



Statistics

21世纪统计学系列教材

Social Statistics

社会统计学

(第2版)

尹海洁 李树林 编著

中国人民大学出版社



Statistics 21世纪统计学系列教材



▲ Social Statistics

社会统计学 (第2版)

尹海洁 李树林 编著

中国人民大学出版社
· 北京 ·

图书在版编目 (CIP) 数据

社会统计学/尹海洁, 李树林编著. —2 版. —北京: 中国人民大学出版社, 2018. 1
21 世纪统计学系列教材
ISBN 978-7-300-25217-9

I. ①社… II. ①尹…②李… III. ①社会统计-教材 IV. ①C91-03

中国版本图书馆 CIP 数据核字 (2017) 第 298881 号

21 世纪统计学系列教材

社会统计学 (第 2 版)

尹海洁 李树林 编著

Shehui Tongjixue

出版发行 中国人民大学出版社

社 址 北京中关村大街 31 号

电 话 010-62511242 (总编室)

010-82501766 (邮购部)

010-62515195 (发行公司)

网 址 <http://www.crup.com.cn>

<http://www.ttrnet.com> (人大教研网)

经 销 新华书店

印 刷 北京市鑫霸印务有限公司

规 格 185 mm×260 mm 16 开本

印 张 25.75 插页 1

字 数 570 000

邮政编码 100080

010-62511770 (质管部)

010-62514148 (门市部)

010-62515275 (盗版举报)

版 次 2013 年 9 月第 1 版

2018 年 1 月第 2 版

印 次 2018 年 1 月第 1 次印刷

定 价 55.00 元

版权所有 侵权必究

印装差错 负责调换

再版前言

这本《社会统计学》教材出版后，我们向使用教材的部分教师、学生征询了他们对教材的感受和意见，也收到了多位读者的来信。总的看来，读者的感受与我们编写教材的初衷是一致的。通过淡化数理推导过程，加深统计思想的理解，强化统计工具的应用，达到了让文科学生能够学会统计学、学好统计学的目的。2013年9月面世的教材，2016年即已售罄。鉴于第一版教材中存在一些计算错误、印刷错误，这里予以纠正。也感谢多位读者帮我们发现了这些问题，并来信一一指出。

为使教材能够紧密联系现实，教材中取自统计年鉴的数据、依据数据绘制的图表均已更新，例题、习题部分也做了必要的更改。

本教材的第1章至第15章主要由尹海洁编写，第16章、全部习题、附表主要由李树林编写。

尹海洁

第一版前言

在二十多年的教学生涯中，笔者一直致力于把学生培养成为社会统计的高手。然而，对于社会学专业的学生而言，这是一门颇有难度的课程。需要找到一种途径，让学生既能够很好地理解统计原理，正确地使用统计分析方法，又不被那些既严谨又极晦涩的数学公式和数理逻辑的推演绊倒。在这个过程中笔者悟出了两个要点：一是要在教学中突出社会统计学的应用性特征。社会统计学是一种分析工具。作为应用者，最重要的是能够正确使用统计方法来分析问题。过于数学化或数理统计化的内容会淡化它的应用特征，所以不应要求应用者一定要搞清数学工具是如何制造出来的，那是数学家的工作。二是要透过数学公式，挖掘构建数学工具的统计思想。任何一个统计公式不过是一种统计思想实现的数学手段而已，而统计思想都是具体、明确且容易理解的。笔者以这两个要点作为指导思想，对教学内容的选取、各章节之间的逻辑安排都做了较大的调整，还从自己的研究中提取了具有实际应用背景的例子。经过多年的探索和积累，取得了一定的经验，可以使“社会统计学”成为学生乐于学，也能学得好的一门课程。

笔者将自己多年的教案进行整理，于2006年着手编写本教材。撰写一本严谨的教材需要花费很多的精力与时间，幸好有哈工大经济系教授经济统计课程的李树林老师和笔者共同完成这项工作。我们详细地讨论了每一章的内容，甚至对每一个符号、每一个角标的使用都进行了仔细的斟酌。初稿完成于2008年。在后来的教学过程中，每次发现问题，我们都会对书稿进行修改。到2013年，书稿已经增删五次。集笔者多年教学经验与对统计学教学的研究于一体，本书至少有以下特点。

