

高等财经教育精品课和精编版教材

统计学

(第二版)

Statistics

主编 徐建邦 李培军

近些年来，存在于我国统计学界的社会经济统计学派和数理统计学派，经过几十年的激烈争论，总算基本达成了共识，即共同接受了大统计思想。

“统计学”并不是作为一级学科的大统计学，而相当于大统计学原理。主要由数理统计学和社会经济统计学原理两部分内容构成的。但不是板块的简单拼凑，而是努力将其有机地融合于统计学的体系之中。

高等财经教育精品课和精编版教材

统 计 学

(第二版)

主 编 徐建邦 李培军

 **东北财经大学出版社**
Dongbei University of Finance & Economics Press

大 连

© 徐建邦 李培军 2006

图书在版编目(CIP)数据

统计学 / 徐建邦, 李培军主编. —2 版. —大连: 东北财经大学出版社, 2006. 7

高等财经教育精品课和精编版教材

ISBN 7-81084-851-8

I. 统… II. ①徐…②李… III. 统计学-高等学校-教材
IV. C8

中国版本图书馆 CIP 数据核字(2006)第 025438 号

东北财经大学出版社出版

(大连市黑石礁尖山街 217 号 邮政编码 116025)

总编室: (0411)84710523

营销部: (0411)84710711

网 址: <http://www.dufep.cn>

读者信箱: dufep@dufe.edu.cn

大连业发印刷有限公司印刷 东北财经大学出版社发行

幅面尺寸: 148mm × 210mm 字数: 485 千字 印张: 16 3/8

印数: 41 001—51 000 册

2006 年 7 月第 2 版

2006 年 7 月第 9 次印刷

责任编辑: 谭焕忠

责任校对: 孙冰洁 龚小晖

封面设计: 张智波

版式设计: 钟福建

定价: 25.00 元

出版者的话

为贯彻落实党的十六大精神，实践“三个代表”重要思想，以“科学的发展观”为指导，切实推进教育创新，深化教学改革，促进现代信息技术在教学中的应用，共享优质教学资源，全面提高教育教学质量，造就数以千万计的专门人才和一大批拔尖创新人才，提高我国高等教育的综合实力和国际竞争能力，教育部在全国高等学校中启动高等学校教学质量与教学改革工程精品课程建设工作。

精品课程是具有—流教师队伍、—流教学内容、—流教学方法、—流教材、—流教学管理等特点的示范性课程。精品课程建设是高等学校教学质量与教学改革工程的重要组成部分。

教材作为体现教学内容和教学方法的知识载体，是进行教学的基本工具，因此，加强高校精品课教材配套建设也是急迫的、重要的。精品课程教材应是系列化优秀教材，可以是精品课程主讲教师自行编写的或相关教材，也可以是国内高水平优秀教材和国外高水平原版教材。

另外，我社成立二十年来，以“专业性、开放式、国际化”为宗旨，以“服务教育、兼顾市场”为方针，出版了一大批财经类教材。这些教材符合教学规律，反映教学前沿，注重教学效果，深受广大读者欢迎，有一些经多个教学循环，多次修订更新，在内容和形式上逐渐提升，在竞争激烈的市场中占有一个逐渐增大的空间。

《高等财经教育精品课和精编版教材》，是在我社原有教材体系的基础上，以精品课（国家级、省级、校级）主讲教师为组织者编写的有关教材，同时对经教学实践检验或获省部级以上奖励的优秀教材进行修订再版，并努力建设一体化设计、多种媒体有机结合的立体化教材。

我们期望，本套教材的出版能对我国高等教育教学质量的提高，为

培养更多更好适应社会经济发展和社会主义市场经济新形势的人才作出一定的贡献。

东北财经大学出版社

2006 年 4 月

前言

统计学是关于数据的科学,而数据每时每刻都在产生,遍布于自然界与人类社会的每个角落,因此,统计学是每个领域、每个人都可能用到的科学,不论你是否接受或是否意识到,统计学事实上已经渗透到我们的工作、学习和生活中。在大学,统计学是许多专业的研究生、本专科生必修的课程,实际上,不仅是大学生,对于任何人而言,学习一些统计学知识,掌握基本的统计分析工具,都是有益处的。

