



普通高等教育“十二五”应用型本科规划教材

应用统计学

主编 陈焱晗





普通高等教育“十二五”应用型本科规划教材

应用统计学

主 编 陈焱晗

中国人民大学出版社
• 北京 •

图书在版编目 (CIP) 数据

应用统计学/陈焱晗主编. —北京: 中国人民大学出版社, 2015. 7
普通高等教育“十二五”应用型本科规划教材
ISBN 978-7-300-21717-8

I. ①应… II. ①陈… III. ①应用统计学-高等学校-教材 IV. ①C8

中国版本图书馆 CIP 数据核字 (2015) 第 167258 号

普通高等教育“十二五”应用型本科规划教材

应用统计学

主 编 陈焱晗

Yingyong 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

邮政编码 100080

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)

印 刷 北京昌联印刷有限公司

规 格 185 mm×260 mm 16 开本

版 次 2015 年 9 月第 1 版

印 张 14.5

印 次 2015 年 9 月第 1 次印刷

字 数 331 000

定 价 33.00 元

版权所有 侵权必究

印装差错 负责调换

内容简介

应用统计学是一门运用统计学的方法原理对各个领域内的定量性质的问题进行数据的收集、整理和分析的科学，是文理兼类专业一门重要的专业基础课。本书即围绕这门课程的教学工作而编写，所面对的读者包括高校文理兼专业的本科生、人文社科及其他专业的学生，也可以作为管理人员和研究人员的参考书。

本书主要特色有以下几点：（1）以突出实际应用并淡化统计原理的数学推导为原则，尽量通俗化地介绍了各章统计方法的基本思想，用引例、实例、例题解析等形式做辅助来完成引出问题、说明问题、深入分析和解决问题的各环节教学工作。（2）虽然要淡化统计原理的推导环节，但并不意味着完全省略，并且对基本原理的理解也正是统计教学与学习中的难点所在。因此，本书尝试性地在参数估计和假设检验两章的原理介绍中，用一种简单而逻辑清晰的说明性示意图形式完成了对各自原理的阐述，替代并改良了往日统计教材中为避免数学推导而进行的大篇幅缺乏逻辑性与系统性的文字说明。这一改进将在很大程度上增强统计原理介绍环节的简洁性、完整性、系统性、逻辑性和易懂性。（3）本书对易混淆的统计知识做了大量的归纳、整理和对比工作，最终以表格、小结或简图的方式呈现，有效提高学习者对这些易混淆知识的辨识度，降低滥用错用率。（4）所举实例和所用数据的时效性均较强，没有很久远的资料，且多为经济管理领域内的数据或实际问题。（5）每章末有高度概括性的本章小结及习题。

前言

近些年，由我国教育者编写的各类统计学教科书不断出版，应该说是国内统计教材建设中的一大进步。这些教材编写的基本原则都是强调统计分析方法的应用性而淡化对分析原理的数学推导。这无疑是正确的，是适合各领域对统计方法大量实际应用需要的。然而，一个基本的事实我们又都赞同，即理解是应用的必经之路。自如灵活的应用永远跨不过对基本原理的正确理解这一关。于是，使学生能够正确地理解统计分析原理的任务就成为教材编写者一份不可推卸的责任。由此可见，上文所述统计教材编写要淡化原理推导的环节并不意味着完全省略对原理的必要解释。那么，在淡化数学推导的情况下实现对统计原理的有效解释就成为广大教材编写者和统计教学工作所必须面对的一项职责。

其实，国外优秀统计教材的编写理念对我们有着一定的借鉴意义，尤其是在充分考虑初学者的认知规律，遵循其认知的逻辑顺序，有条不紊、完整系统地完成对统计知识或原理的阐述方面，国外教材做得要比我们好很多。客观地说，一方面，有些国内教材的编写者难免从自己的表达习惯和知识水平出发来阐述统计原理，而非从学习者的认知规律和知识水平出发来做这一工作。因此在这个问题上，国内教材的编写工作还存在很大的探讨和改进空间。另一方面，目前国内几乎所有的教材都采用大篇幅文字说明的方式来完成对原理解释，而多年的统计教学实践表明：这种单一的文字方式并不适合对复杂抽象的分析原理实现有效的解释。它既难以突出核心概念或重点要素，又在逻辑关系上难以贯穿始终，不利于学习者形成一种系统完整、逻辑清晰的理解思路。所以，国内教材中关于统计原理部分的编写方式至少需要克服以上两方面问题才可能较好地完成。经过反复的思考与探索，编者尝试着对那些不易直接领会的统计原理（主要在推断统计学部分）各绘制一个逻辑推导简图以帮助学习者理解其中的要义。多轮的教学实践结果表明：这种简图的辅助理解作用是明显的，它在一定程度上替代和弥补了纯文字说明形式的不足，有效降低了学习者对抽象的统计分析原理的理解难度。

