

 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

◎ 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



统计学与SPSS应用

● 张小山 主编



华中科技大学出版社
<http://www.hustp.com>

中国·武汉

图书在版编目(CIP)数据

社会统计学与 SPSS 应用/张小山 主编. —武汉: 华中科技大学出版社, 2010. 8
ISBN 978-7-5609-6225-2

I. 社… II. 张… III. 社会统计-统计分析-软件包, SPSS IV. C91-03

中国版本图书馆 CIP 数据核字(2010)第 085923 号

社会统计学与 SPSS 应用

张小山 主编

责任编辑: 朱建丽

封面设计: 刘 卉

责任校对: 刘 竣

责任监印: 周治超

出版发行: 华中科技大学出版社(中国·武汉)

武昌喻家山 邮编: 430074 电话: (027)87557437

录 排: 华中科技大学惠友文印中心

印 刷: 湖北鑫浪潮印务有限公司

开 本: 710mm×1000mm 1/16

印 张: 24.75

字 数: 493 千字

版 次: 2010 年 8 月第 1 版第 1 次印刷

定 价: 35.00 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

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,通过系统合理地编排和深入浅出地讲解,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。

总 序

社会学在中国内地恢复重建已将近30年了。伴随着中国社会的改革开放,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“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。即希望经过 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、研究者加入到这一工作中来,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

目 录

第一编 导 论

第一章 社会统计学.....	(2)
第一节 社会统计学的基本含义.....	(2)
第二节 社会统计学简史.....	(5)
第三节 社会统计学在社会研究中的地位与作用	(10)
第二章 随机现象与基础概率	(21)
第一节 随机现象及其特征	(21)
第二节 概率的定义	(22)
第三节 概率的加法定理	(25)
第四节 概率的乘法定理	(26)
第五节 概率与二项分布	(31)
第三章 统计准备的准备与统计软件应用	(39)
第一节 资料处理	(39)
第二节 SPSS 软件简介	(42)
第三节 SPSS 数据的基本操作	(45)

第二编 单变量统计

第四章 数据的组织与展示	(62)
第一节 数据的特点与类型	(62)
第二节 定类数据的组织与展示	(65)
第三节 定序数据的组织与展示	(69)
第四节 定距数据的组织与展示	(71)
第五节 数据组织与展示的 SPSS 应用	(75)
第五章 集中趋势与离散趋势测量	(93)
第一节 集中趋势测量	(93)
第二节 离散趋势测量.....	(101)
第三节 单变量描述分析的 SPSS 应用	(108)
第六章 正态分布及其他常用分布.....	(114)
第一节 正态分布的含义及性质.....	(114)

第二节	正态分布曲线下的面积·····	(116)
第三节	标准正态分布·····	(117)
第四节	标准正态分布表的使用 f ·····	(120)
第五节	其他常用的统计学分布·····	(124)
第六节	运用 SPSS 检验正态分布·····	(128)
第七章	参数估计 ·····	(132)
第一节	抽样分布·····	(132)
第二节	参数的点估计·····	(137)
第三节	单总体均值与比例的区间估计·····	(141)
第四节	两总体均值差异和比例差异的区间估计·····	(148)
第五节	样本容量的确定·····	(153)
第六节	单总体均值与比例区间估计的 SPSS 应用·····	(156)
第八章	单总体假设检验 ·····	(167)
第一节	假设检验概述·····	(167)
第二节	单个总体均值和比例的假设检验·····	(175)
第三节	单均值和单比例假设检验的 SPSS 应用·····	(180)

