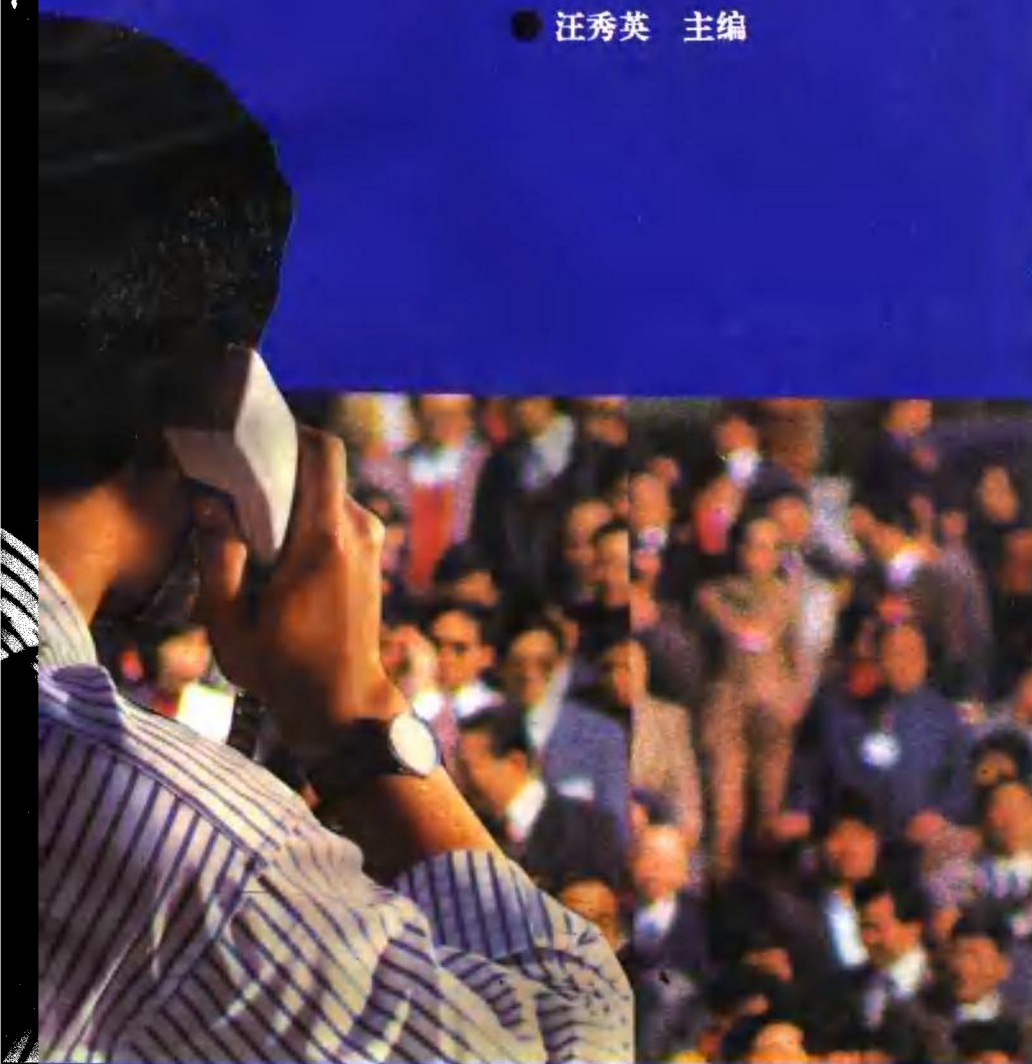


公共关系学

● 汪秀英 主编



中国商业出版社

46
C8
Z1
高等院校统计学专业规划教材

描述统计学

袁卫 吴喜之 贾俊平 编著

中国统计出版社



(京)新登字 041 号

图书在版编目 (CIP) 数据

描述统计学/袁卫等编著. —北京: 中国统计出版社, 1996. 1

高等院校统计学专业规划教材

ISBN 7-5037-1940-0

I. 描…

II. 袁…

III. 描写—统计学—高等学校—教材

IV. C8

中国统计出版社出版

(北京三里河月坛南街 38 号 100826)

