



21世纪社会学系列教材
Textbooks of Sociology in the 21st Century

'91国优教材



社会统计学

(第四版)

Social Statistics (fourth edition)



卢淑华 ⊙ 编著



北京大学出版社
PEKING UNIVERSITY PRESS





91 国优教材

社会统计学

(第四版)

Social Statistics (fourth edition)

卢淑华 ⊙ 编著



北京大学出版社
PEKING UNIVERSITY PRESS

图书在版编目(CIP)数据

社会统计学/卢淑华编著.—4版.—北京:北京大学出版社,2009.11
(21世纪社会学系列教材)

ISBN 978-7-301-09199-9

I.社… II.卢… III.社会统计—高等学校—教材 IV.C91-03

中国版本图书馆CIP数据核字(2009)第198701号

书 名: 社会统计学(第四版)

著作责任者: 卢淑华 编著

责任编辑: 张盈盈 刘金海

标准书号: ISBN 978-7-301-09199-9/C·0341

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>

电 话: 邮购部 62752015 发行部 62750672 编辑部 62753121

出版部 62754962

电子邮箱: ss@pup.pku.edu.cn

印刷者: 北京山润国际印务有限公司

经销者: 新华书店

730×980毫米 16开本 33印张 595千字

1989年8月第1版

2009年11月第4版 2009年11月第1次印刷(总第20次印刷)

定 价: 49.00元

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究

举报电话:010-62752024 电子邮箱:fd@pup.pku.edu.cn

第四版前言

今年距本书首次印刷已经 20 年了,回顾 20 年来,从最初经过了 6 年,才有了第二次印刷,到目前每年都要两次印刷,说明了读者对本书的肯定。例如有的读者说:“自己原本不是社会学出身,对报考社会学研究生感到胆怯,但当读到这本统计学时,有了信心,从而决心报考社会学研究生,如今我已经走上与社会学有关的工作岗位了。”有的社会学毕业生说:“这是我走上人生职场,立即就能用的一门学问。”有的老师说:“这是目前国内介绍社会统计学知识(本科层次)最完整,也最符合社会学专业本科教学特点的教材。”由于读者群的扩大,每每在不同场合,遇到我不认识的人,送上一句“我读过你的书!”时,我都一再感受到本书存在的价值和责任。我想,本书受欢迎的原因,是把高深的统计知识,浅显地介绍给了读者。所谓“深了讲,容易。浅了讲,难!”当然,前提是,讲解本身是正确、不走样的。

本次修订,参考了一些新近出版的有关书籍和杂志,特别是有关统计包应用方面的书籍。正如作者一贯强调的,本书是一本专业基础教材,内容是基本的统计知识,虽然有过两次改版,但总体结构没有大的改变。本次也不例外,下面是本次修改的内容及说明:

1. 第二章增加了多项选择的讨论。虽然从制作统计表来说,似乎没有什么新内容要增加,但调查中确实经常会遇到多项选择问题,SPSS 统计包也设有专门的处理技术加以说明。为此,本次重印作了讨论。此外,在集中趋势的比较中,增加了对出现双峰可能原因的说明。

2. 第三章增加了概率数值大小,偏态、峰态的例题,说明它们实际的含义。逆概公式一节,增加了和人们健康紧密相关的体检例题,以加深对统计知识实用性的理解。

3. 第八章又增加了一例第二类纳伪概率的计算,以便说明样本容量在推论中的重要性。

4. 第十一章增加了 1984 年美国社会统计学家 Hubert M. Blalock 来华讲学,有关研究单位层次间推论的一段话。

5. 第十三章增加了 SPSS 统计包中提供的,用以检验分布正态性的 Normal P-P 图法。

在此修订之际,我要感谢我系的阮桂海和林彬两位教授和我所作的有益讨论。三亚学院的陈飞强老师,虽然素昧平生,但出于对教学认真负责的态度,寄来了他教学中发现的印刷错误,以及同学容易混淆的地方,谨此表示感谢!

