



新编大学社会学教材

总主编 风笑天

Social Statistics and  
the Application of SPSS



# 统计学与SPSS应用

(第二版)

◎ 张小山 主编



华中科技大学出版社

<http://www.hustp.com>

新编大学社会学教材

总主编 风笑天

Social Statistics and  
the Application of SPSS

社会

# 统计学与SPSS应用

(第二版)

● 张小山 主编



华中科技大学出版社  
<http://www.hustp.com>

中国·武汉

## 内 容 提 要

社会统计学是社会科学进行定量分析与研究的基本工具,也是社会学及相关专业的必修课程。本书在扼要介绍社会统计学的核心概念、发展简史、学科背景之后,着重介绍了该门学科的基本原理与方法,尤其是在社会研究领域里有关数据资料的整理、分析和推理的具体技术,主要包括单变量和双变量的描述性统计、基础概率与常用统计分布、参数估计和假设检验、方差分析与回归分析和 SPSS 软件的应用等。

本书依托作者多年的教学和科研实践,借鉴了国内外的研究成果,精心挑选与实际生活密切相关的大量例题和习题,通过系统合理的编排和深入浅出的讲解,力图让数学基础薄弱的读者也能既轻松又扎实地掌握在社会研究中常用的统计方法及其实际操作技术。全书体系规范、取舍得当、叙述清晰、通俗易懂,较好地满足了广大学生学习社会统计基础知识的需要。

本书适合作为高等院校社会学及相关专业本科生的教材或教辅,也可供对社会研究统计分析感兴趣的各类读者阅读与参考。

### 图书在版编目(CIP)数据

社会统计学与 SPSS 应用/张小山主编.—2 版.—武汉:华中科技大学出版社,2018.8

新编大学社会学教材

ISBN 978-7-5680-4232-1

I. ①社… II. ①张… III. ①社会统计-统计分析-软件包-高等学校-教材 IV. ①C91-03

中国版本图书馆 CIP 数据核字(2018)第 181817 号

### 社会统计学与 SPSS 应用(第二版)

张小山 主编

Shehui Tongjixue yu SPSS Yingyong

策划编辑:钱 坤

责任校对:章 红

责任编辑:朱建丽 钱 坤

责任监印:周治超

封面设计:刘 卉

出版发行:华中科技大学出版社(中国·武汉)

电话:(027)81321913

武汉市东湖新技术开发区华工科技园

邮编:430223

录 排:华中科技大学惠友文印中心

印 刷:武汉市籍缘印刷厂

开 本:787mm×1092mm 1/16

印 张:22.25

字 数:549 千字

版 次:2018 年 8 月第 2 版第 1 次印刷

定 价:48.00 元



本书若有印装质量问题,请向出版社营销中心调换  
全国免费服务热线:400-6679-118 竭诚为您服务  
版权所有 侵权必究

## 总 序

社会学在中国内地恢复重建已将近 30 年了。伴随着中国社会改革开放的深入,社会学学科也在不断探索和研究中国社会的过程中一天天发展进步。正是急剧的社会变迁和纷繁复杂的社会生活,为中国社会学的发展壮大提供了最大的舞台。社会学也在这个舞台上大显身手,在帮助人们认识社会规律、制定社会政策、解决社会问题、创建和谐社会等诸多方面发挥着越来越大的作用。

自 20 世纪 80 年代初期进入社会学领域以来,笔者就深深地被这一学科的独特视角、研究方法以及丰富多彩的研究领域所吸引,20 多年来一直在大学醉心于社会学专业的教学与科研工作。80 年代中期在北京大学攻读硕士和博士学位的 5 年中,笔者就从一些国外的和国内早期的社会学教材中获得了许多的收益。90 年代中期至新世纪初在华中理工大学(现改名为华中科技大学)社会学系任教的几年中,曾出版了《现代社会调查方法》(独著,1996 年初版,2001 年第 2 版,2005 年第 3 版)和《社会学导论》(主编,1997 年)两本专业基础课教材。出版社反馈的信息表明,这两本教材受到了广大教师和学生的普遍欢迎,《现代社会调查方法》已发行 12 万册。