新华书店经销

友谊印刷经营公司印刷

*

850×1168 毫米 32 开本 10.75 印张 27 万字

1996 年 1 月第 1 版 1996 年 1 月北京第 1 次印刷

印数: 1—5000 册

*

ISBN 7-5037-1940-0/C · 1130

定价: 10.20 元

(版权所有, 不得翻印)

出版说明

随着社会主义市场经济体系的逐步建立,我国的统计观念、统计体制、统计标准、统计制度和统计调查方法都在发生着重要的变化。为了适应建立社会主义市场经济体制和统计现代化建设的需要,全国统计教材编审委员会制定了《1991—1995年全国统计教材建设规划》,并根据《规划》的要求研究制定了统计学专业主要专业课程的教学大纲。在此基础上,编委会采取邀请、委托等方式组织全国有关院校的专家、学者编写了这批统计学“规划教材”。

这批教材力求做到紧密联系社会主义市场经济和统计现代化建设的实际,充分反映当前我国统计改革的新发展,积极吸收国内外统计科学研究的新成果。相信通过这批教材的出版,对我国的统计教学改革将起到积极的推动作用,对我国的统计教材建设也将起到较好的示范、导向作用。

限于水平和经验,这批教材的编审、出版工作,还会有缺点和不足之处,诚恳欢迎教材的使用单位、广大教师 and 同学们提出批评和建议。

全国统计教材编审委员会

1994年11月

前 言

《描述统计学》是1993年6月全国统计教材编审委员会（青岛会议）讨论确定的探索性教材。这本教材试图将原社会经济统计学原理、国外统计教材中的描述统计部分以及探索性数据分析结合在一起，使高等院校统计学专业（包括数学类的“统计与概率”专业和经济类的“统计学”专业）及其他相关专业学生在不应用概率论和其他高等数学的条件下，能够根据研究的目的，有效地获取数据并对数据进行适当的整理和显示，从而充分提取数据中的信息，利用简单直观的统计方法，描述并探索出数据内在的数量规律性，掌握统计思想，培养学生学习统计学的兴趣，为继续学习推论统计方法及应用各种统计方法解决实际问题打下必要而又坚实的基础。

本教材试图体现如下特点：首先，以各种科学研究、生产和生活中的实际数据及探索数据中数量规律性为主线并贯穿全书；第二，重点介绍各种描述性统计方法的应用条件、方法的正确使用及误用的经验和教训，以及方法背后的统计思想；第三，通过大量实例培养学生探索数据数量规律性及解决实际问题的能力；第四，充分利用简单、直观、易懂的图表，结合计算机软件的应用，培养学生进一步学习的兴趣，激励他们的求知欲；最后，力求简明易懂，基本不涉及微积分、线性代数和概率论的知识，使得具有初等数学基础的读者就能够掌握书中方法，便于自学。因而，这本书不但可作为大专院校统计课程的入门教材，也可作为广大实际工作者的参考书。

本书第一、二、三、四章由袁卫（中国人民大学）执笔并任

主编，第五、六章由吴喜之（南开大学）执笔，第七、八、九章由贾俊平（中国人民大学）执笔。在本书的编写和出版过程中，国家统计局干部培训中心的温明和李毅同志给予了大量的支持和帮助，中国人民大学统计学系和其他院校的教师对于编写大纲提出了许多宝贵意见，书中若干例题选自所列参考书目，在此一并表示感谢。由于我们的水平有限，加之该教材又是在探索之中，书中错误及不当之处在所难免，欢迎读者批评指正，以便我们再加修改。

