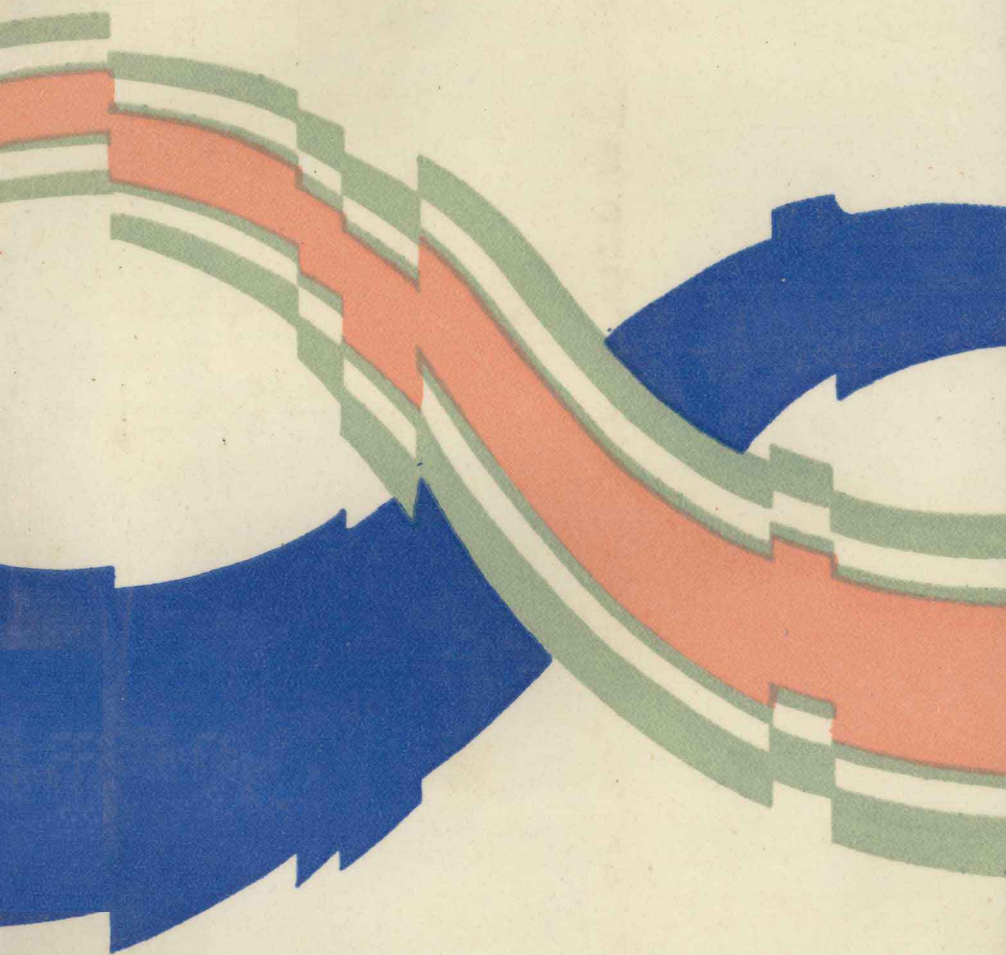


社会统计学

SHE HUI TONG JI XUE

张彦 宋林飞 徐憬 编著



南京大学出版社

社 会 统 计 学

张 彦 宋林飞 徐 慷 编著

南 京 大 学 出 版 社

1990 · 南京

社 会 统 计 学

张 彦 宋林飞 徐 慷 编著

南京大学出版社出版

(南京大学校内)

江苏省新华书店发行 江苏丹阳新华印刷厂印刷

开本: 850×1168 1/32 字数: 169 千

印张: 6.5

印数: 1—4000

1990 年 6 月第 1 版

1990 年 6 月第 1 次印刷

ISBN 7-305-00745-5/F·99

定价: 2.95 元

目 录

第一章 导 论	1
一、社会统计的发展及意义	1
二、社会统计的对象与作用	3
三、社会统计的方法	5
四、社会统计中几个基本概念	7
五、社会统计数字	11
六、社会统计工作的程序	12
第二章 社会统计资料的搜集	14
一、社会统计资料的分类	14
二、搜集资料的方法	15
三、统计调查的组织形式	15
四、统计调查的误差	18
第三章 社会统计资料的整理	20
一、分组原则与标准	20
二、列表法	22
三、图示法	28
四、统计汇总	32
五、统计表	34
第四章 集中趋势的量度：平均指标	41
一、算术平均数	41
二、中位数	49
三、众数、几何平均数、调和平均数	53
四、各种平均数的关系	57
第五章 离中趋势的量度：变异指标	59
一、全距	60
二、平均差	61

