



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

'91国优教材

社会统计学

(第三版)

Social Statistics

卢淑华 编著



北京大学出版社
PEKING UNIVERSITY PRESS



'91国优教材

ZWJC
2005
C8
26

社会统计学

(第三版)

Social Statistics



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. —北京:北京大学出版社, 2001.6
ISBN 7-301-00106-1

I.社… II.卢… III.社会统计-理论 IV.C91-03

中国版本图书馆 CIP 数据核字(2001)第 78502 号

书 名: 社会统计学(第三版)

著作责任者: 卢淑华 编著

责任编辑: 刘金海

标准书号: ISBN 7-301-00106-1/C·008

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://cbs.pku.edu.cn>

电 话: 邮购部 62752015 发行部 62750672 编辑部 62752032

电子信箱: zpup@pup.pku.edu.cn

排 版 者: 北京高新特打字服务社 51736661

印 刷 者: 三河市新世纪印务有限公司

经 销 者: 新华书店

730×980 毫米 16 开本 32.5 印张 595 千字

1989 年 8 月第 1 版

2005 年 5 月第 3 版 2005 年 5 月第 1 次印刷(总第 11 次印刷)

定 价: 41.00 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,翻版必究

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 and 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,作为第一本社会统计学教材,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 | (40) |
| 第三节 离散趋势测量法 | (48) |
| 习 题 | (56) |
| 第三章 概率 | (60) |
| 第一节 基础概率 | (60) |
| 第二节 概率分布、均值与方差 | (86) |
| 习 题 | (105) |
| 第四章 二项分布及其他离散型随机变量的分布 | (108) |
| 第一节 二点分布 | (108) |
| 第二节 排列与组合 | (110) |
| 第三节 二项分布 | (113) |
| 第四节 多项分布 | (120) |
| 第五节 超几何分布 | (128) |
| 第六节 泊松分布 | (132) |
| 习 题 | (137) |
| 第五章 正态分布、常用统计分布和极限定理 | (138) |
| 第一节 什么是正态分布 | (138) |
| 第二节 标准正态分布 | (145) |
| 第三节 标准正态分布表的使用 | (151) |
| 第四节 常用统计分布 | (155) |

| | | |
|-------------|--|-------|
| 第五节 | 大数定理与中心极限定理 | (161) |
| 习 题 | | (177) |
| 第六章 | 参数估计 | (179) |
| 第一节 | 统计推论 | (179) |
| 第二节 | 名词解释 | (181) |
| 第三节 | 参数的点估计 | (183) |
| 第四节 | 抽样分布 | (192) |
| 第五节 | 正态总体的区间估计 | (200) |
| 第六节 | 大样本区间估计 | (210) |
| 习 题 | | (221) |
| 第七章 | 假设检验的基本概念 | (222) |
| 第一节 | 统计假设 | (222) |
| 第二节 | 统计检验中的名词 | (226) |
| 第三节 | 假设检验的步骤和两类错误 | (231) |
| 第八章 | 单总体假设检验 | (237) |
| 第一节 | 大样本假设检验 | (237) |
| 第二节 | 小样本假设检验 | (243) |
| 习 题 | | (248) |
| 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) | (250) |
| 第一节 | 引言 | (250) |
| 第二节 | 大样本二总体假设检验 | (252) |
| 第三节 | 小样本二总体假设检验 | (257) |
| 第四节 | 配对样本的比较 | (263) |
| 习 题 | | (266) |
| 第十章 | 列联表(定类变量—定类变量) | (268) |
| 第一节 | 什么是列联表 | (268) |
| 第二节 | 列联表的检验 | (277) |
| 第三节 | 列联强度 | (285) |
| 习 题 | | (306) |
| 第十一章 | 等级相关(定序变量—定序变量) | (308) |
| 第一节 | 斯皮尔曼等级相关系数 | (308) |
| 第二节 | Gamma 等级相关 | (317) |