最后对长期关心和帮助本书出版的读者、教师、出版社的同仁表示深深的敬意!没有大家的辛勤参与,本书是不可能持续至今的。

作者

2009年8月于北京大学社会学系

关于修改例题部分内容的说明

本书自 1989 年出版,自今已累计印刷达 16 次。其间虽有两次修改版,但书中例、习题基本保留。考虑到近二十年我国社会生活变化巨大,书中不少例题已经过时或不符合当前的情况,而作为通过社会生活介绍统计知识的应用统计学,虽然全书的统计知识不需修改,但例、习题有必要作适当修改。修改中,作者仍然本着一贯的原则,即例题是为介绍和掌握统计知识服务的,只要能说明问题,例、习题应越贴近生活易懂越好,绝不采用生硬难懂的专业术语,以便广大的读者群,包括非社会学专业的,都可正确的理解统计的知识。

最后须要强调的,本书自 2006 年底,大部分图表都做成了多媒体,欢迎老师免费下载(详见本书第三版第 5 次以后各次印刷版本的末页)。网址是:
<http://www.pup.cn>。

2007 年岁末 作者于北京大学社会学系

2005 年版前言

本书自 1989 年问世,2000 年重排后,如今总印刷已达 10 次了。其间除作者在北京大学社会学系历届使用外,还得到了兄弟院校相关专业的使用,从而累积了不少有益的交流和值得修改的地方。但本书是一本专业基础课的教材,一方面基本内容和体系是不变的,另一方面篇幅也不宜过长。因此这次把要增添的内容只作扼要介绍,有的则放在注释中给出,未做太大的展开。本次修改有如下几点:

1. 进一步统一了术语,如频数、频次都统一为频次。
2. 第三章至第六章、第八章、第十三章都有部分内容改为选读,这样既保持了统计知识体系的完整与水准,同时也便于多层次读者的使用。
3. 改写了部分例题的内容,以便更贴近当前社会生活的实际。
4. 根据读者的询问,某些公式在注释里给出了出处。
5. 根据多年教学的经验,某些内容的顺序作了适当调整,如第十章第二节的部分内容。
6. 增加了某些内容和讨论,如配对与随机化设计、生态学谬误、实验性与非实验性研究、PPS 抽样,这些都在相应的第九章、第十一章、第十二章、第十五章有介绍。

最后,在此仅向关心本书、给本书提出过宝贵意见的师生、广大读者以及北大出版社的同仁致意,没有他们的辛勤劳动,本书是不可能出版的。

2005 年 4 月作者于北京大学

重排本前言

社会统计学自 1989 年问世以来,如今已印刷了 4 次,累计印数达 2 万册。本次决定重新排版的原因,是经多次印刷后原版损毁严重,有的字迹模糊不清。所以从内容上讲,重排本与以往的版本基本相同。只在以下几处有些小的改动:

1. 第二章一增加了组距式统计表的名称。
2. 第二章二(三)增加了离散型定距变量采用组距式统计表的解释。
3. 第二章第一节四增加了累计图表的应用实例:洛伦茨曲线和基尼系数。
4. 第十章第三节三增加了对 λ 系和 τ 系数的讨论,并引出了几率比概念。
5. 本书曾由北大出版社向海外推荐,准备在台湾出版。为此,作者将书中的主要专业术语都补充了英语译名。现借重新排版之际一并推荐给内地读者,相信这是很有益的。

另外,新增加的附录是统计专业术语的中英文对照表。该表是按章汇总,词汇是按在各章出现的先后次序排列。

本书在出版的十余年中曾获北京大学科研成果奖、国家教育部高等学校优秀教材二等奖以及由北京市社科院所主持的、对全国开设社会学系或专业共 16 所大学全部硕士生、博士生以及他们的导师、就社会学恢复以来主要课程教材的问卷调查。本书获同类教材综合评分第一名(《社会学研究》1994 年第 5 期)。在此修订之际,谨向读者表示深深的谢意。长江后浪推前浪,相信在现有教材、抛砖引玉的基础上,必将有更优秀的教材问世。

2000 年 10 月作者于北京大学

初 版 前 言

社会统计学是根据 1985 年 11 月国家教委在广州召开文科教学总结会上所确定的课程。它是社会学专业十门必修课中的一门。它与另一门必修课“社会学方法与社会调查研究”结合起来,完整地介绍了当代社会调查研究的科学方法和资料处理技术。其中社会统计学则是侧重介绍资料的收集、整理、分析和推论的处理技术。