(1) 以类别、序列和尺度三个测量层次的变量为维度安排教材内容。在统计学中，变量的测量层次不同，相应的分析方法也不同。与其他应用统计学相比，社会统计学最突出的特点是以变量的测量层次为维度来组织内容。本书彰显了这一优点。一般认为，变量分为定类、定序、定距、定比四个层次。但考虑到定距变量和定比变量在统计方法的应用上差异不大，统计软件中给出的是类别、序列和尺度三种变量，我们也以这三种变量为维度来展开内容，以与统计软件保持一致。

(2) 阐明每一种分析方法的统计思想。学生之所以感到统计学难学，是因为对以

数理逻辑为基础的数学概念、公式推导接受起来有一定的困难。实际上,统计分析方法是在应用中产生和构建的,经过高度抽象以后才成为数学模型。只要把高度抽象的数学语言转换成为具体的、清晰的统计思想,学生接受起来是没有太大障碍的。为使社会学专业的学生能够很好地理解统计学原理,我们在讲解每一种方法之前都先阐明该方法的统计思想,然后说明其数学实现过程,进而用实例讲解其应用特点。由于统计思想很容易理解,因此在此基础上学习统计方法相对比较容易。

(3) 梳理和构建了描述统计的知识体系。应用统计根据数据是否来源于随机样本可以分为描述统计和推论统计两大部分。描述统计要说明的是变量的分布特征或者变量之间的关系。以往的统计学教材中只介绍单变量的描述统计。实际上,对多变量之间关系的分析也属于描述统计。我们从单变量分布的描述与两个变量关系的描述两个维度,运用图表描述与特征值描述两种方法归纳和梳理了描述统计的内容,使描述统计成为既相对独立又较为完整的知识体系。

(4) 将描述统计与推论统计分开讲解。以往的统计学教材在进入对两个变量之间关系的分析阶段后,都不再区分描述统计与推论统计。这样容易让学生产生误解,认为只要分析变量之间的关系就必须根据显著性水平来得出结论。这实际上是对推论统计的误用。对总体数据或非随机样本数据,只应用描述统计即可。只有对随机样本数据才应用推论统计的方法来推论总体。将这两部分分开讲解能够防止学生对推论统计的误用。

(5) 做到了与社会统计软件的对接。目前,大多数社会学和社会工作专业都开设了有关 SPSS 统计软件的课程。本书编写中所使用的专业术语尽可能与统计软件相一致。教材内容也与统计软件中统计分析部分的内容相呼应。读完本书,再学习统计软件应用课程,可使学生将统计理论与统计软件的应用很好地结合起来。

(6) 增加了非参数检验的内容。非参数检验对总体的分布没有任何要求,因此它比参数检验有更广泛的应用,尤其是在总体分布不满足参数检验的要求时,只能应用非参数检验的方法。但在社会统计学的应用中,非参数检验方法没有得到应有的重视,究其原因,是在教学中重视得不够。本书对常用的非参数检验方法均有较为详细的介绍,弥补了以往社会统计学教材在这方面的遗漏。

(7) 增加了时间序列分析。目前统计学教材介绍的都是由一次调查所获取的截面数据的分析方法,对描述社会现象动态变化的时间序列数据的分析方法鲜有介绍。在统计年鉴以及各种专门的年鉴中都有大量的时间序列数据,对这些数据进行分析也是社会学定量研究的重要内容。时间序列分析方法与截面数据的分析方法差异较大。本书引入了这部分内容,有利于学生掌握利用时间序列数据分析事物发展变化规律的方法。

本书在介绍推论统计之前,选取了概率论与数理统计中与推论统计关系最密切的内容,编撰为第6章和第7章。这部分内容是推论统计的理论基础。我们也着重从社会学研究应用的角度来阐述概率论与数理统计的内容。有了这部分内容做铺垫,即便是没有学过概率论与数理统计的学生,也能够较好地掌握推论统计的内容。

在编写本书的过程中,我们既参考了国内应用较多的社会统计学教材,也参考了



经济统计的教材,还参考了国外统计学教材。尤其是非参数检验部分,因为国内教材介绍得较少,所以主要参考了国外的教材。完成本书的教学内容大约需要80个学时,教师可以根据学生的基础和教学安排进行合理的取舍。

