

世纪统计学系列教材

描述统计

贾俊平 编著



中国人民大学出版社

21 世纪统计学系列教材

描述统计

贾俊平 编著

中国人民大学出版社

图书在版编目(CIP)数据

描述统计/贾俊平编著.
北京:中国人民大学出版社,2003
(21世纪统计学系列教材)

ISBN 7-300-04576-6/F·1403

I. 描…
II. 贾…
III. 统计学-教材
IV. C8

中国版本图书馆 CIP 数据核字 (2003) 第 013681 号



贾俊平 编著

出版发行:中国人民大学出版社
(北京中关村大街31号 邮编100080)
邮购部:62515351 门市部:62514148
总编室:62511242 出版部:62511239
本社网址:www.crup.com.cn
人大教研网:www.ttrnet.com

经 销:新华书店
印 刷:北京东方圣雅印刷有限公司

开本:787×965毫米 1/16 印张:11.25
2003年4月第1版 2003年4月第1次印刷
字数:204 000

定价:13.00元
(图书出现印装问题,本社负责调换)

《21 世纪统计学系列教材》编委会

编委会主任 易丹辉

编委会委员 (按姓氏笔画排序)

尹德光 冯士雍 张尧庭

陈希孺 吴喜之 赵彦云

柯惠新 袁 卫 倪加勋

顾 岚 袁寿庄 耿 直



总 序

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，一直坚持着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

1-A113/13

的运作，将有利于提升我国应用统计科学研究的水平，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

2000年8月

于中国人民大学



前 言

现代社会中，我们每天都会面对大量的数据，看电视，看报纸，数据无处不在。要使这些数据变为对你有用的信息，就需要对这些数据作一些处理和分析，使它们能帮助你作决策。假定你参与股票市场的投资，你就会面对各种各样的信息，比如，上市公司的经营信息，股票价格指数的信息，股评信息。如果你面对大量的信息束手无策，就不知道该如何选择股票；如果你对这些信息处理或使用不当，会对你的投资造成损失。这些信息就是我们所说的统计数据。怎样处理和分析数据，这正是统计的用武之地。描述统计就是一套处理和分析数据的基本方法和技术。

本书是一本统计入门读物，目的是让读者掌握一些处理数据的基本统计方法和技术。对于一个没有接触过统计的人来说，可能会对统计产生某种误解，比如，一提到统计，就会马上想到统计工作，这种理解是不全面的，更准确地说是错误的。本书所面对的读者或许从未接触过统计，我们的目的是尽可能让他们对统计有一个正确的理解，在此基础上，能利用所学的统计方法作一些简单的统计处理和分析。

描述统计的内容大体上包括数据的收集、整理、显示和基本分析等。本书基本上就是按照这些内容来组织的。全书共包括6章内容，第一章介绍统计的一些

基本概念和思想，目的是让我们对统计和统计数据有一个基本的了解。第二章介绍统计数据的收集方法，使读者了解通过什么方法去获得所需的统计数据。第三章介绍数据的整理和显示方法，数据整理是对数据的初步加工，以便使它们能满足分析的需要；数据的显示是一种常用的统计技术，通过显示来观察数据的初步特征和规律。第四章介绍数据特征的统计描述方法，通过对数据特征的概括使我们对所研究的问题有一个初步的了解。第五章介绍数据变换，主要讨论一些常用的线性变换和幂变换方法及其应用。第六章介绍统计指数，指数是一种用于经济分析的特殊方法，如何编制各种价格指数以反映事物的综合变动是本章研究的主要内容。

现代统计分析中的计算问题全部依赖于计算机，目前已有多种统计软件用于这一目的，如 SAS、SPSS、STATISTICA、MINITAB 等。为大多数人所熟悉的 Excel 也具有部分的统计功能。本书涉及的统计计算、绘图等均是利用 Excel 来完成的。

此教材的编写得到了统计学系全体教师的支持和帮助。初稿完成后，编委会的专家张尧庭教授十分认真地审阅了本书的初稿，并提出了许多宝贵意见，在此一并表示感谢。

贾俊平

2003 年 1 月于中国人民大学



目 录

第一章 统计与统计数据	1
第一节 统计与统计学	1
一、什么是统计?	1
二、什么是统计学?	2
三、统计规律	3
四、统计的应用领域	5
第二节 数据的计量与类型	6
一、数据的计量尺度	7
二、数据的类型	9
第三节 统计中的几个基本概念	12
一、总体和样本	12
二、参数和统计量	13
三、变量	14
本章主要术语	15
思考与练习	16