此外，统计学中存在大量统计知识或规律让学习者很容易产生混淆从而影响其正确辨识和应用，无论描述统计还是推断统计部分都有，尤其后者。本书编者投入了相当精力来对这些知识或规律进行整理、归纳和对比，最终以文字小结或大量对比性表格的形式呈现给读者，使其不再困扰于林林总总、极易混淆的同类统计知识之中，而是一目了然、拿来即用。

现今的国内教材大多为了突出应用性会在章节中穿插与所介绍的统计分析方法相对应

的软件功能实操内容，不同教材在该环节所安排的篇幅即介绍的详略程度也有差异。而我国的高教制度中对非统计专业本科层次的培养方案里，关于统计学这门课通常规定的课时是 48~64 学时。多年从事普本生统计教学的老师们会深有体会：即便是 64 学时的统计课，要把大纲规定的各章节内容（基础知识、分析原理和例题解析）都扎扎实实地讲完，时间上也是相当紧张的，更不用说再加上各章节对应的软件实操环节了。如果加上后者，恐怕得需要将近 100 学时的时间才能在质量上不打折扣地完成教学和实操任务。因此，在国内现有的学时制度下，将普本层次的统计教材中穿插进大量的软件实操环节，其现实性并不强。目前，国内很多高校采取另开设一门“统计软件的使用”课程的形式来解决有限课时与实操能力培养之间的矛盾。编者认为很值得提倡，因为这样会比较系统完整、从容扎实地培养学生软件实操和输出结果解读的能力，效果要强于在统计教材中穿插进实操环节的做法。

最后，感谢燕晓飞老师等本校各级领导的支持与协助，出版社相关工作人员的努力，以及参编人员在图表制作、公式录入、排版调整和格式统一等方面的积极参与与配合。本书的两位主编人员为该书的编写工作付出了大量辛勤的劳动与努力，其中徐军委老师主要负责第 1、2、3、4、5、9 章的编写，陈焱晗老师主要负责第 6、7、8、10、11、12 章的编写以及全书的选题和统稿工作。另外，参编人员有：简仁祥、张亦欢、谢竹青、郑丽君、张宏亮、李伟征、张静怡等。在此，对为本书的编辑出版工作做出过自己贡献的各方人士表达最诚挚的谢意！

欢迎广大统计教学工作者以及读者们批评指正。

陈焱晗

2015 年 5 月

教师信息反馈表

为了更好地为您服务，提高教学质量，中国人民大学出版社愿意为您提供全面的教学支持，期望与您建立更广泛的合作关系。请您填好下表后以电子邮件或信件的形式反馈给我们。

您使用过或正在使用的我社教材名称		版次	
您希望获得哪些相关教学资料			
您对本书的建议（可附页）			
您的姓名			
您所在的学校、院系			
您所讲授的课程名称			
学生人数			
您的联系地址			
邮政编码		联系电话	
电子邮件（必填）			
您是否为人大社教研网会员	☐ 是，会员卡号：_____ ☐ 不是，现在申请		
您在相关专业是否有主编或参编教材意向	☐ 是 ☐ 否 ☐ 不一定		
您所希望参编或主编的教材的基本情况（包括内容、框架结构、特色等，可附页）			

我们的联系方式：北京市西城区马连道南街 12 号
 中国人民大学出版社应用技术分社
 邮政编码：100055
 电话：010-63311862
 网址：<http://www.crup.com.cn>
 E-mail：rendayingyong@163.com

目 录

第一章 总 论	1
1.1 统计学的历史起源	1
1.2 学科性质及定义	4
1.3 研究对象及应用领域	7
1.4 统计学的分类	9
1.5 统计学与其他学科的关系	11
第二章 数据资料	13
2.1 数据的计量与类型	14
2.2 数据的搜集	16
2.3 数据的整理	19
第三章 描述统计分析	33
3.1 测度数据分布的集中趋势	33
3.2 测度数据分布的离散趋势	40
3.3 测度数据分布的偏度与峰度	45
第四章 概率与概率分布	50
4.1 概率基础	50
4.2 随机变量及其概率分布	58
4.3 大数定律与中心极限定理	68
第五章 抽样与抽样分布	72
5.1 抽样调查中的基本概念	73
5.2 抽样分布	78
第六章 参数估计	86
6.1 参数估计的基本原理	86

