



学系列教材

“十二五”普通高等教育本科国家级规划教材

教育部推荐教材

国家统计局优秀统计教材

统计学

(第7版)

贾俊平 何晓群 金勇进 编著

Statistics

(Seventh Edition)



21世纪统计学系列教材



“十二五”普通高等教育本科国家级规划教材
教育部推荐教材
国家统计局优秀统计教材

统计学

(第7版)

贾俊平 何晓群 金勇进 编著

Statistics
(Seventh Edition)

中国人民大学出版社
· 北京 ·

图书在版编目 (CIP) 数据

统计学/贾俊平等编著. —7 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 978-7-300-25351-0

I. ①统… II. ①贾… III. ①统计学-高等学校-教材 IV. ①C8

中国版本图书馆 CIP 数据核字 (2017) 第 326776 号

“十二五”普通高等教育本科国家级规划教材
教育部推荐教材
国家统计局优秀统计教材
21 世纪统计学系列教材
统计学 (第 7 版)
贾俊平 何晓群 金勇进 编著
Tongjixue

出版发行	中国人民大学出版社	邮政编码	100080
社 址	北京中关村大街 31 号		
电 话	010-62511242 (总编室)	010-62511770 (质管部)	
	010-82501766 (邮购部)	010-62514148 (门市部)	
	010-62515195 (发行公司)	010-62515275 (盗版举报)	
网 址	http://www.crup.com.cn		
	http://www.ttrnet.com (人大教研网)		
经 销	新华书店	版 次	2000 年 9 月第 1 版
印 刷	北京密兴印刷有限公司		2018 年 1 月第 7 版
规 格	185 mm×260 mm 16 开本	印 次	2018 年 1 月第 1 次印刷
印 张	23.25 插页 1	定 价	42.80 元
字 数	540 000		

版权所有 侵权必究 印装差错 负责调换

改革开放以来,高等统计教育有了很大的发展。随着课程设置的不断调整,有不少教材出版,同时也翻译引进了一些国外优秀教材。作为培养我国统计专门人才的摇篮,中国人民大学统计学系自1952年创建以来,走过了风风雨雨,一直坚持理论与应用相结合的办学方向,培养能够理论联系实际、解决实际问题的高层次人才。随着新知识经济和网络时代的到来,我们在教学科研的实践中深切地感受到,无论是自然科学领域、社会科学领域的研究,还是国家宏观管理和企业生产经营管理,甚至人们的日常生活,信息需求量日益增多,信息处理技术更加复杂,作为信息技术支柱的统计方法,越来越广泛地应用于各个领域。

面对新的形势,我们一直在思索,课程设置、教材选择、教学方式等怎样才能使学生适应社会经济发展的客观需要。在反复酝酿、不断尝试的基础上,我们决定与统计学界的同仁共同编写、出版一套面向21世纪的统计学系列教材。

这套系列教材聘请了中科院院士、中国科技大学陈希孺教授,上海财经大学数量经济研究院张尧庭教授,中国科学院数学与系统科学研究所冯士雍研究员等作为编委。他们长期任中国人民大学的兼职教授,一直关心、支持着统计学系的学科建设和应用统计的发展。中国人民大学应用统计科学研究中心2000年已成为国家级研究基地,这些专家是首批专职或兼职研究人员。这一开放性研究基地的运作,将有利于提升我国应用统计科学研究水平,也必将进一步促进高等统计教育的发展。

这套教材是我们奉献给新世纪的,希望它能促进应用统计教育水平的提高。这套教材力求体现以下特点:

第一,在教材选择上,主要面向经济类统计学专业。选材既包括统计教材也包括风险管理与精算方面的教材。尽管名为统计学系列教材,但并不求大、求全,而是力求精选。对于目前已有的内容较为成熟、适合教学需要、公认的较好的教材,并未列入本次出版计划。

第二，每部教材的内容和写作，注意广泛吸收国内外优秀教材的成果。教材力求简明易懂、内容系统和实用，注重对统计方法思想的阐述，并结合大量实际数据和实例说明统计方法的特点及应用条件。

第三，强调与计算机的结合。为着力提高学生运用统计方法分析解决问题的能力，教材所涉及的统计计算，要求运用目前已有的统计软件。根据教材内容，选择使用 SAS, SPSS, TSP, STATISTICA, EViews, MINITAB, Excel 等。