尽管已经过五次修改,但对目前的书稿,笔者仍不甚满意。一本好的教材不仅要融入教师多年的教学经验以及作者对内容的深刻理解和体会,还要能经得起教学的检验。笔者认为本书在这些方面还有完善的空间,将在以后的教学中不断发现和改进教材中的不足。笔者怀着忐忑的心情将本书提供给同行学者和广大学生,希望能对提高文科学生统计分析的水平和能力有所帮助,也希望使用本书的教师能够不吝赐教,就书中存在的问题与我们联系。Email: yhjje@hit.edu.cn.

尹海洁

目 录

第 1 章 导 论	1
第 1 节 社会学研究的过程及统计学的应用	1
第 2 节 统计分析方法在社会学研究中的作用	4
第 3 节 抽样方法与统计分析方法的选择	7
第 4 节 变量的层次及统计分析方法的选择	9
习 题	11

上篇 描述统计

第 2 章 单变量的描述统计分析	15
第 1 节 单变量的分布及其描述方法	15
第 2 节 集中趋势	25
第 3 节 离散趋势	30
习 题	36
第 3 章 两个类别变量关系的描述统计	39
第 1 节 列联表分析	39
第 2 节 分类图	44
第 3 节 列联相关系数	46
第 4 节 等级相关系数	53
习 题	59
第 4 章 两个尺度变量关系的描述统计	62
第 1 节 相关分析	62
第 2 节 回归分析	67
习 题	75

第5章 类别变量与尺度变量关系的描述统计	77
第1节 平均值比较分析	77
第2节 相关比率	79
习 题	82

中篇 概率论基础

第6章 概率与随机变量的概率分布	87
第1节 概率及其计算	87
第2节 随机变量的描述统计	95
第3节 几种常用离散型随机变量分布的特征	106
第4节 几种常用连续型随机变量分布的特征	111
习 题	129
第7章 大数定律、中心极限定理与抽样分布	131
第1节 大数定律	131
第2节 中心极限定理	134
第3节 抽样分布	147
习 题	155

下篇 推论统计

第8章 参数估计	159
第1节 总体特征值的点估计	159
第2节 总体特征值的区间估计	164
习 题	173
第9章 假设检验的基本原理	176
第1节 假设检验的原理	176
第2节 假设检验的基本方法	178
第3节 假设检验的类型	184
第4节 假设检验的两类错误	187
习 题	194
第10章 总体均值与方差的假设检验	195
第1节 单个总体的检验	195
第2节 两个总体的检验	204
习 题	217

第 11 章 两个类别变量关系的假设检验	220
第 1 节 χ^2 检验	220
第 2 节 等级相关系数的检验	227
习 题	233
第 12 章 两个尺度变量关系的假设检验	234
第 1 节 一元线性回归方程的检验	234
第 2 节 相关系数的检验	244
习 题	245
第 13 章 类别变量与尺度变量关系的假设检验——一元方差分析	249
第 1 节 一元方差分析的原理	249
第 2 节 方差分析方法	251
习 题	255
第 14 章 非参数检验	258
第 1 节 单个总体分布特征的检验	258
第 2 节 两个总体分布一致性的检验——两个独立样本的检验	266
第 3 节 两个总体分布一致性的检验——两个相关样本的检验	278
第 4 节 多个总体分布一致性的检验——多个独立样本的检验	287
第 5 节 多个总体分布一致性的检验——多个相关样本的检验	291
习 题	294
第 15 章 抽 样	299
第 1 节 误差与抽样误差	299
第 2 节 样本容量的确定	312
习 题	316
第 16 章 时间序列分析	318
第 1 节 时间序列概述	318
第 2 节 时间序列的描述性分析	321
第 3 节 时间序列的预测方法	326
习 题	356
附表 1 标准正态分布表	360
附表 2 标准正态分布分位数表	362
附表 3 χ^2 分布分位数表	365
附表 4 t 分布分位数表	367