6.2	一个总体参数的区间估计	91
6.3	两个总体参数的区间估计	95
6.4	抽样容量的确定	101
第七章	假设检验	104
7.1	假设检验的基本原理、相关概念及问题	105
7.2	一个总体参数的检验	113
7.3	两个总体参数的检验	118
第八章	方差分析	127
8.1	方差分析的基本问题和基础知识	128
8.2	单因素方差分析	130
8.3	双因素方差分析	135
第九章	相关与一元线性回归分析	144
9.1	相关分析与回归分析的基本概念	145
9.2	一元线性回归分析	151
第十章	多元线性回归分析	165
10.1	多元线性回归模型及假定	165
10.2	多元线性回归模型的估计	167
10.3	多元线性回归模型的检验	170
第十一章	时间序列的分析	175
11.1	时间序列的概念和分类	176
11.2	时间序列的基本分析指标	178
11.3	时间序列的成分和基本模型	181
11.4	时间序列的预测	184
第十二章	统计指数	206
12.1	指数的概念与分类	207
12.2	总指数的编制方法	208
12.3	几种常用的经济指数	214
参考文献		221

第一章 总 论

引例

关于《红楼梦》后 40 回作者是谁的统计学判断

1981 年,首届国际《红楼梦》研讨会在美国召开,会上一共有四个针对红楼梦一书的问题被讨论分析,其中之一,就是大家争议《红楼梦》后 40 回的作者是不是曹雪芹本人。而威斯康星大学讲师陈炳藻独树一帜,宣读了题为《从词汇上的统计论〈红楼梦〉作者的问题》的论文。他从字、词出现的频率入手,首先对全书的字词做了统计,共有 72.960 4 万字。把字按出现的次数做了频数分析。发现频数最高的字,不是“宝”也不是“玉”,而是“了”、“的”、“不”等虚词,然后通过计算机进行统计、处理、分析,对《红楼梦》后 40 回系高鹗所做这一流行看法提出了异议,认为 120 回均系曹雪芹所做。会上有些人支持陈炳藻的观点,他们认为续书人再怎样刻意模仿,也做不到各种性质的词的使用频率与原作如此相似。

上面引例使我们对如何从统计学的角度来分析实际问题略见一斑。其实在现实的生产生活中,统计学的应用也是相当广泛的。比如,天气预报、股市预测、产品检验、民意测验、经济形势分析等等都会用到统计学的理论与知识。那么读者会问:统计学是怎样的一门学科呢?它是用来解决什么样的问题呢?本章就针对上面的问题予以回答,逐一对统计学的发展历史、研究范畴、学科特点与分类、应用领域以及统计学与其他学科的关系等内容进行介绍。

1.1 统计学的历史起源

统计学是一门很古老的科学,一般认为其学理研究始于古希腊的亚里士多德时代,迄今已有 2 300 多年的历史。统计的英文“statistics”一词源于法语“stataus”,在中世纪逐渐演变成 state(国家、状态),意指对国家的状态做调查研究。17 世纪,“statistics”一词开始被人们应用,用来指政府部门记录人们出生和死亡信息的工作。

1.1.1 统计学的创立

17 世纪中叶至 18 世纪中叶是统计学的创立时期,这一时期出现了统计学的雏形并形成了一定的学术派别,主要分为国势学和政治算术两个学派。

国势学派又称记述学派，产生于17世纪的德国。由于该学派是对国家的人口、版图、政体、财政、军备等方面做文字性记述，故又称记述学派。其主要代表人物是海尔曼·康令和阿亨华尔。康令是第一个在德国黑尔姆斯太特大学以“国势学”为题讲授政治活动家应具备的知识。阿亨华尔在格丁根大学开设“国家学”课程，其主要著作是《近代欧洲各国国势学纲要》，书中讲述“一国或多数国家的显著事项”，主要用对比分析的方法比较各国实力的强弱，为德国的君主政体服务。“国势学”与“统计学”明显的不同在于它不用数字资料。