感谢中国人民大学出版社的同志们，他们怀着发展我国应用统计科学的热情和提高统计教育水平的愿望，经过反复论证，使这套教材得以出版。感谢参与教材编写的同行专家、统计学系的教师。愿大家的辛勤劳动能够结出丰硕的果实。我们期待着与统计学界的同仁共同创造应用统计辉煌的明天。

易丹辉
于中国人民大学

《统计学》第7版是在第六版的基础上修订而成的。本次修订在广泛吸取读者意见的基础上，对第六版中的部分内容及不当之处进行了修改。

第7版在内容体系上与第六版相比没有大的变动，除了对部分章节在内容上做了更新，还对部分内容和练习题作了较大篇幅的删减。在软件使用上也有所变化，使用了 Excel 2013 和 SPSS 19.0（中文版）两款软件。

第7版的各章执笔人是：第1章、第3章、第4章、第7章、第10章、第11章、第12章、第13章由贾俊平编写；第5章、第6章由何晓群编写；第2章、第8章、第9章、第14章由金勇进编写。

感谢为本次修订提出宝贵意见的教师和读者。感谢中国人民大学出版社对本书出版的大力支持。

由于作者水平所限，书中的错误和疏漏之处在所难免。敬请读者提出宝贵意见，以便进一步修订和改进。

贾俊平

第1章 导论 1

1.1 统计及其应用领域 2

1.2 统计数据类型 4

1.3 统计中的几个基本概念 6

思考与练习 9

第2章 数据的搜集 11

2.1 数据的来源 12

2.2 调查方法 14

2.3 实验方法 23

2.4 数据的误差 27

思考与练习 33

第3章 数据的图表展示 34

3.1 数据的预处理 35

3.2 品质数据的整理与展示 41

3.3 数值型数据的整理与展示 49

3.4 合理使用图表 61

思考与练习 63

第4章 数据的概括性度量 67

4.1 集中趋势的度量 68

4.2 离散程度的度量 77

4.3 偏态与峰态的度量 85

思考与练习 89

第5章 概率与概率分布 92

5.1 随机事件及其概率 93

5.2 离散型随机变量及其分布 96

5.3 连续型随机变量的概率分布 108

思考与练习 114

第6章 统计量及其抽样分布 115

6.1 统计量 116

6.2 由正态分布导出的几个重要分布 117

6.3 样本均值的分布与中心极限定理 121

思考与练习 124



第7章 参数估计	126
7.1 参数估计的基本原理	127
7.2 一个总体参数的区间估计	132
7.3 两个总体参数的区间估计	138
7.4 样本量的确定	148
思考与练习	150
第8章 假设检验	155
8.1 假设检验的基本问题	156
8.2 一个总体参数的检验	163
8.3 两个总体参数的检验	170
8.4 检验问题的进一步说明	181
思考与练习	184
第9章 分类数据分析	187
9.1 分类数据与 χ^2 统计量	188
9.2 拟合优度检验	189
9.3 列联分析: 独立性检验	191
9.4 列联表中的相关测量	195
9.5 列联分析中应注意的问题	199
思考与练习	202
第10章 方差分析	204
10.1 方差分析引论	205
10.2 单因素方差分析	210
10.3 双因素方差分析	220
思考与练习	229
第11章 一元线性回归	233
11.1 变量间关系的度量	234
11.2 一元线性回归	242
11.3 利用回归方程进行预测	255
11.4 残差分析	258
思考与练习	261
第12章 多元线性回归	265
12.1 多元线性回归模型	266
12.2 回归方程的拟合优度	269

12.3	显著性检验	271
12.4	多重共线性	274
12.5	利用回归方程进行预测	277
12.6	变量选择与逐步回归	278
	思考与练习	282
第 13 章	时间序列分析和预测	286
13.1	时间序列及其分解	287
13.2	时间序列的描述性分析	289
13.3	时间序列预测的程序	293
13.4	平稳序列的预测	298
13.5	趋势型序列的预测	303
13.6	复合型序列的分解预测	309
	思考与练习	314
第 14 章	指数	318
14.1	基本问题	319
14.2	总指数编制方法	321
14.3	指数体系	328
14.4	几种典型的指数	332
14.5	综合评价指数	338
	思考与练习	340
附录一	术语表	344
附录二	用 Excel 生成概率分布表	351
参考文献		361

