



北京市高等教育精品教材立项项目

统计学

贾俊平 金勇进 编著

STATISTICS



统 计 学

贾俊平 金勇进 编著

[illegible]

 中国人民大学出版社

图书在版编目 (CIP) 数据

统计学/贾俊平, 金勇进编著.
北京: 中国人民大学出版社, 2004
北京市高等教育精品教材立项项目

ISBN 7-300-05445-5/F·1709

I. 统…

II. ①贾…②金…

III. 统计学-高等学校-教材

IV. C8

中国版本图书馆 CIP 数据核字 (2004) 第 027742 号

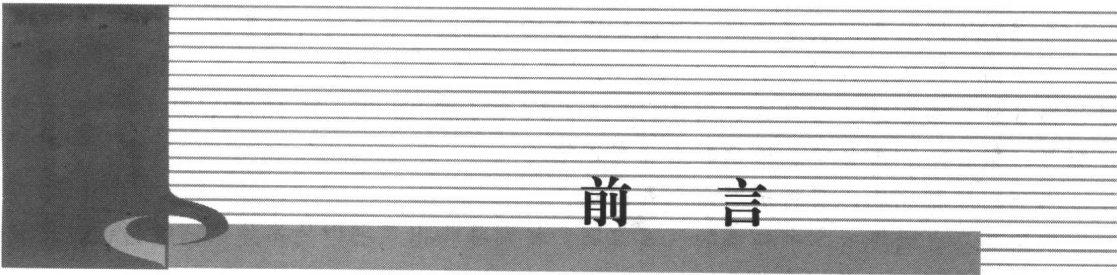
北京市高等教育精品教材立项项目

统计学

贾俊平 金勇进 编著

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)	010-62511239 (出版部)	
	010-82501766 (邮购部)	010-62514148 (门市部)	
网 址	http://www.crup.com.cn		
	http://www.ttrnet.com (人大教研网)		
经 销	新华书店		
印 刷	北京东方圣雅印刷有限公司		
开 本	787×1092 毫米 1/16	版 次	2004 年 6 月第 1 版
印 张	28.75	印 次	2004 年 6 月第 1 次印刷
字 数	586 000	定 价	36.00 元 (含光盘)

版权所有 侵权必究 印装差错 负责调换



前 言

《统计学》一书是为高等院校非统计学专业的学生而编写的一本公共课程教材，使用对象主要包括财经管理类各专业以及非财经类专业，如社会学专业、法学专业、新闻学专业、政治学专业、档案学专业等。本书也可作为各实际部门工作者的参考书。

本书大体上包括描述统计、推断统计和经济管理中常用统计方法三部分内容。第1章至第4章为描述统计部分，第5章至第10章为推断统计部分，第11章至第14章为实际中常用的一些统计方法。本书在写作上力求突出以下特点：

第一，将统计计算与计算机有机结合。根据教材内容的特点，我们选用 Excel 作为计算工具，并且将 Excel 的应用与教材内容完全结合。大部分数据表格和计算结果也都以 Excel 形式给出。

第二，注重统计方法思想的讲述，淡化统计方法的数学推导。考虑到非统计学专业学生学习统计的目的主要是应用，在统计方法的讲述上注重从实际问题入手，而对方法本身则不做更多的推导。

第三，在每一章的后面都给出了本章的中英文对照主要术语，以便读者复习和掌握一些关键性术语。

第四，每章后面的习题注重类型的多样化，并较多地采用了实际数据。

第五，书后附有教学和学习辅助光盘，可作为教师教学和学生自学使用。光盘内容包括教学和学习用 ppt，教材每章后的练习题以及教材中的例题、书后练习题的数据实现了与 Excel 的完全链接，使教学和学习都更加方便。



考虑到不同院校的教学特点,对部分内容可根据实际情况进行选讲。

本书由中国人民大学统计学院的贾俊平和金勇进编写,其中第1章、第3章、第4章、第5章、第6章、第7章、第10章、第11章、第12章、第13章、第14章由贾俊平编写,第2章、第8章、第9章由金勇进编写。

本书的顺利完成,要特别感谢下列同志:

感谢吴喜之教授对本书初稿的认真审阅,特别是指出了书中的多处原则性错误,作者对这些错误进行了纠正,受益匪浅。

感谢易丹辉教授的大力支持以及对时间序列分析和预测一章提出的修改建议。

感谢曾讲授过该门课程的王燕老师、王星老师、张云老师、姚嘉秋老师、赵以立老师、陈龙老师、刘文卿老师、李静萍老师、黄向阳老师、杜子芳老师等。他们多年的教学经验为本书写作积累了宝贵财富。

感谢中国人民大学出版社对本书出版的大力支持。感谢编辑同志对书稿的认真校对。

编写一本好的教材并不容易,本书也难以让读者和同行满意。您的建议始终是我们进一步修改的基础。

贾俊平

2004年2月于中国人民大学

目 录

第 1 章 导论	1
1.1 什么是统计学	1
1.2 统计数据类型	5
1.3 统计中的几个基本概念	7
本章主要术语	10
思考与练习	11
第 2 章 数据的搜集	12
2.1 数据的来源	12
2.2 调查数据	15
2.3 实验数据	25
2.4 数据的误差	30
2.5 数据文件	36
本章主要术语	37
思考与练习	38
第 3 章 数据的整理与显示	39
3.1 数据的预处理	39

hek 53/02



3.2	分类和顺序数据的整理与显示	44
3.3	数值型数据的整理与显示	51
3.4	统计表	66
	本章主要术语	68
	思考与练习	68
第4章	数据分布特征的测度	73
4.1	集中趋势的测度	73
4.2	离散程度的测度	84
4.3	偏态与峰态的测度	94
	本章主要术语	97
	思考与练习	98
第5章	概率与概率分布	102
5.1	事件及其概率	102
5.2	离散型概率分布	116
5.3	连续型概率分布	127
	本章主要术语	142
	思考与练习	143
第6章	抽样分布	147
6.1	三种不同性质的分布	147
6.2	一个总体参数推断时样本统计量的抽样分布	149
6.3	两个总体参数推断时样本统计量的抽样分布	157
	本章主要术语	160
	思考与练习	160
第7章	参数估计	163
7.1	参数估计的一般问题	163
7.2	一个总体参数的区间估计	169
7.3	两个总体参数的区间估计	176
7.4	样本容量的确定	186
	本章主要术语	189

思考与练习	190
第 8 章 假设检验	194
8.1 假设检验的基本问题	194
8.2 一个正态总体参数的检验	203
8.3 两个正态总体参数的检验	212
8.4 假设检验中的其他问题	224
本章主要术语	228
思考与练习	228
第 9 章 列联分析	231
9.1 分类数据与列联表	231
9.2 拟合优度检验	234
9.3 独立性检验	240
9.4 列联表中的相关测量	242
9.5 列联分析中应注意的问题	246
本章主要术语	249
思考与练习	250
第 10 章 方差分析与试验设计	252
10.1 方差分析引论	252
10.2 单因素方差分析	258
10.3 方差分析中的多重比较	268
10.4 双因素方差分析	270
10.5 试验设计初步	281
本章主要术语	285
思考与练习	286
第 11 章 一元线性回归	289
11.1 变量间关系的度量	289
11.2 一元线性回归	297
11.3 利用回归方程进行估计和预测	311
11.4 残差分析	315



本章主要术语	320
思考与练习	321
第 12 章 多元线性回归	326
12.1 多元线性回归模型	326
12.2 回归方程的拟合优度	330
12.3 显著性检验	332
12.4 多重共线性	335
12.5 利用回归方程进行估计和预测	339
12.6 虚拟自变量的回归	340
本章主要术语	350
思考与练习	351
第 13 章 时间序列分析和预测	355
13.1 时间序列及其分解	355
13.2 时间序列的描述性分析	357
13.3 平稳序列的平滑和预测	362
13.4 有趋势序列的分析和预测	368
13.5 复合型序列的分解	380
本章主要术语	386
思考与练习	387
第 14 章 指数	392
14.1 引言	392
14.2 加权指数	393
14.3 指数体系	399
14.4 几种常用的价格指数	402
14.5 多指标综合评价指数	406
本章主要术语	410
思考与练习	411
附录一 各章练习题答案	414
附录二 常用统计表	431