政治算术就是对社会经济现象进行数量性描述和分析比较。它的兴起完全是受当时英国社会形势影响的结果。当时鼠疫流行，伦敦市民开始对死亡、出生等含有大量数字的报告关心起来，一个叫约翰·格朗特（John Graunt）的学者发现了不少规律性现象：出生婴儿中男婴比例稍高、婴儿死亡率较大、男性在各年龄组中死亡率高于女性等。约翰·格朗特的一个朋友叫威廉·配第（William Petty），继承并发展了格朗特的工作，他用计量和比较的方法，将英国国力与法国、意大利、荷兰等国家进行比较，目的是证明英国比其他各国强大，学者Petty被认为是政治算术的创立者，他提出了一套较系统的方法对社会经济现象进行数量性的描述和分析比较。代表作有《政治算术》，但并未采用“统计学”一词。马克思在《资本论》中评价配第是“政治经济学之父”。

1.1.2 统计学的发展

自17世纪中叶，经过几代统计学家的努力，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。描述统计是用来描绘（describe）或总结（summarize）观察量的基本情况的统计总称。描述统计学研究如何取得反映客观现象的数据，并通过图表形式对所收集的数据进行加工处理和显示，进而通过综合概括与分析得出反映客观现象的规律性数量特征。

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，逐渐形成了现代统计学的基本框架。

统计研究过程的起点是统计数据，终点是探索出客观现象内在的数量规律性。在这一过程中，如果搜集到的是总体数据（如普查数据），则经过描述统计之后就可以达到认识总体数量规律性的目的；如果所获得的只是研究总体的一部分数据（样本数据），要找到总体的数量规律性，则必须应用概率论等相关理论并根据样本信息对总体进行科学的推断。

描述统计是整个统计学的基础，推断统计则是现代统计学的主要内容。由于在对现实问题的研究中，所获得的数据主要是样本数据，因此，推断统计在现代统计学中的地位和作用越来越重要，已成为统计学的核心内容。从描述统计学发展到推断统计学，既反映了统计学发展的巨大成就，也是统计学发展成熟的重要标志。

在统计学的发展过程中，做出突出贡献的科学家有很多，主要可概括为三类。

1.1.2.1 与概率论的发展相关的代表人物

概率，最初来自赌博（gambling），每一个赌者都想准确地估测掷色子的某一结果出现的可能性。作为数理统计的基础，概率论最早出现在法国，遗憾的是当时法国概率论与英国的政治算术之间没有发生过任何交流，否则统计学的发展进程会进一步加快。对概率论的发展做出重要贡献的数学家有以下几位：

伯努利（J. Bernoulli, 1654—1705），瑞士人。20世纪初，主要探讨使中心极限定理

成立的最广泛的条件，伯努利是第一个研究这一问题的数学家，他于1713年首先提出后人称之为“大数定律”的极限定理，并系统地论证了“大数定律”。即样本容量越大，样本统计数与总体参数之差越小。最大贡献就是“伯努利大数法则”，但概率仅限于离散值。简单地讲，大数定律就是“当试验次数足够多时，事件发生的频率无穷接近于该事件发生的概率”。该描述即伯努利大数定律。

拉普拉斯 (P. S. Laplace, 1749—1827)，法国人。1812年发表了《概率分析理论》一书，在该书中总结了当时整个概率论的研究，论述了概率在选举、审判、调查、气象等方面的应用。在分布的极限定理方面走出了根本性的一步，证明了二项分布的极限分布是正态分布。拉普拉斯改进了他的证明并把二项分布推广为更一般的分布。他还系统地把概率论方法应用到统计学研究中，建立了严密的概率数学理论，并广泛应用到人口统计、天文学等领域。拉普拉斯也是一位伟大的天文学家，在研究天体问题的过程中，创造和发展了许多数学方法，在科学技术的各个领域有着广泛的应用。他对统计学的突出贡献主要体现在两方面：一是提出由部分推断总体的抽样调查方法这一重大的思想方法，二是将微积分引入到概率计算上，这是个创举，成功地解决了连续型随机变量的概率计算问题。

高斯 (Gauss, 1777—1855)，德国人。18岁的高斯发现了质数分布定理和最小二乘法。通过对足够多的测量数据进行处理后，可以得到一个新的概率性质的测量结果。在此基础上，他专注于曲面与曲线的计算，成功得到高斯钟形曲线（正态分布曲线）。其函数被命名为标准正态分布（或高斯分布）。

