变论及其他新的边缘学科的出现为统计学的进一步发展提供了新的科学方法和思想。将一些尖端科学成果引入统计学,使统计学与其交互发展将成为未来统计学发展的趋势,统计学也将会有一个令人振奋的前景。今天已经有一些先驱者开始将控制论、信息论、系统论以及图论、混沌理论、模糊理论等方法 and 理论引入统计学,这些新的理论和方法的渗透必将会对统计学的发展产生深远影响。

统计学产生于应用,并在应用过程中发展壮大。随着社会经济的发展、各学科相互融合趋势和计算机技术的迅速发展,统计学的应用领域、统计理论与分析方法也将不断发展,在所有领域展现它的生命力和重要作用。

1.2 统计调查基础

统计调查是按照统计研究的目的和任务,运用科学的组织形式和方法,有组织、有计划地搜集统计资料的工作过程。统计资料的搜集是统计工作过程的一个重要阶段,它是统计工

作定量认识的开始,是整个统计研究的基础,也是进一步整理和分析的前提。只有在占有了丰富客观的统计资料的基础上,才能通过进一步的整理和分析达到对社会经济现象的正确认识,从而更好地发挥统计的信息、咨询和监督职能。

1.2.1 统计调查简述

通过统计调查搜集的数据资料,可以分为如下两种形式。

- 原始资料:是指向总体单位搜集的反映个体特征的、有待于进一步加工整理才能反映总体特征的统计资料。原始资料是总体单位的标志表现。
- 次级资料:是指对原始资料已经经过某种加工、在一定程度上可以反映总体特征的统计资料。归根到底,各种次级资料都来源于原始资料。

统计调查必须达到准确性和及时性两个基本要求。准确性是指调查资料客观反映现象和过程本质的程度,及时性则指搜集资料完成的时间符合该项调查所规定的要求。统计调查的准确性和及时性是衡量统计调查工作质量的重要标志。

统计调查的一个重要问题,是根据不同的调查目的和调查对象的特点,选择合适的调查方法。统计调查有各种各样的分类,根据被研究对象总体的范围,可分为如下两种。

- 全面调查:对构成调查对象总体的所有个体逐一进行调查登记的一种调查方式。
- 非全面调查:对构成调查对象总体的一部分个体进行调查登记的一种调查方式。

社会经济现象的数量特征从其发展变化的性质看,分为两种类型:一种是其发展变化表现为一段时间内活动过程的累积结果,称为时期现象;另一种是其发展变化表现为一定时点上的瞬间状态的结果,称为时点现象。时期现象必须通过连续调查才能取得反映社会经济现象在一段时间内发展变化的累积结果;时点现象则通过不连续调查取得反映社会经济现象发展变化到某一时点的数量特征。

按搜集资料方法的不同,统计调查的方式可分为:直接调查、凭证调查、派员调查、问卷调查和统计调查方案等。

统计调查是一项高度统一、具有科学性和群众性的实践活动,应该有计划、有组织地进行。从统计调查工作的长期实践看,无论采取哪种调查方式,都要事先根据实际需要与操作可能性,对被调查现象进行调查研究的,设计制定出切实可行、行之有效的统计调查方案。

调查方案是进行某项调查活动达到预期目的的指导性文件。一个完整的统计调查方案应包括以下基本内容:调查目的、调查对象、调查项目、调查表、调查时间和调查的组织工作。下面分别加以说明。

1. 调查目的

确定调查目的是任何一项统计调查方案首先要解决的问题。任何社会经济现象和过程都可以根据实际需要不同方面、不同角度来搜集资料。调查目的就是指通过某项特定的调查,需要弄清的情况和解决的问题。调查目的不同,调查的内容和范围也就不同。目的不明,就无法确定向谁调查、调查什么,以及用什么方式、方法进行调查,其后果必然是调查活动的盲目性。