编 者

1995年3月

目 录

第一章 绪论	1
第一节 统计数据与统计学	1
第二节 统计学的产生和发展	4
第三节 统计学的分科	6
第四节 统计学与其他学科的关系	10
第二章 统计数据的搜集与整理	12
第一节 数据的计量尺度	12
第二节 统计数据的来源	14
第三节 统计调查方法与方案	18
第四节 统计数据的质量	24
第五节 统计数据的整理	26
习 题	38
第三章 统计数据的描述	40
第一节 绝对数与相对数	40
第二节 分布集中趋势的测度	42
第三节 分布离散程度的测度	57
第四节 分布偏态与峰度的测度	64
第五节 数据的线性变换	67
习 题	69
第四章 统计数据的显示	72
第一节 统计表	72

第二节	统计图	75
第三节	统计图的误用与充分利用	90
习 题		101
第五章 探索性数据分析 (一)		106
第一节	引言	106
第二节	茎叶图	109
第三节	字母值	116
第四节	箱线图与批比较	124
第五节	数据变换	138
习 题		150
第六章 探索性数据分析 (二)		153
第一节	拟合直线的分组法	153
第二节	残差分析	159
第三节	二维数据的中位数平滑法	164
第四节	二维数据的一般处理及交互作用	170
习 题		179
第七章 时间数列分析		180
第一节	时间数列分析基础	180
第二节	长期趋势分析	192
第三节	季节变动分析	220
第四节	循环波动与不规则波动分析	229
第五节	时间数列综合分析	236
习 题		244
第八章 相关与回归		249
第一节	简单线性相关	249
第二节	一元线性回归	260
第三节	非线性回归和多元回归	267
习 题		276

第九章 指数	279
第一节 指数编制的基本问题	279
第二节 简单指数	282
第三节 加权指数	288
第四节 指数体系与经济分析	306
第五节 几种常用的重要指数	316
第六节 指数理论的发展	327
习 题	329
参考书目	332

第一章 绪 论

第一节 统计数据与统计学

在我们的日常工作与生活中，到处都有统计数据。例如，开会时会议主持人要统计一下出席会议的人数；球类比赛时解说员总要统计竞赛双方的进攻次数和成功率；学生考试后非常关心自己的考试成绩和名次；企业管理人员每天要掌握生产销售情况和利润的数字；报刊和电视中常听到国民生产总值、物价指数和证券指数的数字等等。日常工作与生活中的这些数字就是我们所关心的统计数据。

统计学是一门收集、整理、描述、显示和分析统计数据的科学，其目的是探索数据内在的数量规律性。正是因为统计学总是在和数据打交道，因而我们也可称统计学为“数据的科学”。

怎样理解统计学是数据的科学呢？我们先看看统计学的英文名词，再举几个例子。

统计学的英文是“*statistics*”。在英文字典中它有两个含义：当它以单数名词出现时，表示一门科学的名称——“统计学”；当它以复数名词出现时，表示“统计数据”或“统计资料”。*statistics* 这一英文的解释至少可以告诉我们两件事情。

第一，由于统计数据在英文中是以复数形式出现的，表明统计数据不是指个别的单个数字，而是指同类的较多数字。因为单个数字如果不和其他数据进行比较，是不能说明问题的。例如某个学生在某次考试中得了85分，如果仅凭这一个数字我们很难对这个学生的知识和能力作出判断和评价。因为这个85分可能是班

上的最高分，可能是中等水平的分数，也可能是较低的分数。如果还知道其他学生的考分或者知道这次考试的平均分数，我们就可以对这个学生的成绩做出评价了。在生产和生活实际问题里，我们总是可以收集到较多的数据，因而在现实问题中的数据都可以称为统计数据。