正是在这样一种基础上,出版社盛情邀请笔者组织编写一套“新编大学社会学教材”。这对笔者来说,无疑是一项十分艰巨的任务。因为一方面自己才疏识浅,对社会学的许多领域了解不多,难以胜任这一工作;另一方面,自己日常的教学、科研以及指导研究生的任务也比较繁重,时间和精力上也有一定困难。虽几次推托,但终究经不住出版社的盛情邀请,只好勉为其难地担当起这套教材总主编的重任。

客观地说,目前国内的社会学教材已有不少。但除了“社会学概论”、“社会学研究方法”、“西方社会学理论”、“社会心理学”、“社会统计学”等主干课程的教材版本相对较多外,其他分支社会学的教材往往很少。同时,许多教材由于编写和出版的时间较早,部分内容已不适应当前教学的需要。另外,相对于美国等西方发达国家中社会学教材种类繁多的状况,国内社会学教材可选择的余地实在太小。这些状况为这套教材的编写提供了现实的需要。

这套“新编大学社会学教材”的编写目标,主要体现在以下三个方面。

一是全面性。希望经过 5~10 年的努力,编写出覆盖目前社会学系所开设的 30 门左右的专业课程的教材(其中,专业基础课教材 10 种左右,专业选修课教材 20 种左右),以增加社会学教师选择教材的空间,同时改变目前少数专业选修课找不到教材的局面。

二是规范性。这种规范性一方面体现在全套教材的整体规划和单本教材的具体设置都是依据社会学学科的内容结构;另一方面也体现在对每一种教材的编写要求是以现有的、成熟的、社会学界普遍采用的体系、框架和知识点为依据。

三是本土化。尽管社会学教材中的许多内容都会涉及西方社会、西方社会学家以及西方社会学理论,但我们更加强调在教材的编写中要将社会学的基本原理应用于中国社会的现实,要以中国的社会、中国的材料来向中国的学生介绍社会学的基本概念、基本理论和研究领域。

笔者深知,要达到上述目标,需要全国社会学界同行的大力支持。我们真诚希望有更多的有经验的社会学教师、研究者加入这一工作中来,成为这套教材的编写者。

如果有更多的学生从这套教材中获得了对社会学知识的了解,打开了认识中国社会的窗口,那么我们的工作和努力就获得了回报。我们期待着!

风笑天

2007年11月于南京大学

## 修订版前言

据出版社反馈,我和几位青年教师编写的《社会统计学与 SPSS 应用》出版以来,已经印刷六次,得到广大读者充分肯定。不过,我们在使用过程中陆续发现书中还存在不少问题和差错,许多热心的读者也提出了一些中肯的意见和建议。作为主编,深感责任重大,有义务将全书仔细修改完善,不辜负广大读者的厚望。在出版社的大力支持下,此项工作得以顺利完成。修改后的新版在不改变原书总体框架的前提下,在如下方面做了较大改进。

第一,充实完善了部分内容。比如第十四章多元回归分析中,补充了对整个回归方程的检验。又如第十二章方差分析中,补充讲解了回归方程的方差分析:一是有利于更好地理解方差分析的原理及其应用,二是为第十四章多元回归分析中相关内容的进一步讲解奠定了良好的基础,增添了一些重要公式的推导及扩展性的说明。

第二,增加了许多有助于读者理解的注释。细心的读者会发现,新版中增加了许多文中注释和脚注,以帮助读者透彻地理解某些原理、公式和术语。同时,基本概念和术语都附上了英文原文。新版对一些容易弄混淆的概念如数学期望、方差、抽样分布、显著性水平、自由度等做了更加清晰的说明。

第三,纠正了不少模糊甚至错误的表述。比如原书第八章单总体假设检验中的例 8-1,相关概率的计算是不对的,新版已经纠正。又如第九章两总体假设检验中的独立样本的  $t$  检验公式有误,式中分别与两个样本方差  $S_1^2$ 、 $S_2^2$  相乘的系数应该是  $(n_1 - 1)$  和  $(n_2 - 1)$  而不是原来的  $n_1$  和  $n_2$ 。更改后的公式更加准确,并且与前面第七章第四节的小样本区间估计的公式保持一致。新版的语言也经过仔细润色,使之更加准确流畅。

第四,补充了一些有价值的参考文献。本书是社会统计学方面的入门性教材,旨在帮助读者在相关领域打下坚实的基础。书后的参考文献补充了一些富有启发的拓展性的中英文读物,以利于读者更好地领会社会统计学的精髓,开展更加深入的学习与研究,掌握更加复杂实用的统计方法与技巧。