理解统计对每个人都是必要的

统计在许多领域都有应用。在日常生活中，我们也会经常接触到各种统计数据，比如，媒体报道中使用的一些统计数据、图表等。下面就是统计研究得到的一些结论：吸烟对健康是有害的；不结婚的男性会早逝 10 年；身材高的父亲，其子女的身材也较高；第二个出生的子女没有第一个聪明，第三个出生的子女没有第二个聪明，依此类推；两天服一片阿司匹林会减少心脏病第二次发作的概率；如果每天摄取 500 毫升维生素 C，生命可延长 6 年；怕老婆的丈夫得心脏病的概率较大；学生在听了莫扎特钢琴曲 10 分钟后的推理测试会比他们听 10 分钟娱乐节目或其他曲目做得更好。这些结论是正确的吗？你相信这些结论吗？要正确阅读并理解这些数据，就需要具备一些统计学知识。

理解并掌握一些统计学知识对普通大众是有必要的。每天我们都会关心生活中的一些事情，其中就包含统计知识。比如，在外出旅游时，需要关心一段时间内的天气预报；在投资股票时，需要了解股票市场价格的信息，了解某只特定股票的有关财务信息；在观看世界杯足球赛时，需要了解各支球队的技术统计等。

理解和掌握一些统计知识，对政治家或制定政策的人来说更为重要，在他们做决策时，如果不懂统计可能会闹出笑话来。比如，一个统计办公室的主管是一位行政事务官，一次与统计学者开会，统计学者抱怨从其他部门得到的一些估计值没有给出标准误差（估计时的误差大小，表示估计的精度），这个主管马上问道：“误差也有标准吗？”一个统计顾问提交给茶叶委员会的报告中，含有标题为“饮茶人数的估计值（含标准误差）”的附表。不久，一封信被送到这个统计学者手中，问什么是人们喝红茶时的“标准误差”。健康部门的一位官员看到一个统计学者提供的报告，报告中提到去年由于某种疾病，平均 1 000 人中死亡 3.2 人，这位官员对这个数字产生了兴趣。他问他的私人秘书，3.2 人是如何死法？他的秘书说：“先生，当一个统计学家说死了 3.2 人时，意味着三个人已经死了，两个人正要死。”

本章将介绍统计学的一些基本问题,包括统计学的含义、统计数据及其分类、统计中常用的基本概念等。

1.1 统计及其应用领域

1.1.1 什么是统计学

统计是处理数据的一门科学。人们给统计学下的定义很多,比如,“统计学是收集、分析、表述和解释数据的科学”;“统计是一组方法,用来设计实验、获得数据,然后在这些数据的基础上组织、概括、演示、分析、解释和得出结论”。综合地说,统计学(statistics)是收集、处理、分析、解释数据并从数据中得出结论的科学。

统计学是关于数据的科学,它所提供的是一套有关数据收集、处理、分析、解释并从数据中得出结论的方法,统计研究的是来自各领域的的数据。数据收集也就是取得统计数据;数据处理是将数据用图表等形式展示出来;数据分析则是选择适当的统计方法研究数据,并从数据中提取有用信息进而得出结论。

数据分析所用的方法可分为描述统计方法和推断统计方法。描述统计(descriptive statistics)研究的是数据收集、处理、汇总、图表描述、概括与分析等统计方法。推断统计(inferential statistics)是研究如何利用样本数据来推断总体特征的统计方法。比如,要了解一个地区的人口特征,不可能对每个人的特征一一进行测量;对产品的质量进行检验往往是破坏性的,也不可能对每个产品进行测量。这就需要抽取部分个体即样本进行测量,然后根据获得的样本数据对所研究的总体特征进行推断,这就是推断统计要解决的问题。

1.1.2 统计的应用领域

统计方法是适用于所有学科领域的通用数据分析方法,只要有数据的地方就会用到统计方法。随着人们对定量研究的日益重视,统计方法应用到自然科学和社会科学的众多领域,统计学已发展成为由若干分支学科组成的学科体系。可以说,几乎所有的研究领域都要用到统计方法,比如政府部门、学术研究领域、日常生活中、公司或企业的生产经营管理中都要用到统计。下面将给出统计在工商管理中的一些应用。