三、标准差	63
四、均互差	69
五、离散系数	70
第六章 相对指标	72
一、相对数的种类	72
二、结构相对数	75
三、比较相对数	77
四、强度相对数	79
第七章 时间数列	81
一、时间数列的概念和种类	81
二、时间数列的分析指标	84
三、时间数列变动的分析	91
第八章 统计指数	103
一、统计指数的概念和种类	103
二、总指数的计算方法	104
三、指数体系与因素分析	111
四、平均指标的因素分析	115
第九章 相关与回归分析	119
一、相关分析的概念与意义	119
二、回归分析	123
三、相关分析	132
第十章 正态分布	137
一、正态曲线的一般形式	138
二、正态曲线下的面积	140
三、正态分布表的使用	143
第十一章 概率论	146
一、概率的基本意义	146
二、概率的数学性质	148
三、排列组合	154
四、独立性与随机抽样	158
第十二章 假设检验：二项分布	160
一、二项分布	161

二、统计检验的步骤	163
三、例题与应用	168
第十三章 单样本检验：中心极限定理	172
一、抽样分布	172
二、中心极限定理	174
三、含均值和成数的单样本检验	181
第十四章 参数估计	188
一、点估计	188
二、区间估计	190
三、其他类型的置信区间	196
四、抽样误差	197
五、确定样本的容量	199

第一章 导 论

一、社会统计的发展及意义

社会统计的实践已有几千年的历史，是最早产生的统计。中国大约在公元前一千多年的殷代、西周，就有人口登记的记载。东周及其以后的各个朝代，关于土地、人口的数字不绝于史册。古希腊据说在公元前四百年就进行过人口普查；古罗马据说在公元前四百年就建立了出生、死亡登记制度。从奴隶社会直到封建社会这段漫长的时期，社会统计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，这是应奴隶主或封建王朝实现征税或服兵役和服劳役的需要而进行的。自 17 世纪后，随着资本主义生产关系的产生和发展，简单的人口、土地、资源统计已不能满足社会发展的需要，因而产生了以工业、农业、贸易、交通等方面统计为主的社会经济统计。与此同时，属于对社会总体观察的社会统计内容也得到相应提高。如 1662 年格朗特发表了第一本关于人口统计的著作《自然和社会观察》，对当时伦敦人口的出生率、死亡率和性比例等进行了综合的社会统计。在不断丰富的统计实践经验的基础上，比较系统的统计理论知识逐步形成。

回溯统计学的发展史，早期主要有三大学派：一为记述学派；二为政治算术学派；三为数理学派。数理学派在统计上的历史贡献，是把概率论正式地引进统计学，从而开创了统计学的新阶段。统计史学家一般把 17 世纪中叶出现的初始的统计学称为古典统

计学，把引进概率论之后的统计学称为近代统计学。以凯特勒为代表的近代统计学是可以通用于社会现象和自然现象的方法论科学。

社会统计学融合了比较记述和政治算术的观点，又继承和发扬了凯特勒强调的研究社会现象的传统，并把政府统计与社会调查融合起来，进而形成社会统计学。社会统计学在 20 世纪虽然有一个从实质论向方法论转变的趋势，但仍强调以事物的质为前提和认识质的必要性。

二次大战后，世界的政治发生了重大变化，科学技术迅速进步，经济发展成为全人类面临的共同问题。但在注重经济增长的同时，在工业发达国家里，出现了一系列难以解决的社会问题，诸如环境污染、犯罪率和离婚率上升、失业和贫困加剧等等。在发展中国家里，经济的增长反而带来政治的不稳定性，社会动荡和国内贫富差距拉大。这一切使人们逐渐认识到，经济增长并不意味着社会发展，经济高涨不等于社会进步，经济效益好不等于社会效益佳，因而感到经济指标并不具备反映社会生活状况全貌的能力。这就要求从全社会发展的观点出发，应用社会调查和各种统计方法，收集大量的、更全面的事实来描述、分析、研究社会发展状况和发展趋势，从而监测社会发展和采取相应的措施，达到统计为社会服务的目的。