第五,本书讲解 SPSS 时运用的、部分练习题目所提到的原始数据,都进行了统一编号,它们和练习题的详细解题过程、部分原始问卷及其他参考资料均可到华中科技大学社会学院网站的“人才培养\>\>本科生教育”中进行免费下载。

另外,许多年轻教师反映本课程讲授难度较大,希望我们能编写课件予以帮助。这次我们专门组织人员,制作了比较详细的课件,有需要的教师可向出版社申请。联系方式:qq3098247382,李永。

当然,尽管作出了很大努力,新版仍可能存在某些不足甚至错谬之处,欢迎读者批评指正。

张小山

2018年6月28日

# 目 录

## 第一编 导 论

第一章 社会统计学概述	(2)
第一节 社会统计学的基本含义	(2)
一、相关概念	(2)
二、社会统计学的含义	(3)
第二节 社会统计学简史	(4)
一、古典时期	(5)
二、近代时期	(6)
三、现代时期	(7)
第三节 社会统计学在社会研究中的地位与作用	(9)
一、社会研究的含义及体系	(9)
二、社会统计学的知识背景与哲学基础	(13)
三、社会统计学的作用	(15)
第二章 随机现象与基础概率	(18)
第一节 随机现象及其特征	(18)
第二节 概率的定义	(19)
一、随机事件	(19)
二、随机事件的概率	(20)
三、古典概率类型	(20)
第三节 概率的加法定理	(21)
一、特殊情况	(21)
二、一般情况	(21)
第四节 概率的乘法定理	(22)
一、特殊情况	(22)
二、一般情况	(23)
三、概率论在日常生活中运用的几个例子	(25)
第五节 概率与二项分布	(26)
一、随机变量的含义	(26)
二、随机变量的概率分布	(27)
三、二项分布	(30)
第三章 统计准备的准备与统计软件应用	(34)
第一节 资料处理	(34)
一、资料的审核、复查	(34)

二、问卷编码和变量设置 .....	(34)
第二节 SPSS 软件简介 .....	(36)
一、SPSS 的产生与发展 .....	(36)
二、SPSS 与其他常用统计软件比较 .....	(37)
三、SPSS 的安装、启动与界面 .....	(37)
第三节 SPSS 数据的基本操作 .....	(39)
一、数据的输入 .....	(39)
二、数据文件的调用和保存 .....	(41)
三、数据的清理 .....	(44)
四、数据文件的编辑 .....	(46)
五、数据文件的整理 .....	(47)
六、变量的变换和计算 .....	(50)

## 第二编 单变量统计

第四章 数据的组织与展示 .....	(54)
第一节 数据的特点与类型 .....	(54)
一、数据的概念与特征 .....	(54)
二、数据的类型 .....	(54)
第二节 定类数据的组织与展示 .....	(57)
一、定类数据的组织 .....	(57)
二、定类数据的图示 .....	(58)
第三节 定序数据的组织与展示 .....	(60)
一、定序数据的组织 .....	(60)
二、定序变量的图示:累积频数分布图 .....	(60)
第四节 定距数据的组织与展示 .....	(62)
一、数据分组 .....	(62)
二、分组的步骤 .....	(63)
三、定距数据的图示:直方图 .....	(64)
第五节 数据组织与展示的 SPSS 应用 .....	(65)
一、频数分布 .....	(65)
二、条形图和饼图 .....	(67)
三、累积频数分布图 .....	(68)
四、定距数据的分组 .....	(72)
五、直方图 .....	(77)
第五章 集中趋势与离散趋势测量 .....	(81)
第一节 集中趋势测量 .....	(81)
一、定类变量:众数 .....	(81)
二、定序变量:中位数 .....	(83)
三、定距变量:均值 .....	(85)