第二，作为单数的统计学和作为复数的统计数据在英文中都用同一名词，其间显示有密切的关系。统计学是由收集、整理、描述、显示和分析统计数据的方法组成的，这些方法来源于对统计数据的研究，目的也在于对统计数据的研究。离开了统计数据，统计方法乃至统计学就失去了它存在的意义。这正如我们俗语所说“巧媳妇难为无米之炊”。这里的“巧媳妇”就是掌握统计方法的统计学家或统计工作人员，“米”就是统计数据，“炊”就是我们统计研究或统计工作的目的，即探索数据内在的数量规律性。显然，没有统计数据或没有较好的统计数据，就是再高明的统计方法或统计学家也难有所作为。

那么，什么是数据内在的数量规律性？我们用下面几个简单的例子来说明。

【例 1.1】 每个家庭新生婴儿的性别可能是男的，也可能是女的。在过去没有实行计划生育时，有的家庭几个孩子都是男孩，也有的都是女孩。从表面上看，新生儿的性比例似乎没什么规律可循。但如果对新生儿的性别进行大量观察，即观察成千上万个或者更多，就会发现男孩稍多于女孩，大致为每生育 100 个女孩，就有 107 个男孩。这个性比例 107 : 100 就是新生儿性别的数量规律性，因为古今中外都大致相同，它是由人类社会长期遗传和发展所决定的。因为人类社会要发展，就要保持男女人数的大致相同。那么有人会问，新生儿男多于女，不是性别不平衡了吗？是的，新生儿时男多于女不平衡，但男孩的死亡率高于女孩，到了中青年时，男女人数就大致相同了。进入中老年后，男性的死亡率仍然高于女性，导致男性的平均预期寿命比女性短，老年男性

反而少于老年女性。从一个国家乃至全人类看，婴幼儿时男多于女，中青年时大致相同，老年时女多于男。在中青年结婚生育时性别大致平衡，总人口上性别也是大致平衡的，保证了人类社会的进化和发展。对人类性比例的研究是统计学的起源之一，也是统计方法探索的最早的数量规律性之一。

【例 1.2】 我们都做过掷硬币和掷骰子的游戏，都知道随机地掷一次硬币或骰子是不能事先确定出现正面反面或某个点数，也就是说个别的游戏或试验中充满了不确定性或偶然性。赌徒们正是利用了这种偶然性进行赌博。但当我们进行大量观察，即不断做重复试验时，就会发现掷一枚均匀硬币出现正面或反面次数都接近相同，即比值接近 $1/2$ 。试验的次数越多，就越接近 $1/2$ 这一稳定的数值。同样，在掷骰子时，出现 1—6 点的比率也逐渐接近 $1/6$ 。这里的 $1/2$ 和 $1/6$ 就是掷硬币和掷骰子出现某一特定结果的概率，也就是我们探索的数量规律性。

【例 1.3】 在进行农作物试验时，如果其他试验条件相对固定，我们会发现某种粮食作物的产量会随某种施肥量的增加而增加。当开始增加施肥量时，产量增加较快。以后增加同样的施肥量，粮食产量的增加量逐渐减少。当施肥量增加到一定数值时，产量不再增加。这时如果再增加肥料，产量反而会减少。这一施肥量与粮食产量的数量关系就是我们要探索的数量规律性。当我们从大量试验数据中用统计方法探索到施肥量与产量之间的数量关系，就可以考虑肥料的费用并选择最佳的施肥效果了。

【例 1.4】 某企业开发出一种新的化妆品，在正式投产生产之前，需要根据市场需求情况制定其价格和销售量等策略。该企业委托某市场调查公司在全市 200 万户家庭中抽取 1000 户家庭作为样本，免费赠送给这 1000 户家庭试用，然后了解该化妆品的市场前景。如何科学地从该市 200 万户家庭中抽出 1000 户家庭进行调查，并由 1000 户家庭反馈的市场信息科学地推断出全市 200 万户家庭对该化妆品的购买意愿这一数量规律是统计工作者的任