1. 企业发展战略

发展战略是一个企业的长远发展方向。制定发展战略一方面需要及时了解和把握整个宏观经济的状况及发展变化趋势,了解市场的变化;另一方面,还要对企业进行合理的市场定位,把握企业自身的优势和劣势。所有这些都离不开统计,需要统计提供可靠的数据,利用统计方法对数据进行科学的分析和预测,等等。

2. 产品质量管理

质量是企业的生命,是企业持续发展的基础。质量管理中离不开统计的应用。在一些

知名的跨国公司，六西格玛准则成为一种重要的管理理念。质量控制成为统计学在生产领域的一项重要应用。各种统计质量控制图广泛用于监测生产过程。

3. 市场研究

企业要在激烈的市场竞争中取得优势，首先必须了解市场，要了解市场，则需要做广泛的市场调查，取得所需的信息，并对这些信息进行科学的分析，以便作为生产和营销的依据，这些都需要统计的支持。

4. 财务分析

上市公司的财务数据是股民投资的重要参考依据。一些投资咨询公司主要是根据上市公司提供的财务和统计数据进行分析，为股民提供投资参考。企业自身的投资也离不开对财务数据的分析，其中要用到大量的统计方法。

5. 经济预测

企业要对未来的市场状况进行预测，经济学家也常常对宏观经济或某一方面进行预测。在进行预测时要使用各种统计信息和统计方法。比如，企业要对产品的市场潜力作出预测，以便及时调整生产计划，这就需要利用市场调查取得数据，并对数据进行统计分析。经济学家在预测通货膨胀时，要利用有关生产价格指数、失业率、生产能力利用率等统计数据，通过统计模型进行预测。

6. 人力资源管理

企业利用统计方法对员工的年龄、性别、受教育程度、工资等进行分析，作为企业制定工资计划、奖惩制度的依据。

当然，统计并不是仅仅为了管理，它是为自然科学、社会科学的多个领域发展起来的，为多个学科提供了一种通用的数据分析方法。从某种意义上说，统计仅仅是一种数据分析的方法。与数学一样，统计是一种工具，是一种数据分析的工具。表 1-1 列出了统计的一些应用领域，目的是让我们通过简单浏览形成这样一个概念：统计学非常有用！

表 1-1

统计的应用领域

actuarial work (精算)	auditing (审计学)
agriculture (农业)	crystallography (晶体学)
animal science (动物学)	demography (人口统计学)
anthropology (人类学)	dentistry (牙科学)
archaeology (考古学)	ecology (生态学)
econometrics (经济计量学)	management science (管理科学)
education (教育学)	marketing (市场营销学)
election forecasting and projection (选举预测和策划)	medical diagnosis (医学诊断)
engineering (工程)	meteorology (气象学)
epidemiology (流行病学)	military science (军事科学)
finance (金融)	nuclear material safeguards (核材料安全管理)
fisheries research (水产渔业研究)	ophthalmology (眼科学)
gambling (赌博)	pharmaceutics (制药学)
genetics (遗传学)	physics (物理学)
geography (地理学)	political science (政治学)

geology (地质学)	psychology (心理学)
historical research (历史研究)	psychophysics (心理物理学)
human genetics (人类遗传学)	quality control (质量控制)
hydrology (水文学)	religious studies (宗教研究)
industry (工业)	sociology (社会学)
linguistics (语言学)	survey sampling (调查抽样)
literature (文学)	taxonomy (分类学)
manpower planning (劳动力计划)	weather modification (气象改善)

利用统计方法可以简化繁杂的数据，比如，用图表展示数据，建立数据模型。有人认为统计的全部目的就是让人看懂数据，其实这仅仅是统计的一个方面，统计更重要的功能是对数据进行分析，它提供了一套分析数据的方法和工具。不同的人对数据分析的理解会大不一样，曲解数据分析是常见的现象。在有些人的心目中，数据分析就是寻找支持：他们的心目中可能有了某种“结论”性的东西，或者说他们希望看到一种符合他们需要的某种结论，而后去找些统计数据来支持他们的结论。这恰恰歪曲了数据分析的本质，数据分析的真正目的是从数据中找出规律，从数据中寻找启发，而不是寻找支持。真正的数据分析事先是没有结论的，通过对数据的分析才能得出结论。统计不是万能的，它不能解决你所面临的所有问题。统计可以帮助你分析数据，并从分析中得出某种结论，但对统计结论的进一步解释，则需要你的专业知识。比如，吸烟会使患肺癌的概率增大，这是一个统计结论，但要解释吸烟为什么能引起肺癌，这就不是统计学家所能做到的，需要有更多的医学知识才行。