| | | |
|-------------|-------------------------------|-------|
| 第三节 | 其他等级相关系数 | (325) |
| 习 题 | | (331) |
| 第十二章 | 回归与相关(定距变量—定距变量) | (333) |
| 第一节 | 回归研究的对象 | (333) |
| 第二节 | 回归直线方程的建立与最小二乘法 | (336) |
| 第三节 | 回归方程的假定与检验 | (340) |
| 第四节 | 相关 | (349) |
| 第五节 | 用回归方程进行预测 | (362) |
| 习 题 | | (366) |
| 第十三章 | 方差分析(定类变量—定距变量) | (369) |
| 第一节 | 引言 | (369) |
| 第二节 | 一元方差分析 | (370) |
| 第三节 | 二元方差分析 | (383) |
| 第四节 | 多元方差分析 | (409) |
| 习 题 | | (413) |
| 第十四章 | 非参数检验(定类变量—定序变量) | (415) |
| 第一节 | 非参数检验 | (415) |
| 第二节 | 符号检验 | (416) |
| 第三节 | 符号秩检验 | (423) |
| 第四节 | 秩和检验 | (426) |
| 第五节 | 游程检验 | (428) |
| 第六节 | 累计频次检验 | (431) |
| 第七节 | 两个以上样本的非参数检验 | (436) |
| 习 题 | | (439) |
| 第十五章 | 抽样 | (442) |
| 第一节 | 引言 | (442) |
| 第二节 | 抽样调查方法 | (443) |
| 第三节 | 抽样误差 | (448) |
| 第四节 | 样本容量的确定 | (463) |
| 习 题 | | (468) |
| 附 表 | | (469) |
| 附表 1 | 随机数字表 | (471) |
| 附表 2 | 二项分布表 | (473) |



| | | |
|-------|-------------------|-------|
| 附表 3 | 泊松分布表..... | (477) |
| 附表 4 | 标准正态分布表..... | (479) |
| 附表 5 | t 分布表..... | (480) |
| 附表 6 | χ^2 分布表..... | (482) |
| 附表 7 | F 分布表..... | (485) |
| 附表 8 | 相关系数检验表..... | (494) |
| 附表 9 | 符号秩检验表..... | (495) |
| 附表 10 | 秩和检验表..... | (496) |
| 附表 11 | 游程数检验表..... | (497) |
| 附录 | 统计术语中英文对照..... | (499) |
| 参考书目 | | (505) |

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,人们很自然的就会想到社会调查。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。

或许会产生这样的问题：对于某些社会学的概念，在现实的社会中，是否会出现无法找到作为间接测量的替代物或操作化定义？回答是否定的，正如恩格斯在《自然辩证法》一书中说到，人类的概念不过是一种简称，利用这种简称，把许多不同的、可以从感觉上感知的事物，依照其共同的属性把握住^①。可见，社会学的每个概念都是对这种或那种现实存在的现象的概括和抽象，因此必然在社会生活中可以找到与之对应的经验层次的替代物，也就是操作化定义。而社会学工作者的任务，就在于发现这些概念的替代物，并使之不断完善。一般来说，一个效度高的操作化定义并不是很容易设计出来的。如果说统计分析技术还可以借助于统计工作者协作的话，那么，操作化定义的设计必须由课题研究人员自己来完成。而操作化定义设计的好坏则取决于研究人员对课题理解的深度、情况的掌握以及研究的素质和艺术。

概念及其操作化定义，在数量上或质量上往往有所变化。或者说概念的表现形式往往具有多值性。因此，概念可以看做为变量，而各种表现形式就是变量的不同取值。例如性别是概念，也是变量，而男和女则是变量的两种可能取值。下面列举一组变量及其可能的取值(表 1-2)。

表 1-2

| 变 量 | 变量的可能取值 |
|--------|-------------|
| 性 别 | 男；女 |
| 家庭子女数 | 1个；2个；3个…… |
| 重 要 性 | 非常重要；一般；不重要 |
| 文化 程 度 | 大学；中学；小学；文盲 |