2. 调查对象和调查单位

确定调查对象和调查单位,是为了解决向谁调查、由谁具体提供统计资料的问题。

调查对象是所要调查研究的社会经济现象的总体,它由许多性质相同的个别单位组成,即统计总体。调查对象由调查目的决定,通过对被调查对象地认真分析,掌握其主要特征,从而科学规定调查对象的涵义,明确调查对象的范围,即明确哪些现象该调查,哪些现象不该调查。调查单位也就是总体单位,它是调查对象的组成要素,即调查对象所包含的具体单位,是调查项目的承担者。

另外,还需将调查单位和填报单位加以区分。填报单位也叫报告单位,它是统计调查中负责向上级报告调查内容、提交统计资料的单位。在调查中,调查单位和填报单位两者可能一致,也可能不一致。

总之,确定调查单位可使我们知道从哪里取得有关标志的具体资料,它可以是人、企事业单位和物;确定填报单位可使我们知道由谁提供具体资料,它只能是某个人或某个单位。

3. 调查项目和调查表

调查项目又称调查纲要,是依附于调查单位的基本标志,它完全由调查目的、任务和调查对象的性质特点决定,包括品质标志和数量标志所构成的标志体系。通俗地说,调查项目就是一份在调查过程中应该获得答案的各种问题的清单。

确定调查项目时,应以调查目的为依据,结合被调查对象的性质特点,同时考虑到国家管理、科学研究等方面对统计资料的需要。在拟定调查项目时,应遵循“少而精”的原则,并注意以下4个问题。

- 所选择的调查项目必须能够取得确切的资料。
- 调查的每一个项目应该有确切的涵义和统一的解释,避免各自按不同的理解进行调查,使资料无法按统一的标准汇总。
- 应紧紧围绕调查目的和任务确定调查项目。
- 注意同类调查中确定的调查项目的前后连贯性,以满足对社会经济现象动态对比分析的需要。

调查项目确定之后,就要进一步使用调查表。把诸多调查项目按照一定顺序排列在表格上,就形成了调查表,使调查登记资料规范化、标准化。

调查表有单一表和一览表两种形式。单一表是在一张调查表上只登记一个调查单位的项目,适用于调查项目较多、比较详细的调查,便于进行分类整理。一览表是在一张调查表上登记若干调查单位的项目,容纳的调查项目要少一些,它便于比较各个调查单位的情况,便于相互核对和汇总计算。

调查表的格式一般由表头、表体和表脚3部分组成。表头在调查表的上方,包括调查表的名称、填报单位的名称、地址、性质、隶属关系等,它们是核实和复查各调查单位资料时必不可少的内容。表体是调查表的主要部分,包括调查项目和项目的具体表现、栏号、计算单位等,便于在整理调查资料时或编写填表说明时使用。

调查表拟定好后,必须附有简明扼要的填报说明和调查项目解释。前者用来说明填表时应注意的事项;后者则是对调查表所列各个调查项目、统计指标的涵义、包括范围、计算方法等进行统一的规范说明,保证调查资料的准确性和标准化。

4. 调查时间和时限

调查时间是指调查资料所属的时间,即客观时间。如果调查的是时期现象,调查时间就

是资料所反映的起止日期；如果调查的是时点现象，调查时间就是规定的统一标准时间。

调查时限是指进行调查工作的期限，包括搜集资料和报送资料整体工作所需的时间，即主观时间。为保证调查资料的及时性，任何调查都应尽可能缩短调查时限。

5. 调查的组织工作

在调查方案中，还必须研究确定调查的组织工作计划，使调查工作有组织上、措施上的保证。组织工作计划包括调查机构、调查地点、选择调查方法等问题。此外，调查组织工作计划中，对于调查前的准备工作，包括宣传教育、调查员培训、文件印刷、调查资料报送办法、调查经费的预算和开支办法、提供或公布调查成果的时间等，均应做具体规定。

对于大规模统计调查，确定调查方案前往往需要做试点调查。通过试点，检验调查方案是否可行，以便加以修改和补充。此外，还要积累实施调查方案的经验，提高调查人员的业务技能，确保圆满完成调查任务。

我国现已明确提出统计调查方法改革的目标模式，即建立一个以必要的周期性普查为基础，以经常性的抽样调查为主体，同时辅之以重点调查、科学推算和有限的全面报表综合运用的统计调查方法体系。因此，统计报表、抽样调查和普查是最基本的统计调查方式。