1.2 统计数据 的类型

统计数据是对现象进行测量的结果。比如，对经济活动总量的测量可以得到国内生产总值（GDP）；对股票价格变动水平的测量可以得到股票价格指数；对人口性别的测量可以得到男或女这样的数据。下面从不同角度说明统计数据的分类。

1.2.1 分类数据、顺序数据、数值型数据

按照所采用的计量尺度的不同^①，可以将统计数据分为分类数据、顺序数据和数值型数据。

分类数据（categorical data）是只能归于某一类别的非数字型数据，它是对事物进行分类的结果，数据表现为类别，是用文字来表述的。例如，人口按照性别分为男、女两

^① 数据的测量尺度有四种：第一，分类尺度（nominal scale）。按照事物的某种属性对其进行的平行的分类，数据表现为类别。第二，顺序尺度（ordinal scale）。对事物类别顺序的测度，数据表现为有序类别。第三，间隔尺度（interval scale）。对事物类别或次序之间间距的测度，没有绝对零点，数据表现为数字。第四，比率尺度（ratio scale）。对事物类别或次序之间间距的测度，有绝对零点，数据表现为数字。

类；企业按行业属性分为医药企业、家电企业、纺织品企业等，这些均属于分类数据。为便于统计处理，对于分类数据，可以用数字代码来表示各个类别，比如，用1表示“男性”，0表示“女性”；用1表示“医药企业”，2表示“家电企业”，3表示“纺织品企业”；等等。

顺序数据（rank data）是只能归于某一有序类别的非数字型数据。顺序数据虽然也是类别，但这些类别是有序的。比如将产品分为一等品、二等品、三等品、次品等；考试成绩可以分为优、良、中、及格、不及格等；一个人的受教育程度可以分为小学、初中、高中、大学及以上；一个人对某一事物的态度可以分为非常同意、同意、保持中立、不同意、非常不同意；等等。同样，顺序数据也可以用数字代码来表示。比如，1——非常同意，2——同意，3——保持中立，4——不同意，5——非常不同意。

数值型数据（metric data）是按数字尺度测量的观察值，其结果表现为具体的数值。现实中所处理的大多数是数值型数据。

分类数据和顺序数据说明的是事物的品质特征，通常是用文字来表述的，其结果均表现为类别，因而也可统称为定性数据或品质数据（qualitative data）；数值型数据说明的是现象的数量特征，通常是用数值来表现的，因此也可称为定量数据或数量数据（quantitative data）。

1.2.2 观测数据和实验数据

按照统计数据的收集方法，可以将其分为观测数据和实验数据。**观测数据**（observational data）是通过调查或观测收集到的数据，这类数据是在没有对事物人为控制的条件下得到的，有关社会经济现象的统计数据几乎都是观测数据。**实验数据**（experimental data）则是在实验中控制实验对象而收集到的数据。比如，对一种新药疗效的实验数据，对一种新的农作物品种的实验数据。自然科学领域的大多数数据为实验数据。

1.2.3 截面数据和时间序列数据

按照被描述的现象与时间的关系，可以将统计数据分为截面数据和时间序列数据。**截面数据**（cross-sectional data）是在相同或近似相同的时间点上收集的数据，这类数据通常是在不同的空间获得的，用于描述现象在某一时刻的变化情况。比如，2010年我国各地区的国内生产总值就是截面数据。**时间序列数据**（time series data）是在不同时间收集到的数据，这类数据是按时间顺序收集到的，用于描述现象随时间变化的情况。比如2010—2012年我国的国内生产总值就是时间序列数据。

图1-1给出了统计数据分类的框图。

区分数据的类型是十分重要的，因为对不同类型的数据需要采用不同的统计方法来处理和分析。比如，对分类数据，我们通常计算出各组的频数或频率，计算其众数和异众比率，进行列联表分析和 χ^2 检验等；对顺序数据，可以计算其中位数和四分位差，计算等级相关系数等；对数值型数据，可以用更多的统计方法进行分析，如计算各种统计量，进行参数估计和检验等。

