新编社会学系列教材

社会统计学讲义

郑卫东 编

Social Statistics

本书充分吸取了社会统计理论与技术发展的最新成果，难度适中，既考虑了学生的学习理解能力，又努力满足社会学等专业的本科生在学习及科研中对社会统计知识应用的需求，同时在数理推导与实践应用之间寻求一种平衡。本书强调社会统计知识的实践运用，对的数学推导过程也有介绍，以单变量分析与双变量分析为重点内容，同时对常用的多因素方差分析、多元线性回归分析与二项回归分析亦有涉及。

应社会学、社会工作、社会保障等专业本科生教学与学习的需求，又可供其他相关专业师生和普通读者学习参考之用。

北京大学出版社
PEKING UNIVERSITY PRESS

社会统计学讲义

郑卫东 编

Social Statistics

北京大学出版社
PEKING UNIVERSITY PRESS

图书在版编目(CIP)数据

社会统计学讲义/郑卫东编. —北京:北京大学出版社,2013.8
(新编社会学系列教材)
ISBN 978-7-301-22137-2

Ⅰ.①社… Ⅱ.①郑… Ⅲ.①社会统计-高等学校-教材 Ⅳ.①C91-03

中国版本图书馆 CIP 数据核字(2013)第 025012 号

书　　名:社会统计学讲义
著作责任者:郑卫东　编
责 任 编 辑:陈相宜
标 准 书 号:ISBN 978-7-301-22137-2/C·0874
出 版 发 行:北京大学出版社
地　　址:北京市海淀区成府路 205 号　100871
网　　址:http://www.pup.cn　新浪官方微博:@北京大学出版社
电 子 信 箱:ss@pup.pku.edu.cn
电　　话:邮购部 62752015　发行部 62750672　编辑部 62753121
　　出版部 62754962
印 刷 者:北京飞达印刷有限责任公司
经 销 者:新华书店
　　730 毫米×980 毫米　16 开本　18 印张　294 千字
　　2013 年 8 月第 1 版　2013 年 8 月第 1 次印刷
定　　价:36.00 元

序

笔者从2006年开始给社会学专业本科生讲授“社会统计学”与“社会统计软件应用”课程，至今已经有六个年头。其间使用的教材与参考书主要包括：香港中文大学李沛良教授著《社会研究的统计应用》、杰克·莱文与詹姆斯·艾伦·福克斯合著《社会研究中的基础统计学》、北京大学卢淑华教授编著《社会统计学》、中国人民大学薛薇教授编著《SPSS统计分析方法及应用》、北京大学郭志刚教授主编《社会统计分析方法——SPSS软件应用》及上海财经大学张彦教授主编的《社会统计学》等。这些都是很有影响的教材，在各高校使用率较高，笔者和学生在使用这些教材的过程中同样获益匪浅。每一本教材都比较完整而系统地介绍了社会统计学的内容体系，注意理论与实践相结合，深入浅出、循循善诱地引导读者学习。同时，它们又各有特点：有些教材偏重介绍统计知识的操作应用，并不特别关注统计理论的数学推导过程，这对于那些社会统计学的初学者特别有吸引力；有些教材在统计知识的数学推理方面介绍得非常详细，对那些喜欢深入研究社会统计数理基础的同学，帮助很大，而且有助于他们应对研究生入学考试；有些教材在内容体系的设置上偏重介绍单变量分析与双变量分析，而对于多变量分析（诸如多因素方差分析、多元线性回归分析、Logistic回归分析等）或者一笔带过或者基本没有介绍，这样的内容体系已经不能满足当下本科生的学习要求；有的教材则把编写重点放在多因素方差分析、多元线性回归分析、Logistic回归分析、因子分析、聚类分析等，这样的内容体系对于本科生来讲已经偏难，超过了他们的理解与接受能力；等等。

在充分体验不同教材的特点与优势的同时，笔者也深切感受到几乎每一本教材的优势同时又可能是它的劣势。这主要是因为授课对象数学基础参差不齐，以及他们对这门课程的需求各异所致，所谓众口难调。而社会统计学课程的教学目的应该是在传授基本教学内容体系的基础上，尽量考虑并满足学生的多元化需求。可是，我们深深体会到，上面所介绍的教材中的任何一本都满足不了这种要求。因此，在教学实践中，笔者不得不大量参阅各种社会统计学教材，取长补短，以编写适应当下本科生特点的教案。这样