附表 5 F 分布分位数表	369
附表 6 单个样本 K-S 检验统计量分布的临界值表	397
附表 7 秩和检验表	398
附表 8 游程检验 ($n_1 \geq n_2$)	399
附表 9 符号秩检验表	401
参考文献	402

第1节 社会学研究的过程及统计学的应用

社会学是一门研究社会现象、社会问题发生发展以及变化规律的科学。说它是一门科学，是因为它大量使用实证的方法进行研究。也就是说，社会学这门学科的科学性不是体现在它的研究对象或研究内容上，而是体现在研究方法上。在现代社会学研究中，统计调查、定量分析是应用最为广泛的研究方法。社会统计学也是社会学定量研究中的支柱性技术。实际上，统计学不是仅仅作为一门技术，在收集完资料，进行统计分析时才进入社会学研究过程的，而是作为一种思想贯穿社会学研究的始终。我们可以通过对社会学研究过程的阐释来明确统计学在这个过程中所起的重要作用。

一、确定研究课题

社会学研究的大部分课题都来源于社会现实，有很多是需要解决的实际问题。因此社会学的研究具有很强的时代感，其研究结果可以直接为社会服务。研究者应本着重要性、创造性和可行性的原则来选择研究课题。但是，如何选题，选择什么样的课题，与研究者的学科基础和方法论倾向有很大的关系。不同的问题可能就界定了不同的方法。当然，同一问题也可以使用不同的方法进行研究。一般来说，研究课题确定之初，研究类型已被基本界定。如果是定量研究，就意味着必须用统计学的方法对数据资料进行分析；即使是定性研究，也可能涉及用量化的方法来描述事物。例如，小群体研究属于定性研究，可以使用参与研究的方法，但对群体成员之间的关系网络特征却可以采用定量的方法来描述。对于一个既可以进行定性研究又可以进行定量研究的课题，具有良好统计学基础的学者更倾向于选择定量研究的方法。这是因为统计调查的大样本对总体具有很好的代表性，而且用数据语言来描述事物具有更高的精确

性，假设检验的研究逻辑也使得研究结论更具有可信性。也就是说，统计分析作为一种思想，在研究者选题的过程中起到一定的作用。

二、探索性调查

研究课题确定以后，通常要进行文献检索，阅读大量文献的目的是了解该课题所属领域中相关的研究成果、研究的最新动态、最前沿的问题和研究的最高水平，以使自己的研究在前人研究的基础上有所创新。接下来应该进行探索性研究，目的是了解研究对象，并对所研究问题的状况、特点、存在的原因、可能的发展趋势进行初步的分析，寻求研究的切入点、重点，确定研究方法。探索性研究大多采用参与观察、非结构性访谈的方式收集资料，属于定性研究。但这种定性研究的目的却是为理论假设的提出和概念的操作化打下基础，同时也为大范围的统计调查打下基础。

三、理论假设的提出与概念的操作化

社会问题的研究大体上可以分为两类：一类是描述性研究；一类是解释性研究。描述性研究是对事物状况或水平的研究，解释性研究则是对事物之间相互关系、社会问题存在以及发展变化原因的研究。解释性研究需要提出理论假设。理论假设是对所研究的问题做出的尝试性解答，用来指导研究设计并确定研究方向。但这种尝试性的解答并不是盲目提出的，而是在研究者前期研究的基础上做出的，或是基于通过探索性研究发现的事物所具有的某些人们尚不知晓的特征或事物之间的某些关系提出的。但探索性研究是特定的小规模、小范围研究，在这样的研究中发现的一些特征或事物之间的关系在大的总体中是否存在还需要进一步的检验，因此这些发现只能作为理论假设提出，并通过后续的、大规模的调查予以检验。

不论是描述性研究还是解释性研究都会涉及一些抽象概念，比如社会地位、学术水平、社会声望，等等。这些抽象概念的实际含义虽然都能够为人们所理解，但在实证研究中却很难对它们进行具体的描述。这就需要将抽象概念具体化，这个过程就是操作化的过程。操作化就是将抽象概念与具体的社会现象相联系，将抽象概念转化为具体可测量的变量或指标的过程。描述性研究就是对这些操作化的结果进行定量的描述和分析，而解释性研究所涉及的至少是两个抽象概念之间关系的命题。将抽象概念操作化以后，还要根据理论假设对变量之间的关系提出具体的研究假设，这些研究假设可以转化为具体的统计假设，从而可以用统计调查的结果加以检验。可以说，抽象概念的操作化是定量分析的基础和前提。