表 1 正态曲线下的面积	431
表 2 t 统计量的临界值	433
表 3 χ^2 统计量的临界值	434
表 4—1 F 统计量的临界值: $F_{0.10}$	436
表 4—2 F 统计量的临界值: $F_{0.05}$	438
表 4—3 F 统计量的临界值: $F_{0.025}$	440
表 4—4 F 统计量的临界值: $F_{0.01}$	442
表 5 随机数表	444
参考书目	447

在日常生活中，我们经常会接触到“统计”这一术语，在有关媒体中也经常会看到一些报导经常使用一些统计数据、图表等。难怪很多人会将“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，想到我们的政府统计机构、企业中从事统计工作的人，等等。这种理解是自然的，但又不全面的。在本章中，我们将讨论有关统计学的一些基本问题，包括统计的含义、统计数据及其分类、统计中一些常用的基本概念等。

1.1 什么是统计学

一、统计学与统计规律

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。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。如图 1—1 所示，可将统计研究的过程描述为：

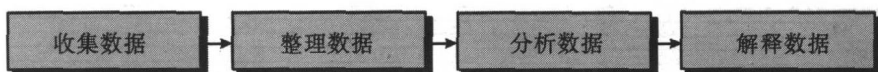


图 1—1 统计研究的过程

数据收集也就是取得统计数据。数据整理是将数据用图表等形式展示出来。数据分析则是通过统计方法研究数据，其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 (inferential statistics) 主要研究如何根据样本信息来推断总体的特征。数据解释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”。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，统计学也就失去了存在的意义。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。

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儿进行观察，就会发现新生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 就是新生儿性别比的数量规律，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，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 $1/2$ 。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 点~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。

二、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)



续前表

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)

统计学是一门应用性很强的学科,由于几乎所有的学科都要研究和分析数据,因而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并进而研究各学科内在的规律,还需要由各学科的研究来完成。比如,古老的大量观察法已经发现了新生婴儿的性别比是 107:100,但为什么会是这样的比例?形成这一比例的原因则应由人类遗传学或医学来研究和解释,而非统计方法所能解决的了的。再如,利用统计方法对吸烟和不吸烟者患肺癌的数据进行分析,得出吸烟是导致肺癌的原因之一,但为什么吸烟能导致肺癌?这就需要医学进行解释了。

由此可以看出,统计学能做什么和不能做什么。可以这样说,统计方法仅仅是一种有用的、定量分析的工具,它不是万能的,不能解决你想要解决的所有问题。能否用统计方法解决各学科的具体问题,首先要看使用统计工具的人是否能正确选择统计方法,其次还要在定量分析的同时进行必要的定性分析,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的结果做出合理的解释和分析,才能得出令人满意的结论。尽管各学科所需要的统计知识不同,所使用统计方法的复杂程度大不相同,统计学也不能解决各学科的所有问题,但统计方法在各学科的研究中会发挥越来越重要的作用。

三、历史上著名的统计学家

在统计学的产生与发展过程中,出现了一些著名的统计学家,他们为统计学的发展做出了卓越贡献。下面列出了历史上著名的一些统计学家。

- Jacob Bernoulli (伯努利) (1654—1705)



- Edmond Halley (哈雷) (1656—1742)
- De Moivre (棣美佛) (1667—1754)
- Thomas Bayes (贝叶斯) (1702—1761)
- Leonhard Euler (欧拉) (1707—1783)
- Pierre Simon Laplace (拉普拉斯) (1749—1827)
- Adrien Marie Legendre (勒让德) (1752—1833)
- Thomas Robert Malthus (马尔萨斯) (1766—1834)
- Friedrich Gauss (高斯) (1777—1855)
- Johann Gregor Mendel (孟德尔) (1822—1884)
- Karl Pearson (皮尔森) (1857—1936)
- Ronald Aylmer Fisher (费歇) (1890—1962)
- Jerzy Neyman (尼曼) (1894—1981)
- Egon Sharpe Pearson (皮尔森) (1895—1980)
- William Feller (费勒) (1906—1970)