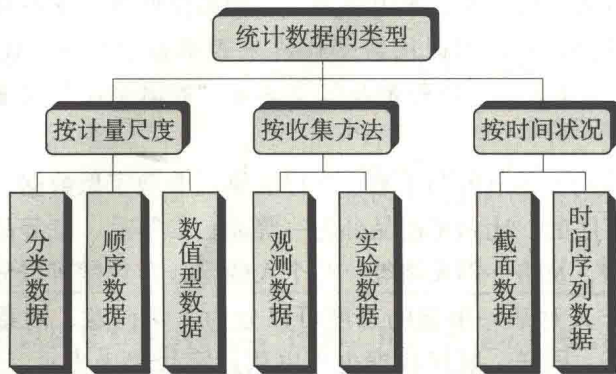


图 1-1 统计数据的分类

1.3 统计中的几个基本概念

统计中的概念很多，其中有几个概念是经常要用到的，有必要单独加以介绍。这些概念包括总体和样本、参数和统计量、变量等。

1.3.1 总体和样本

1. 总体

总体 (population) 是包含所研究的全部个体 (数据) 的集合，它通常由所研究的一些个体组成，如由多个企业构成的集合，多个居民户构成的集合，多个人构成的集合等。组成总体的每个元素称为个体，在由多个企业构成的总体中，每个企业就是一个个体；由多个居民户构成的总体中，每个居民户就是一个个体；由多个人构成的总体中，每个人就是一个个体。

总体范围的确定有时比较容易。比如，要检验一批灯泡的使用寿命，这批灯泡构成的集合就是总体，每个灯泡就是一个个体，总体的范围很清楚。但在有些场合总体范围的确定则比较困难，比如，对于新推出的一种饮料，要想知道消费者是否喜欢，首先必须弄清哪些人是消费的对象，也就是要确定构成该饮料的消费者这一总体，但事实上，我们很难确定哪些消费者购买该饮料，总体范围的确定十分复杂。当总体的范围难以确定时，可根据研究的目的来定义总体。

总体根据其所包含的单位数目是否可数可以分为有限总体和无限总体。有限总体是指总体的范围能够明确确定，而且元素是有限可数的。比如，由若干个企业构成的总体就是有限总体，一批待检验的灯泡也是有限总体。无限总体是指总体所包括的元素是无限的、不可数的。例如，在科学实验中，每个实验数据可以看做总体的一个元素，而实验则可以无限地进行下去，因此由实验数据构成的总体就是一个无限总体。

总体分为有限总体和无限总体主要是为了判别在抽样中每次抽取是否独立。对于无限总体，每次抽取一个单位，并不影响下一次的抽样结果，因此每次抽取可以看做是独立的。对于有限总体，抽取一个单位后，总体元素就会减少一个，前一次的抽样结果往往会影响到第二次的抽样结果，因此每次抽取是不独立的。这些因素会影响到抽样推断的结果。

最后，再对总体的概念作进一步的说明。如前所述，要检验一批灯泡的寿命，这批灯泡构成的集合就是总体。在统计问题中，我们只是关心每个灯泡的寿命，而不是灯泡本身，所以也可以把这批灯泡的寿命集合作为总体，这个总体是一些实数构成的集合。一般而言，有限总体就是有限个实数的集合。如果不是针对一批特定的灯泡，而是全面地考察某企业生产的灯泡寿命，可能的寿命是多少呢？答案是 $[0, +\infty)$ 这样一个区间。或者这样看这个问题，随机从该企业生产的灯泡中拿出一个，这个灯泡可能的寿命是多少？答案只能是“非负实数”，当然这个“非负实数”在实际检验前是未知的。这时称该企业生产的灯泡寿命总体是取值于 $[0, +\infty)$ 区间上的一个随机变量，这是一个无限总体。在统计推断中通常是针对无限总体的，因而一般把总体看做随机变量。通常情况下，统计上的总体是一组观测数据，而不是一群人或一些物品的集合。

2. 样本

样本 (sample) 是从总体中抽取的一部分元素的集合，构成样本的元素的数目称为样本量 (sample size)。抽样的目的是根据样本提供的信息推断总体的特征。比如，从一批灯泡中随机抽取 100 个，这 100 个灯泡就构成了一个样本，然后根据这 100 个灯泡的平均使用寿命去推断这批灯泡的平均使用寿命。