做无形中降低了选用教材的地位，增加了学生的学习难度。学生要理解教案的内容，单纯看选用教材是不够的，他们必须及时查阅教材与多本教学参考书，才能找到教案中的全部内容。实际上，很多同学很难做到这一点，他们往往课堂上听听，理解个大概，课后却懒得查阅参考书，不求甚解，以至于问题越积越多，最后失去了学习社会统计学的兴趣。所以，笔者认为非常有必要根据学生的特点和教学大纲的要求，编写一本能够博取众家之长、避开众家之短的《社会统计学》教材，以方便教学与学习。这本新编教材应该内容全面，充分吸取社会统计理论与技术发展的最新成果；同时，难度适中，既要充分考虑学生的学习理解能力，还要满足社会学等专业的本科生在学习及科学研究中对社会统计知识的需求；而且要在数理推导与实践应用之间做好平衡。

为了实现上述目标，笔者基于六年来的教学体会以及学生的反馈，对上述教材及教学参考书的内容进行了提炼、汇编与整合，同时还参考了林彬教授的“社会科学方法论”课堂讲义、郭志刚教授的《社会统计学》课件、周灵飞教授的《统计学原理》课件等，当然还有更多人的教材、论文与课件等，在此无法一一说明。编写教材的指导思想是注重社会统计知识的实践运用，不特别专注统计知识的繁复的数学推算，但对于基本的数学推导过程也作介绍；在内容体系方面，单变量分析与双变量分析是重点，同时对科学研究中常用的多因素方差分析、多元线性回归分析与二项 Logistic 回归分析等也作介绍，以提高学生分析解决比较复杂的现实问题的能力。希望这本教材能够做到体系完整、内容全面、难度适中，适应社会学、社会工作、社会保障等专业本科生的教学与学习要求。

既然是编写的教材，本书中的绝大部分内容均来自成说，笔者把参考文献目录附于书后，再次向这些前辈、专家致以诚挚敬意。本书的出版得益于华东政法大学社会发展学院的教材出版计划，也得益于北京大学出版社编辑老师们的辛勤奉献，在此向他们致以诚挚谢意！

郑卫东

2012 年 9 月 15 日

目 录

第一篇 导论

第二篇 描述统计

第三篇 从描述统计到推断统计

第四篇 相关与回归分析

第一篇　导论

第一章　社会研究的统计应用

第一节　社会统计学的产生与发展

一、社会统计应用的方法论基础

美国芝加哥大学华裔社会学教授赵鼎新曾经总结，社会科学研究有两个传统：一曰解释传统，一曰解读传统。解读传统的目的不在于寻找事物内在的逻辑关系，而在于理解和厘清特定人类活动在特定文化条件下的内在含义或意义；而解释传统的目的则是寻找具体事物或事件的内在机制以及与之相应的因果、辩证、对话型(dialogical)或历史性关系。

德国社会哲学家 W. 狄尔泰是解读传统的先驱，他认为社会科学的性质及研究对象与自然科学有着本质的不同，对社会和人类行为的研究离不开价值判断，社会科学的任务不仅是客观描述社会，还必须涉及伦理、宗教、艺术等价值观念。此外，人具有自由意志，人类行为既没有规律性也无法预测。社会历史现象都是独特的、偶然的，对此无法采用自然科学的“规范”方法加以研究，而只能用人文科学的方法加以阐述和记录。与解读传统不同，解释传统的先驱 A. 孔德主张，社会学应当是一门类似于自然科学的、以研究社会发展规律为目的的学科，这门学科应当采用建立在观察基础之上的实证主义方法。法国社会学家迪尔凯姆发展了孔德的实证主义，他把社会现象界定为普遍存在于群体间的、由外界的强制力施加于个人所引起的社会行为、社会思想和社会感受，是一种集体的行为和观念，是可以观察的社会事实。西方社会学家大多在某种程度上坚持由孔德和迪尔凯姆所开创的实证主义方向，把自然科学方法论作为自己的基本原则，把寻找社会现象的内在规律视为自己的研究使命，把以数学为基础的统计技术作为社会研究的基本工具。在他们看来，社会科学研究在理论构建、证据收集、证据分析与评判、理论检验等方面所运用的方法，与自然科学方法并无本质区别。也就是说，统计技术在社会科学研究的解释传统中得到了充分的运用和发展。

