

教育部普通高等教育“十五”国家级规划教材
教育部研究生工作办公室推荐研究生教学用书
全国统计教材编委会推荐教材

Statistics
in Survey
Research

调查研究中的 (第2版)
统计分析法

柯惠新 编著
沈浩

中国传媒大学 出版社

调查研究中的统计分析法

柯惠新 沈浩 编著

中国传媒大学出版社

图书在版编目 (CIP) 数据

调查研究中的统计分析法/柯惠新,沈浩著. -- (第2版), -- 北京:中国传媒大学出版社,2005.4

ISBN 7-81085-261-2

I. 调… II. ①柯… ②沈… III. 统计调查-统计分析-方法 IV. C812

中国版本图书馆 CIP 数据核字 (2005) 第 022342 号

调查研究中的统计分析法 (第 2 版)

作 者: 柯惠新 沈 浩

责任编辑: 愚 言

封面设计: 北京灵麒时代广告有限公司

出版发行: 中国传媒大学出版社

社 址: 北京市朝阳区定福庄东街 1 号 邮编: 100024

电 话: 65450532 或 65450528 传真: 010-65779140

网 址: <http://www.cucp.com.cn>

经 销: 新华书店总店北京发行所

印 刷: 北京中科印刷有限公司

开 本: 787×1092 毫米 1/16

印 张: 38.75

版 次: 2005 年 8 月第 2 版 2005 年 8 月第 1 次印刷

ISBN 7-81085-261-2/K·119 定价: 76.00 元

版权所有 翻印必究 印装错误 负责调换

第二版前言

《调查研究中的统计分析法》自 1992 年出版至今，已经过去了十多个年头。其间被全国不少高等学校相关专业所采用，把它作为本科生或研究生的教材或参考书；也广泛地被从事市场调查和社会调查的研究者和实际工作者所使用，把它作为一本提供调查方法的主要参考书。对于这本书的框架、内容和风格，各方面使用者的反映良好。

在这十多年中，本书经由中国传媒大学出版社（原北京广播学院出版社）多次印刷，其间虽然封面变换了三次，个别内容也作过一些改动，但是由于我本人时间和精力所限，对它的内容却一直未能系统地进行修订。

2000 年，本书被列为全国统计教材编委会第五轮推荐使用的教材（统编办字[2000]15 号），2001 年，被列为教育部研究生工作办公室推荐的研究生教学用书（教研厅[2001]1 号），2002 年被列为教育部“十五”规划国家级教材选题（教高函[2002]17 号）。我所在学校中国传媒大学（原北京广播学院）有关的领导和学校出版社对本教材的再版也很重视。这些都促使我下决心对本教材进行一次系统的修订。

为了做好修订工作，我曾认真地向多位长期使用本教材的高校教师、学生和其他相关领域的研究人员征求意见，在此基础上，形成了本教材修订的基本思路。据我们的了解，被征询意见的使用者对本教材的框架、基本内容和风格都是充分肯定的，希望修订时能予以保留；同时认为书中所使用的部分案例和相应的软件版本已经陈旧，希望能进行更新。

综合国家教育部对规划教材的指导意见以及诸多读者、使用者的建议，我们为本教材的再版进行了以下的修订工作：

1. 保留原教材二十一章的框架、基本内容和风格，对原文中部分阐述不够准确之处和明显的印刷错误进行了修改；
2. 在第二章《描述性统计学》中增加了“其他描述分布的统计图和统计量”的内容；
3. 在第三章《概率分布》中增加了“正态分布的常用性质”的内容；
4. 将原附录 II《LISREL 的使用说明》全部改写为目前更为常用的《AMOS 5.0 软件使用说明》；
5. 更换了大部分应用实例，同时充实了应用实例的内容，即，除了结合实例介绍软

件 SPSS 11.5 for Windows 的使用说明之外,还对统计方法应用的注意点做了必要的说明;

6. 删除了原附录 I SPSS|PC+的软件介绍,将 SPSS 11.5 for Windows 的使用说明具体融合到各章的应用实例中;
7. 将贯穿原教材的《1990 年亚运会广播电视宣传效果调查》的案例更新为《2004 年雅典奥运会媒介传播效果研究》和《2000—2001 北京奥运申办媒介传播效果研究》的调查案例;
8. 增加了关于市场调查研究方面的案例,其中《1998 年郑州大瓶装水市场调查》的案例也贯穿了教材的多个章节;
9. 增加了有关新媒体研究、语言学等领域的例子;
10. 对大部分章节的习题进行了修订,更换了一些时效性较强的习题,但仍保留了那些虽然比较旧、但启发性较强的“经典”习题。所有的计算题都给出了参考答案;
11. 将原教材中手工绘制的图形全部更换成计算机软件绘制的图形。