社会统计学,正如一切其他领域的应用统计学,它的渊源都是来自概率论和数理统计学,因此都带有浓厚的数理统计学特色。但它和数理统计学教材或其他领域的应用统计学相比,有如下的特点:

1. 从内容阐述来讲,多取材于社会现象,并多采用直观、浅显的讲解,尽量减少数学上的难度,以适合文科学生的程度和需要。

2. 社会统计学根据变量的四个层次(定类、定序、定距和定比)组织与展开教材的内容,这是社会统计学所特有的,很适合社会现象的研究和学生对统计技术的掌握,这在国内其他应用统计中还未见过。

3. 根据社会现象的特点,社会统计学扩大了定类变量的讨论,特别是对列联表作了详尽的讨论。

4. 根据社会研究的需要,讨论了适合各种变量层次的统一相关强度定义方法和测量。这些量度的定义和公式,只有在社会统计学学科中才有系统的介绍。对这些公式的解释亦具有社会学研究的特色。

社会统计学课程是北京大学建立社会学系以来,在袁方教授鼎力倡导加强现代社会调查方法的研究而开设的。1984 年以来,起初本教材内容在研究生中讲授。1986 年以后,逐渐扩大到社会学系本科生。目前本教材内容已作为社会学系大学生教材。研究生则以介绍多变量分析技术为主,这部分内容在本教材中没有编入,因为这样会更适合大学本科生的程度。

本教材由于已将数理统计学和社会研究中的应用溶合在一起,并对概率和数理统计知识的讲授也有一定的深度。因此,学生无须先进行概率论或其统计知识的学习,只需有高等数学的知识就可以了。本教材的讲授约需 80 学时。

此外,由于社会统计学的习题都是作者自己编写的,这些习题,有的已作为书中例题给出。有一些习题,集中编在各章之后。

在编写本教材过程中,作者参考了大量国外社会统计学书籍,也吸收了国内外数理统计学和其他领域应用统计学的内容。就书中例题的内容来说,多结合近几年北大社会学系的研究课题和国内有关的课题。学生通过本教材的学习,可以直接学到统计技术在社会研究中的运用。

需要说明的是国内统计学中的名称不够一致,经济类也有一些社会统计学,但就其内容来说与本书是不同的。本书是以数理统计学为基础,并注意与国外社会统计学名称与讲授内容的一致性。

这本教材已被国家教委批准作为高等学校社会学专业的统编教材。由于社会学目前在我国还处在恢复初建阶段,教材建设亟待完善,作为第一本社会学教材,错误不当之处在所难免,切望国内同行和广大读者批评指正。

本教材在审编过程中,承蒙北大经济管理系胡健颖教授细心审阅,北师大数学系李占柄教授对书中公式作了审查,北大出版社王禹功副编审和苏勇副编审以及中国社科院社会学所唐军同志为本书的出版付出了辛勤劳动,特此致以谢意。并向一切关心过本书出版的社会学系师生员工表示致意。

1989年5月作者于北京大学

目 录

第一章 社会学研究与统计分析	(1)
第一节 社会学研究的科学性	(1)
第二节 社会调查资料的特点和统计学的运用	(7)
第三节 怎样选用统计分析方法	(16)
第二章 单变量统计描述分析	(21)
第一节 分布 统计表 统计图	(21)
第二节 集中趋势测量法	(43)
第三节 离散趋势测量法	(51)
习 题	(59)
第三章 概率	(63)
第一节 基础概率	(63)
第二节 概率分布、均值与方差	(90)
习 题	(110)
第四章 二项分布及其他离散型随机变量的分布	(113)
第一节 二点分布	(113)
第二节 排列与组合	(115)
第三节 二项分布	(118)
第四节 多项分布	(125)
第五节 超几何分布	(133)
第六节 泊松分布	(137)
习 题	(142)
第五章 正态分布、常用统计分布和极限定理	(143)
第一节 什么是正态分布	(143)
第二节 标准正态分布	(150)
第三节 标准正态分布表的使用	(156)

