

高等学校教材

# 社会统计学

卢淑华 编著

北京大学出版社

高等学校教材

社会统计学

卢淑华 编著

北京大学出版社

书名题字：王君华

封面设计：张自忠

责任编辑：王禹功

## 社 会 统 计 学

卢淑华 编著

\*

北京大学出版社出版

(北京大学校内)

北京大学印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

\*

850×1168毫米 32开本 19印张 480千字

1989年8月第一版 1989年8月第一次印刷

印数：0001—7,000册

ISBN 7-301-00106-1/C·008

定价：4.65元

## 内 容 简 介

本书系统地介绍了社会统计学的基本内容：统计描述和统计推论，并以变量的四个层次（定类、定序、定距、定比），系统地阐述了单变量和双变量的各种统计技术。本书内容多取材于社会现象，作者采用直观、浅显的叙述，将数理统计和社会研究中的应用溶合在一起。本书为高等学校统编教材，社会学专业必修课，同时可供文科有关专业师生参考。

# 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，完整的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，正如一切其它领域的应用统计学，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它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4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，这在国内其它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，只须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它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。

# 目 录

<b>第一章 社会学研究与统计分析</b> .....	(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.....	(8)
第三节 怎样选用统计分析方法.....	(20)
<b>第二章 单变量统计描述分析</b> .....	(26)
第一节 分布 统计表 统计图.....	(26)
第二节 集中趋势测量法.....	(47)
第三节 离散趋势测量法.....	(57)
习 题.....	(66)
<b>第三章 概率</b> .....	(71)
第一节 基础概率.....	(71)
第二节 概率分布、均值与方差.....	(104)
习 题.....	(127)
<b>第四章 二项分布及其它离散型随机变量的分布</b> .....	(131)
第一节 二点分布.....	(131)
第二节 排列与组合.....	(133)
第三节 二项分布.....	(138)
第四节 多项分布.....	(145)
第五节 超几何分布.....	(155)

第六节	泊松分布	(160)
	习 题	(166)
<b>第五章</b>	<b>正态分布、常用统计分布和极限定理</b>	<b>(168)</b>
第一节	什么是正态分布	(168)
第二节	标准正态分布	(175)
第三节	标准正态分布表的使用	(184)
第四节	常用统计分布	(189)
第五节	大数定理与中心极限定理	(195)
	习 题	(217)
<b>第六章</b>	<b>参数估计</b>	<b>(218)</b>
第一节	统计推论	(218)
第二节	名词解释	(220)
第三节	参数的点估计	(223)
第四节	抽样分布	(235)
第五节	正态总体的区间估计	(243)
第六节	大样本区间估计	(256)
	习 题	(269)
<b>第七章</b>	<b>假设检验的基本概念</b>	<b>(270)</b>
第一节	统计假设	(270)
第二节	统计检验中的名词	(274)
第三节	假设检验的步骤和两类错误	(281)
<b>第八章</b>	<b>单总体假设检验</b>	<b>(288)</b>
第一节	大样本假设检验	(288)
第二节	小样本假设检验	(295)
	习 题	(301)



<b>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)</b> .....	(303)
第一节 引言 .....	(303)
第二节 大样本二总体假设检验 .....	(305)
第三节 小样本二总体假设检验 .....	(311)
第四节 配对样本的比较 .....	(318)
习 题 .....	(322)
<b>第十章 列联表 (定类变量—定类变量)</b> .....	(324)
第一节 什么是列联表 .....	(324)
第二节 列联表的检验 .....	(334)
第三节 列联强度 .....	(345)
习 题 .....	(368)
<b>第十一章 等级相关 (定序变量—定序变量)</b> .....	(371)
第一节 斯皮尔曼等级相关系数 .....	(371)
第二节 Gamma 等级相关 .....	(381)
第三节 其它等级相关系数 .....	(391)
习 题 .....	(397)
<b>第十二章 回归与相关 (定距变量—定距变量)</b> .....	(400)
第一节 回归研究的对象 .....	(400)
第二节 回归直线方程的建立与最小二乘法 .....	(404)
第三节 回归方程的假定与检验 .....	(409)
第四节 相关 .....	(419)
第五节 用回归方程进行预测 .....	(435)
习 题 .....	(439)

**第十三章 方差分析 (定类变量—定距变量) .....(442)**

第一节 引言.....(442)

第二节 一元方差分析.....(443)

第三节 二元方差分析.....(459)

第四节 多元方差分析.....(489)

习 题.....(494)

**第十四章 非参数检验 (定类变量—定序变量) .....(497)**

第一节 非参数检验.....(497)

第二节 符号检验.....(498)

第三节 符号秩检验.....(506)

第四节 秩和检验.....(510)

第五节 游程检验.....(513)

第六节 累计频次检验.....(516)

第七节 两个以上样本的非参数检验.....(522)

习 题.....(526)

**第十五章 抽样.....(530)**

第一节 引言.....(530)

第二节 抽样调查方法.....(531)

第三节 抽样误差.....(536)

第四节 样本容量的确定.....(555)

习 题.....(562)

**附 表.....(564)**

附表 1 随机数字表.....(566)

附表 2 二项分布表.....(568)

附表 3 泊松分布表.....(572)

附表 4	标准正态分布表.....	(574)
附表 5	t 分布表 .....	(575)
附表 6	$\chi^2$ 分布表.....	(577)
附表 7	F 分布表.....	(580)
附表 8	相关系数检验表.....	(589)
附表 9	符号秩检验表.....	(590)
附表10	秩和检验表.....	(591)
附表11	游程数检验表.....	(592)
<b>参考书目 .....</b>		<b>(594)</b>

# 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，是直接为四个现代化服务的。例如，为配合我国当前人口政策的贯彻执行，社会学工作者进行了大量生育意愿的调查和研究。为配合经济体制的改革，对改革的社会心理反应，进行了大量的调查和研究。这些都为有关领导的决策提供了可靠的依据。

### 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、社会流动、社会角色等等。这些术语在使用时都必需给出作者的定义，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）：

概 念	、抽 象 定 义	操 作 化 定 义
都市化	现代都市的生活形态。	如女就业人数；子女数；交通；通讯手段等等。
个人现代化	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；子代过日生与长辈是否过日生的比较；花在子代闲暇时间的多少等等。

（表1—1）

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唯一的。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唯一的。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<sup>①</sup>。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, 概念可以看作是变

<sup>①</sup> 参阅《马克思恩格斯全集》第20卷, 第579页。

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。下面例举一组变量及其可能的取值（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

（表1—2）

## 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，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 with 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。