如果你对统计学有兴趣,到书店去看一下,就会发现统计学教材种类很多,而且内容也往往有所不同。传统的统计学教材大致可以分为两类:一类是属于数理统计类的,内容侧重于对统计方法原理的阐述和公式推导证明,对于大多数人而言,显得难于学习和运用;另一类属于社会经济统计原理类,虽称作“原理”,但往往偏重于介绍只在经济领域使用的统计方法,对于通用的统计分析工具则介绍得不够全面,也未更深入地解释统计学的原理。中国统计学界曾就统计学的学科属性等问题激烈地争论了几十年,争论的结果是多数人接受了“大统计”思想,承认统计学应当是一个包含彼此存在差异的多个分支学科的综合体。正是由于统计学的分支学科存在差异,同时,统计学界的不同学派也存在认识上的差异,从而造成了统计学教材内容体系和学术风格方面的不同。不过,随着对统计科学认识的不断深化,统计学教材间的差异正在缩小,特别是前述的社会经济统计学原理类教材,已经加强了对通用的统计学方法的介绍,而其中许多内容原来是属于数理统计范畴的。

从“大统计”的角度讲,任何一部统计学教材都不可能涵盖统计学的全部内容,因此,虽然很多教材都称为《统计学》,但所介绍的只是统计学的一部分。本书与许多同类名称的教材一样,只是介绍了一些基本的统计方法,更确切一些的名称似乎应当是“统计学基本方法”、“基础统计学”或者是“初级统计学”,但这些名称都很难准确地描述书中的内容,所以,我们还是用了大多数同类教材共同使用的名称。但我们要提醒读者

注意,你在本书中所看到的,并不是统计学的全貌,考虑到读者的情况以及将本书作为教材使用时的课时安排,许多统计学方法特别是多元统计方面的内容并没有收入本书。如果你想了解和掌握更多、更全面的统计学知识,还应当阅读相关的书籍。

与同类教材及本书的旧版本相比,本书试图在以下几个方面有所突破和创新:

第一,坚持“大统计”思想,按“描述统计”、“推断统计”、“核算统计”框架构造内容体系。描述统计与推断统计是得到普遍认同的对统计学方法的分类,如果从广义的角度理解,本书第二、三、四、十、十一章基本上属于描述统计范畴,而第五、六、七、八、九、十二、十三章则可以归入或大致归入推断统计范畴。核算统计在有些人看来不属于统计学范畴,但相关的方式方法在描述社会经济现象的总量、结构、动态等方面确实是不可替代的,因此,本书将其纳入总体框架。第十四章属于核算统计范畴。

第二,结合统计软件介绍统计方法。随着计算机技术的不断发展和统计软件的逐渐普及,结合统计软件介绍统计学正逐步成为更多统计学教材的选择。统计软件本身并不能帮助我们理解统计思想和统计学方法,但它确确实实在统计计算及处理大规模数据方面为我们提供了极大的便利。目前有许多统计软件或数据处理软件被不同需求的人群使用,最终,从普及程度方面考虑,我们选择了 Excel 软件,在主要的章节介绍了如何利用 Excel 得到相关的统计分析结果。

第三,尽量减少对原理的阐述和公式推导,便于学习和使用。目前,统计学教材越来越注重实用性,因为对于大多数学习者而言,学习统计学的目的不在于了解和掌握它到底是什么,而是希望能够学会利用这种工具来处理所遇到的问题。出于这样一种考虑,我们在教材中适当减少了对某种统计方法的原理介绍和公式推导,希望就此降低学习的难度。当然,必要的介绍和推导还是要有的,特别是有助于正确理解和使用统计学工具的内容。

本书适合经济管理类非统计学专业本科生使用,也可以作为统计学专业本科生、其他专业研究生的参考教材使用,亦可作为培训教材使用。

本书由徐建邦、李培军任主编。其中,第一、二章由徐建邦编写,第三、四章由刘沈忠编写,第五、六、七、八、九章由冯叔民编写,第十章由赵

振伦编写,第十一章由孙玉环编写,第十二章由冯力编写,第十三章由庄连平编写,第十四章由陈梦根编写。

编者

2006年4月

目 录

第一章 绪论	1
第一节 关于统计学.....	1
第二节 关于学习统计学.....	9
第三节 统计学与计算工具	17
小结	24
关键术语	24
第二章 数据	25
第一节 数据及相关概念	26
第二节 数据来源	35
第三节 数据调查误差	39
第四节 数据汇总	41
第五节 数据显示	48
小结	56
关键术语	57
Excel 操作提示.....	58
练习题	58
第三章 单变量描述统计	60
第一节 频数与累计频数	60
第二节 集中趋势的度量	62
第三节 离散程度的度量	76
第四节 相对位置和异常值的检验	83
第五节 偏态与峰态的测定	87
第六节 探索性数据分析	90
小结	93
关键术语	93
重要公式	94