凯特勒 (Quetelet, 1796—1874)，比利时人。他被称为“近代统计学之父”，提出了著名的“平均人”概念，即人类的身高、体重、智商、情商等诸多测量值都围绕平均数波动且符合正态分布，因此正态分布又叫“凯特勒分布”。他认为统计学不仅要记述各国的国情，研究社会现象的静态，而且要研究社会生活的动态，研究社会现象背后的规律性。凯特勒的这一思想为近代统计学的科学化奠定了基础。其主要著作有：《概率论书简》，他主张用研究自然科学的方法研究社会现象，正式把古典概率论引进统计学，使统计学进入一个新的发展阶段。

1.1.2.2 与描述统计学的发展相关的代表人物

高尔顿 (F. Galton, 1822—1911)，英国人。高尔顿 (Galton) 通过测量父亲和儿子的身高，发现了“向均数回归”这一现象，并提出了回归的思想。在研究父子身高关系时，提出了用“相关系数”度量二者之间的关系，并给出了明确的公式及计算方法。这样，高尔顿首次提出了“相关”这一名词。1901年，高尔顿、皮尔逊、韦尔登共同创办《生物统计》杂志，并为该杂志提供了丰厚的资金支持，使得该杂志成为世界上第一本印有全彩图片的期刊，至今该杂志仍然是统计学领域的卓越期刊。其中戈赛特的 t 检验的提出，就发表在这一杂志上。

卡·皮尔逊 (K. Pearson, 1857—1936)，英国人。皮尔逊是早期统计学的创始人之一，在他年轻的时候，受高尔顿的影响，对统计学开始感兴趣。她对统计学的贡献主要可归纳为四个方面：(1) 当高尔顿离开统计学领域，转而研究其他问题的时候，皮尔逊接替了他的工作，接管了高尔顿的生物统计实验室，而且后来成了《生物统计》期刊的唯一编辑（共同合伙人高尔顿和威尔登已去世）。(2) 皮尔逊发现了“偏态分布”（这之前拉普拉斯发明了正态分布），皮尔逊声称，只要有均数、标准差、峰度、对称度，就可以去完整

地描述一个分布。当然这些东西在以后被证实存在很多问题。(3) 皮尔逊最大的成就之一就是创造出了拟合优度检验,也就是卡方拟合优度检验。这一检验即使在今天依然有其重要性,它可以模拟现实中不同的数学模型,然后利用拟合优度检验来确定哪一个更好。(4) 皮尔逊完善了高尔顿提出的相关系数,并证明了回归中的复相关系数,是复相关系数的提出者。

1.1.2.3 与推断统计学的发展相关的代表人物

歌赛特(W. S. Gosset, 1777—1855),英国人。据说他同时拥有数学和化学两个学位。与许多学者一样,歌赛特当时并没有直接从事统计学的研究,当时统计学甚至还不上一门学科。他当时从事的是啤酒酿造行业,然而正是在这一似乎与统计无关的行业里,他做了一项统计相关的研究,想弄清楚发酵时需要添加多少酵母最为合适。他最重要的一个贡献就是提出了小样本的检验思想,现在看来这似乎并无任何出奇之处。但在当时,统计学几乎就是大样本的科学,当时皮尔逊几乎所有工作都是基于大样本的假设。然而歌赛特根据自己的经验认为,有些情况下,大样本对于研究者来说太过于奢侈了,必须专注于小样本。但一旦采用小样本分析,就无可避免地会牵扯到误差的问题。在大样本情况下,你可以假定没有误差或者误差很小甚至可以忽略不计,而小样本必须考虑到这一问题。那么小样本情况下,误差究竟有多大呢?歌赛特通过自己的不断努力,最终于1908年发表了一篇极为重要的文章 the probable error of the mean,提出了 t 分布,这是至今仍在广泛应用的 t 检验(也叫 student t 检验)的基础。

费雪(R. A. Fisher, 1890—1962),英国人。费雪的一生主要在统计学和生物学两个领域从事科学研究。他的著作中不乏在生物学方面应用统计方法的例子。正是由于他在统计学方面的深厚造诣,才使他在生物学方面能做出突出的贡献。费雪对统计学的贡献大致可分为两方面:(1) 关于统计推理的基础理论;(2) 统计方法论与技术。