1.3.2 参数和统计量

1. 参数

参数 (parameter) 是用来描述总体特征的概括性数字度量，它是研究者想要了解的总体的某种特征值。研究者所关心的参数通常有总体平均数、总体标准差、总体比例等。在统计中，总体参数通常用希腊字母表示。比如，总体平均数用 μ (mu) 表示，总体标准差用 σ (sigma) 表示，总体比例用 π (pi) 表示，等等。

由于总体数据通常是不知道的，所以参数是一个未知的常数。比如，我们不知道某一地区所有人口的平均年龄，不知道一个城市所有家庭的收入的差异，不知道一批产品的合格率，等等。正因为如此，所以才进行抽样，根据样本计算出某些值，然后估计总体参数。

2. 统计量

统计量 (statistic) 是用来描述样本特征的概括性数字度量。它是根据样本数据计算出来的一个量，由于抽样是随机的，因此统计量是样本的函数。研究者所关心的统计量主要有样本平均数、样本标准差、样本比例等。样本统计量通常用英文字母来表示。比如，样本平均数用 $\bar{x}$ (读作 x -bar) 表示，样本标准差用 s 表示，样本比例用 p 表示，等等。

由于样本是已经抽出来的，所以统计量总是知道的。抽样的目的就是要根据样本统计量去估计总体参数。比如，用样本平均数 ($\bar{x}$) 去估计总体平均数 (μ)，用样本标准差 (s) 去估计总体标准差 (σ)，用样本比例 (p) 去估计总体比例 (π)，等等。

总体、样本、参数、统计量的概念可以用图 1-2 来表示。

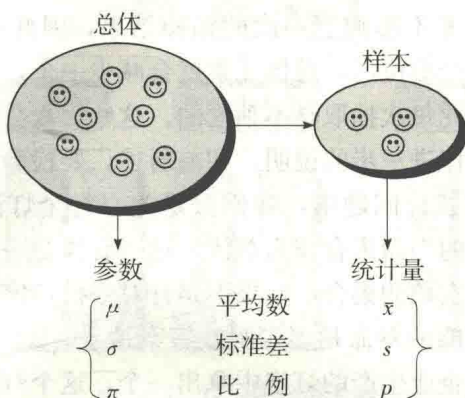


图 1-2 总体和样本、参数和统计量

除了样本均值、样本比例、样本方差这类统计量，还有一些是为统计分析的需要而构造出来的统计量，比如用于统计检验的 z 统计量、 t 统计量、 χ^2 统计量、 F 统计量等，它们的含义将在后面相关的章节中介绍。

1.3.3 变量

变量 (variable) 是说明现象某种特征的概念，其特点是从一次观察到下一次观察结果会呈现出差别或变化。如“商品销售额”“受教育程度”“产品的质量等级”等都是变量。变量的具体取值称为变量值。比如商品销售额可以是 20 万元、30 万元、50 万元等，这些数字就是变量值。统计数据就是统计变量的某些取值。变量可以分为以下几种类型。

1. 分类变量

分类变量 (categorical variable) 是说明事物类别的一个名称，其取值是分类数据。如“性别”就是一个分类变量，其变量值为“男”或“女”；“行业”也是一个分类变量，其变量值可以为“零售业”“旅游业”“汽车制造业”等。

2. 顺序变量

顺序变量 (rank variable) 是说明事物有序类别的一个名称，其取值是顺序数据。如“产品等级”就是一个顺序变量，其变量值可以为“一等品”“二等品”“三等品”“次品”等；“受教育程度”也是一个顺序变量，其变量值可以为“小学”“初中”“高中”“大学”等；一个人对某种事物的看法也是一个顺序变量，其变量值可以为“同意”“保持中立”“反对”等。

3. 数值型变量

数值型变量 (metric variable) 是说明事物数字特征的一个名称，其取值是数值型数据。如“产品产量”“商品销售额”“零件尺寸”“年龄”“时间”等都是数值型变量，这些变量可以取不同的数值。数值型变量根据其取值的不同，又可以分为离散型变量和连续型变量。离散型变量 (discrete variable) 是只能取可数值的变量，它只能取有限个值，而且其取值都以整位数断开，可以一一列举，如“企业数”“产品数量”等就是离散型