随着社会学学科的发展，社会统计技术在社会研究中的应用不断向广

度与深度扩展。社会研究是指以经验的方式，对社会世界中人们的行为、态度、关系，以及由此形成的各种社会现象、社会产物所进行的科学的探究活动。社会学研究则是运用科学的方法来收集和分析社会事实，以理解社会现象之间的关系，尤其是两个社会现象之间的因果关系。社会统计学就是运用统计的一般原理，对社会各种静态结构与动态趋势进行定量描述或推断的一种专门方法与技术。

二、社会统计学的产生与发展

统计学是一门收集、整理和分析数据的方法科学，其目的是探索数据的内在数量规律性，以实现对客观事物的科学认识。从原始人结绳记事起统计就已萌芽，但一般认为，作为一门学科，统计学产生于17世纪中叶。而在统计学正式产生之前，统计技术在现实生活中的应用已经有非常悠久的历史了。

从古代中国来看：(1)中国是世界上最早进行人口统计的国家之一，同时也是世界上唯一有长期不间断人口资料记录的国家。据《后汉书》记载，早在公元前2200年，大禹就“平水土，分九州，数万民”，其中的“数万民”就是统计人口。夏禹王朝人口为1355万人，这是我国最早的人口调查资料。我国历史上第一次完整地记载全国各州、郡的户数和人口，是在公元2年(西汉平帝元始2年)。据《汉书·地理志》记载，当时有1223.3万户、5959.4万人。

(2)战国时期的“上计”制度。西周建立了统计报告制度。战国七雄的一个共同点是各国均继承并发展了自西周王朝以来中央政府实行的岁计(一年一次的会计资料汇总)、大计(三年一次的会计资料汇总)制度，并将其正式命名为“上计”制度。

(3)统计被认为是治理国家的重要手段。管子曰：“不明于计数而欲举大事，犹无舟楫而经于水，险也。”(《管子·七法》)秦商鞅的“强国知十三数”：“竟内仓、口之数，壮男、壮女之数，老弱之数，官、士之数，以言说取食者之数，利民之数，马、牛、刍藁之数。欲强国，不知国十三数，地虽利，民虽众，国愈弱至削。”(《商君书·去强》)为有效统治国家，秦始皇建立编户制，东汉曾进行全国田地测量，唐代计口授田，宋明有田亩鱼鳞册的土地调查地图。

(4)明朝的户帖、黄册、鱼鳞册制度。明初创建“户帖制”，由户部制作“户帖”(即登记表)，统一格式和内容，逐级下发，让每户居民翔实填写人口，逐户登记在册。主要内容有：户主姓名、籍贯及丁口数(男子成丁、不成丁，姓名，年龄；女子大口、小口，姓名，年龄)，还分别记载全家其他成员的“花名册”，包括姓名、性别、年龄及与户主的亲属关系；另有“事产”一项，详载该户

土地、房屋、山林、河塘、耕畜、船只等不动产和动产基本情况；最后一项是“户别”，即属于“军户”、“民户”或“匠户”等。黄册和鱼鳞册是明代赋役征发的主要依据，乃是登记天下人口和土地的档案，其中登记人口及其财产状况的叫黄册，绘制全国土地田亩的叫鱼鳞册。

从世界范围来看：(1)早在公元前4500年，巴比伦王国就开展了全国性调查，按族登记人口。(2)古埃及在公元前27世纪，为建金字塔和大型农业灌溉系统，曾进行全国人口和财产调查。(3)公元前15世纪，犹太人为了战争曾经对以色列进行男丁调查。(4)《旧约》中记载，公元前10世纪前后，犹太国王大卫和所罗门对全国进行了比较完整的人口和财产调查。(5)公元前6世纪，罗马帝国以国势调查作为治理国家的有效手段，规定每五年进行一次人口、土地、牲畜、家奴的调查。