四、众数、中位数和均值的比较 .....	(87)
第二节 离散趋势测量 .....	(88)
一、定类变量:异众比率 .....	(88)
二、定序变量:四分位差 .....	(89)
三、定距变量:标准差与方差 .....	(91)
四、相对离散程度:离散系数 .....	(93)
第三节 单变量描述分析的 SPSS 应用 .....	(94)
一、单变量描述分析的基本过程 .....	(94)
二、单变量描述分析的实例分析 .....	(96)
第六章 正态分布及其他常用分布 .....	(100)
第一节 正态分布的含义及性质 .....	(100)
一、频数分布与正态曲线 .....	(100)
二、正态分布的数学表达式 .....	(101)
第二节 正态分布曲线下的面积 .....	(102)
第三节 标准正态分布 .....	(103)
一、标准值与标准正态分布 .....	(103)
二、标准值的实际意义 .....	(104)
第四节 标准正态分布表的使用 .....	(105)
第五节 其他常用的统计学分布 .....	(108)
一、 $t$ 分布 .....	(109)
二、 $\chi^2$ 分布 .....	(109)
三、 $F$ 分布 .....	(110)
第六节 运用 SPSS 检验正态分布 .....	(112)
一、直方图 .....	(112)
二、P-P 图 .....	(113)
第七章 参数估计 .....	(116)
第一节 抽样分布 .....	(116)
一、总体与样本 .....	(116)
二、概率抽样的几种类型 .....	(117)
三、统计量 .....	(117)
四、抽样分布 .....	(117)
五、样本均值的抽样分布 .....	(118)
第二节 参数的点估计 .....	(120)
一、总体均值与方差的点估计 .....	(120)
二、总体比例的点估计值 .....	(121)
三、点估计值的评价标准 .....	(121)
第三节 单总体均值与比例的区间估计 .....	(123)
一、单总体均值的区间估计 .....	(124)
二、大样本总体比例的区间估计 .....	(128)

第四节	两总体均值差异和比例差异的区间估计	(130)
一、	两总体均值差的区间估计	(130)
二、	两个总体比例差的区间估计	(133)
第五节	样本容量的确定	(134)
一、	估计总体均值时样本容量的确定	(135)
二、	估计总体比例时样本容量的确定	(135)
三、	估计两个总体均值差时样本容量的确定	(136)
四、	估计两个总体比例之差时样本容量的确定	(137)
第六节	单总体均值与比例区间估计的 SPSS 应用	(137)
一、	单均值的区间估计	(137)
二、	单总体比例的区间估计	(140)
第八章	单总体假设检验	(147)
第一节	假设检验概述	(147)
一、	统计推论与假设检验	(147)
二、	假设检验的基本思想	(147)
三、	假设检验的基本概念	(148)
第二节	单个总体均值和比例的假设检验	(154)
一、	单个总体均值的检验	(154)
二、	单个总体比例的检验	(158)
第三节	单均值和单比例假设检验的 SPSS 应用	(158)
<b>第三编 双变量统计</b>		
第九章	两总体假设检验	(164)
第一节	均值差异的假设检验	(164)
一、	大样本的 $Z$ 检验	(164)
二、	小样本的 $t$ 检验	(165)
第二节	比例差异的假设检验	(168)
第三节	均值差异比较与检验的 SPSS 应用	(169)
一、	Means 过程	(170)
二、	独立样本的 $t$ 检验	(174)
三、	配对样本的 $t$ 检验	(178)
第十章	交互分类与 $\chi^2$ 检验	(183)
第一节	交互分类和交互分类表	(183)
一、	交互分类的作用	(183)
二、	交互分类表的表现形式	(186)
三、	交互分类表的形式要求	(189)
第二节	$\chi^2$ 检验	(190)
一、	$\chi^2$ 检验的基本假定和原假设	(190)
二、	$\chi^2$ 检验的计算公式和检验步骤	(190)