修订后的再版教材适用的范围与原教材相同,可用作高等学校人文科学相关专业的本科生和研究生的教材,也可用作各种社会调查、受众调查、市场调查等方面的参考书或培训教材。课程的安排可根据专业的实际和需要分为三个层次:

1. 一学年课程:72 学时(4/周一学期)的课程学习,加上 36 学时(2/周一学期)结合 SPSS 软件使用 and 实际调查数据处理分析的课程学习。内容:全部二十一章,包括应用实例和附录。目标:熟练掌握各种统计分析方法及相应的统计软件;
2. 一学期课程:72 学时(4/周一学期)的课程学习。内容:二十一章的大部分章节,不包括上机练习。目标:基本掌握各种常用统计方法及相应的统计软件。
3. 半学期课程:36 学时(4/周半学期或 2/周一学期)的课程学习。内容:前三十三章的大部分章节,不包括上机练习。目标:掌握统计学入门的基本概念和方法。

本教材修订的分工为:我本人负责全书第一章至第二十章正文的修订和全书终稿的审定;沈浩负责各章的应用实例、附录和二十一章的修订;我的研究生(01-04 级)协助进行了原教材的扫描、电子版的整理、公式和图表的绘制以及各章习题和部分例题的修订。

具体帮助整理或录入原教材电子版的同学主要有:朱川燕、郑丽、田卉、南隼、陈洲、韩涛、黄刚、陈晓华、魏思华、崔蕴芳、曾兴、黄可、谢婷婷、王锡苓、姚逸晨、柴岳强、范欣珩、杨守睿。

具体帮助绘制电子版图表的同学主要有：黄鸣、范欣珩。

具体参与各章部分习题修订的同学主要有：齐之娉（第一章）、黄可（第二章）、范欣珩（第三、四章）、赵巍（第五章）、何业文（第六章）、王仕（第七章）、杨克清（第八、十二章）、王静（第九章）、黄鸣（第十章）、施德俊（第十一、十六章）、董瑞峰（第十四章）、李馨（第十五章）、秦福贵（第十七章）、吕飞（第十八章）、朱川燕（第十九章）。

具体参与各章应用实例修订的同学主要有：黄鸣（第九、十七章）、姚逸晨（第十八、二十章）、何业文（第六章）、刘文明（附录）。

具体参与部分例题修订的同学主要有：孙江华（第十六章），吕飞、王静、王宁（第二十章）。

具体帮助校对的同学主要有：范欣珩、吴志国、葛晶晶、李维、吴璟、蒲璐璐。

本教材的修订得到了中国传媒大学校领导、教务处、研究生处和新闻传播学院领导的大力支持和帮助；我的博士研究生肖明、马广斌、黄刚、孙江华对修订工作提供了具体的建议和素材；中国传媒大学出版社的闵惠泉主编、吴三军、杜丽华和欣雯老师为此书的修订付出了大量的努力。对于上述领导、老师和同学们的诚挚帮助，我们在此特表示衷心的感谢。

此外，还要特别感谢 SPSS 公司为我们提供了正版的 SPSS 11.5 for Windows 软件用于本教材的修订；感谢央视-索福瑞公司资助我们完成了《2004 年雅典奥运会媒介传播效果研究》的课题，并得以将其用于教材的案例中；感谢郑州市中美纯水公司、《中国妇女报》、陕西人民广播电台、中国社会科学院哲学研究所等单位允许我们在本教材的修订中使用他们的部分调研数据。同时，还要感谢所有在本教材所列参考书目中的作者，在本教材的编写和修订的过程中，从他们的著作吸取了不少的营养。

鉴于我们的水平所限，修订后的教材仍会有不少错误和不妥之处，恳请广大读者提出宝贵意见。希望这本教材能够继续为我国调查研究数据的统计分析水平的提高尽微薄之力。

柯惠新

2005 年 1 月 25 日

原版序言

半个世纪以来，统计学有了长足的进步。从描述到分析，从单项分析到多元分析，从手算到电子计算机的运用，使统计从一个不起眼的课题发展成为一门与多学科（尤其是应用科学、生物科学和人文科学）相互影响，以致内容深广的科学。它对许多实用科学和软科学的数量化起着中心的作用，所以统计方法对于现代社会的各个部门的管理、运筹和发展都是不可缺少的重要工具。

调查研究是工商、文教、卫生、经济、心理、传播等许多部门和学科的重要手段。而统计方法运用得好，可以使调查研究的设计与分析更科学、更深入并且更节约。

从需要到掌握，这中间还要有个过程。有人说“统计还不容易”，可能他只停留在（满足于）描述性统计的低阶段。更多的人说“统计不容易学”，这是因为他从未遇到一位好老师或一本好书。