本世纪 60 年代以来，西方发达资本主义国家先后都制订与实施了社会发展计划，法国较早强调计划的社会方面。在 1962—1965 年的战后第四个计划中，法国政府把计划的名称改为“经济与社会发展计划”，原来计划的名称是“经济现代化与投资计划”。法国的这一“经济与社会发展计划”规定，要关心经济增长成果的分配和区域发展，在计划中和很高的总增长率（每年 5.5%）并列的，还有社会投资（住房、教育和卫生）每年增长 10% 的计划目标。其它资本主义发达国家在计划化的过程中，也越来越重视计划的社会方面。垄断资本主义国家制订与实施社会发展计划的活动直接推

动了社会统计的发展。60年代初首先在美国掀起了一个颇有声势的“社会指标运动”。1966年美国的雷蒙·布埃尔出版了《社会指标》一书。70年代,一些区域性或世界性的组织开始颁发统一使用的社会统计指标体系。1976年经互会为各成员国颁发了《社会统计基本指标体系》。同年,联合国经济合作与发展组织(OECD)编制了《社会生活质量的计量》的社会统计指标体系。

从50年代到80年代初期,我国的统计工作、统计理论与方法,都侧重于经济统计方面。1982年,我国把国民经济发展计划改为国民经济与社会发展计划,从而形成了一个社会综合考察、综合治理、综合发展的计划,使统计工作更加全面、完善与深入,也更加困难了。为了适应这一状况,我国的统计工作发生了一系列变化。

首先,是统计组织机构与统计制度的变化。1982年,国家统计局成立了社会统计司,并开始建立与完善社会统计制度。过去,偏重于经济的统计工作与统计方法的发展,是为了适应恢复生产与发展社会主义计划经济的需要。现在,经济统计与社会统计同时受到重视,是为了更好地完成我国经济与社会协调发展的战略任务。

其次,是统计指标与制度化报表的变化。1983年7月全国统计工作会议审查通过《社会统计指标体系》(草案)。从此,反映社会统计指标体系的各种月、季度、年度报表,进入了各行各业,成为各级政府了解情况、检查工作与制定规划所不可缺少的工具。1985年,由国家统计局社会统计司编写的《中国社会统计资料》首次公开出版。

二、社会统计的对象与作用

社会统计学是搜集与分析社会现象的数量特征的一门科学。社会统计学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。它是现代社会中社会学、政治学、历史学、法学等社会科学进行具体社会研究时的基本手段。自从我国于 80 年代初期恢复社会学这门学科以来,各个社会学系(专业)为建构社会统计学理论与方法,形成具有中国特色的社会统计学教材,做了不少有益的工作。目前,社会统计学是社会学专业学生的必修课程。同时,其他一些社会科学学科也开始把这门课程列入教学计划。

社会统计学的研究对象,是社会生活及其条件、社会关系及其结构状况的数量特征与数量关系,以及隐藏在这些数量特征与数量关系中的社会发展变化的客观规律。

社会统计学与社会经济统计学既有联系又有区别。社会统计在研究社会生活领域内的问题时必然涉及到物质生产领域。生活水平的高低,社会需求是否得到满足,社会发展的快慢等等都必须应用经济指标来观察和衡量。但从社会总体来看,经济只是社会的一部分,单纯用经济指标不能全面反映社会的发展状况,需要建立包括一部分经济指标在内的社会指标体系来对社会状况作系统的、全面的综合描述和反映。此外,在经济统计中,大多数现象可以直接计量,具体化为各种指标,而在社会统计中,有些社会现象则不可能直接计量,只能用间接性指标加以反映。

社会统计和数理统计是两门既有区别又有联系的不同性质的科学。数理统计以概率论为基础研究随机现象数量关系的规律性。概率论和数理统计学中关于大数定律的数学定律,对必然性与偶然性的关系作了科学的表述,适用于研究一切随机现象的数量关系。社会现象的数量特征具有高度的复杂性和不可重复性,要从数量方面描述社会状况、社会发展和社会问题,应用数理统计作为描述工具是很自然的,这将大大扩展社会统计的研究领域。社会现象是人类有意识参与的结果,这一点只是改变概率的应用条件,并不改变社会现象的随机性质。在社会现象中,即使是相同的意识作用也完全有可能不确定的结果,这就提供了概率论应用的可