三、 $\chi^2$ 检验的其他运用 .....	(193)
第三节 关系强度的测定 .....	(195)
一、 $\phi$ 系数和Q系数 .....	(196)
二、V 系数 .....	(197)
三、C 系数(列联系数) .....	(197)
第四节 交互分类与 $\chi^2$ 检验的 SPSS 应用 .....	(198)
一、进行交互分类的基本过程 .....	(198)
二、实例分析 .....	(201)
第十一章 相关分析 .....	(209)
第一节 相关的基本性质 .....	(209)
一、相关关系的程度 .....	(209)
二、相关关系的方向 .....	(211)
三、相关关系的类型 .....	(211)
四、相关的对称性 .....	(211)
五、消减误差比例的意义 .....	(213)
第二节 相关测量法 .....	(214)
一、两个定类变量: $\lambda, \tau_y$ .....	(214)
二、两个定序变量:Gamma, $d_y$ .....	(217)
三、两个定距变量:简单线性回归( $b$ )和积矩相关( $r$ ) .....	(223)
四、定类变量与定序变量: $\lambda, \tau_y$ .....	(227)
五、定类变量与定距变量:相关比率 $E^2$ .....	(228)
六、定序变量与定距变量:相关比率 $E^2$ .....	(229)
七、小结 .....	(230)
第三节 相关系数的假设检验 .....	(231)
一、两定类变量: $\chi^2$ 检验 .....	(231)
二、两定序变量:Z 检验、 $t$ 检验 .....	(232)
三、两定距变量:F 检验 .....	(234)
四、定类/定距变量:F 检验(或单因素方差分析) .....	(235)
五、总结:相关测量法与检验法 .....	(237)
第四节 相关测量和检验的 SPSS 应用 .....	(238)
一、Crosstabs 中的相关分析 .....	(238)
二、Correlate 中的相关分析 .....	(242)
第十二章 方差分析 .....	(249)
第一节 方差分析的概念与基本原理 .....	(249)
一、什么是方差分析 .....	(249)
二、方差分析的基本思想 .....	(250)
三、方差分析的基本假定 .....	(250)
第二节 单因素方差分析 .....	(251)
第三节 多因素方差分析简介 .....	(255)

一、基本思路 .....	(255)
二、注意事项 .....	(256)
第四节 方差分析的 SPSS 应用 .....	(257)
一、单因素方差分析的基本过程 .....	(257)
二、单因素方差分析的实例分析 .....	(258)
三、单因素方差分析的多重比较检验 .....	(261)
<b>第四编 多变量统计</b>	
<b>第十三章 详析分析与偏相关</b> .....	(266)
第一节 详析分析 .....	(266)
一、因果分析 .....	(266)
二、阐明分析 .....	(268)
三、条件分析 .....	(269)
第二节 偏相关分析 .....	(270)
一、偏相关系数 .....	(270)
二、偏相关系数的检验 .....	(273)
第三节 偏相关分析的 SPSS 运用 .....	(274)
一、偏相关的操作过程 .....	(274)
二、案例分析 .....	(275)
<b>第十四章 多元回归分析</b> .....	(279)
第一节 一元线性回归 .....	(280)
一、一元线性回归方程 .....	(280)
二、标准回归方程 .....	(280)
三、一元线性回归方程的检验 .....	(281)
第二节 复相关 .....	(283)
一、复相关与决定系数 .....	(283)
二、复相关系数的检验 .....	(284)
第三节 多元线性回归 .....	(285)
一、多元线性回归方程 .....	(285)
二、多重共线性及其诊断参数 .....	(288)
三、回归方程的检验 .....	(289)
四、逐步回归分析 .....	(291)
五、虚拟变量 .....	(292)
第四节 对数回归分析 .....	(294)
一、从多元线性回归到对数回归 .....	(295)
二、对数回归(logistic regression) .....	(295)
三、多项对数回归(multinomial logistic regression) .....	(298)
第五节 线性回归的 SPSS 应用 .....	(299)
一、线性回归的分析过程 .....	(300)

---

二、奇异值、影响点的确定 .....	(305)
三、实例分析 .....	(306)
第六节 对数回归分析的 SPSS 应用 .....	(312)
一、binary logistic 的分析过程 .....	(312)
二、实例分析 .....	(315)
附录 .....	(321)
附录 A 随机数字表 .....	(321)
附录 B 标准正态分布表 .....	(322)
附录 C $Z$ 检验:常用的显著度( $\alpha$ )与对应的临界值( $ Z_0 $ ) .....	(324)
附录 D $r$ 值化为 $Z$ 值 .....	(325)
附录 E $t$ 分布表 .....	(326)
附录 F $\chi^2$ 分布表 .....	(328)
附录 G $F$ 分布表 .....	(330)
部分练习题答案 .....	(333)
参考文献 .....	(337)
后记 .....	(340)

---

# 第一编 导 论

---

# 第一章

## 社会统计学概述

社会统计学在社会研究与日常生活中具有广泛的用途,掌握社会统计学的基本原理与方法成为现代人应该具备的基本素质之一。本章主要介绍社会统计学的相关概念、发展简史、学科背景和重要作用。