四、测量工具的设计

社会测量是社会现象数量化的过程。社会现象本身并不是以数量的形式存在的。事物的数量特征是人类为了精确地认识事物，使用一定的测量工具对事物进行测量，

并把测量结果作为一种特征赋予客观事物而形成的。要想用数学的方法对事物进行精确的定量分析,首先要将社会现象数量化,这就是社会测量的过程。社会测量的对象必须是具体的事物,而不是抽象的概念。也就是说,社会测量只有在操作化的基础上才能进行。通过社会测量,为原本没有数量表现的社会现象赋予数量化的特征,将操作化的概念转化成具有不同取值的变量,这使得对社会现象和社会问题进行定量分析成为可能。社会测量除了要明确测量对象之外,还要有测量工具。对自然现象的测量工具都是实物性的,如天平、米尺、仪表等。由于社会测量的水平很低,还没有制造出仪器或仪表类的工具,而且测量内容有很多是非物质形态的,如人的思想、态度,因此,目前社会测量还主要依赖量表、问卷等工具对社会现象进行比较粗略的量化。社会测量是定量研究的前提。

在统计调查中,测量工具也就是资料收集的工具。使用什么样的工具来收集资料是由资料收集方法决定的。如果采用自填式问卷或结构性访谈的方法收集资料,问卷或结构性访谈提纲就是收集资料的工具。如果采用结构式观察的方法收集资料,观察卡片就是收集资料的工具。这些测量工具中的每一个问题就是一个变量,而问题的答案便是变量的取值。一个个体对某个问题的回答或选择便是对这个个体的测量结果。有些事物比较简单,只用一个问题便可以获得测量结果,这就是单变量问题。有些事物比较复杂,需要从不同角度使用多个问题才能够进行很好的描述,这就是多变量问题。如果描述一个事物的多个问题回答方式相同,回答结果可以汇总,则这样一组问题可以被视为量表。用量表测量事物比用单变量测量事物更准确、更细致。

五、调查的实施

调查的实施过程就是使用测量工具对被测事物进行测量从而获取数据资料的过程。这个过程在社会学研究中表现为由调查对象来填答问卷,或者由访问员将询问的结果记录在问卷上。调查实施过程中的一个重要问题就是确保资料的准确性。这对调查员的责任心有着极高的要求,同时要求有完善的检查督导体系。如果调查对象填答不认真,或者调查员不负责任,甚至有极个别的作假,将会造成很大的工作误差,甚至制造出垃圾数据。这不仅会使后续的统计工作失去意义,而且有可能得出错误的结论。

六、审核、录入与录入误差的消除

问卷回收以后,研究人员必须逐份审核,目的在于发现填答中的问题。对于有问题的问卷,能回访的回访,不能回访的作为无效问卷处理。审核是一项极其枯燥乏味的工作,却是保证数据质量必不可少的重要环节。审核合格的问卷可交由录入人员进行计算机录入。数据录入阶段也是容易产生误差的阶段。耐心细致的人,录入误差会小一些;粗心大意的人,录入误差会大一些。人的身体状况不好、心情不好等都会造成较大的录入误差。平均来说,录入误差的比例在5%左右。在进行统计分析之前,应设法将录入误差消除。传统的方法是校对。但是在计算机数据文件的单元格与问卷的答案之间进行校



对是一项耗时的工作，甚至比录入花费的时间还要多，而且一次校对以后还剩下多少错误数据也是不得而知的。比较好的方法是并行录入对比纠错。即将原始数据按照相同的格式和结构录入两遍，然后将对应的变量和单元格内容进行对比。如果对应数据相等，则认为录入正确；如果对应数据不等，则至少有一个人的录入是错误的。具体操作方法见尹海洁、刘耳编著的《社会统计软件 SPSS 15.0 FOR WINDOWS 简明教程》（社会科学文献出版社，2008）。