能性。而事实上，概率论的实际应用是从人口调查和社会保险等事业上先开始，然后才推广到生物、物理等自然现象中去的。

社会统计是了解社会状况、监测社会动态、制订社会改革与社会发展规划的一个基本工具。社会统计从其对象出发的定量描述，往往准确地测定与显示社会生活全貌或某一方面的数量特征，从而为各级政府与实际工作部门的决策和日常管理活动提供必要的数 据。俗话说：“情况明，决心大”。数据就是情况，没有数据情况就无法明了。社会统计是有计划地建设社会主义社会的过程中不可缺少的一项工作。

社会统计是各门社会科学提高其科学性与加强应用的一个重要基础。现在，大家都对社会科学不太满意。社会科学首先是一种科学，科学就是正确地认识客观对象。以往，我们社会科学工作者在科学的定量研究方面，成果还不甚显著。这里有主观与客观两方面的原因。如今，随着社会统计实践的发展，积累起越来越多的社会统计资料。这势必为各门社会科学的发展奠定扎实的基础。同时，也为社会科学更好地为社会主义建设事业服务而提供了条件。

社会统计是统计工作的一个重要组成部分。过去，我国各级政府综合统计和劳动、文教、政法等主管部门的统计机构与统计人员在社会统计方面已经做了不少有益的工作。近几年来，统计部门的实际工作者与社会学研究者共同努力，建立起越来越多的社会指标，搜集了大量社会统计资料。这也必然将推动整个统计工作与统计学的发展。

三、社会统计的方法

第一是描述方法，即搜集与显示社会情况的方法。第二是归纳方法，即根据总体中抽取的样本对总体进行概括推论，根据反复调查研究找出带有普遍性的社会现象规律。早期统计均属描述性统

计。所谓描述性统计就是讨论范围仅以搜集资料本身为限，而不予以扩大。描述性统计是统计学的基础。人们在对社会进行研究的过程中常常遇到这种情况：由于资料过多而无法恰当地吸收所需的信息内容，有必要将这些资料压缩整理，使研究者从中提炼出最有用的信息。于是学习统计学，首先就需要掌握处理统计总体的一系列计量、分析的基本方法，经过分组并计算一些统计指标，如平均数、标准差及相关系数等，就可以掌握压缩资料和提炼信息的基本方法。需要指出的是，当调查者需要处理两个以上变量的相互关系时，描述性统计往往特别有用。

所谓归纳性统计就是依据有限资料对于总体性质作归纳或推论，从而使统计的功能大为扩充。由于这种统计往往不以归纳推理为限，也运用演绎过程进行推估，所以也称统计推论。统计推论需要的统计资料相对而言较少，所以在人力、物力及时间等方面比较经济；研究和应用的范围也拓宽了许多。本书较深的内容是社会统计推论，通过学习概率论，我们就有了一门纯演绎的数学分支来为归纳推理提供理论根据。

大量观察法、分组法、综合指标法、平均法、动态法、指数法、相关法等奠定了社会统计学的科学基础。以此为工具，社会统计的任务是结合我国的实际情况，通过多种方法调查和收集社会统计资料，全面、客观、深入地反映我国建国以来的社会条件、社会结构、社会关系和人民物质文化生活的各个方面。同时，把社会调查方法和社会学中的定量与定性相结合，动态与静态相结合，单因素与多因素相结合的理论分析不断提高到新水平。

社会统计科学要透过大量的社会现象，通过数量分析来认识社会的现状及其发展状况，就必须以马克思主义方法论为指导。例如，社会统计学必须把社会作为一个整体，从整体出发去研究和认识社会诸方面的问题，这就是一条重要马克思主义基本原理。根据事物的质和量相互联系、相互制约的辩证关系，运用统计方法来研究社会现象时，必须在确定社会现象质的规定性基础上，研究其

量的差别。从统计学的角度来讲，就是透过统计数字来看事物的本质和规律。根据对立统一规律以及从发展中研究事物的原理，就要运用综合指标法、动态分析法、指数分析等，从事物的相互联系，相互制约中，从事物的运动中分析事物之间的关系、矛盾和发展变化。

四、社会统计中几个基本概念

统计学所研究的对象具有大量性、变异性和历史性的特点，这就决定了统计研究的方法必须采用大量观察法。所谓大量观察，即是就总体中足够多的单位进行调查并综合分析，用以反映社会总体的数量特征。