第二章 数据的收集	18
第一节 统计数据的来源	18
一、统计数据的间接来源	18
二、统计数据的直接来源	19
第二节 调查方案设计	24
一、确定调查目的	25
二、确定调查对象和调查单位	25
三、设计调查项目和调查表	25
四、方案设计中的其他内容	27
第三节 调查问卷设计	27
一、问卷的基本结构	27
二、提问项目的设计	30
三、回答项目的设计	32
四、问题顺序的设计	35
第四节 统计数据的质量	36
一、统计数据的误差	36
二、统计数据的质量要求	37
本章主要术语	37
思考与练习	37
第三章 数据的整理与显示	39
第一节 数据的预处理	39
一、数据的审核与筛选	39
二、数据的排序	42
第二节 分类和顺序数据的整理与显示	43
一、分类数据的整理与显示	43
二、顺序数据的整理与显示	48
第三节 数值型数据的整理与显示	51
一、数据的分组	51
二、数值型数据的图示	55
第四节 统计表	68
一、统计表的构成	68
二、统计表的设计	69

本章主要术语	71
思考与练习	71
第四章 数据分布特征的描述与分析	76
第一节 集中趋势的测度	76
一、分类数据：众数	77
二、顺序数据：中位数和分位数	79
三、数值型数据：均值	83
四、众数、中位数和均值的比较	89
第二节 离散程度的测度	91
一、分类数据：异众比率	91
二、顺序数据：四分位差	92
三、数值型数据：方差和标准差	93
四、相对离散程度：离散系数	98
第三节 偏态与峰态的测度	99
一、偏态及其测度	99
二、峰态及其测度	101
第四节 时间序列的描述性分析	103
一、一般增长率	104
二、平均增长率	104
三、年度化增长率	105
四、增长率的应用	106
本章主要术语	108
思考与练习	109
第五章 数据变换	115
第一节 为什么要进行数据变换	115
一、什么是数据变换？	115
二、数据变换的效果	116
三、何时需要变换	117
第二节 线性变换	119
一、原点的变换	119
二、尺度的变换	120

第三节 幂变换·····	121
一、幂变换的基本形式·····	121
二、为稳定散布而变换·····	122
本章主要术语·····	132
思考与练习·····	132
第六章 统计指数 ·····	134
第一节 指数编制的基本问题·····	134
一、指数的性质·····	134
二、指数的分类·····	135
三、指数编制的基本问题·····	136
第二节 加权指数·····	137
一、权数的确定·····	137
二、加权综合指数·····	138
三、加权平均指数·····	141
第三节 指数体系·····	143
一、总量指数与指数体系·····	143
二、指数体系的分析与应用·····	144
第四节 几种常用的价格指数·····	148
一、零售价格指数·····	148
二、消费价格指数·····	151
三、股票价格指数·····	153
第五节 多指标综合评价指数·····	154
一、多指标综合评价指数的构建·····	154
二、几种常用的综合评价指数·····	157
本章主要术语·····	161
思考与练习·····	162
附录：单号练习题答案 ·····	164
参考文献 ·····	169



第一章

统计与统计数据

第一节 统计与统计学

一、什么是统计？

在日常生活中，我们会经常接触到“统计”这一术语。很多人可能会将统计一词与统计工作联系起来，一提到统计首先想到的是统计工作，这种理解是不全面的。统计作为一种社会实践活动已有悠久的历史，可以说，自从有了国家就有了统计实践活动。最初，统计只是一种计数活动，为统治者管理国家的需要而收集资料，通过统计计数以弄清国家的人力、物力和财力，作为国家管理的依据。然而在今天，“统计”一词已被人们赋予多种含义，因此很难给出一个简单的定义。在不同场合，统计一词可以具有不同的含义，通常，统计一词包含以下三种含义：

一是统计工作，它是指统计数据的收集活动。我们的各级政府机构基本上都有统计部门，如统计局，它们的职能主要就是从事统计数据的收集。大多数企业也都有专门从事统计工作的人员，负责企业生产和销售数据的记录、积累以及向上级部门报送数据的任务。一个从事统计工作的人，他所提到的统计一词也许就