本书著者柯惠新博士是位重视统计应用和普及的新秀。不仅在理论上有很好的素养，在应用上身体力行，而且在教学方面也很热心认真。她是在“跑遍书店也未找到合用教材”时下定决心动笔的。

读了本书原稿，我为大家感到庆幸。因为本书不仅顾及了广大非数理专业读者当前的实际需要，而且注意到扩大读者视野，并带向国际先进水平。“深入浅出，引人入胜”正是读完本书的第一印象。

汤旦林

1992年3月12日

于北京

原版前言

各种类型的大规模调查研究的成败在很大程度上取决于问卷的设计是否科学、所抽取的样本是否有代表性、调查的实施是否可以信赖、数据的处理是否准确无误、数据中所包含的大量信息是否最大限度地得到了提炼和分析等等。针对社会科学、行为科学、管理科学特别是各类调查中所面临的种种挑战性问题,需要有一本既比较全面又针对性强的关于统计分析方法以及计算机统计软件使用方法的教材或工具书。目前,国内市面上出售的文科用统计学方面的教材远远不能满足这一要求;而理工科用的统计学教材虽然统计方法比较全面,但是没有很好地针对这些学科中许多复杂的实际问题,而且需要用到高等数学知识,使社会科学工作者、统计工作者和文科学生难以接受。为此我们下决心编写这本教材兼统计软件使用说明。本书主要根据我们多年来对中国传媒大学(原北京广播学院)新闻、广告、管理类的本科生和研究生讲授的统计学讲义、对广播电视受众调研工作者培训的讲义,以及近年来参加广播电视方面的受众调查研究的数据处理分析方面的实践体会,再加上参阅大量国内外有关教材和论著而编写的。

本书的特点是实用性强,内容广泛,并有所侧重,讲解深入浅出。对各种方法的理论背景进行通俗的、描述性的说明,不做严格的数学推导。强调对统计思想和方法的理解与应用,强调对使用现有统计软件能力的培养。即使读者一时还不能对软件运用自如,也可以学会如何从计算机的输出结果中找到自己所需的信息。读者只要具有高中程度,即可读懂本书。除最后一章(加*号)的内容有些难度外,其余各章都是不难学会应用的。从事数理统计研究和教学的同志从本书的应用实例和计算机统计方法中或许也能得到一些启迪。书中关于 LISREL 模型和方法的介绍与应用在国内出版物中还属首次。

本书分五大部分,共有二十一章。计算机统计软件 SPSS | PC+和 LISREL 的使用说明在附录中给出。本书的前五部分由柯惠新编写,SPSS | PC+的使用说明由黄京华编写,LISREL 的使用说明由沈浩编写。书中应用实例的计算机输出结果也全部由黄京华和沈浩核对后给出。

在本书的编写过程中,北京广播学院的赵玉明教授、王纪言副教授、甘章泉教授、李鉴增副教授等都曾给予各方面的鼓励和支持。原北京广播学院出版社在经费困难的情况下出版了本书。几年来,中国现场统计研究会的许多老师和同志们也一直在关心和支持着我们的工作。汤旦林教授在百忙之中审阅了全稿,提出了许多宝贵中肯的意见,谨此一并表示衷心的感谢。

限于编者的学识和水平,书中错误及不妥之处在所难免,恳请广大读者不吝赐教。

作者

1992年1月21日

绪言 统计学发展史简介

此简介并非统计学发展的全貌，也不准备逐一介绍统计史上有关伟人的业绩。在此仅是向初次接触数理统计学的读者简单地就我们身边比较熟悉的事情，片断地说明有史以来到今天为止，统计学是沿着怎样的路径发展而来的，从而让读者有一个初步大概的印象。

一、统计学的起源

statistics（统计学）一词起源于法语 status（状态）。该词自中世纪以来逐渐演变成含政治意味的 state（国家、状态）。因此，统计学原来包含的意义是指对国家的状态做调查研究。古代的中国和埃及都有过对国家的大事做统计调查的历史。到了希腊、罗马时代，社会机构日益复杂化，对于从政者来说，掌握国家全面情况的统计知识就变得越来越有用了。亚里斯多德所著《国家论》一书中，对很多国家的政治、学问、宗教、艺术和风俗等进行了详细的记述，该书可以认为是后来发展起来的所谓“国势学”的先驱之作。