大量观察法之所以成为统计上特有的方法，是与大数规律的作用分不开的。大数规律是随机现象出现的基本规律，它的一般意义是：

观察过程中每次取得的结果不同（因为具有偶然性），但大量重复观察结果的平均值却几乎接近确定的数值。

社会现象是复杂多变而又互相联系的，其中的个体单位由于必然的、主要的因素和偶然的、次要因素交互作用，彼此间存在着不同程度的数量差异，但在综合大量社会现象的数量特征时，个体单位的偶然数量上的差异可以相互抵销，使大量社会现象的数量特征借助平均数形式，显示出必然的规律性。

举例来说，当我们观察个别家庭或少数家庭的新生儿的性别时，男女的比例可能极为参差不齐——有的生男不生女，有的生女不生男，有的是女多男少，有的是女少男多。然而经过大量观察，男婴、女婴的出生数则趋于接近，也就是说男婴或女婴的出生数各占新生儿总数的 $1/2$ 。新生儿的性别比经过大量观察趋于一特定值，这就是我们在第十章中将要详细阐述的概率的一个实例。大数规律和概率论实质上是讨论同一性质的问题，只是概率论侧重

于有概率的各种计算方法的叙述，而大数规律侧重于统计规律性的理论说明，这是我们一开始学习描述性统计时就要加以认识和理解的。同时，我们有必要掌握下面几个统计学的基本概念。

1. 总体和个体

统计是大量观察的结果。统计所研究的不是个别的数量关系而是总体的数量关系，而没有大量个体的数量特征当然就谈不上总体的数量表现。所以总体和个体是统计学的基本概念。所谓总体，就是作为统计研究对象的、由许多具有共性的单位构成的整体。显然，总体具有两个特征：大量性、同质性。构成总体的每一个基本单元称为个体，也称为(总体)单位或个案。

大量观察法是从总体中的个别单位开始的，然而把观察得来的个别数量综合起来之后，结果所表明的却是整个总体运动的规律性。这就是说，统计所能表示的总体的数量关系虽然代表着这个集合体中的每个个体，但又不同于每个个体特征。用日本统计学家横山雅男的话来讲：“一勺之水无色可辨，而大海则蔚然而紺碧；少量之气无迹可寻，而太空则穹然而青苍”。

统计学上关于总体和个体关系的观点是完全符合系统论的原理和原则的。凯特勒曾经对个体与总体的关系作过一个精辟的比喻，他说：“假定从最近距离来注视画在平面上的大圆线的一小部分，这时无论圆线是怎样用心被画出来的，但是看起来总象是很多的点，奇妙地、随便地排列着，并且好象是把它们胡乱地结合在一起似的。可是，如果把眼睛移开一些来看，那就好象多数的点，在一定幅度的弓状形上面很规律地分布着似的。如果再离远一些看，那么，各个点就从眼界中消失了，它们便合而为一，形成一个规律分明的整体，于是也就可以认识到所画圆线的性质了”。

2. 标志和变量

关于统计观察的实施，首先要识别和确定被观察现象的标志。大量观察法就是从总体中的个别单位开始，一个一个地进行观察

和登记。总体的每个单位都具有许多属性和特性。说明总体中单位的属性或数量特征的名称在统计上称为标志。如在学生总体中每个学生的性别、籍贯、年龄、文化程度、民族等特征，就是该总体的单位标志。凡能用数量的多少来表示的标志，称为数量标志，如年龄。凡不能用数量的多少来表示而只能用属性上的差别来表示的标志，称为品质标志，如性别。变量则是可变的数量。如年龄因人而异并可用数值来表示，就是变量。变量的数值大小称为变量值(亦称标志值)。

事实上，统计调查都离不开对数量标志进行观察和计量，汇集得来的某一数量标志的一系列数值，在统计上又称数据。按照数据连贯衔接情况，变量可以分为连续变量和离散变量(即不连续变量)。前者是在任意两个相邻数值之间还有多个不同的数值，后者是指数值只能按整数计，不可能出现小数。

变量按其性质又可以分为确定性变量和随机变量。所谓随机变量，在相同的条件下进行观测，其可能实现的值(或观测值)不止一个。例如学生的身高、工人的收入、城市的人口密度、某病的发病率、某地的年平均气温等都是随机变量。随机变量正是统计处理的对象。