五、设计问卷

问卷是指一组与研究目标有关的问题。这些问题则是根据概念操作化所提出的。问卷包括的内容一般有：

(一) 事实：被访人的年龄、性别、职业、文化程度等等。这些事实属于基本资料，在分析过程中，往往被当作自变量来考虑。

(二) 态度与看法：被访人对某种行为、政策是否赞成，对某种职业的评价等等。

(三) 行为趋向：行为趋向具有假设性。要了解的往往是在某一种情况下，

^① 参阅《马克思恩格斯全集》第20卷，第579页。

被访者会有什么样的行为。

(四) 理由:了解被访人采取某种态度和行为趋向的原因。

问卷的回答有两种方式:固定答题式和自由答题式。固定答题式一般是多种答案选择,这种问卷在大规模调查中经常使用。固定答题中答案的设计,取决于研究人员对问题实际情况了解的程度。为此,在探索性研究阶段,不妨采用自由答题式,以便收集到更多的活思想、新情况。

六、试填问卷

把问卷发给研究对象中的少数人试填,以便使问卷设计不周或遗漏之处,尽量在试填阶段予以纠正。否则,当大规模调查一经开始,纠正起来将相当困难,甚至不可能。这点凡具有实际经验的人,都知道它在调查研究中的实际价值和不可缺少性。

七、调查实施(抽样调查)

设计好问卷,下一步就可以开展大规模社会调查收集资料了。社会学研究,很少采用全面调查,一般都是从研究总体中,科学的抽取一部分进行研究,然后从局部推论到全体。但即便是抽样社会调查,一般调查人数也在数百人以上。因此对于大型抽样社会调查,培训访问员协助调查是必须的。但研究人员在问卷调查中,自己也要参加一部分实地调查,以便及时发现问题,指导访问员和对访问员进行质量检查。

八、校核与登录

问卷回收之后,应立即逐份进行校核,看看是否有填错或明显不合理的部分。例如初婚年龄过小或家庭成员之间关系前后填写有矛盾,或编码不合理等等。情况发现愈早、纠正起来愈方便。否则,事过境迁,无论访问员或被访者回忆起来,都将十分困难。问卷校核之后,可以将资料录入计算机保存起来,以便进一步处理。

九、统计分析与命题的检验

问卷回收的资料还只是原始的数据,它必须经过整理、归纳与分析,才能作为研究命题或假设的凭据。而资料的整理、归纳、分析以及如何收集资料,正是统计分析所要讨论的基本内容。

进行统计分析,其计算量一般都很大。目前这些大量重复、令人困倦的计

算工作,一般都委以计算机去完成。例如国外常用的社会科学软件包有:SPSS; SAS; LISREL 以及 STATPAC 等等,这些都有专门的软件包课程讲授。

通过统计分析,可以进行以下几方面工作:检验最初研究阶段的命题或假设是否得以证实或部分证实,并在此基础上对研究内容提出建议和确定进一步的研究方案。可见,社会研究方法是遵循了人们认识真理的基本规律的。那就是理论来源于实践,而又必须受到实践的检验。人们在理论联系实践的多次循环认识中,逐步掌握了事物的客观规律。

总结以上所谈的步骤,可以看出社会研究包括了不可缺少的两个层次(或两个层面)的结合(表 1-3)。

表 1-3 科学研究方法

| 抽象层 | | 概念 | 命题 | 理论 |
|------|----|----------|--------------------------|----------|
| 经验层 | | | | |
| 研究设计 | 原则 | 观察 量度 | 测定 分析 | 测定 分析 |
| | 数据 | 搜集 | 使用计算机对数据记录、 储存、管理与分析。 | |

抽象层次包括前面列举的一、二、三步骤,主要目的在于确定课题、概念以及概念与概念之间假设的关系(命题)或一组命题。但作为科学研究,仅此还不够,还必须得到经验层次支持与证实。为此,必须搜集数据。而概念与搜集之间,必须通过观察与量度才能使研究得以量化。这就是研究设计的概念操作化。作为研究,一般都要分析社会现象与现象之间的关系,因此,在对概念进行操作化,搜集数据的同时,还要对所假设的命题或理论进行测定、分析和检验。而社会学工作者的能力则表现在自由穿梭于两个层面、富于想象而又清晰、明朗,这就是要求具有良好的社会学想象力。

下面用一个简化的循环图来说明社会学研究称之为科学研究的进程,它又称科学环(图 1-1)。

其中,社会学理论和假设是指导我们应该收集那些资料,它是研究的基础,是定性的研究。而实践和经验概括则要解决资料如何收集,如何整理,如何分析和如何推论。