概括来说，统计学产生前的社会统计实践活动主要局限于对事物进行原始的调查登记和简单的计算汇总。而现代意义上的统计学的发展基本上是沿着两条主线展开的：一是以“政治算术学派”为开端形成和发展起来的、以社会经济问题为主要研究对象的社会经济统计；二是以概率论的研究为开端并以概率论为基础形成和发展起来的，以方法和应用研究为主的数理统计。

1. 统计学的创立时期

统计学的萌芽产生在欧洲。17世纪中叶至18世纪中叶是统计学的创立时期。在这一时期，统计学理论初步形成了一定的学术派别，主要有国势学派和政治算术学派。

(1) 国势学派。国势学派产生于17世纪的德国，其主要代表人物是康令(1606—1681)和阿亨瓦尔(1719—1772)。康令是第一个在德国黑尔姆斯太特大学以“国势学”为题讲授政治活动家应具备的知识的人。阿亨瓦尔在格丁根大学开设“国家学”课程，其主要著作是《近代欧洲各国国势学纲要》，主要用对比分析的方法研究了解国家组织、领土、人口、资源财富和国情国力，比较各国实力的强弱。1749年，阿亨瓦尔根据拉丁文“Status”、意大利文Stato和Statista及德文Statisti等字根创造出“Statistik”这个新词，原意指“国家显著事项的比较和记述”，后来正式命名为“统计学”。该学派在进行国势比较分析中，主要以文字记述国家的显著事项，而不注重数量对比和数量计算，故亦称记述学派。

(2) 政治算术学派。政治算术学派产生于17世纪中叶的英国，创始人是威廉·配第(1623—1687)，代表作是他于1676年完成的《政治算术》一书。

这里的“政治”是指政治经济学,“算术”是指统计方法。在这部书中,他利用实际资料,运用数字、重量和尺度等统计方法对英国、法国和荷兰三国的国情国力,作了系统的数量对比分析,为以后经济统计的发展开拓了道路。政治算术学派的另一个代表人物是约翰·格朗特(1620—1674)。他以1604年伦敦教会每周一次发表的“死亡公报”为研究资料,在1662年出版了《关于死亡公报的自然和政治观察》的论著。书中分析了60年来伦敦居民死亡的原因及人口变动的关系,首次提出,通过大量观察,可以发现新生儿性别比例具有稳定性和不同死因的比例等人口规律;并且第一次编制了“生命表”,对死亡率与人口寿命作了分析。约翰·格朗特为人口统计的发展开拓了道路。

2. 统计学的发展时期

18世纪末至19世纪末是统计学的发展时期。在这一时期,各种学派的学术观点已经成形,并且形成了两个主要学派,即数理统计学派和社会统计学派。

(1) 数理统计学派。在18世纪,概率理论日益成熟,为统计学的发展奠定了基础。19世纪中叶,凯特勒(1796—1874)首先将概率论原理引入社会现象的研究。在《社会物理学》等书中,他认识到人类的社会活动服从于一定规律,并发现这种规律只有通过大量观察才能被人们所认识。凯特勒被称为现代统计学之父。1867年,一门兼有数学和统计学双重性质的学科被命名为“数理统计学”。

(2) 社会统计学派。凯特勒的另一个重要贡献,是把政治经济学、数学和当时政府统计工作的方法结合在一起,建立了一个专门研究社会现象的统计学派。后来这个学派传到德国,就出现了以克尼斯(1821—1898)、梅尔(1841—1925)和恩格尔(1821—1896)为代表的德国社会统计学派。该学派产生于19世纪后半叶,它融合了国势学派与政治算术学派的观点,沿着凯特勒的“基本统计理论”向前发展,但在学科性质上认为统计学是一门社会科学,是研究社会现象变动原因和规律性的实质性科学,以此同数理统计学派通用方法相对立。