是指统计工作。

二是统计数据，它是指统计工作的结果。我们经常会看到专门出版统计数据的出版物，如《统计年鉴》，在报纸、杂志、网络及其他媒体上都会见到大量的统计数据。这些数据就是统计工作成果的体现。当你看到或听到“据统计……”这样的说法时，这里的统计一词就是指统计数据。

三是统计学，它是指分析统计数据的方法和技术。一个从事统计研究或统计教学的人，他心目中的统计一词多数是指“统计学”。他们在日常的表达中，习惯将统计学简称为统计。

正确理解“统计”概念十分必要，一提到“统计”就想到统计工作的思维习惯是狭隘的，应予以纠正。

二、什么是统计学？

统计学是随着人类社会的发展和社会管理的需要而发展起来的。随着社会管理的日趋复杂，仅仅用数字计量客观现象已不能满足社会的需要，人们开始试图对客观现象进行定量分析，与此相适应，研究统计计量和分析方法的统计学开始诞生。“政治算术学派”的创始人威廉·配第和约翰·格朗特，首先在其著作中使用了统计数字和图表等方法来分析研究社会、经济和人口现象，这不仅为人们进一步认识社会提供了一种新的方法和途径，也为统计学的发展奠定了基础。

目前，随着统计方法在各个领域的应用，统计学已发展成为具有多个分支学科的大家族。统计学家们给统计学下的定义繁多，比较有代表性的是不列颠百科全书的定义：“统计学是收集、分析、表述和解释数据的科学。”这一定义揭示了统计学是一套处理数据的方法和技术。

统计学的定义告诉我们，统计离不开数据。统计研究的过程首先要有数据，在拿到数据后，为满足分析的需要，还要对数据进行一定的整理，而后再对数据进行分析 and 解释。因此可将统计研究的过程描述如下（见图 1—1）：

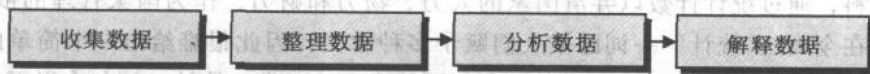


图 1—1 统计研究的过程

收集数据是取得统计数据的过程，它是进行统计分析的基础。离开了统计数据，统计方法就失去了用武之地。如何取得较为可靠的统计数据是统计学研究的内容之一。数据的收集方法可归为两大类：一是观察方法，二是实验方法。观察方法是通过调查或观测而收集数据；实验方法是在实验中控制实验对象而收集数据。

整理数据是对统计数据的加工处理过程，目的是使统计数据系统化、条理化，符合统计分析的需要。整理数据是介于收集数据与分析数据之间的一个必要环节。整理数据的内容包括：数据的筛选，纠正数据中的错误，找出符合分析需要的数据；数据分组，将数据按需要进行分门别类；数据的表述，将数据用图表等形式展示出来，以便找出数据的初步特征，或者是方便别人看懂数据所要表达的问题。

分析数据是统计学的核心内容，是通过统计方法探索数据内在规律的过程，也是统计研究的目的所在。分析数据所用的方法可分为描述统计方法和推断统计方法。

解释数据是对分析的结果进行说明，说明结果为什么是这个样子，结果所隐含的事物的特征是什么，从数据中得出了哪些规律性的结论，等等。

可见，统计学是一门有关统计数据的科学，统计学与统计数据有着密不可分的关系。在英文中，“statistics”一词有两个含义：当它以单数名词出现时，表示作为一门科学的“统计学”；当它以复数名词出现时，表示“统计数据”或“统计资料”。从中可以看出，统计学与统计数据之间有着密不可分的关系。统计学是由一套处理统计数据的方法所组成，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，统计方法就失去了用武之地，统计学也失去了它存在的意义。而统计数据不用统计方法去分析也仅仅是一堆数据而已，不能得出任何有益的结论。

从统计方法的构成来看，可以将统计学分为描述统计和推断统计。这种划分一方面反映了统计方法发展的前后两个阶段，同时也反映了应用统计方法探索客观事物数量规律性的不同过程。描述统计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特征。内容包括统计数据的收集方法、数据的加工处理方法、数据的显示方法、数据分布特征的概括与分析方法等。推断统计则是研究如何根据样本数据去推断总体数量特征的方法，它是在对样本数据进行描述的基础上，对统计总体的未知数量特征作出以概率形式表述的推断。