务。

为什么统计方法能够通过对数据的大量观察和处理而研究和探索出其内在的数量规律性呢？这是由客观事物本身的特点和统计方法的特性共同决定的。从客观事物方面来说，根据辩证法的基本原理，任何客观事物都是必然性与偶然性的对立统一。同样，任何一个数据，也都是必然性与偶然性共同作用的结果，是二者作用的对立统一。必然性反映了事物本质的特征和联系，是比较稳定的，因而它决定了事物的内在本质是有规律可循的。偶然性反映了该事物每个表现形式的差异。如果客观事物只有必然性一个方面的特征，事物的表现形式就会比较简单，就可以比较容易地把握它的规律性。正是由于偶然性的存在，造成了事物的表现形式与必然性和规律性发生偏移，从而形成了表面形式的千姿百态，形成数据表现形式的千差万别。这样，必然性的数量规律性就被掩盖在表面的差异之中了。我们在上面举出的四个例子，本身都存在必然性的数量规律，但就每个新生儿的性别、每次掷硬币的结果、每次施肥带来的增产、每个家庭的购买意愿看都是不同的、有差异的，表现形式是充满偶然的。但每个例子本身都是有规律可循的，应用统计方法就可以从表面的偶然性中探索到内在的、本质的数量规律；从统计方法来看，统计学提供了一系列的方法，专门用来收集数据、整理数据、显示数据的特征进而分析和探索（或推断）出事物总体的数量规律性。当然，如果事物本身的规律比较简单，所用的统计方法也就相对容易；如果事物本身的规律错综复杂，所用的统计方法也就相对复杂。那么显然，用什么统计方法，取决于我们有什么样的数据和要解决什么问题。

第二节 统计学的产生和发展

统计学产生于 17 世纪中叶，是从几个不同的领域开始的。统计学的一个源头是来自英国威廉·配第（William Petty）的《政治

算术》(1676年)。配第在书中用大量的数字对英、法、荷三国的经济实力进行比较,用数字、重量、尺度等定量的方法进行分析和比较,表达他的思想和观点。马克思在《资本论》中评价配第“是政治经济学之父,在某种程度上也可以说是统计学的创造人。”

统计学的另一个创始人是英国的约翰·格朗特(John Grannt)。格朗特在1662年出版了《关于死亡表的自然观察与政治观察》。他通过大量观察的方法,研究并发现了人口与社会现象中重要的数量规律性。如新生儿的性比例稳定在14:13;男性在各年龄组中死亡率高于女性;新生儿和大城市的死亡率较高;一般疾病与事故的死亡率较稳定而传染病的死亡率波动较大等等。在研究中,格朗特不但探索了人口变化和发展的一些数量规律,而且还对伦敦市总人口数做出了较科学的估计。如果说配第是政府统计的创始人,则格朗特可被认为是人口统计的创始人。

统计学的第三个源头是古典概率论,奠基人包括法国的帕斯卡尔(B. Pascal)和费马特(Pierre de Fermat)。早在他们之前,已有一些数学家在研究赌博中的数量规律性了。意大利诗人但丁早在15世纪就讨论过掷三颗骰子可能出现的各种点数。16世纪中,意大利科学家伽利略讨论了掷三骰子出现10点次数多于9点次数的原因。在数学家们对机会游戏研究的基础上,帕斯卡尔和费马特通过通信的方式,将赌博中出现的各种具体问题,归纳为一般的概率原理,为后来概率论和统计学的发展奠定了重要的基础。

自17世纪中叶上述几位统计学家从不同角度开始了统计学研究后,经过数代统计学家的努力,历经两个半世纪,到19世纪末建成了古典统计学(主要是描述统计学)的基本框架。

20世纪初,大工业的发展对产品质量检验问题提出了新的要求,即只抽取少量产品作为样本对全部产品的质量好坏做出推断。因为大批量产品生产要做全面检验,既费时、费钱又费人力,加之有些产品质量的检验要做破坏性检验,全部检验已不可能。1907