费雪建立了样本分布、区间估计以及假设检验等综合理论并且确立了数理学的教学框架,创立了参数估计和假设检验的推断统计学理论体系,为现代统计学核心内容即推断统计学的发展做出了卓越的贡献。

1.2 学科性质及定义

1.2.1 统计的含义

在现实生活中,提到“统计”一词时常有不同理解。比如“根据统计”中的“统计”,一般指统计资料;“我是搞统计的”中的“统计”,一般指统计工作;“大学课程中包括统计”中的“统计”,一般指统计学。所以,统计(statistics)有三种含义,即统计工作、统计资料和统计学。

统计工作(statistical operation)是指具体从事的统计设计、资料收集与整理、分析预测,以及提供各种统计资料的实践活动的总称。例如,计算反映A、B两班成绩的各种指标,并且进行分析评价的过程就是统计工作。

统计资料(statistical data)是指在统计工作过程中所获得的以资料数据表现的信息资料。例如,反映A、B两班成绩的各种指标就是统计资料。

统计学 (statistics) 是阐述正确指导统计活动科学原理和方法的学科体系。从用文字来描述、分析问题发展为用数据来描述、分析和认识问题。也就是说, 统计学已经发展成为一门运用数据来观察、描述、探究、分析、把握、归纳、总结和认识问题的学科。从广义上讲, 统计学是包括自然科学和社会科学在内的统计科学理论的总和。本书认为: 统计学是一门收集、整理、分析数据的方法论科学, 目的是探索数据的内在规律性以科学地认识客观事物。主要论述对社会经济现象如何进行统计设计、统计调查和统计整理及分析统计资料的理论和方法。

那么什么是统计数据的内在规律呢? 下面用几个简单的例子来说明一下。

【例 1—1】 一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, 新生婴儿的性别比例似乎没什么规律可循。但如果对新生婴儿的性别进行大量观察, 即观察成千上万个或者更多时, 就会发现性别比例还是有规律的, 即婴儿总数中男孩要多于女孩, 大致为每生育 100 个女孩, 就有 105 个左右的男孩。这个 $105:100$ 的比例就是新生婴儿男女性别的数量规律性, 古今中外都大致相同, 它是人类社会长期遗传与发展的结果。因为人类社会要延续、要发展, 就要保持男女人数的大致相同, 那么有人会问, 新生婴儿男多于女, 不是性别不平衡了吗? 是的, 新生儿时男多于女会出现性别不平衡, 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, 导致男性的预期寿命平均比女性短, 长寿的男性要少于女性。从一个国家看, 如果没有人为的 B 超检查后堕胎等干扰, 其规律是婴幼儿时男性略多于女性, 中青年时男女人数大致相同, 老年时女性又略多于男性。这样既保证人类在中青年结婚生育时性别的大致平衡, 也使得在人口总数上男女也大体相当, 有利于人类社会的进化和发展。对人类性别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之一。

【例 1—2】 我们都玩过掷色子或者抛硬币的游戏, 都知道随机地掷一次色子出现的点数和随机抛一次硬币出现正面或者反面的情况是不能事先知道的, 也就是说这样的游戏或者实验充满了不确定性或者偶然性。机会游戏或者赌博正是利用了这种不确定性和偶然性才能够吸引人。但是当我们进行大量的试验和观察时, 即不断做重复的实验, 就会发现掷一枚硬币出现正面反面的比率大体相同, 都是 $0.5:0.5$, 实验的次数越多, 出现正面和反面的可能性就接近 $1/2$ 这一稳定的数值。同样, 重复掷色子的次数越多, 出现 1 至 6 中任意点数的可能性就越接近 $1/6$ 。这里的 $1/2$ 和 $1/6$ 就是掷硬币和掷色子出现的某一特定结果的概率, 这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, 如果其他实验条件相对固定, 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, 粮食产量增加量逐渐减少。当施肥量增加到一定数值, 产量不再增加。这时如果再增加肥料, 产量反而会减少。这一施肥量与粮食的数量关系就是我们要探索的规律性。当我们从大量实验数据中用统计方法探索到施肥量与粮食之间的数量关系时, 就可以考虑肥料的费用并选择最佳的施肥效果。

【例 1—4】 与上面例 1—3 相似但又不同的是商品广告费用与销售额的关系。一般而言, 随着广告费用的增加, 商品知名度和销售额会相应增加。对于不同的商品, 投入相同的广告费带来的销售额显然是不尽相同的。即使是同一种商品, 投入的第一个 50 万元广