6. 周期性普查

普查是专门组织的，一般用来调查属于一定时点上社会经济现象数量的全面调查，它能为国家制定重大方针政策、编制长远规划提供重要依据。普查有如下一些特点：普查的对象主要是时点现象，它的数量在短期内往往变动不大，不需要连续登记，只要间隔一段较长的时间进行调查；普查与其他任何调查相比，更能提供全面、系统的国情国力方面的基本统计资料。

7. 抽样调查

抽样调查是一种专门组织的非全面调查，它是按照随机原则从总体中抽取部分单位组成样本，对样本单位进行测定并计算样本指标，据此对总体的数量特征进行估计或作出判断的一种调查方式。抽样调查是现代推断统计的核心。因为无论是对总体的参数估计或假设检验，都是以测定样本得到的样本指标——统计量为依据的。

与其他非全面调查相比，抽样调查具有如下 3 个重要特点。

(1) 按照随机原则从总体中抽取样本单位，随机原则就是机会均等原则。调查者不带有主观倾向，完全凭借偶然性抽取样本单位，保证了总体中每一单位都有同等的中选可能性。随机原则是抽样技术的基本原则，只有按照随机原则抽取样本，才能使样本的结构同总体结构相似，由此获得的样本指标才可以作为随机变量，以概率论为基础对总体进行估计或检验。

(2) 抽样调查是以样本指标（即统计量）为依据推断总体参数或检验总体的某种假设，在于对总体数量特征作出估计或某种判断，而且它是以概率论阐明的有关分布规律为依据的估计，可以计算其可靠性和精确度。例如，抽取部分产品检验其质量，计算出样本的合格率和抽样误差，然后推算全部产品的合格率，并指明这种估计的可靠程度和精确度。

(3) 抽样调查的误差可以事先计算并加以控制。用样本指标推算总体，不可避免地会产生误差，即抽样误差。抽样误差是随机误差，它是抽样调查所固有的，是对抽样推断精确度

的一种度量,其分布具有一定的规律性,可以依据这种分布的规律和具体的抽样条件计算抽样误差的大小。

8. 重点调查

重点调查是专门组织的一种非全面调查,即在所要调查的现象总体的全部单位中选择一部分重点单位进行调查,借以了解总体的基本情况。所谓重点单位,是着眼于现象数量方面,尽管这些单位在全部单位中只是一部分,但是它们的某一标志的标志总量在总体标志总量中占有绝大比重,通过对重点单位的调查,就能掌握总体的基本情况。

重点调查实质上是范围较小的全面调查,其目的是反映现象总体的基本情况。一般地,当调查任务只要求掌握调查对象的基本情况,而在总体中确有部分单位又能比较集中地反映所研究的项目和指标时,采用重点调查比较适宜。同时,由于重点单位比较少,在调查过程中调查项目可多一些,获得的统计资料就较丰富、详细。重点单位虽然对总体来说最具有代表性,但是不可能完整地反映现象总体,也不具备推断总体总量的条件。

9. 典型调查

典型调查也是专门组织的一种非全面调查,它是根据调查研究的目的和任务,对所研究的现象总体进行初步分析的基础上,有意识地选择若干具有代表性的单位进行调查和研究,借以认识事物发展变化的规律。

所谓有代表性的典型单位,是指那些最充分、最集中体现总体某方面共性的单位。

典型调查的特点在于,调查单位是根据调查目的和任务,在对现象总体进行全面分析的基础上,有意识地选择出来的。显然,典型调查单位的确定与其他非全面调查相比较,更多地取决于调查者主观的判断与决策。若要社会经济现象的复杂性和期望判断结果尽可能准确些,就要求被选择的各个典型应该在总体所要研究的特征中最具有代表性。也就是说,这些单位的数量标志表现最能反映总体各个单位的一般水平。