年,英国的戈赛特(W. S. Gosset)提出了小样本 t 统计量理论,使统计学进入了现代统计学(主要是推断统计学)的阶段。以后经过著名统计学家费歇(R. A. Fisher)给出了 F 统计量、极大似然估计、方差分析等方法和思想,内曼(J. Neyman)和皮尔逊(E. S. Pearson)的置信区间估计和假设检验,沃尔德(A. Wald)的序贯抽样和统计决策函数,到本世纪中基本构筑了现代统计学的基本框架。

从本世纪50年代以后,统计理论、方法和应用进入了一个全面发展的新阶段。统计学受计算机科学、信息论、混沌理论等现代科学技术的影响,新的研究领域层出不穷,例如多元统计分析、现代时间序列分析、贝叶斯统计、非参数统计、线性统计模型等。另一方面,统计方法的应用领域不断扩展,几乎所有科学研究都离不开统计方法。因为不论是自然科学、工程技术、农学、医学、军事科学还是社会科学都离不开数据,要对数据进行研究和分析就必然要用到统计方法。因而可以说统计方法与数学、哲学一样成为所有学科的基础。

第三节 统计学的分科

统计学的内容十分丰富,研究和应用的领域非常广泛。从统计教育的角度,统计学大致有以下两种分类:

一、描述统计学(*Descriptive Statistics*)和推断统计学(*Inferential Statistics*)

统计学分为描述统计学和推断统计学,一方面反映了统计发展的前后两个阶段,另一方面也反映了统计方法研究和探索客观事物内在数量规律性的先后两个过程。下面我们给出统计方法探索客观现象数量规律性过程的框图:

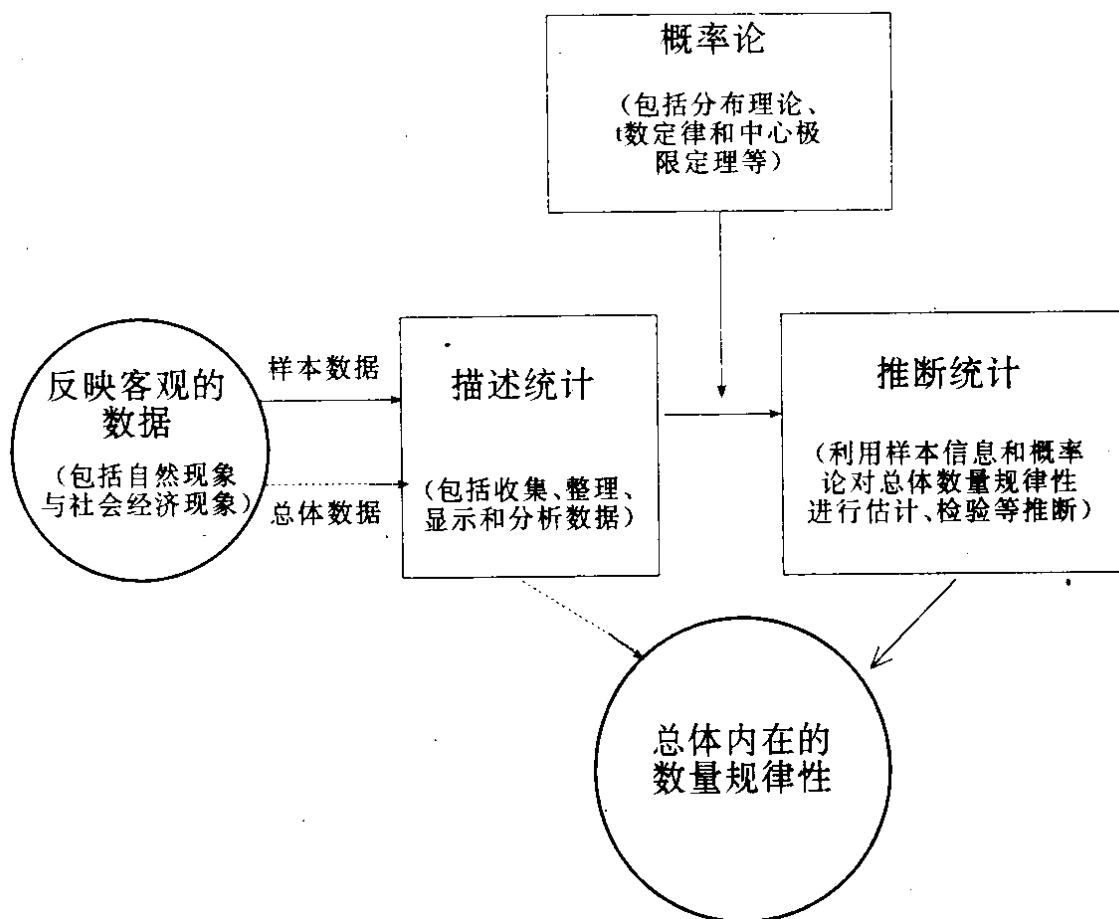


图 1.1 统计学探索客观现象数量规律性过程的框图

由图 1.1 我们看到，统计研究过程的起点是数据，终点是探索到客观事物总体内在的数量规律性。要达到统计研究的目的，如果我们收集到的是总体数据（如普查），则经过描述统计之后就可以达到探索内在数量规律性的目的了；但如果我们所获得的数据只是研究总体的一部分数据，要探索到总体的数量规律性，必须应用概率论的理论并根据样本整理出的信息对总体做出科学的推断。显然，描述统计是整个统计学的基础和统计研究工作的第一步。它包括对客观现象的度量、调查方案的设计、及时、快速、经济地收集数据、整理数据、用图表显示数据、分析和提取数据中的有用信息以最终推断总体；推断统计是现代统计学的核心和统计研究工作的关键环节，因为统计最终能否科学准确地探索到总

体内在的数量规律性与选用何种统计量、选用什么推断方法、如何进行推断有着直接的联系。一个出色的统计工作者的能力和技巧在推断统计中将得到充分的体现和检验。但如果没有描述统计收集可靠的数据并提供有效的样本信息，即使再高明的统计学家和再科学的推断方法也难于得出准确的结论。因而，推断统计对描述统计又有很强的依赖性。

应该认识到，尽管描述统计可以在获得总体数据时直接探索出总体数量规律性来，但这种情况在实际工作中很少见到。因为客观事物的总体有许多是无限的或近似无限的，例如统计物理研究中要弄清楚分子运动的规律，而分子又是无穷多的，不可能全部观察和试验；社会经济现象的总体显然多是有限的，但由于总体单位很多，加之要考虑时间，费用和人力的投入，也难于一一调查，而多采取抽样的办法。例如全国的人口数量和变化、耕地面积、企业个数和经营情况等，虽然可以通过普查得到全部数据，但普查要投入大量的人力、财力和物力，而且要很长的时间才能收集、整理出所要的数据，不是每年都能搞的，因而我国定为每 10 年搞一次人口普查、每 10 年搞一次工业普查和农业普查，其他各年均以抽样调查数据进行推断。再有就是大量的管理和研究工作不可能组织普查，例如城市居民家庭每月的收入支出调查、某种商品的市场调查、某个事件的民意测验等都只能通过抽样调查方法然后对总体数量规律性进行科学的推断。因而，在描述统计中收集、整理和分析的多是样本数据。这样，科学地整理样本数据、显示样本数据的特征和规律、提取样本数据中的有用信息就显得格外重要了。本书作为描述统计学的教材，重点是介绍整理、显示和分析数据的方法，科学有效地提取样本数据中的信息，并通过对样本数据的分析初步探索出数据内在的数量规律性，为进一步进行科学的统计推断打下坚实的基础。由于设计调查方案和进行科学的抽样有专门的教材进行介绍，在本书中未作为重点。