Excel 操作提示	96
练习题	97
第四章 两变量描述统计	100
第一节 相对指标	100
第二节 协方差和相关分析	103
第三节 交叉表与散点图	109
小结	111
关键术语	112
重要公式	112
Excel 操作提示	112
练习题	113
第五章 概率与概率分布	114
第一节 概率	114
第二节 大数定律与中心极限定理	143
小结	145
关键术语	145
重要公式及公式推导	146
Excel 操作提示	147
练习题	152
第六章 参数估计	153
第一节 参数估计中的一些基本概念	153
第二节 一个总体参数的区间估计	157
第三节 两个总体参数的区间估计	167
小结	174
关键术语	175
重要公式	175
Excel 操作提示	175
练习题	176
第七章 统计假设检验	177
第一节 假设检验的基本思想	177
第二节 假设检验的基本形式	185

小结	200
关键术语	201
重要公式	201
Excel 操作提示	202
练习题	202
第八章 方差分析	204
第一节 单因素方差分析	205
第二节 无交互作用的双因素方差分析	211
第三节 有交互作用的双因素方差分析	220
小结	230
关键术语	231
重要公式及公式推导	231
Excel 操作提示	232
练习题	235
第九章 抽样调查	238
第一节 简单随机抽样	238
第二节 类型随机抽样	256
第三节 等距抽样	266
第四节 整群抽样	271
小结	279
关键术语	279
重要公式	279
练习题	279
第十章 时间序列分析	281
第一节 时间序列的对比分析	281
第二节 长期趋势测定	291
第三节 季节变动的测定	305
小结	313
关键术语	315
重要公式	316
Excel 操作提示	317

练习题.....	317
第十一章 统计指数分析	320
第一节 统计指数的一般问题.....	320
第二节 综合指数.....	322
第三节 平均指数.....	326
第四节 指数体系与因素分析法.....	332
第五节 几种常用经济指数的编制.....	342
小结.....	347
关键术语.....	348
重要公式.....	349
Excel 操作提示	349
练习题.....	351
第十二章 回归分析	354
第一节 简单线性回归模型.....	355
第二节 判定系数与相关系数.....	366
第三节 回归方程的显著性检验.....	377
第四节 回归预测.....	389
第五节 残差分析.....	397
第六节 多重线性回归分析.....	408
小结.....	435
关键术语.....	435
重要公式.....	437
Excel 操作提示	439
练习题.....	440
第十三章 列联分析	449
第一节 列联表.....	449
第二节 一致性检验.....	456
第三节 独立性检验.....	459
第四节 列联表的相关测量.....	461
小结.....	465
关键术语.....	465

Excel 操作提示	465
练习题	466
第十四章 国民经济核算	469
第一节 国民经济核算概述	469
第二节 国民经济核算的基本问题	474
第三节 主要国民经济总量指标介绍	489
小结	494
关键术语	494
练习题	495
附表	497
附表 1 正态曲线下的面积	497
附表 2 t 分布的临界点	498
附表 3 χ^2 分布的临界点	499
附表 4 F 分布	501
附表 5 检验等级相关系数 $\rho_i = 0$ 的临界值(r_i)表	507
附表 6 肯达尔一致性系数中 S 的临界值表	508