告费、第二个 50 万元广告费和第三个 50 万元广告费等所带来的销售额也是不同的，统计方法就是要从广告费与销售额的数据中找出其内在的数量规律，从而确定最佳广告策略。

【例 1—5】 天津汽车制造厂与丰田公司合资生产的威驰（Vios）汽车安装的是 8A 型发动机，排量为 1.342L，公司声称其手动挡汽车每百公里油耗（等速情况下）不超过 5L。要检验购买的车辆是否达到了说明书上的节油标准，就需要计算在高速公路上行驶的平均油耗。假定你的汽车平均每百公里油耗为 5.5L，那么你的汽车是属于质量达标汽车中恰巧由于运气不好碰上的个例呢，还是该批产品本身就不合格呢？统计方法可以做出检验并给予回答。

【例 1—6】 某企业开发出一种新的化妆品，在正式投产前，需要根据市场需求情况制定价格和销售策略等。该企业委托某市场调查公司在全市 200 万户家庭中抽取 1 000 户家庭作为样本，免费赠送给这 1 000 户家庭试用，然后了解该化妆品的销售前景。如何科学地从该市 200 万户家庭中抽取 1 000 户进行调查，并由这 1 000 户家庭反馈的市场信息科学地推断出全市 200 万户家庭对化妆品的购买意愿这一数量规律，是统计工作者的任务。

为什么统计方法能从大量的数据观察和处理探索出其内在的数量规律性呢？这是由于统计方法的特性以及客观规律本身的特点共同决定的。任何客观事物都是必然性与偶然性的对立统一，同样，任何一个事物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。前面的六个例子，本身都存在必然性的数量规律，但它们的结果都是有差异的，其表现形式都是充满偶然性的。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（或推理）出事物总体的数量规律性。

统计一词的 3 种含义是密切联系、缺一不可的。统计工作与统计资料是工作过程与成果的关系，统计学与统计工作是理论与实践的关系。因此，统计一词是统计工作、统计资料、统计学的综合概括，是统计的过程与结果、理论与实践的辩证与统一。

统计的含义及其内部联系如图 1—1 所示。

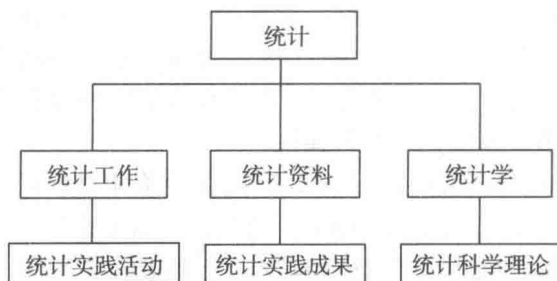


图 1—1 统计学的含义及其内部关系

1.2.2 统计学的学科性质

统计学是一门认识方法论科学。统计活动的主要内容是数据的收集、整理和分析，所以统计学是研究如何收集、整理和分析数据，以及如何对现象总体进行抽样推断、趋势分

析的认识方法论科学。

从方法论科学的角度来看，统计学主要研究客观现象总体所表现出来的统计绝对数、统计相对数和统计平均数，以及它们在时间、空间变动及相互关系的计算分析方法。统计意义上的时间变动是个动态概念，空间变动是个静态概念。也就是说，统计学既要研究现象数值静态变动的计算分析方法，又要研究现象数值动态变动的计算分析方法。

客观现象总体所反映的数量特征和数量关系，与统计学的研究方法存在着相互依存、相互促进的关系。一方面，统计方法的提炼和检验来自统计实践；另一方面，只有经过漫长的历史时期，统计实践才会有较大的发展，统计学才可能形成。

早期的统计学研究总是和实际问题的数量分析结合在一起，往往就事论事，因此被称为以客观现象为主体的实质性统计学。直到 19 世纪，比利时统计学家凯特勒将概率论引入了统计学，论证了社会经济生活中的随机现象也有一定的规律性，统计学才步入了学科自身的发展轨道。