第四节	常用统计分布	(160)
第五节	大数定理与中心极限定理	(166)
习 题	(182)
第六章	参数估计	(184)
第一节	统计推论	(184)
第二节	名词解释	(186)
第三节	参数的点估计	(188)
第四节	抽样分布	(197)
第五节	正态总体的区间估计	(205)
第六节	大样本区间估计	(215)
习 题	(226)
第七章	假设检验的基本概念	(227)
第一节	统计假设	(227)
第二节	统计检验中的名词	(231)
第三节	假设检验的步骤和两类错误	(236)
第八章	单总体假设检验	(242)
第一节	大样本假设检验	(242)
第二节	小样本假设检验	(249)
习 题	(254)
第九章	二总体假设检验(二分变量—二分变量; 二分变量—定距变量)	(256)
第一节	引言	(256)
第二节	大样本二总体假设检验	(258)
第三节	小样本二总体假设检验	(263)
第四节	配对样本的比较	(269)
习 题	(272)
第十章	列联表(定类变量—定类变量)	(274)
第一节	什么是列联表	(274)
第二节	列联表的检验	(283)
第三节	列联强度	(291)
习 题	(312)

第十一章 等级相关(定序变量—定序变量)	(314)
第一节 斯皮尔曼等级相关系数	(314)
第二节 Gamma 等级相关	(323)
第三节 其他等级相关系数	(332)
习 题	(337)
第十二章 回归与相关(定距变量—定距变量)	(339)
第一节 回归研究的对象	(339)
第二节 回归直线方程的建立与最小二乘法	(342)
第三节 回归方程的假定与检验	(346)
第四节 相关	(355)
第五节 用回归方程进行预测	(368)
习 题	(372)
第十三章 方差分析(定类变量—定距变量)	(375)
第一节 引言	(375)
第二节 一元方差分析	(376)
第三节 二元方差分析	(390)
第四节 多元方差分析	(416)
习 题	(420)
第十四章 非参数检验(定类变量—定序变量)	(422)
第一节 非参数检验	(422)
第二节 符号检验	(423)
第三节 符号秩检验	(430)
第四节 秩和检验	(433)
第五节 游程检验	(435)
第六节 累计频次检验	(438)
第七节 两个以上样本的非参数检验	(443)
习 题	(446)
第十五章 抽样	(449)
第一节 引言	(449)
第二节 抽样调查方法	(450)
第三节 抽样误差	(455)

第四节	样本容量的确定	(470)
习题	(475)
附表	(476)
附表 1	随机数字表.....	(477)
附表 2	二项分布表.....	(479)
附表 3	泊松分布表.....	(483)
附表 4	标准正态分布表.....	(485)
附表 5	t 分布表	(486)
附表 6	χ^2 分布表	(488)
附表 7	F 分布表.....	(491)
附表 8	相关系数检验表.....	(500)
附表 9	符号秩检验表.....	(501)
附表 10	秩和检验表	(502)
附表 11	游程数检验表	(503)
附录	统计术语中英文对照	(505)
参考书目	(511)
增加多媒体教学软件的说明	(512)

第一章

社会学研究与统计分析

第一节 社会学研究的科学性

社会学是一门研究不断变化着的社会生活的科学,提起社会学研究,人们很自然地就会想到社会调查。社会学工作者凭借社会这样一个大工厂,对社会的资料进行收集、整理和分析,以便对社会学的假设、理论进行谨慎的求证。社会对于社会学来说,犹如实验室、工厂对于自然科学一样,它是进行科学研究的源泉和手段。通过社会调查进行社会学研究,其成果与自然科学通过科学实验获取成果具有同等的科学价值。社会调查在社会学的现代化科学研究中,扮演了重要的角色。下面让我们来回顾一下社会研究方法教程中所列举的社会调查研究的大致步骤。

一、确定课题

社会现象,包罗万千,如何从中确定研究的课题呢?首先,课题的研究必须具有社会价值;其次,还要考虑人力、物力的可能。社会学研究课题除了少部分来源于社会学理论外,大部分都是来源于当前的社会现实和要解决的实际问题。因此,社会学的研究具有强烈的时代感,是直接为国家现代化服务的。例如,为配合我国当前构建和谐和谐社会问题,社会学工作者进行了大量关于社会各阶层利益的调查、效率与公平的调查,这些都为国家的有关决策提供了可靠的依据。可以不夸张地说,社会学在社会信息全面的提供上起到了其他学科不可替代的作用。