第一章 绪 论

第一节 关于统计学

一、statistics 的含义

统计学的英文单词是 statistics, 这个单词具有多种意义。它可以是单数, 意思是“统计学”, 如“Statistics is a fun course(统计学是一门有趣的课程)”;也可以是复数, 意思是“数字”或“数据”, 如“*These statistics indicate an upswing in the economy(统计表明经济正在起飞)*”。据考证, statistics 源于拉丁语 status 一词, 意为各种现象的状态和状况, 由这个词根可以组成多个名词和形容词, 其中, stato 表示国家的概念, 同时也表示关于国家和国情方面的知识, 而通晓这方面知识的人称为 sataista。17 世纪该词转化为德语, 产生了一个形容词 statistische, 意为“统计的”、“统计学的”, 此后, 这个德文单词又被翻译成英语 statistics 沿用至今。另一种说法是, statistics 保留了 state 这个词根, 意为“国家”或“城邦”, 而统计或统计学最初即是关于国家的各类知识的总称, 直至 17 世纪中叶才逐渐被“政治算术”这个名词所替代, 并且很快被演化为“统计学(Statistics)”。

从外文单词的起源和演化情况看, 统计是应管理国家的要求产生和发展的。世界各国早期的历史著作中, 都记载着关于土地、人口、财产等方面的数字资料。例如, 我国的夏朝时期就已经有了人口数目和土地数目的登记; 古希腊的亚里士多德撰写“城邦纪要”, 共撰写了 150 余种纪要, 其内容包括各城邦的历史、行政、科学、艺术、人口、资源和财富等社会和经济情况的比较、分析等等。随着社会经济的发展, 统计在国家事务中的地位也越来越重要, 时至今日, 统计仍然是世界各国各个层次政府机构的支柱, 关于经济的、社会的、自然的统计已经成为国际组织、政府、企业、社会团体乃至个人重点关注的对象。此外, 我们还可以看出, 统计的基本特征是数字。statistics 的一种解释, 就是以数字表述事实, 从古至今, 统计所反映的始终是社会各个领域的数量表现与特征。

与英文单词 *statistics* 相似,中文的“统计”一词通常也可以有多种含义。例如“企业统计”,它可以指一门课程,其中的“统计”意为统计学或统计方法;它也可以指企业的统计机构及相关的工作,则其中的“统计”就是统计工作的意思;如果我们说“企业统计表明,某地工业生产形势不容乐观”,那其中的“统计”就是指统计资料或统计数据。

二、统计学的起源

一般认为,作为一门独立的、完整的、具有现代意义的科学,统计学大约产生于 17 世纪末(也有人认为统计学的产生要早得多,可以追溯到古希腊的亚里士多德撰写“城邦纪要”的时代,但这种观点不为大多数人所认可,因为自古希腊或古代中国开始的统计实践活动与现代统计学有着本质的差别)。

与统计是应管理国家的需要而产生的一样,统计学或称现代统计学的产生与发展,也源于社会经济的发展需求。自 16 世纪中期以来,在心理学、医学、人体测量学、遗传学、人口学、天文学、农学等诸多领域,以及传统的国家管理领域,人们对如何收集数据以及对数据进行分析进行了长期的、大量的研究与探索,诸如大数定律、随机误差和误差分布、概率和概率分布、中心极限定理、假设检验、回归、列联表、正交设计等等统计理论与方法不断被提出和改进发展,传统统计领域的方法也不断得到完善。可以说,是社会经济的需求造就了现代统计学,而统计学的发展也极大地促进了各个领域科学技术的发展与进步。经过 300 余年的建设与发展,统计学已经成为具有多个分支和广泛应用领域的科学,并且许多新的分支或交叉学科诸如决策论、时间序列、多元分析、经济计量学、博弈论、临床试验、非参数推断、序贯分析、数学生物分类学、可靠性分析等,正不断发展完善,统计学的应用领域也在继续扩大。

在现代统计学的形成与发展过程中,数学起着重要的作用,并且自 20 世纪 50 年代以来,统计变得越来越数学化,模型在统计学中得到了广泛的应用;测度论、群论、数论、组合理论等数学方法被应用到统计学中。电子计算机也给统计学带来了巨大的变化。由于使用计算机,大量的、动态的数据收集和筛选以及一些极其繁杂的分析计算成为可能,计算机还大大地简化了统计计算的难度,节省了大量人力,方便了统计学的普及和应用。一些新的统计分支学科如数据挖掘,也随着计算机的应用而产生。

在统计学的形成及发展过程中,曾有过多种学派,如早期的政治算术学派、国势学派,以及其后的数理统计学派、社会统计学派等,各学派相互论争、相互借鉴,有效地促进了统计学科的发展。许多学者为统计学的形成和完善作出了杰出的贡献。在各类关于统计学起源与发展历史的书籍中,我们可以找到他们具体的业绩内容。目前,许多统计学理论或方法仍然以这些学者的名字命名,如贝叶斯统计、马尔可夫链、皮尔逊积矩相关系数、F 检验等等。