1.2 统计数据的类型

统计数据是对现象进行计量的结果。比如,对经济活动总量的计量可以得到国内生产总值(GDP)数据;对股票价格变动水平的计量可以得到股票价格指数的数据,对人口性别的计量可以得到男或女这样的数据等等。由于使用的计量尺度不同,统计数据可以分为不同的类型。下面我们从不同角度说明统计数据的分类。

一、分类数据、顺序数据、数值型数据

按照所采用的计量尺度不同,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)是对事物进行分类的结果,数据则表现为类别,是用文字来表述的。它是由分类尺度计量形成的。例如,人口按照性别分为男、女两类;企业按照经济性质分为国有、集体、私营、合资、独资企业等,这些均属于分类数据。为便于统计处理,对于分类数据我们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表示国有企业,2表示集体企业,3表示私营企业,等等。

顺序数据(rank data)也是对事物进行分类的结果,但这些类别是有顺序的。它是由



顺序尺度计量形成的。比如将产品分为一等品、二等品、三等品、等外品等；考试成绩可以分为优、良、中、及格、不及格等；一个人的受教育水平可以分为小学、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，对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) 是使用自然或度量衡单位对事物进行计量的结果，其结果表现为具体的数值。现实中我们所处理的大多数都是数值型数据。

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称品质数据 (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)。

二、观测数据和实验数据

按照统计数据的收集方法，可以将其分为观测的数据和实验的数据。观测数据 (observational data) 是通过调查或观测而收集到的数据，这类数据是在没有对事物人为控制的条件下而得到的，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) 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，对一种新的农作物品种的实验等，自然科学领域的大多数数据都为实验数据。

三、截面数据和时间序列数据

按照被描述的对象与时间的关系，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，它所描述的是现象在某一时刻的变化情况。比如，2003 年我国各地区的国内生产总值数据就是截面数据；时间序列数据 (time series data) 是在不同时间上收集到的数据，它所描述的是现象随时间而变化的情况，比如 1996—2003 年我国的国内生产总值数据就是时间序列数据。

下面我们给出统计数据分类的框图，如图 1—2 所示。

区分数据的类型是十分重要的，因为对不同类型的数据，我们将采用不同的统计方法来处理和分析。比如，对分类的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；对顺序的数据，我们可以计算其中位数和四分位差，计算等级相关系数等；对数值型数据还可以用更多的统计方法进行处理，如计算各种

统计量、进行参数估计和检验等。

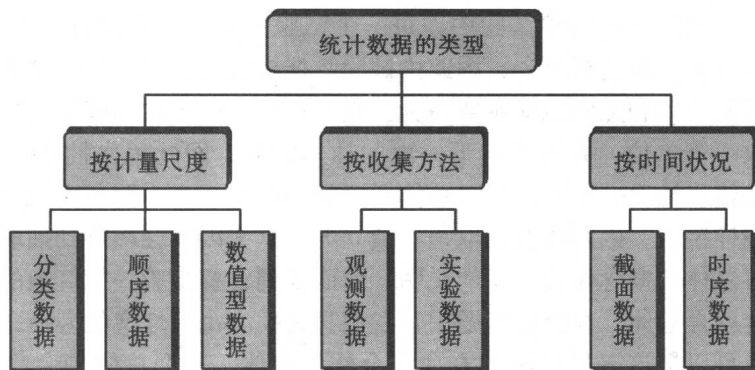


图 1—2 统计数据的分类

1.3 统计中的几个基本概念

统计中的概念众多，其中有几个概念是我们经常要用到的，有必要单独作一介绍。这些概念包括总体和样本、参数和统计量、变量等。

一、总体和样本

1. 总体 (population)。总体是包含所研究的全部个体的集合。它通常是由具有某种共同性质的许多个体组成，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，等等。组成总体的每一个元素称为个体，在由多个企业构成的总体中，每一个企业就是一个个体，由多个居民户构成的总体中，每一个居民户就是一个个体，由多个人构成的总体中，每一个人就是一个个体。

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，这一批灯泡构成的集合就是总体，每一个灯泡就是一个个体，总体的范围很清楚。但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消费该饮料，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，而且元素的数目是有限可数的。比如，由若干个企业构成的总