二、国势学

创立国势学体系的可以说是德国的 H. Conring（1606~1681）。他对许多国家的状况进行了记述，并在各大学进行讲演。Conring 的国势学和我们现在所说的统计学是明显不同的。他虽然对人口、版图、政体、财政、军备等方面进行了文字性的记述，但几乎不用数字资料。到了 18 世纪，G. Achenwall（1719~1772）对统计学的性质、意义及范围明确定义为“把国家的显著事项全部记述下来的学科”，并称此学科为 statistik（德文：统计学）。他对 Conring 的业绩给予了很高的评价，称他为“统计学之父”。不过后人一般都称 Achenwall 为“统计学之父”。他的主要著作作为《近代欧洲各国国势学概论》。Conring 和 Achenwall 都很少做数量方面的观察，没有触及到统计资料的实质。十七八世纪，国势学派的学者们常常在各地的大学内进行讲演，因此又被称为“德国大学派统计学”。他们完全不用数量测定值，对图形表格、数字式子十分蔑视，这与英国的政治算术学派是很不相同的。

德国的地理学家 Büsching（1724~1793）把统计学看作是地理学的一部分。他收集了各国的资料，并对各国的资料进行了分类比较。丹麦的地理、历史、语言学家 Anchersen（1700~1765）把 15 个国家的状况进行了分类整理，给出了十分容易理解的一览表，他因此被称为“尚表学派之祖”。使用数字、图表的尚祖学派和 Achenwall 的后继者们之间发生了争论。此后，以 Lüder（1760~1819）和 Knies（1812~1898）为首的学者对 Conring—Achenwall 派的统计学进行了激烈的批判。对“统计学作为一门学科其意义到底何在”展开了热烈的讨论。Knies 把当时已在英国发展起来的政治算术叫做统计学，认为它是收集、整理和表示资料的科学，即是一种方法论的科学，

而把 Conring-Achenwall 派的统计学叫做国势学。

三、政治算术

17 世纪, 政治算术学派统计学在英国兴起, 这完全是由于当时英国的社会形势影响所产生的结果。在 1348 年、1563 年、1592 年、1603 年和 1665 年鼠疫流行, 伦敦市民对于死亡、出生、结婚、洗礼等含大量数字的报告变得关心起来。1662 年, 伦敦的一个商人 J. Graunt (1620~1674) 著书《关于伦敦死亡表的观察》, 成为政治算术学派的鼻祖。他利用寺院提供的关于死亡和洗礼的资料, 首先制作了死亡表, 并指出某些疾病的死亡人数占全部死亡人数的比例是稳定的。他还发现了不少规律性的现象, 例如: 男女人数占总人口数的比例大致相等; 新生儿中男婴的比例稍高; 婴幼儿的死亡率较大; 都市的死亡率大于地方的死亡率等等。Graunt 所给出的各种结论在当时来说显得过于大胆和武断, 但是 20 年后, 关于人口方面的事实表明确实存在一些规律性的东西。认识到社会现象中存在规律性, 这是 Graunt 的一个伟大功绩。他所制作的死亡表, 直至 18 世纪末期都被用于计算人类的死亡率。在法国、荷兰等国, 考虑年金(抚恤金、养老金等)时也是以他的死亡表为基础的。

不过, 政治算术学派的代表人物一般被认为是英国的经济学家配弟(W. Petty, 1623~1687)。他是 Graunt 的朋友, 他继承和发展了 Graunt 的研究工作。根据对人口、土地、财政、经济等各方面的大量观察, 配弟完成了《政治算术》一书。在书中他第一次用计量和比较的方法, 将英国的国力与法、意、荷等国进行了比较研究, 目的是要论证英国比其他各国强大。他所感兴趣的是与政治有关的问题, 只是从数量方面来处理分析。他提出了一套较为系统的方法, 用于对社会经济现象进行数量性的描述和数量性的分析比较, 创立了政治算术学派的统计学。这是与排斥数量只讲观念的国势学派的统计学很不相同的。

属政治算术学派的天文学家 E. Halley (1656~1742) 在配弟对全球人口数目估算研究工作的基础上, 进一步做了更合理、更精确的人口估算, 这是在人口统计方面的极大贡献。特别是他通过对死亡率的研究制作了死亡表, 并根据该死亡表对人寿保险年金进行了精确的计算, 这是关于生命保险理论的最早的科学研究。以后各国都仿照 Halley 的方法进行了这方面的研究, 为生命保险理论打下了基础。遗憾的是, 虽然英国后来建立了生命保险公司, 但当时对 Halley 的算法似乎不很理解。以 Halley 的死亡表为根据来计算保险金额的新的保险公司一直到 18 世纪中期才出现。

英国政治算术学派的影响传播到了整个欧洲大陆, 此后涌现了一批(包括在德国的)学者, 特别是在人口统计方面提出了一系列以大量观测数据为基础的研究方法。

下面再谈谈数理统计学的基础概率论的发展。

四、概率论的起源与发展

最早涉及原始概率论问题的数学书的作者名叫 L. Pacioli (1445? ~1510)。他提出了赌博中的一个问题: 假设力量不平均的两个竞争者进行比赛, 但比赛中途停止, 没有最后结果, 这时, 赌金应该如何分配才合理? 这一类的问题叫做得分问题(Problem of points)。在那以后的 200 年间, 对得分问题的研究一直没有中断。在三次方程式的解法上曾引起极大争议的 G. Cardano (1501~1576) 是精通赌博的, 为