### 第一节 社会统计学的基本含义

#### 一、相关概念

要理解社会统计学的基本含义,必须首先弄清与统计相关的几个基本概念。

##### (一) 统计

统计(statistics)一词在日常生活中使用的频率相当高,它与人们的关系也非常密切。许多人都会关注统计部门公布的各类经济和社会统计数据,那些热衷于投资理财的“股民”和“基民”就更是如此,他们甚至时刻焦急地期盼反映国家经济运行状况的动态数据的公布。一般来说,统计一词具有三层含义,它可以泛指统计工作、统计资料和统计学。第一,统计工作是对各种现象在数量方面进行收集、整理、分析的工作过程,如各级统计部门所做的大量事情。第二,统计资料是反映各种现象的数字(或数据)资料(包含各种与数字有关的图形与表格),是统计工作所取得的成果,如公布出来的各种统计数据:生育率、性别比、人均收入、物价指数等。第三,统计学是用来研究和评价数据资料的一种科学程序,是研究收集、整理和分析数据资料的理论和方法的科学。本书主要在第三种意义上使用统计一词。统计学与统计工作之间是理论和实践的关系。注意,在统计学中,英文“statistics”还有一个含义,即指样本的统计量,它可以反映样本的特征和属性,与该词对应的是反映总体特征与属性的“参数”(parameters)。

##### (二) 统计学

统计学(statistics)是一门系统地收集、整理、计算、分析和解释数据资料的科学,它通常被看成应用数学的一个分支。英文“statistics”源于拉丁词“status”,该词在中世纪拉丁语中有“国家”的意义。最初,统计学是用文字来描述一个国家的制度和情况,但到19世纪,该词逐渐形成狭隘的意义:“用数字的方法说明国家的特征”。后来,虽然它仍旧被用来指称描述

这些特征的实际数据,却已扩大到其他科学方面的数据;最后该词被用来指称基于对这类数据的分析并从中引出结论的方法。统计学在当今社会具有非常广泛的用途,工业、农业、医疗、卫生、教育、金融、保险、气象、军事等领域都离不开统计学和统计方法。事实上,将统计学的原理和方法创造性地运用于各个现实领域(或学科领域),就会产生各个分支统计学:工业统计学、经济统计学、医疗统计学、教育统计学、军事统计学、体育统计学等。简言之,哪里存在大量的数据,哪里就有统计学的用武之地。我们现在生活在数据的海洋之中,如何理解这些大量的数据,发现其中的模式与意义并做出合理的判断与决策,成为现代人必须具备的基本素质与能力。

一般来说,统计学可划分为描述性统计(descriptive statistics)和推论性统计(inferential statistics)两大类。①描述性统计旨在用最简单的概括形式反映出大量数据资料所容纳的基本信息,主要包括图示法、集中趋势及离散程度分析等。它是统计学中比较简单的部分,可为推论性统计学打基础。②推论性统计(推断性统计)旨在用从样本中所得到的数据信息来推断总体的基本情况,即由样本的统计值来猜测、评判总体的参数值,主要包括参数估计和假设检验等。推论性统计是统计学的核心,被一些学者视为真正的统计学。如《简明不列颠百科全书》给出的统计学的定义便是“根据从总体中随机取出的样本里所获得的信息来推断关于总体的性质的一门学科”<sup>①</sup>。

### (三) 数理统计

数理统计是数理统计学(mathematical statistics)的简称,就是关于统计的数学理论。它从量的方面分析事物之间的相互联系和相互作用,试图通过对事物量的规定性的分析来把握事物质的规定性,使人们从数据中(尤其是大量数据中)提取有意义的信息。统计运用的是数字语言,它要研究客观事物的数量,主要包括数量状态、数量关系和数量变化规律等。而且,统计学主要探讨的是总体数量而不是个体数量,即要综合反映大量同类现象的数量方面。单个现象的发生或许是不确定的,但大量同类现象的发生却可能存在某种确定的规律性,这就是著名的“大数法则”或“大数定理”。不妨说,数理统计是一切其他应用性统计学的基础,这些应用于不同领域的分支性统计学主要借助的是数理统计的基本原理与技术。数理统计的基础是概率论(probability theory),即研究随机现象中数量规律的数学理论。事实上,没有概率论,就没有推论性统计学,因为后者是以概率抽样及概率分布为基础的。虽然与几何、算术等不同,概率论是一门相对较新的数学分支,但是现在它恐怕是所有数学学科中应用最广的一门分支,已经极其有效地渗透到我们的文化之中<sup>②</sup>,从天气预报到人寿保险,从游戏到投资,从选举预测到粮食估产……都离不开概率与统计的思想与方法。