3. 20世纪社会统计学的迅速发展期

20世纪初以来,科学技术迅猛发展,社会发生了巨大变化。社会经济的发展,要求统计学提供更多的统计方法;社会科学本身也不断地向细分化和定量化发展,也要求统计学提供更有效的调查整理、分析资料的方法。社会统计学进入了快速发展时期。

第一次世界大战前后，随着社会统计学派的中心逐步向英、美等国转移，社会统计学与社会学的关系日益明确。1900 年，马约·史密斯出版了《统计学和社会学》一书。1920 年，史特威·恰平的《实地调查与社会研究》出版。二次大战后，社会统计学的实践意义逐步得到了人们的认可。但是，华盛顿大学统计学和社会学教授 Adrian E. Raftery 指出，二战以前，在美国可供社会学研究的数据都显得支离破碎，统计方法也比较简单，仅仅停留在描述性统计的层次上。社会统计学的真正迅速发展是在二次大战之后，这时研究用的数据变得越来越复杂，同时统计方法也在不断发展，以适应数据分析的需要。Adrian E. Raftery 将战后统计学方法在社会学中的应用过程分为三个时期。第一代统计方法起于 20 世纪 40 年代晚期，研究者主要运用交互表（cross-tabulations）的方法，同时对关联测量（measures of association）和对数线性模型（log-linear models）倾注了许多心血。第二代统计方法始现于 20 世纪 60 年代，这一时期的研究者主要面对的是个体层次的调查数据，同时他们将注意力集中在具有线性结构关系（LISREL）的因果模型和事件史分析（event history analysis）上。第三代统计方法在 20 世纪 80 年代晚期就已经初现端倪，研究者所处理的数据已经不能简单地归入上述的任何一个范畴。一方面是因为这些数据具有与众不同的形式，比如文本和口述，另一方面是因为在与空间的和社会网的数据联系时，依赖性已经成为一个至关重要的方面。尽管有许多新的挑战，但用统计学方法研究这一领域的条件已经成熟。

中国的统计学是在 20 世纪前期形成和发展起来的。20 世纪前半期，传入我国的主要是来自两个不同学派的统计思想和理论：一是德国和日本的社会统计学，主张探寻社会发展规律的实质论；二是英、美的数理统计学，坚持通用方法论。国外统计学家对社会经济课题的关注，对早期中国统计界产生了深远的影响，引发了中国学者在理论和实践两个方面进行开创性研究。一批学者展开了科学意义上的统计调查和研究，涉及人口、物价、统计、农业和工业等专门课题。另一些学者则取得了被国际学术界公认的理论研究成果。在 20 世纪 30 年代前后，形成了中国统计学的第一次发展高潮。新中国成立后，在引入苏联计划经济模式的同时，亦引入了苏联统计理论和统计模式。苏联统计理论被奉为马克思主义统计理论的圭臬，影响乃至改变了一代学人的统计思想。在计划经济时期，我国建立了现代国家意义上的、比较完整的统计制度和组织机构。由千百万人参加的统计实务使得 20 世纪 50 年代成为统计学在中国的第二次发展高潮。但之后，中国的统计学过早

地走向平庸化。70年代末，在改革开放浪潮冲击下，我国统计学面临来自理论和实际两个方面的挑战。随着制度变迁、社会分化和经济关系重组，传统的苏式统计理论与中国的改革实践之间出现了尖锐矛盾，依附于计划经济的社会经济统计学在逐步成长的市场之下失去了存在的基础。由此造成传统调查方式的失效以及数据汇总中的遗漏等一系列问题。80年代初，随着社会学等学科的恢复，以及与国外同仁的交流合作，社会统计学重新焕发了生机。1982年，国家统计局成立社会统计司；1983年，《中国社会统计资料》首次公开出版。20世纪90年代以来，随着国内社会学学科发展以及越来越多的海外留学者归国就业，高校科研院所等机构汇聚了一批熟悉社会统计技术、了解定量研究前沿话题的青年学者，他们通过教学与科研，极大地促进了中国社会统计学的发展。中国社会统计学进入了快速发展时期。

