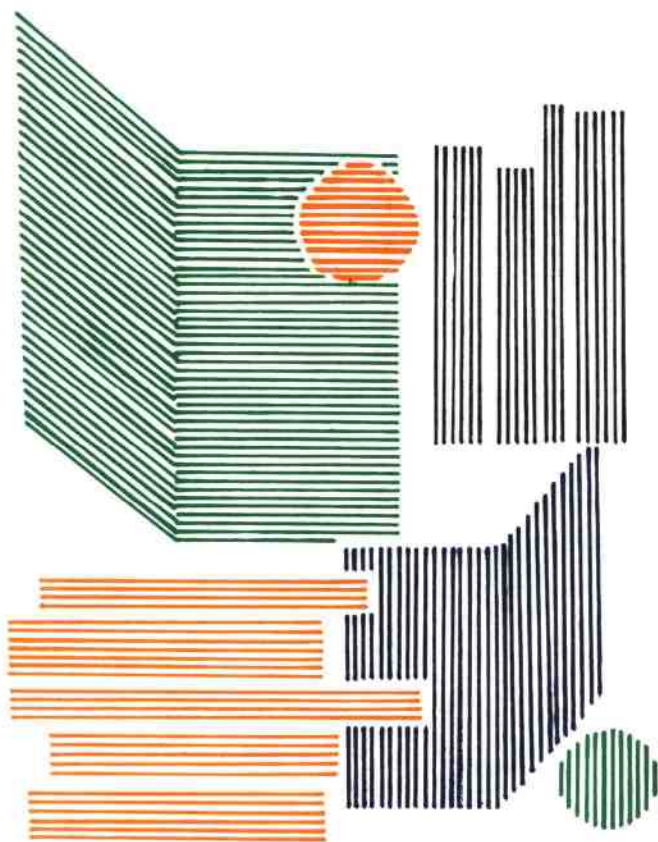


统计学

(修订版)

袁 卫 贾俊平 金勇进 编著
何晓群 易丹辉 高敏雪



中国统计出版社

统 计 学

(修订版)

袁 卫 贾俊平 金勇进 编著
何晓群 易丹辉 高敏雪

中国统计出版社

(京)新登字 041 号

图书在版编目(CIP)数据

统计学/袁卫等编著.

—北京中国统计出版社,1996.12 修订

ISBN 7-5037-0894-8

I. 统…

II. 袁…

III. 统计学—教材

IV. C8

中国统计出版社出版

(北京三里河月坛南街 75 号 100826)

新华书店经销

北京顺义兴华印刷厂印刷

*

850×1168 毫米 32 开本 12.375 印张 32 万字

1992 年 9 月第 1 版

1996 年 12 月第 2 版 2003 年 1 月北京第 16 次印刷

印数:77100—81100

*

定价:18.00 元

(版权所有 不得翻印)

前 言

《统计学》修订版是在原《统计学》(袁卫等编著,中国统计出版社1992年版)的基础上编写而成的。这次修订基本上保持了原教材的内容体系和框架,但在内容和写作方法上有较大变动。在内容上,对有关章节做了必要的调整和补充;在写法上,力求从实际问题入手阐明统计方法的基本思想和原理,做到深入浅出,通俗易懂,使学生通过自学就能基本上掌握教材内容。

本教材是高等院校财经各专业统计学课程的必修教材,同时也是自学考试的指定用书。全书由11章组成,内容上大体可分为四部分:一、描述统计(第一、二章);二、推论统计(第三、四、五、六章);三、经济与管理中常用的统计方法(第七、八、九章);四、国民经济核算基础知识(第十、十一章)。学习时,统计专业的学生要求学习本教材的全部内容;非统计专业和参加“文凭考试”的学生学习内容不包括教材中带有“※”号的部分。