## 二、社会统计学的含义

社会统计学(social statistics)是一门系统地收集、整理、计算、分析和解释有关社会现象的数据资料的科学,旨在对社会现象内部的各种联系和关系进行定量分析和统计处理。广义的社会统计学,可以简单地理解为将统计学的基本原理和方法创造性地运用于社会领域。

① 参见中美联合编审委员会的《简明不列颠百科全书》(第7卷),中国大百科全书出版社1986年版,第831页。

② 参见[美]约翰·塔巴克著,杨静译的《概率论和统计学——不确定性的科学》,商务印书馆2007年版,第4页。



社会统计学将所要研究的社会现象看成是一些具有不同取值的变量(variables),依据收集到的数据资料描述和分析这些变量的基本性质,以及特定的变量受其他相关变量的影响情况。变量的具体表现称为变量值(variant-value),也就是那些反映特定变量特征的不同取值,而这些变量值在统计上又称为数据(data)。例如,研究大学生的服饰消费情况,其月消费额就是一个可以有不同取值的变量,可以考察该变量与其他变量(如性别、年级、家庭背景、学科类别等)之间的关系。统计分析的结果可能揭示出:女大学生比男大学生消费高,高年级大学生比低年级大学生消费高,文科生比理科生消费高。注意,统计分析针对的是变量,如果是恒定不变的常数,则无法进行统计分析。在上例中,如果月消费额是一个不受性别、年龄等因素影响的常数,意味着该数值与其他变量根本没有关系,我们也就没有必要展开进一步的统计分析。再如,收集到的调查资料是女大学生的服饰消费情况,这时性别是固定的,即为常数,当然也就不能分析不同性别对大学生服饰消费的影响;同样,如果资料是关于一年级大学生的服饰消费状况的,则不能分析不同年级对大学生服饰消费的影响。

社会统计学和一般数理统计学的最主要的区别是:数理统计所处理的变量主要是数学属性最高的定比层次的变量,而社会统计学则涉及大量的数学属性较低的定类和定序层次的变量<sup>①</sup>,这是因为很多社会现象都难以进行精确的、数字化的测量,其数学属性比较有限。另外,社会统计学的研究对象主要是异质性很大的复杂的人,所以需要采用较大规模的抽样调查,才能保证统计推论比较有说服力。社会统计学为社会科学进行定量分析提供了一个重要的手段和方法。社会学是较早地借用统计分析来研究社会事实的一门社会科学,由于它自身的特性和创始人的倡导,社会学在其一百多年的发展历程中,逐步形成了一套比较系统的、科学的研究社会事实的方法。本书主要介绍的是在社会学研究中最基本、最常用的一些统计技术,当然,这些技术也能广泛地运用于其他社会科学领域。需要指出的是,社会统计学与社会调查(social survey)是密切相关的,都为项完整的社会研究的组成部分。社会统计学重点解决的问题是如何处理和分析所收集到的数据资料,而资料的收集工作则主要由社会调查来完成(相应的课程是社会调查方法)。不过,在社会调查阶段(如问卷设计的过程中),必须事先考虑资料将如何进行统计分析,这样收集的资料才更有价值、更具针对性。另外,需要指出的是,社会统计学也包含描述性统计和推论性统计两部分;而按所涉及的变量数目,它还可以分为单变量统计、双变量统计和多变量统计(多元统计)。本书除导论外,后面三大编就是依据第二种分类来安排的,而各编内部则主要按描述性统计和推论性统计这两类进行组织。

## 第二节 社会统计学简史

统计学可视为人们长期统计实践的理论概括与总结。人类的统计实践可谓源远流长,原始社会从“结绳记事”开始,就有了对自然社会现象的简单计数活动,从而产生了统计的萌芽。“statistics”源于拉丁词“status”,该词自中世纪起逐渐演变成含政治意味的state(国

<sup>①</sup> 关于变量的测量层次,参见本书第四章。