了防备当时在意大利赌场中流行的欺诈行为，他从数学上下功夫研究了各种赌博的方法，写出了一本《赌博者手册》。

不过概率论的真正历史被认为是从17世纪才开始的。在17世纪中期，B. Pascal (1623~1662) 和 P. d. Fermat (1601~1665) 就赌博中的得分问题经常通信交换意见，用数字方法处理这类问题的研究从此开始了。在当时的赌场中，常用投掷骰子、钱币或翻纸牌的办法进行赌博。有经验的赌博者可能知道某一种情况（事件）出现的可能性大一些或小一些。例如，把骰子连续扔三次，出现面朝上的点数之和为10的次数比9的次数要多些。Pascal 和 Fermat 提出了“概率”这一概念，用来描述某一事件发生的可能性。为了计算有关事件发生的可能次数，又发展了排列组合的理论以及集合论的理论。由此，种种的赌博问题也就迎刃而解了。C. Huygens (1629~1695) 所著《骰子赌博的理论》一书，是当时最大的一部力作。

对概率论进行了重要的研究并使之成为数学的一个分支的，是瑞士的大数学家伯努利 (J. Bernoulli, 1654~1705)。在他所著的四部巨作《推论法》(在他去世后的1713年才出版) 一书中，有以他自己的名字命名的法则——“伯努利大数法则”：若在一试验中事件 A 发生的概率为 p ，将此试验重复进行 n 次，设 A 发生的次数为 r_n ，那么当 n 充分大时

$$\left| \frac{r_n}{n} - p \right| > \varepsilon \quad (\varepsilon \text{ 为任意指定的正数})$$

成立的概率可以任意小。例如在投掷骰子时，出现1点的概率 $P = \frac{1}{6}$ 。设投掷 n 次时

1点出现 r_n 次，由于有偶然性， $\frac{r_n}{n}$ 不一定等于 $P = \frac{1}{6}$ 。但是随着 n 的逐渐增大， $\frac{r_n}{n}$ 与 p 的差可以越来越小，也即这个差大于 ε 的可能性越来越小。比较确切地说，当 n 趋于无穷大时， $\frac{r_n}{n}$ (在统计意义上) 趋近于 p 。这是一个十分重要的法则。

进入18世纪后，de Moivre (1667~1754) 所著《偶然论》一书出版，书中有类似于伯努利法则的大数定律，还有更精确的关于概率的数值计算法。

如果上述关于概率问题的研究方法当时能在政治算术上得以利用的话，那么社会统计学的进步可能会早得多、快得多。遗憾的是，在法国出现的概率论和在英国出现的政治算术之间，可以说是什么交流都没有发生过。

此后，在 Monmort、Buffon、D. Bernoulli、Bayes、Legendre、Lagrange 等研究的基础上，19世纪初，拉普拉斯 (Laplace, 1749~1827) 一举完成了《解析概率论》(1812) 这一大作，将概率的定义从有限的情形推广到了连续的情形，并将当时数学界发现的牛顿—莱布尼兹的微积分学应用到了概率的分析理论和计算方法上。虽然拉普拉斯还未给出严格的数学概念，但他很好地应用了高等数学的方法去研究概率论有关的数学理论，这一点得到了高度的评价。他还将概率论应用于统计学，提出了由部分的调查资料去推断全体的抽样统计法，例如由一部分地区的人口和出生率来估计全国人口数。

在 18 世纪,人们注意到了概率论与自然科学特别是力学、天文学等学科的关系。例如在观测天体运动时会有误差产生,虽经多次测量,由于有误差,得到的总是和真值不同的值。因此产生了如何推断真值的问题。虽然在 18 世纪, Cotes、Simpson、Legendre、D. Bernoulli 等也对此进行了大量的研究,但真正解决问题的是高斯(Gauss, 1777~1855)。根据大量的研究和经验,他提出误差值落在 (a, b) 区间的概率等于

该区间上曲线 $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$ 下的面积,这时称误差服从正态分布或高斯分布。正