本教材各章执笔人是:袁卫(绪论)、贾俊平(第一、七、八、九章)、金勇进(第二、五章)、何晓群(第三、四章)、易丹辉(第六章)、高敏雪(第十、十一章)。全书由贾俊平最后总纂定稿。参加本教材编写大纲讨论的有赵彦云、杜子芳、刘文卿、叶青、赵以立等同志,中国统计出版社对该书的出版给予大力支持,在此一并表示感谢。

尽管我们努力想奉献给读者一本简明实用的教材,但仍然有可能达不到这一要求。教材中的不当或疏漏之处在所难免,恳请读者多提宝贵意见。

编 者

1996年6月于中国人民大学

目 录

绪 论	(1)
第一章 统计数据的搜集与整理	(10)
第一节 数据的计量与类型	(10)
第二节 统计数据的搜集	(14)
第三节 统计数据的整理	(22)
习 题	(34)
第二章 统计数据的描述	(35)
第一节 绝对数与相对数	(35)
第二节 集中趋势的描述	(39)
第三节 离散程度的描述	(52)
习 题	(63)
第三章 概率与概率分布	(65)
第一节 概率基础	(65)
第二节 随机变量及其分布	(81)
习 题	(106)
第四章 参数估计	(109)
第一节 统计推断的基本概念	(109)
第二节 参数估计基本方法	(120)
第三节 总体均值和总体比例的估计	(128)

* 第四节 两个总体均值及两个总体比例之差的估计·····	(136)
习 题·····	(141)
第五章 假设检验 ·····	(144)
第一节 假设检验的一般问题·····	(144)
第二节 假设检验方法·····	(151)
第三节 假设检验方法的总结·····	(164)
习 题·····	(170)
* 第六章 方差分析 ·····	(172)
第一节 单因素方差分析·····	(172)
第二节 无交互作用的双因素方差分析·····	(184)
习 题·····	(192)
第七章 相关与回归 ·····	(194)
第一节 简单线性相关·····	(194)
第二节 一元线性回归·····	(204)
* 第三节 多元线性回归·····	(216)
习 题·····	(223)
第八章 时间数列分析 ·····	(226)
第一节 时间数列分析基础·····	(226)
第二节 长期趋势分析·····	(239)
第三节 季节变动分析·····	(253)
第四节 循环波动与不规则波动分析·····	(262)
习 题·····	(269)
第九章 指 数 ·····	(271)
第一节 指数的性质与分类·····	(271)

第二节	加权指数	(274)
第三节	指数体系	(284)
第四节	几种常用的重要指数	(293)
习 题		(301)
第十章	国民经济核算概述	(303)
第一节	国民经济核算的对象	(303)
第二节	国民经济核算常用分类	(306)
第三节	国民经济核算的内容框架	(309)
习 题		(319)
第十一章	国民经济基本总量统计	(320)
第一节	生产范围与生产总量统计	(320)
第二节	国内生产总值及其核算方法	(324)
第三节	收入和财产总量统计	(331)
第四节	经济增长率及其分析	(337)
习 题		(344)
附录一:	各章习题答案	(346)
附录二:	表 1 二项分布表	(355)
	表 2 泊松分布表	(363)
	表 3 标准正态分布表	(365)
	表 4 正态分布分位数表	(368)
	表 5 t 分布表	(371)
	表 6 χ^2 分布表	(373)
	表 7 F 分布表	(376)

绪 论

一、统计数据与统计学

在我们的日常工作与生活中，到处都可以见到统计数据。例如，开会时会议主持人要统计一下出席会议的人数；球类比赛时解说员总要统计竞赛双方的进攻次数和成功率；学生考试后非常关心自己的考试成绩和名次；企业管理人员每天要掌握生产销售情况和利润的数字；报刊和电视中常听到国民生产总值、物价指数和证券指数的数字等等。日常工作与生活中的这些数字就是我们所关心的统计数据。

统计学是一门收集、整理、描述、显示和分析统计数据的科学，其目的是探索数据内在的数量规律性。正是因为统计学总是在和数据打交道，因而我们也可称统计学为“数据的科学”。