七、统计分析与理论假设的检验

数据录入计算机以后便可进入统计分析阶段。如果是描述性研究，则要对关键变量的分布情况进行统计分析，用定量的方法准确地说明事物的特征或状况。如果是解释性研究，则要根据理论假设的内容，对变量之间的关系进行分析，并比较分析结果与理论假设是否一致。如果定量分析结果与理论假设一致，理论假设被证实，则说明理论假设是正确的；如果定量分析结果与理论假设相反，理论假设被证伪，则说明理论假设是错误的；如果理论假设被部分地证实，则需要通过修正理论假设并结合调查结果得出结论。

从上述对社会调查过程与统计分析的关系的阐释中可以发现，统计分析方法是在调查资料收集以后的数据分析阶段才得到直接应用的，但这并不意味着社会学研究者可以不懂统计分析，只需借助统计学家的工作便可完成社会学研究。实际上，统计分析作为一种思想贯穿社会学研究的始终。统计学思想从研究课题的选择及研究类型的确定阶段就进入了研究过程。一定的研究课题，可以进行定性研究，也可以进行定量研究。选择哪种研究类型，取决于研究者的方法论倾向和统计学基础。只有具备良好统计学基础的学者才能够很好地运用统计学方法来进行社会学研究。统计学思想对研究设计也具有一定的指导作用。在理论假设的提出及操作化过程中，研究者必须考虑理论假设可以衍生出什么样的统计假设，用什么样的统计结果来检验假设，将社会现象转化成什么层次的变量可以满足假设检验的要求，以及对于一定的测量结果采用什么样的统计方法进行分析，等等。只有在统计学思想指导下进行的研究设计才能满足后续的使用统计分析方法的要求。因此统计分析绝不是一项阶段性的工作，而是作为一种思想，贯穿社会学研究的始终，决定着社会学研究的类型和特征。

第2节 统计分析方法在社会学研究中的作用

一、统计分析方法应用水平是社会学研究科学性的重要标志

数学方法因其描述事物的精确性、分析事物之间关系的准确性而在众多的学科中得到广泛的应用。在自然科学领域几乎找不到不使用数学方法的学科。马克思对数学

方法的应用给予了高度的重视。保尔·拉法格在《回忆马克思恩格斯》中谈到,马克思认为:“一种科学只有在成功地运用数学时,才算达到了真正完善的地步。”^① 由于社会科学大多是从早期的社会哲学中分化出来的,很多学科的研究方法以思辨为主,加上社会科学研究对象的特殊性,使得其对数学方法的应用比自然科学落后很多。在社会科学中,有一批学科率先使用了数学方法,从而使其研究结论的科学性得到了普遍认可,社会学便是其中之一。社会学的实证研究从学科创始阶段就与统计学建立了密不可分的联系,迪尔凯姆的《自杀论》便是用统计分析的方法进行理论提升的典范。社会学家用统计的方法对社会现象进行定量描述,对事物之间的关系进行定量分析,是社会学研究从思辨走向实证的一个突出特点。尤其是随机抽样方法的应用,使得通过一个一定容量的样本来研究一个庞大的总体成为可能,而这种可能性存在的依据便是数理统计的抽样理论。统计分析方法已经成为社会学家认识和分析社会现象的最有效手段之一。由于统计分析要进行大量的、繁杂的计算,因此一些分析方法的应用受到了很大的限制。计算机技术的发展和普及很好地解决了这一问题,优秀的统计软件可以帮助我们在瞬间完成大量复杂的计算。可以说,统计软件的出现大大促进了统计学方法的应用,也将社会学研究的科学性提高到了一个新的阶段。

二、统计分析方法应用的目的是发现和描述社会现象的统计规律

(一) 社会调查资料的特点

在社会学研究中,调查单位和分析单位往往都是比较小的个体,如个人、家庭等。社会测量的目的是将这些个体的特征转化为数量,调查过程也就是向这些个体收集资料的过程。通过调查收集得到的资料有以下两个特征。