态分布的发现,促使对实测值进行整理分析所用的最小二乘法方面的研究也得以开展,并被应用于实际的统计之中。

在以上这些一流数学家的努力下,概率论在理论、应用两方面都得到了很大的发展。

五、十九世纪的统计——凯特勒的功绩

19 世纪,德国大学派的统计学受到了批判和清算,人们认识到统计学的意义在于它是一门关于统计方法的方法论学科。比利时的凯特勒(A. Quetelet, 1796~1874)对此观念的形成给予了很大的影响。他以机械的社会观对社会的各种现象进行数量性的分析,开创了社会统计的一个新纪元。这一时代是所谓统计万能的时代。在 1830 年以后的 20 年间,统计学唤起了社会的普遍关心。不少国家有了官方统计,开办了统计学的杂志,成立了各种统计协会等等,同时,国势调查也十分盛行。统计学是由德国大学派和英国政治算术派发展而来的,凯特勒的功绩在于把握住了这两者的本质,在发展政治算术,应用数量观察分析方法的同时,将统计方法应用于社会生活的所有方面,可以说是统计学的新纪元。他学习了数学、物理学、天文学,并将这些知识应用于社会学。他的代表作《社会物理学》“是要给政治科学和精神科学附加上一种以观察和计算为基础的方法,而支配着社会现象的法则和方法则是概率论。”他将道德统计加入到统计学之中,强调犯罪现象的规律性。他提出了著名的所谓“平均人”的概念,即所有因素都取平均的“典型的”人,有平均的身长、平均的体重、平均的智能、平均的道德观念等。他认为,“平均人”在社会中犹如物体的重心,各个社会成员都围绕着“平均人”摇摆波动。虽然他考虑社会问题的方法有些机械,论点也有不少是过头的,但是随着近代科学的潮流,他的研究却导致了“统计万能”时代的到来。他是第一个将作为数学分支的法国古典概率论引入社会统计研究的学者,因此常被认为是数理统计学的创始人。他对人体测量学也有很大的贡献,他指出人体的测量值是服从正态分布的,因此有时候也把正态分布叫做凯特勒分布。

六、描述性统计学的发展

作为近代统计学,一方面有和经济学结合以社会复杂的经济问题为研究对象的计量统计学,还有研究生物现象的生物统计学。对生物统计学做出重大贡献并由此创立了描述性统计学的主要有 F. Galton (1822~1911) 和 K. Pearson (1857~1919)。Galton 是创作《物种起源》的达尔文的表弟,出生于知识之家,先在大学学习医学、植物学,后来在剑桥大学热衷于数学研究。他的研究十分广泛,涉及到心理学、人类学、生物学、遗传学、指纹学等等,受到其表兄《物种起源》大作的刺激,全力投入

到对遗传学的数理统计方法研究之中，终于有了伟大的功绩。他研究了平均值的偏差问题和回归问题，这在统计方法上是一大进步。

Pearson 继承和发展了 Galton 的统计思想。他一生致力于生物测量学、优生学和遗传学的统计方法研究，对一般生物现象进行数量描述，极大地丰富了统计学的概念。他创造了许多统计学用语，例如“频度分布”、“频度分布函数”、“回归”、“相关”、“拟合度”等等。可以说，今天的描述性统计学中大部分内容都是由 Pearson 整理出来的，大部分统计学用语也是他命名的。他与 Galton、Weldon 在 1901 年开创了 *Biometrika* 杂志。这是一本很有权威的学术杂志，直至今日，也是世界上级别最高的统计学杂志之一。他所主办的研究所编制了很多数学用表，为从事统计调查、科学实验以及应用数学等领域的工作者提供了重要的工具。

七、推断统计学的诞生

在 1920 年前后，统计学有了重大的转机，这就是推断统计学的诞生。可以说是由 W. S. Gorsset (1876—1937) (笔名叫 Student) 开始研究，最终由 R. A. Fisher (1890—1962) 创立而成的。

当时在英国 Dublin 啤酒公司作技师的 Gorsset 想要解决啤酒质量的检验问题，可是对所涉及的研究对象（比如啤酒中酵母菌的含量、啤酒对小麦的比收获量等等）却很难获取大样本的数据。他苦心钻研统计理论，终于想出了一种小样本的检验方法——t 检验法，并于 1906 年以“Student”的笔名在 *Biometrika* 的杂志上发表了。虽然从量上看这是一件很小的工作，但在思想上却可以说是开创了一个新纪元。他的结果后来经 Fisher 从数学上进行了严格证明，由此诞生了小样本检验的理论。Fisher 是很有声望的统计学家、剑桥大学的名教授。起初他在英国 Rothamsted 试验农场当技师，对农业试验中如何应用数理统计方法的问题进行了深入的研究，导出了分布法则，提出了方差分析法以及各种检验法等。他的伟大功绩在于从统计的意义上明确了推断、检验的含意及其与数学的关系，创立了估计理论和检验理论等统计的理论体系，开拓了统计学的新领域，也就是创立了与过去的描述性统计学所不同的推断统计学。

第二次世界大战以后，推断统计学的研究成为数理统计学的主流，特别是在美国，在理论和应用两方面都有很大的成就。A. Wald 是这些研究者中的代表人物之一，《序贯分析》是他的代表作。