3. 统计指标和统计指标体系

统计所关心的不是个别事物的量，而是要综合说明大量社会现象的数量关系，概括地表明其一般特征和规律性。这主要是通过统计所特有的指标来实现的。统计指标是反映总体(或样本总体)的数量特征的概念。如人口数目、耕地面积、就业率、性比例等等，这些概念用于反映一定统计总体的数量方面时，就是统计指标。统计指标都是综合性的，它是由同质的个别单位构成的集合体。在实际应用时，一个完整的统计指标总是由两部分构成(如表1.1所示)；一个质的规定性就是一个统计总体。统计指标和标志之间的主要区别在于，前者是说明整体特征的；后者是说明整体中各单位特征的；前者的数值是由后者的数值汇总而成。由于统

计指标能够综合说明总体的数量特征，所以一般又称它为综合指标。

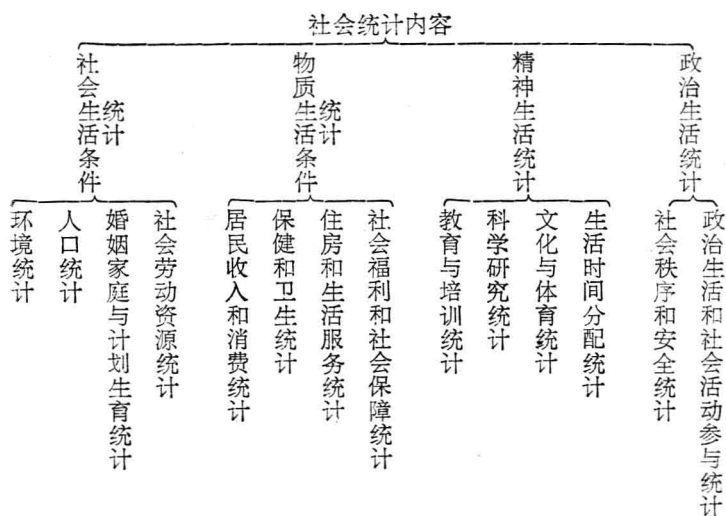
表 1.1 城 镇 待 业 人 员

年 份	城镇待业人员(万人)	待业率(%)	指标名称
1952	376.6	13.2	指标数值
1957	200.4	5.9	
1978	530.0	5.3	
1985	238.5	1.8	

资料来源：《中国社会统计资料—1987》，中国统计出版社，第 32 页。

统计指标按其数值的三种表现形式(即绝对数、相对数和平均数)，有总量指标、相对指标和平均指标(包括变异指标)之分。总量指标一般由总体单位的数目和标志值汇总而来，而相对指标或平均指标一般由总量指标派生而来，用以从统计资料提炼出更多的有用信息。根据反映总体特征的不同要求，相对指标和平均指标的形式有多种变化，所以在描述性统计中占有很大篇幅。

社会统计的观察、研究领域极其广泛，涉及到社会生活的各个方面。事实证明，要全面衡量和评价社会发展水平，仅靠经济统计分析及经济指标不行；仅靠少数社会生活指标也不行。例如，不能单纯凭按人口平均的国民收入或犯罪率指标来评价一个社会的发展状况，而要运用自然的、社会的、政治的、物质的、精神的等各个方面的指标进行综合分析和说明。一系列有内在联系的社会统计指标的集合体叫做社会指标体系。自 1983 年以来，国家统计局制定并多次修订指标项目，共计 1100 多个指标，包括我国的自然资源、环境保护、人口与家庭、劳动、居民生活、文化、教育、体育、社会秩序与安全、政治与社会参与等；内容概括如下：



从60年代起,社会指标运动逐步发展到世界规模。联合国科教文组织建议各国建立《社会 and 人口统计体系(SSDS)》,计有12个方面,200多个指标;美国社会统计有255个指标;日本有1000多个指标,其广泛性可想而知。社会指标虽然在国际和国内都得到重视和研究,但因为开发的时间较短,在社会指标的定义、对象与范围、类型和社会综合评价指标等方面,还有待继续深入研究。

五、社会统计数字

所有统计资料,都是数量性的,即使质的资料,也须数量化后才能处理。统计数据有时为精确值,如球队的队员数,教室中上课的学生数;有时为近似值,如江苏省1985年人口6213万人。就统计数据而言,大多数是近似值,这就存在着一个近似值有效数字的问题。要正确识别和书写有效数字,有下列规则必须掌握:

(1) 近似值整数末尾的零不算有效数字。例3200米,有效数字两位,准确到百位,可能误差为50米。这就是说,真值在3200