三、统计规律

统计学提供了探索数据内在规律的一套方法。那么，什么是统计数据的内在数量规律性，为什么统计方法能通过对数据的分析找出其内在的数量规律性，我们用下面的几个例子来说明。

就单独的一个家庭来观察，每个家庭的新生婴儿的性别可能是男性，也可能

是女性。如果不对生育人口进行任何限制，有的家庭的几个孩子可能都是男孩，而有的家庭的几个孩子则可能都是女孩。从表面上看，新生婴儿的性别比例似乎没有什么规律可循，但如果对大量的家庭新生婴儿进行观察，就会发现新生婴儿中男孩略多于女孩，大致为每出生 100 个女孩，相应地就有 107 个男孩出生。这个性别比例 107:100 就是新生婴儿性别比的数量规律，因为古今中外这一比例都大致相同，这是由人类自然发展的内在规律所决定的。人类社会要发展，就要保持男女人数上的大致相同。尽管从新生婴儿来看，男性婴儿略多于女性，似乎并不平衡，但由于男性婴儿的死亡率高于女性，到了中年时，男女人数就大体相同了。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，老年男性反而少于女性。生育人口在性别上保持大体平衡，保证了人类社会的进化和发展。对人口性别比例的研究是统计学的起源之一，也是统计方法所探索的数量规律性之一。

我们都知道投掷硬币和掷骰子的游戏，随机地投掷一次硬币或骰子出现正面、反面或某个点数是不确定的，完全是偶然的。但我们进行多次的重复投掷达到相当数量时，就会发现投一枚均匀硬币出现正面和反面的次数大体相同，即比值接近于 $1/2$ ，而且投掷的次数越多，就越接近于 $1/2$ 这一稳定的数值。同样，在掷骰子时，出现 1~6 点的比例也逐渐接近于 $1/6$ 。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，也就是投掷硬币或骰子时所呈现的数量规律性。

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，这时如果再增加施肥量，产量反而会减少。粮食产量与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。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，以求得最大的效益。

上述例子说明，就一次的观察或试验来说，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。因为客观事物本身是必然性与偶然性的对立统一，必然性反映了事物的本质特征和规律，偶然性反映了事物表现形式上的差异。如果客观事物只有必然性一个方面的特征，它的表现形式就会比较简单，我们可以很容易地把握它的规律性。正是由于偶然性的存在，才使事物的表现形式与必然的规律性之间产生偏

移，从而形成了表面形式上的千差万别，使得必然性的数量规律性被掩盖在表面的差异之中了。统计数据作为客观事物的一种数量表现，是事物必然性与偶然性共同作用的结果。偶然性使得对同一事物的多次观察或试验得到的统计数据有所差异，而必然性则隐含在统计数据本身之中，是我们要利用统计方法去寻找的。在上面的几个例子中，尽管每个新生婴儿的性别、每次投掷硬币或骰子的结果、每次施肥量带来的产量增加等都是不同的、有差异的，但它们本身都存在必然的数量规律，通过统计方法可以尽可能去掉数据所呈现的偶然性，而找出统计数据中所隐含的内在规律性。

四、统计的应用领域

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用到统计方法。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)	hydrology (水文学)
agriculture (农业)	Industry (工业)
animal science (动物学)	linguistics (语言学)
anthropology (人类学)	literature (文学)
archaeology (考古学)	manpower planning (劳动力计划)
auditing (审计学)	management science (管理科学)
crystallography (晶体学)	marketing (市场营销学)
demography (人口统计学)	medical diagnosis (医学诊断)
dentistry (牙医学)	meteorology (气象学)
ecology (生态学)	military science (军事科学)
econometrics (经济计量学)	nuclear material safeguards (核材料安全管理)
education (教育学)	ophthalmology (眼科学)
election forecasting and projection (选举预测和策划)	pharmaceutics (制药学)
engineering (工程)	physics (物理学)
epidemiology (流行病学)	political science (政治学)
finance (金融)	psychology (心理学)
fisheries research (水产渔业研究)	psychophysics (心理物理学)
gambling (赌博)	quality control (质量控制)
genetics (遗传学)	religious studies (宗教研究)
geography (地理学)	sociology (社会学)
geology (地质学)	survey sampling (调查抽样)
historical research (历史研究)	taxonomy (分类学)
human genetics (人类遗传学)	weather modification (气象改善)