八、推断统计学的应用

近代统计学即推断统计学虽然起源于统计方法在农业试验中的应用，但其重要意义远远超出这一范围。在自然科学、管理、工农业生产、社会科学、医学、心理学、行为科学、商业、气象……等几乎任何领域，推断统计学都是适用的。以下仅就其在管理部门和社会调查中的应用做一简要的介绍。

首先介绍在大量生产过程中的质量管理。当大量地反复地生产某种产品时，使用完全相同的原料、完全相同的工具和按照完全相同的生产过程实际上是不可能的，因此无法生产出完全相同的产品。对产品表现出的不均一性，有些原因是可以查明的，对这些影响均一性的因素一定要去掉。不过即使将查明了原因的影响因素全部排除掉，产品之间也还会残存有差异。这种差异（变动）被认为是由偶然性（随机性）造

成的。对这种偶然性的处置就成了我们统计学的研究对象。也就是说，首先要研究如何查找非偶然性的影响因素，如何制定产品的质量和形状等的规格；其次，要检验实际产品是否符合这些规格，在检验时要用什么样的方法等。以上问题的解决都要用到推断统计学，这样实行的管理就叫做统计质量管理（SQC）。这种管理最早是由英国实行，二次大战期间，美国通过这种统计质量管理，提高了美国军工产品的质量，给厂家带来了巨额的利润。以后，一些工业发达国家纷纷仿效。20世纪60年代，日本在学习美国统计质量管理的基础上，结合日本的国情，进一步发展完善，成为全面质量管理（TQC），对日本的国民经济发展起到了重要的推动作用。目前，日本的全面质量管理已引起世界上越来越多国家的重视，并在许多国家中得到了应用。我国也在试行全面质量管理。

其次介绍在社会调查中的应用。所谓社会调查，曾被认为是要对全部对象（总体）进行全面调查，从而为将来制定计划和决策准备基础资料。但是，全面调查从时间和经济上来说都是十分困难的。即使只对全面调查数据进行统计处理，等做出决策时总体中的某些因素可能已经发生了变化。如果我们对总体的某些分布情况有一定的把握，就不必进行全面调查，可以进行部分调查即抽样调查，这从时间和经济上来说都是合算的。根据抽样调查的结果，可以对总体的状况进行估计和推断。由部分推断全部，概率论和数理统计理论起着重要的作用，这是近代统计学的主要特征。关于抽样调查，有很多方法，对此也在进行着大量的研究。实际上各国在经济统计、国势调查、社会调查、收视率调查、民意测验上等等，采用的几乎都是抽样调查。

部分（抽样）调查的方法不仅在社会统计中也在科学实验中应用，对所有的实验结果也是按部分资料来看待的，故在处理实验资料时要利用统计推断。一般情况下实验数据都不多，因此，要想从中得到一些结论，必须严格地按概率论和数理统计理论的有关准则行事。

最后举一个选举预测的例子，从中可以体会到正确进行抽样调查的重要性。

美国的盖洛普（Gallup, 1901~1984）是新闻学界所熟悉的、毕生从事民意调查事业的人物。他在1935年创建了“美国民意学会”。20世纪30年代，美国的一些报纸杂志也开始纷纷成立专门的机构进行民意测验，也进行选举预测。在1936年美国总统大选的预测中，《文学文摘》（Literary Digest）等杂志社由于预测结果与实际得票率相差太多而被迫停刊，而盖洛普的“美国民意学会”却由于预测的成功而声名大振。

《文学文摘》的失败主要是由于抽样的偏差，他们无视一般的劳动阶层（上流阶层的人大多数投兰登（Randon）的票，而结果却是罗斯福（Roosevelt）当选）；盖洛普却采用配额方法（quota method）来抽样，即考虑地区、年龄、性别、支持的党派、收入等种种情况，按一定的比例配额来决定抽样数。实际上，盖洛普的抽样法还不是我们所指的那种随机抽样，他们只是从比例上去控制，而调查员实际上还是根据自己的意愿怎么方便就怎么去调查的，即还是属于一种“有意选择法”，还满足不了随机性的要求。因此，1936年盖洛普预测的成功还只是一种粗糙的结果。

在1948年总统大选的预测中，不但Crossley、Roper等杂志全部失败，连盖洛普也失败了。据分析，其原因主要有抽样和访问调查中的问题以及浮动票的预测问题。

由于这次失败，Crossley 和 Roper 从此将力量转向市场调查。但盖洛普却对失败进行了认真的反省，根据概率论的原理全面研究了抽样的方法，终于在 1950 年的中期选举中获得了 15 年间全国选举预测的最佳结果。此后，盖洛普的名字便成了民意测验的同义词。