二、了解情况

在确定研究课题之后,通过查阅文献和向有经验、有知识的人,了解本课题已有的进展。同时,更重要的还要向社会进行了解。运用个案调查、典型调查进行探索性研究,了解人们现实的想法与动态,以便取得第一手资料。

三、建立假设

在前两步的基础上,明确研究的范围,并在初步探索的基础上,提出一定的想法和建立假设。

举例说,如果我们确定的课题是有关生育意愿的问题,那么,通过探索性研究,发现生育的意愿是因人而异的。具体说,文化程度高的人,希望生育的子女数,就比文化程度低得要少些。城市的居民又比农村的居民希望生育的子女数要少些。这里我们不仅讨论孤立的社会现象,诸如:希望生育的子女数;文化程度;地区等等。同时还要研究社会现象之间的联系。例如:希望生育的子女数与文化程度之间的联系;希望生育的子女数与地区之间的联系。这称作命题或模型的研究。模型的表述有两种方式:

(一) 差异式:差异式仅说明现象与现象之间存在关系。例如上面所谈的地区不同、生育意愿亦不相同就是差异式模型。

(二) 函数式:函数式不仅表现了现象与现象之间存在关系,而且还表现了两者间变化的方向:

A 高则 B 高(正比);

A 高则 B 低(反比)。

例中文化程度愈高,则生育意愿愈低就是反比关系。

需要强调的是,以上模型只是一种初步想法,或称假设。最终能否确立,还必须通过实践的检验。

四、确立概念和测量方法

通过上例可以看出,在社会学的研究中,需要采用适当的术语或概念来描述研究的对象。例如上例中就运用了“生育意愿”这样一个概念。概念是进行研究的基本单位,它犹如建筑大厦中的基石。通过它,才能建立起整体间的联系。概念一般具有抽象的属性,没有时间和空间的限制。抽象层次越高,所概括的现象越广。

概念不仅包括人们习惯了的一些术语,例如:性别、职业、文化程度,同时还

包括根据研究需要所设计出来的术语,又称构念。例如社会化、都市化、社会地位、社会适应、职业流动、社会角色等等。这些术语在使用时都必须给出作者的定义,以免混淆。例如“角色”这个概念,其含意既可指客观对行为者的要求,也可指行为者主观上对自己的要求,还可指行为者的实际行动等等。这些在使用中都必须有明确的界定。其次,上面所说的定义还只是抽象性的。为了进一步开展定量研究,还必须对所运用的概念建立定量测量的方法。犹如自然科学中仅仅知道了温度的定义和公式还不够,还必须有人发明温度计来测量温度一样。社会学研究中所涉及的概念,也必须用一串可以观察、可以测量的指标来模拟它。这称作概念的操作化定义。试比较以下一组概念的抽象定义和操作化定义(表 1-1)。

表 1-1

概念	抽象定义	操作化定义
都市化	现代都市的生活形态。	妇女就业人数;子女数;交通;通讯手段等等。
个人现代化观念	一个人由于经济、工业等现代工业因素的影响所产生的内部变化。	对时间、效率、家庭、亲属、消费、自信等具体问题的看法。
子代偏重	经济上、感情上和生活上对子代过分的偏重。	子代抚育费与家庭平均消费的比较;子代过生日与长辈是否过生日的比较;花在子代闲暇时间的多少等等。

可见,抽象概念通过操作化定义得到了测量和量化。所谓操作化定义(operational definition)或许是翻译不够传神,它的意思是通过概念的操作化,非量化的概念得以运算(operation)、得以测量操作了!可见,概念的操作化使定性和定量研究之间架起了桥梁,它是定量研究的飞跃和艺术所在。任何一门学科定量研究的重要,正如拉法格在《回忆马克思》中所写道,马克思认为:“一种科学只有在它成功地运用数学时,才算达到了真正完善的地步。”

操作化定义就其本质来说,只是对抽象概念的间接测量。而间接测量的手段往往不是惟一的。正像通过液体体积变化来间接测量温度一样,其温度计里装的液体既可以是水银,也可以是酒精。同样,操作化定义对于同一个概念也不是惟一的。而一个好的操作化定义应尽量模拟和包含抽象定义的内容。