第三编 双变量统计

第九章	两总体假设检验 ·····	(186)
第一节	均值差异的假设检验·····	(186)
第二节	比例差异的假设检验·····	(191)
第三节	均值差异比较与检验的 SPSS 应用·····	(192)
第十章	交互分类与χ^2检验 ·····	(208)
第一节	交互分类和交互分类表·····	(208)
第二节	χ^2 检验·····	(216)
第三节	关系强度的测定·····	(222)
第四节	交互分类与 χ^2 检验的 SPSS 应用·····	(225)
第十一章	相关分析 ·····	(237)
第一节	相关的基本性质·····	(237)
第二节	相关测量法·····	(242)
第三节	相关系数的假设检验·····	(262)
第四节	相关测量和检验的 SPSS 应用·····	(270)
第十二章	方差分析 ·····	(283)
第一节	方差分析的概念与基本原理·····	(283)

第二节	单因素方差分析	(285)
第三节	多因素方差分析简介	(290)
第四节	方差分析的 SPSS 应用	(292)

第四编 多变量统计

第十三章	详析分析与偏相关	(304)
第一节	详析分析	(304)
第二节	偏相关分析	(308)
第三节	偏相关分析的 SPSS 运用	(313)
第十四章	多元回归分析	(319)
第一节	一元线性回归	(320)
第二节	复相关	(323)
第三节	多元线性回归	(326)
第四节	对数回归分析	(336)
第五节	线性回归的 SPSS 应用	(342)
第六节	对数回归分析的 SPSS 应用	(356)
附录		(368)
附录 A	随机数字表	(368)
附录 B	标准正态分布表	(369)
附录 C	Z 检验:常用的显著度(α)与对应的临界值($ Z_0 $)	(370)
附录 D	r 值化为 Z 值	(371)
附录 E	t 分布表	(371)
附录 F	χ^2 分布表	(373)
附录 G	F 分布表	(374)
部分练习题答案		(378)
参考文献		(382)
后记		(384)

第一编 导 论

第一章

社会统计学

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、几个相关概念

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种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资料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”^①。

(三) 数理统计

所谓数理统计学(mathematical statistics),简称数理统计,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

^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。

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^①,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。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^②,这是因为很多社会现象都难以进行精确的、数字化的

^① 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。

^② 关于变量的测量层次,参见本书第四章。

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地、科学地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),必须考虑资料将如何进行统计分析 with 处理,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(国家、状态)。因此,统计学原先包含的意义主要是指对国家的状态做调查研究。实际上,古代的中国和埃及都有过对国家的大事做统计调查的历史,如中国从公元前 21 世纪的夏朝开始就有户口、土地等方面的记载。而在古希腊、罗马时代,随着社会机构日益复杂化,对于从政者来说,掌握国家全面情况的统计知识就变得越来越有用了。亚里士多德所著《国家论》一书中,对很多国家的政治、宗教、艺术和风俗等进行了详细的记述,该书可以认为是后来发展起来的所谓“国势学”的先驱之作。^①当然,统计的广泛运用和统计学的真正建立则肇始于 17 世纪的欧洲,直接得益于资本主义社会经济、文化和科技的迅速发展,以

^① 参见柯惠新和沈浩编著的《调查研究中的统计分析法》(第 2 版),中国传媒大学出版社 2005 年版,第 1 页。

及社会分工的日益发达。下面将统计学的建立和发展分 3 个时期进行简要介绍。

一、古典时期

这里所谓的古典统计学时期,主要是指 17 世纪中叶到 18 世纪末这一时期,它也是统计学的初创阶段。“统计方法引进社会现象的研究是 17 世纪中一件头等重要的大事,并在以后几个世纪中证明是卓有成效的。”^①这一时期的统计学又分为德国的“国势学”和英国的“政治算术”两大学派,它们也是统计学诞生的两个直接源头,并对以后统计调查的发展产生了巨大的影响。