三、社会统计学与其他学科的关系

1. 社会统计学与统计学的关系

统计学发展至今，已经成为一门比较成熟的学科，包括应用统计学与数理统计学。统计学作为一门独立的学科，它更偏重数理统计的内容，主要通过利用概率论建立数学模型，收集所观察到的系统的数据，进行量化的分析、总结，并进而进行推断和预测，为相关决策提供依据和参考。它被广泛地应用于各门学科，从物理和社会科学到人文科学，甚至被用于工商业及政府的情报决策之上。社会统计学则是运用统计的一般原理，对社会各种静态结构与动态趋势进行定量描述或推断的一种专门方法与技术。

统计学与社会统计学的联系表现在：统计学可以用到几乎所有的学科领域；统计学可以帮助其他学科探索学科内在的数量规律性；但是，统计学不能解决各学科领域的所有问题，对统计分析结果的解释需要各学科领域的专业人员。只有社会研究的专业人员才能更好地理解和分析社会调查数据。

2. 社会统计学与数学的关系

社会统计学与数学既有着密切的联系，又是完全不同的两个学科。二者的联系是：社会统计学运用到大量的数学知识；数学为统计理论和统计方法的发展提供基础。二者的区别是：数学研究的是抽象的数量规律，社会统计学则是研究具体的、实际现象的数量规律；数学研究的是没有量纲或单位的抽象的数，社会统计学研究的是有具体实物或计量单位的数据；社会统计学与数学研究中所使用的逻辑方法不同，数学研究所使用的主要是演绎，社

会统计学则是演绎与归纳相结合，占主导地位的是归纳。

社会统计学更看重统计方法的理解和运用，并不特别关注其中的数理推导过程。随着计算机技术与统计软件应用的推广，数学的推理与计算基本上由计算机自动完成。所以，数学基础差的学习者没有必要气馁，只要努力同样可以学好社会统计学。通过教学，我们希望学习者达到这样的目标：学会正确地组织、整理待处理和分析的数据；弄清相关的统计概念和统计方法的含义，知道统计方法的适用条件；能够准确地判断待解决问题的性质，并能熟练地把现实问题转换为统计分析问题；针对统计分析问题，能够恰当地选择一种或几种统计分析方法探索性地分析数据，并能在计算机上操作实现；能够读懂计算机分析的输出结果，发现规律，得出分析结论。

第二节　社会研究的历程与统计学的应用

一、社会学量化研究的科学环

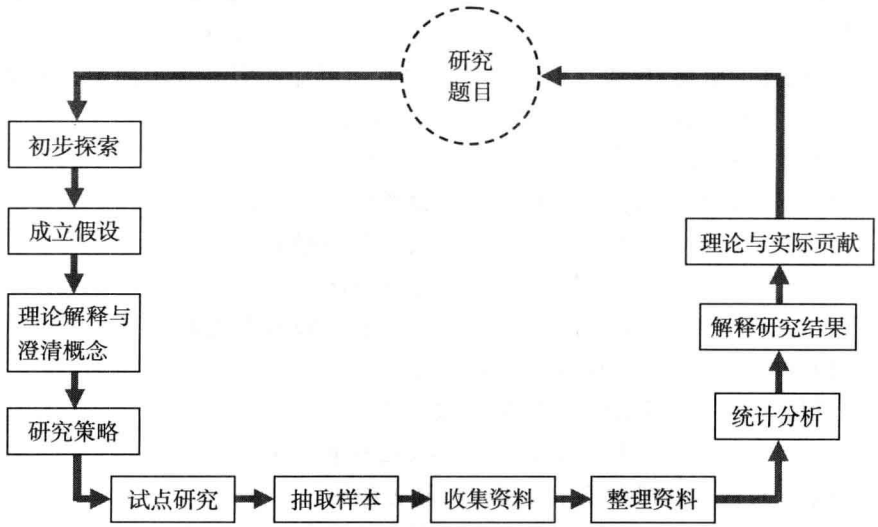

图 1-1　社会学量化研究的科学环

李沛良教授认为，一项完整的社会学量化研究大概要经过图 1-1 所示的所有环节。

1. 确定研究题目

研究者根据个人研究兴趣确定一个研究题域，如有的人对青少年犯