九、现代统计学

现代统计学无论是在数学理论还是在应用范围上，都得到了飞速的发展。特别是随着计算机的发展与普及，在应用上层开了广泛的前景。无论在自然科学、社会科学还是医、农、林等各个领域中，统计学都已经逐步成为不可缺少的工具。根据作者的理解和体会、现代统计学有以下几个特点：

1. 理论和方法不断地被完善和深化。

随着实际应用的需要，对数理统计理论和方法提出了越来越高的要求。从线性到非线性、从低维到高维、从显在到潜在、从连续到离散等等，现在基本上都有了相应的较为完备的理论和方法。新的课题和方向也在继续探索研究之中。

2. 计算机及其相应的统计软件已逐步成为统计工作者不可缺少的工具。

大量优秀统计软件的研制成功为统计的应用开拓了新的广阔前景。国外社会科学工作者常用的软件包如 SPSS、SAS、DATA-TEST、STATA 等等也已逐步引进我国，有些已经被汉化。与此同时，我国的统计工作者也开发研制了各种实用的统计软件。标准统计方法的应用使得计算量再大也变得简易可行，而复杂的、难于从理论上证明的新的统计方法也可以借助计算机模拟进行研究。统计和计算、统计和计算机已经成为不可分割的一体。作为统计学的一个新的分支，计算机统计学（在我国称概率统计计算）正在兴起。

3. 现代统计学正逐步成为一门通用的研究、如何合理地有效获取、整理和分析数据的独立的交叉性学科。

我国的统计学界在传统上将统计学派分成“社会统计学派”和“数理统计学派”，前者认为统计学是一门社会科学，而后者则认为统计学是一门应用数学。我们认为两者都带有一定的片面性。按目前国际上比较流行的看法，统计学是一门独立的学科，是一种方法论。它的原则既适用于自然科学，也适用于社会科学等诸多领域。因此在某种意义上可以说现代统计学是一门独立的交叉性学科。

基于这样的认识，在我国统计教育学界两个学派的共同努力下，1998 年教育部对原有的本科专业目录进行调整修订时，终于将统计学设置成为了一级专业学科。而在 1998 年之前，我国高校本科的统计学专业一直是二级学科（设在经济学学科下的“统计专业”以及设在数学学科下的“概率论与数理统计专业”）。统计学专业的升级，对促进统计学的专业教育，对培养既有数理统计理论基础又有实际应用能力的人才，起到了重要的作用。

可以说，社会科学工作者和数理统计工作者相结合，相互学习，探索社会科学研究领域中的新问题、新方向，无疑是一个可以大有作为的新天地。这也是我们近 20 年来在教学和应用实践中的深切体会，在此献给读者，以结束这篇似乎过于冗长的绪言。

绪言	统计学发展史简介/1
第一部分	基础统计学
第一章	统计学的性质/2
1.1	随机抽样/2
1.2	随机化实验/6
1.3	社会科学中的随机化实验/9
	本章小结/10
第二章	描述性统计学/11
2.1	几个基本概念/11
2.2	频数表与直方图/12
2.3	分布的中心/16
2.4	分布的形状/23
2.5	利用相对频率进行计算/28
2.6	其他描述分布的统计图和统计量/29
	本章小结、应用实例/34
第三章	概率分布/49
3.1	离散型随机变量的概率/49
3.2	概率树/53
3.3	总体的均值与方差/57
3.4	离散型随机变量的概率/60
3.5	连续型随机变量的概率分布/65
3.6	正态分布/67
	本章小结/72

第四章	抽样/75
4.1	随机抽样/75
4.2	蒙特卡罗法/78
4.3	样本均值 \bar{X} 的抽样分布/85
4.4	样本比例 P 的抽样分布/89
	本章小结/95
第二部分	常用统计分析方法
第五章	置信区间/98
5.1	演绎法与归纳法简介/98
5.2	总体均值 μ 的置信区间/99
5.3	两个总体均值之差 ($\mu_1 - \mu_2$) 的置信区间/106
5.4	总体比例 π 的置信区间/113
5.5	单侧置信区间/116
	本章小结/118
第六章	假设检验/121
6.1	利用置信区间进行假设检验/121
6.2	概率值 (单侧的) /124
6.3	经典的假设检验/132
	本章小结、应用实例/140
第七章	回归分析/145
7.1	简单线性回归/145
7.2	回归模型/151
7.3	样本斜率的抽样分布/155
7.4	总体斜率的置信区间和假设检验/157
7.5	自变量为定类变量时的回归/160
7.6	最简单的非线性回归/163
	本章小结、应用实例/165
第八章	方差分析/174
8.1	单因素的方差分析/174
8.2	双因素的方差分析简介/182
	本章小结/187
	应用实例/188
第九章	相关分析/193
9.1	简单 (积矩) 相关/193
9.2	总体相关系数的检验/196
9.3	相关和回归/199
9.4	其他相关系数及相关测量法简介/206