(一) 国势学派

国势学派又称为记述学派,它强调以文字来记述国家显著事项。德国的康令(H. Conring, 1606—1681)被视为国势学派的创立者,他最先在德国的大学开设国势学课程。其后,18 世纪的阿亨瓦尔(G. Achenwall, 1719—1772)继承了国势学的基本主张,将统计学(德文 statistik)定义为“把国家的显著事项全部记述下来的学科”,并对该学科的性质、意义及范围给出了明确的界定。后人都称阿亨瓦尔为“统计学之父”。他的主要著作作为《近代欧洲各国国势学概论》。国势学派在研究各国的显著事项时,主要运用对比分析的方法研究国家组织、人口、军队、领土、财产等,以比较各国实力的强弱。不过,需要指出的是,康令和阿亨瓦尔都很少做数量方面的观察,他们的研究偏重对事物性质的解释,主要用文字进行表述,缺乏数字内容,因而在很大程度上没有触及统计资料的实质。到了 19 世纪,德国学者克尼斯(K. Knies, 1821—1898)把当时已在英国发展起来的政治算术称为统计学,认为它是收集、整理和展示资料的科学,即是一种方法论的科学,而把以康令和阿亨瓦尔为代表的统计学称为国势学。

(二) 政治算术学派

政治算术学派产生于 17 世纪的英国,主要代表人物为约翰·格朗特(J. Graunt, 1620—1674)和威廉·配弟(W. Petty, 1623—1687)。格朗特被誉为现代统计学的开创者,政治算术学派的鼻祖。他是伦敦的一位商人,于 1662 年著书《关于伦敦死亡表的观察》,这也是统计学的第一份经典文献。他利用寺院提供的死亡和洗礼的资料,首先制作了死亡表,并指出某些疾病的死亡人数占全部死亡人数的比例是稳定的。不过,政治算术学派的真正代表人物一般认为是英国的经济学家配弟。配弟是格朗特的朋友,他继承和发展了格朗特的研究工作。根据人口、土

^① 参见[英]沃尔夫的《十六、十七世纪科学、技术和哲学史》,商务印书馆 1985 年版,第 657 页。

地、财政、经济等方面的大量观察,配弟完成了《政治算术》一书。该书以数字、重量和尺度为基础,用计算和对比的方法,配以简明的图表,比较了英、法、荷三国的经济、军事、政治等方面的实力,这些数字资料具有很强的说服力,并已成为现代统计学广为采用的方法和内容。^① 属政治算术学派的天文学家哈雷(1656—1742)在配弟对全球人口数目估算研究工作的基础上,进一步做了更合理、更精确的人口估算,这是在人口统计方面的极大贡献。特别是哈雷通过对死亡率的研究制作了死亡表,并根据该死亡表对人寿保险年金进行了精确的计算,解救了当时荷兰的财政危机,这是关于生命保险理论最早的科学研究^②——人寿保险事业从此诞生。

英国政治算术学派的影响传播到了整个欧洲大陆,此后涌现了一批(包括德国的)学者,特别是在人口统计方面提出了一系列以大量观测数据为基础的研究方法。

二、近代时期

近代统计学时期主要是指18世纪末到19世纪末这100多年的时间,它也是统计学得到迅速发展的一个时期,涌现出一批非常著名的研究者,诞生了许多从事统计工作的相关机构。1795年,拉普拉斯(F. Laplace, 1749—1827)提出用概率论研究社会人口问题,出版了《概率论的解析理论》,从而奠定了现代人口统计理论的基础。概率论主要探讨不确定现象出现的机会大小,它的产生有点戏剧性,很大程度上是源于赌博的需要。欧洲中世纪的一些赌徒为了在赌博中稳操胜券,便向一些数学家求教,而后者逐渐发展出一门应用价值极高的数学理论。真正意义上的概率论始于17世纪。在早期概率论的研究中,除了上面提到的被誉为古典概率研究的集大成者拉普拉斯以外,做出过重要贡献的数学家还有帕斯卡(B. Pascal)、费马(P. Fermat)、莱布尼兹(G. Leibniz)、伯努利(J. Bernoulli)、德·莫佛尔(A. de Moivre)、贝叶斯(T. Bayes)、高斯(C. Gauss)、勒让德(A. Legendre)、贝塞尔(F. Bessel)、德·康多塞(M. de Condorcet)、辛普森(T. Simpson)、布丰(C. de Buffon)、泊松(S. Poisson)等。“政治算术”研究的是简单的确定性数量关系,而概率统计则是研究复杂的、随机性现象,这极大地充实和深化了数量问题研究的内容,并将统计学推向了一个新的发展时期。