19 世纪中叶以后，随着社会经济的发展和科学技术的进步，统计学获得了迅速的发展。一方面，各行各业创立了许多有效的统计方法，例如生物遗产学上的相关和回归分析方法、农艺学的区间设计方法、人口学上的抽样方法、教育心理学的假设检验方法、经济学的物价指数方法等；另一方面，概率论为统计学研究不确定的随机现象、认识统计规律性，提供了理论与方法的数学框架，而计算机科学和信息处理技术的发展则为统计方法论的可操作性提供了支持。在这些因素的共同作用下，以统计方法为中心的方法论统计学才逐步形成。当然，方法论统计学并不排斥从实质性科学独立出来的各专业的统计学方法的并存，如投入统计学、质量控制统计学等。

1.3 研究对象及应用领域

1.3.1 统计学的应用领域

目前，统计方法已被应用到自然科学和社会科学的众多领域，统计学也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采用统计方法。表 1—1 列出了统计学的一些应用领域。

表 1—1 统计学的一些应用领域

精算 (actuarial work)	水文学 (hydrology)
农业 (agriculture)	工业 (industry)
人口统计学 (demography)	文学 (literature)
教育学 (education)	市场营销学 (marketing)
金融 (finance)	制药学 (pharmaceutics)
遗传学 (genetics)	质量控制 (quality control)
.....	

统计学提供了探索数据内在规律的一套方法。例如，在进行农作物试验时，如果其他

试验条件相同，我们会发现某种粮食作物的产量会随着某种肥料施肥量的增加而增加。当最初增加施肥量时，产量增加较快，以后增加同样的施肥量，粮食产量的增加会逐渐减少；当施肥量达到一定的数值时，产量不再增加；这时再增加施肥量，产量反而会减少。粮食产量和施肥量之间的这种数量关系，就是我们所要探索的数量规律性，可以用统计学中的相关关系来加以解释。即如果我们能从大量的试验数据中用统计方法找出产量与施肥量之间的数量关系，就可以确定出最佳的施肥量，以求取得最大的收益。

上述例子说明，就一次观察或试验来说，其结果往往是随机的，但通过多次观察或试验得到大量的统计数据，利用统计方法是可以探索出其内在的数量规律性的。因此，统计学是一门应用性很强的学科，几乎所有的学科都要研究和分析数据，统计学与几乎所有的学科领域都有着或多或少的联系。这种联系表现为：统计方法可以帮助其他学科探索学科内在的数量规律性，而对这种数量规律性的解释以及各学科内在的规律，则需要由各学科的研究来完成。也就是要在用统计方法进行定量分析的基础上，应用各学科的专业知识对统计分析结果做出合理的解释和分析，才能得出令人满意的结论。

1.3.2 统计学的研究对象及其特点

统计学界对统计学研究对象的界定存在混杂现象，科学地界定统计学研究对象是科学定义统计学的前提，只有真正理解统计学研究对象的意义，才能真正认识统计学的思想精髓。

本书认为，统计工作是研究社会经济总体数量方面的调研活动，这就决定了统计学的研究对象是大量社会经济现象在数量方面的问题，即通过社会经济现象在一定时间、地点、条件下的数量表现、数量关系及数量界定来揭示其规律。由于统计学与统计工作是理论与实践的关系，因此，二者的研究对象应该是一致的，即统计工作的对象也应是大量的社会经济现象的数量方面。

统计学的研究对象具有如下特点：

1.3.2.1 数量性

统计发展的历史告诉我们：统计学的首要特点是其数量性，亦即研究社会经济现象的数量方面（数量性）是统计学研究对象的基本特点。因为，数字是统计的语言，数据资料是统计的原料。一切客观事物都有质和量两个方面，事物的质与量总是密切联系、共同决定着事物的性质。没有无量的质，也没有无质的量。但在认识的角度上，质和量是可以区分的，可以在一定的质的情况下，单独地研究数量方面，也可通过认识事物的量进而认识事物的质。因此，事物的数量是我们认识客观现实的重要方面，通过分析统计数据资料，研究和掌握统计规律性，就可以达到统计分析研究的目的。例如，要研究国民生产总值，就要对其数量、构成及数量变化趋势等进行分析，这样才能正确地分析和研究国民生产总值的规律性。

统计研究对象的数量性，是统计区别于其他社会经济调查研究活动的根本特点。社会经济统计不是“纯数量”的研究，这是统计与数学的重要区别。统计反应的数量具有具体性，而数学反应的数量具有抽象性。例如，反映社会经济现象特征的统计数据，都是明确规定了具体时间、地点以及条件下的数量。2013年（时间）我国（地点）初步核算的国内生产总值为568 845亿元（国家统计局，2014-02-24），这个数量就是2013年我国国