以上介绍的各种方法各有特点和作用,在体系中的地位也各不相同,而对复杂多变的社会经济现象,采用一种调查方法来搜集资料往往不能达到预期的目的,必须将多种方法结合运用。如何在实践中将不同调查方法结合运用是一个灵活性很大的实践问题。例如,统计报表和典型调查的结合,既可以搜集到全面的统计资料,又可以掌握各种具体情况,既有助于正确应用统计资料,也有助于深入了解报表存在的问题。对普查取得的全面资料,可以运用抽样调查的结果加以检验和修正,借以评价普查资料的质量,掌握普查结果的可靠度。

总之,在具体实践中应根据不同专业、不同调查目的和不同的调查对象,科学地综合运用多种统计调查方法,不断提高统计调查的效率和效益,达到数据准确、信息灵通、资料丰富的要求,以适应市场经济条件下各级政府、各部门、各企业及社会公众对统计信息不断增长的需求。

1.2.2 资料整理简述

统计资料的整理工作就是根据统计的目的,对调查所得的原始资料进行审核、科学地分类和汇总,或对已经过加工的综合统计资料进行再加工,使其成为可供统计分析使用的描述现象总体综合特征的资料。

通常,调查获得的原始资料是杂乱、分散、零星而无序的,因而还不能反映事物的本质

和变化规律性。要完成统计的任务还必须对统计的原始资料进行整理,使之条理化、系统化,达到去粗存精、去伪存真、由此及彼、由表及里地进行科学加工的目的,并在此基础上进一步进行统计分析,从而认识到事物的本质,掌握事物发展变化的规律性。有时即使是对于某些已经过加工的综合资料,也需要按统计目的的要求重新进行整理。统计资料的整理是统计工作中一个十分重要环节。

统计资料整理的一般步骤如下。

(1) 首先对统计调查得来的资料进行严格的检查和审核,审核的内容如下:

① 资料的完整性。检查预定调查对象的资料是否齐全,调查所规定资料的项目是否完整。

② 资料的及时性。检查所获得的资料是否符合调查的时间上的要求。

③ 资料的正确性。要检查调查资料的有关项目的内容是否合理,不同的项目之间的资料有无矛盾之处。还要检查调查资料在计算上有无错误,例如可以对列联表中的有关合计数字纵横相加,以验证计算是否正确。

(2) 进行分组汇总。在此之前需设计统计资料的汇总方案。汇总方案要能反映最基本的、最能说明问题本质特征的统计指标。分组既要考虑统计指标汇总的要求,也要考虑到原始资料的特点。在确定汇总和分组的办法后,按分组要求进行分组汇总,计算各组的分组指标和综合指标。

(3) 再次审核。为了保证汇总整理的质量,还需要对已整理好的统计资料再次进行审核。审核后编制统计表,简明扼要地表达现象在数量方面的特征和有关联系。

1.2.3 统计分组介绍

统计分组就是根据事物的内在特点和统计任务的要求,对所研究的统计对象按照一定的标志划分为若干个部分或组。把统计对象中那些性质相同的单位集中在一起,把不同性质的单位分在不同的部分或组中。分组后组内各单位性质相同,组与组之间性质不同。

统计分组具有如下一些作用。

(1) 区别现象的类型。把某些标志相同的单位划分在同一组内,把标志不同的单位分在不同的组内,就可以区分现象的类型。对各类型的组计算它们的数量特征就可以对总体的不同组或部分进行分析。

(2) 揭示现象的结构。用某一标志将性质不同的单位进行分组,就可以计算各组的统计指标在总体中所占的比重,就可以反映现象的内部结构。

(3) 反映现象与现象之间的依存关系。按照有关指标对现象进行分组,可以分析该因素对其他因素的影响程度和因果关系。

下面介绍关于统计分组的方法。

(1) 分组标志和分组界限。

分组标志就是将总体区分为不同组别的标准和根据。一个总体一般都具有多种特征,要根据研究的需要,选择适当的标志作为分组标志。对于同一批资料,若采用的分组标志不同则有可能得出相异甚至相反的结论。标志选择不当,分组结果就不能正确地反映总体的性质特征。总体单位的标志有品质标志和数量标志两种,因此分组标志也有品质标志和数量标志两种。

分组界限就是要在分组标志的变异范围内,划定各相邻组间的性质界限和数量界限。任