比利时人凯特勒(A. Quetelet, 1796—1874)是统计学发展史上承前启后的重要人物,也是社会统计学早期的开拓者,被誉为“经验社会学之父”。他一生撰写了

^① 参见张晓庆、王玉良和王景海编著的《统计学》,科学出版社2007年版,第3页。

^② 参见柯惠新和沈浩编著的《调查研究中的统计分析法》(第2版),中国传媒大学出版社2005年版,第2页。

大量的有关统计学方面的著作,代表作有《社会物理学》(1869)^①、《关于应用于道德科学、政治科学的概率论书简》(1846)、《犯罪学》(1831)、《比利时统计研究》(1829)等。凯特勒是将概率论思想引入社会研究的第一人,极大地推动了数学方法在社会科学领域的应用,许多现代社会统计手段都来源于他的思想。凯特勒坚信社会生活具有一定的统计规律,并提出了著名的“平均人”(average man)概念,塑造出一个具有平均身高、平均体重、平均智力和道德品质的典型人物,他主张统计的任务就是关于平均人的比较研究。凯特勒还是一位出色的社会活动家,并且是国际统计学会的创始人之一。

事实上,近代统计学是为满足多门学科(如人口学、天文学、心理学、医学、农学、遗传学和人体测量学等)的需要而发展起来的,而统计方法的运用也确实在这些领域里取得了创造性的成果。一大批著名学者为此做出了重要的贡献,其中包括遗传学家孟德尔(J. Mendel, 1822—1884)、优生学的创始人高尔顿(F. Galton, 1822—1911)等。以高尔顿为例,他是生物统计学的主创者,受凯特勒的影响,他也利用正态法则研究优先学、遗传学等方面的问题,先后提出了百分位数、中位数、四分位差、相关与回归等概念和计算方法。另外,高尔顿的学生皮尔逊(K. Pearson, 1857—1936)系统地发展了老师的相关与回归理论,研究了复相关和偏相关,把物理学上“矩”的概念移植到统计学中,给出了极大似然(或拟然 likelihood)估计方法,导出了重要的 χ^2 分布。以皮尔逊为代表的统计学家所进行的,以大量观察和正态分布为基础的关于总体分布曲线的研究,确立了“大样本”统计理论,奠定了“描述统计学”的框架体系。^②

需要指出的是,随着统计方法的不断改进和完善,统计工作得到有效推广,深入与全面的人口普查得以开展,相关的统计机构也逐步建立起来,这些又推动着统计学的进一步发展。美国于1790年就有了人口统计局(US Census of Population)。从1791年开始,辛克勒陆续出版了21卷巨著《苏格兰统计论述》(1791—1825)。此书问世,在人口普查内容和方法上对欧洲产生了巨大影响。从1801年起,英国就开始进行经常性的全国人口普查。18世纪30年代,英、法等一些国家率先建立了统计学会,成为当时探讨社会问题和发展统计技术的中心。英国于1837年成立了全国统计总局(general register office),其后又制定出规范的统计报告,涉及诸如生育、死亡和犯罪等问题。在1853年第一次国际统计会议后,人口普查更加完善,许多国家纷纷设立统计局,从而把社会调查纳入了职业化的道路,极大地促进了社会调查研究方法和专业社会统计学的形成和发展。^③不同于那些将

① 该书初版于1835年,原名为《论人及其才能的发展》。

② 参见耿修林、谢兆茹编著的《应用统计学》,科学出版社2002年版,第3页。

③ 参见苏拓主编的《社会调查原理与方法》,湖北科学技术出版社1989年版,第31~32页。