大约在 19 世纪,以数理统计为主要特征的统计学传入中国,并得以延续发展。新中国成立之后,统计学的发展受苏联的影响很大,社会经济统计被确立为统计学,而数理统计则被视为数学的分支,学科建设走了很大的一段弯路。改革开放以后,在解放思想、实事求是的精神指导下,我国统计学的学科建设逐步拨乱反正、走上了合乎逻辑的发展轨道,数理统计与社会经济统计相互融合、共同发展,渐渐跟上了国际统计学科发展的潮流。目前,在教育部制定的本科专业目录中,统计学与哲学、数学、经济学、管理学等并列为一级学科。

总体上讲,统计学的起源、形成与发展是一个较为复杂的过程,对于这段历史的描述也是仁者见仁、智者见智,本书并未作具体阐述。一般地说,不了解统计学的起源及发展过程,并不影响学习统计学。如果想深入研究这方面的问题,应当查阅相关的专题文章或论著。

三、统计学的定义及分类

在不同的教科书上,对统计学的理解和表述可能略有差别。总体上讲,一个得到公认的理解是:统计学是关于数据的科学。各种统计学定义大体以此为基础。例如,不列颠百科全书给统计学下的定义是“统计学是收集和分析数据的科学与艺术”;美国辛辛那提大学安德森等人所著《商务与经济统计》一书的表述是“统计学是收集、分析、表述和解释数据的艺术和科学”;美国弗里德曼等人所著《统计学》一书则将统计学定义为“是对令人困惑费解的问题作出数字设想的艺术”。

本书对统计学的定义是:统计学是关于数据的科学,研究如何收集、整理、分析、表述数据,并通过数据得出基本结论。

“数”与“质”是事物的两个基本属性,任何现象都有其数量表现,而这种数量表现又隐含着事物本身品质或规律性方面的信息。统计学就是

研究事物“数量”特征的方法论科学。统计学家们探索、开创各种收集数据和分析数据的方法,提出并通过实践来检验各种理论模型,不断为统计学的应用寻找新的领域。人们可以运用统计学提供的方法收集、筛选、分析各类数据,使之转化为比数字本身更为复杂的知识,进而得出相关的结论,或者依据分析结果作出各种决策。

与经济学、管理学、化学或心理学等实质性科学不同,统计学没有自己的基于实验或观察的经验研究对象。统计学是通用的方法论科学,它为各个学科领域的研究、经济与社会管理以及人们日常生活学习等提供一整套研究对象的方法。如果对这些方法做简要的划分,我们大致可以将其分为三类,即描述统计、推断统计和核算统计,或者也可以将它们称为统计学的三个分支。

(一)描述统计

一般意义上的描述统计是用于汇总数据的表、图和数值方法的统称。

通常,观察到的或通过实验得到的数据是杂乱无章的,我们往往无法从中看出某种特征或规律性的东西,经过统计处理,数据能够以读者易于理解的、更为直观的形式显示出来,数据的特征也会以图形、表格或数值形式展示给读者。在这一过程中我们使用的统计方法就统称为描述统计。具体地说,描述统计包括:(1)变量数列、时间数列及相关的表格形式;(2)直方图、盒形图等各类统计图形;(3)均值、标准差、中位数等概括性数值。描述统计所得到的结论只限于已有的数据范围,不能直接推广到未知的领域。

例如,表1—1列出了2003年上海证券交易所30个上市公司的部分数据^①。相对而言,信息较为杂乱,不易理解。

表1—2列出了30个上市公司3个数值型变量(每股净资产、市盈率、流通股本)主要的概括性数值,简要地描述了由这30个公司数据所组成的数据集的特征。其中,均值、中位数反映变量值的集中趋势或一般水平;标准差、方差则反映了变量值的离散趋势即差异情况;众数表明出现次数最多的是哪一个变量值;极差则说明最大值与最小值的距离有多少(关于这些概括性数值以及下面提到的各种统计图的知识,我们将在

^① 资料来源:<http://business.sohu.com/stock.shtml>