1. 随机性

客观现象可以分为确定性现象和非确定性现象。如在一个大气压下水到零度必然结冰,这是确定性现象。但哪一天刮风、哪一天下雨就是非确定性现象。如果我们研究的现象在一定条件下有多种状态,出现哪一种状态是不确定的,这样的现象就是非确定性现象。非确定性现象也称随机现象。社会现象大多是非确定性现象。如果我们要研究两个现象之间的关系,这种关系也表现为确定性与非确定性。在真空中下落的物体,其下落速度与下落时间的关系是确定性关系,只要下落时间确定了,对应的下落速度也就随之确定了。而身高与体重的关系就是非确定性关系。虽然高个子的平均体重会大于矮个子,但身高相同的人体重却不同。非确定性现象和非确定性关系的存在,使我们不能像预言水到零度必然结冰那样去预言人到一定年龄必然结婚,同样,也不能像通过分析一滴水的成分而知道所有水的成分那样,通过抽查一个人的情况而知道所有人的情况。

由于我们在研究影响现象的某一个或某几个因素时,对该现象起作用的其他因素仍然存在,仍然影响着它的状态,因此在一定条件下,某事物会处于什么样的状态是

^① 拉法格等. 回忆马克思恩格斯. 北京: 人民出版社, 1973: 7.



随机的。随机性主要表现在个体上。某一个体之所以区别于其他个体，是因为其某些特征与其他个体有别。社会统计学中的变量就是指在个体上具有不同取值的特征。任何一个变量都会有多个取值，从研究者的角度看，某个个体在某个变量上的取值是不确定的，它可能是变量取值集中的任何一个，这就是数据资料的随机性，也称不确定性。

2. 统计规律性

随机现象虽然是不确定的，但并不是无法研究的，对大量的随机现象进行统计时，往往可以发现一定的规律性。对于一个孕妇来说，生男孩还是生女孩纯属偶然，但考察大量出生婴儿的性别比例，就可以发现性别比例是一个基本恒定的数，是100女：105男。^① 每个人的身高都有不同的表现，但对大量人的身高进行测量可以发现，中国乡村18岁男性的平均身高为171.0厘米，18岁女性的平均身高则为158.1厘米；中国城镇18岁男性的平均身高为171.9厘米，18岁女性的平均身高则为159.8厘米。^② 这种规律会随着观察数目的增多而凸显出来。通过对大量个体的观察所得到的这些具有稳定性的规律就是随机现象的统计规律性。正因为如此，随机现象是可以研究的，研究的目的是认识个体，而是认识总体。对社会现象进行统计分析的主要目的不是用定量的方法来描述个体，而是通过对大量个体特征的统计分析来描述和分析社会现象的统计规律。

比如，某个孕妇生男孩还是生女孩不是社会学家关注的问题，但观察大量出生婴儿的性别，获取出生婴儿的性别比例却是社会学家关注的重要问题。某个人能挣多少钱不是社会学家关注的问题，对大量人口的收入进行统计，明确人均收入、不同行业的收入、不同阶层的收入却是社会学家关注的重要问题。社会统计学就是通过大量观察记录个体特征并对其结果进行统计分析，以达到了解和把握社会整体的目的。

（二）统计学是发现和彰显统计规律的有效工具

统计规律是客观存在的事实，但它往往又隐藏在大量数据之中，要通过大量的观察、精细的统计分析才能发现。社会现象的统计规律是对资料的简化、浓缩和概括。使用社会统计学的方法可以使庞杂的资料得到简化，将大量的数据浓缩为几个特征值。研究者可以通过计算百分比、平均数、标准差等特征值对单变量进行描述，从而弄清楚变量的状态、水平和分布特征；也可以通过计算相关系数、回归系数、相关比率等描述变量之间关系的特征值对多变量进行分析，从而明确各变量之间的内在联系。也就是说，使用统计分析方法可以帮助研究者从纷繁复杂的资料中提取有用的信息，并得出相应的结论。

社会学研究常常要面临庞大的总体。除了像人口普查那样由国家组织进行的全面调查外，绝大部分社会学研究都不可能访遍或穷尽研究对象，只能利用有限的资料来推及总体。尤其是在研究总体无限或难以界定的情况下，人们根本无法对整个总体进

① 温勇，尹勤. 人口统计学. 南京：东南大学出版社，2006：21.

② 国家体育总局. 第二次国民体质监测报告. 北京：人民体育出版社，2007：131-132.