怎样理解统计学是数据的科学呢？我们先看看统计学的英文名词，再举几个例子。

统计学的英文是“*statistics*”。在英文字典中它有两个含义：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”。“*statistics*”这一英文的解释至少可以告诉我们两件事情。

第一，由于统计数据在英文中是以复数形式出现的，表明统计数据不是指个别的单个数字，而是指同类的较多数字。因为单个数

字如果不和其他数据进行比较，是不能说明问题的。例如某个学生在某次考试中得了 85 分，如果仅凭这一个数字我们很难对这个学生的知识和能力作出判断和评价。因为这个 85 分可能是班上的最高分，可能是中等水平的分数，也可能是较低的分数。如果还知道其他学生的考分或者知道这次考试的平均分数，我们就可以对这个学生的成绩做出评价了。在生产和生活实际问题里，我们总是可以收集到较多的数据，因而在现实问题中的数据都可以称为统计数据。

第二，作为单数的统计学和作为复数的统计数据在英文中都用同一名词，其间显示有密切的关系。统计学是由收集、整理、描述、显示和分析统计数据的方法组成的，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，统计方法乃至统计学就失去了它存在的意义。这正如我们俗话所说“巧媳妇难为无米之炊”。这里的“巧媳妇”就是掌握统计方法的统计学家或统计工作人员，“米”就是统计数据，“炊”就是我们统计研究或统计工作的目的，即探索数据内在的数量规律性。显然，没有统计数据或没有较好的统计数据，就是再高明的统计方法或统计学家也难有所作为。

那么，什么是数据内在的数量规律性？我们用下面几个简单的例子来说明。

每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，新生儿的性比例似乎没什么规律可循。但如果对新生儿的性别进行大量观察，即观察成千上万个或者更多，就会发现男孩稍多于女孩，大致为每生育 100 个女孩，就有 107 个男孩。这个性比例 107 : 100 就是新生儿性别的数量规律性，因为古今中外都大致相同，它是由人类社会长期遗传和发展所决定的。因为人类社会要发展，就要保持男女人数的大致相同。那么有人会问，新生儿男多于女，不是性别不平衡了吗？是的，新生儿时男多于女不平衡，

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，导致男性的平均预期寿命比女性短，老年男性反而少于老年女性。从一个国家乃至全人类看，婴幼儿时男多于女，中青年时大致相同，老年时女多于男。在中青年结婚生育时性别大致平衡，总人口上性别也是大致平衡的，保证了人类社会的进化和发展。对人类性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性之一。

我们都做过掷硬币和掷骰子的游戏，都知道随机地掷一次硬币或骰子是不能事先确定出现正面反面或某个点数，也就是说个别的游戏或试验中充满了不确定性或偶然性。赌徒们正是利用了这种偶然性进行赌博。但当我们进行大量观察，即不断做重复试验时，就会发现掷一枚均匀硬币出现正面或反面次数都接近相同，即比值接近 $1/2$ 。试验的次数越多，就越接近 $1/2$ 这一稳定的数值。同样，在掷骰子时，出现 $1\sim 6$ 点的比率也逐渐接近 $1/6$ 。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，也就是我们探索的数量规律性。

在进行农作物试验时，如果其他试验条件相对固定，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，粮食产量的增加量逐渐减少。当施肥量增加到一定数值时，产量不再增加。这时如果再增加肥料，产量反而会减少。这一施肥量与粮食产量的数量关系就是我们要探索的数量规律性。当我们从大量试验数据中用统计方法探索到施肥量与产量之间的数量关系，就可以考虑肥料的费用并选择最佳的施肥效果了。

某企业开发出一种新的化妆品，在正式投产生产之前，需要根据市场需求情况制定其价格和销售量等策略。该企业委托某市场调查公司在全市 200 万户家庭中抽取 1000 户家庭作为样本，免费赠送给这 1000 户家庭试用，然后了解该化妆品的市场前景。如何科学地

从该市 200 万户家庭中抽出 1000 户家庭进行调查，并由 1000 户家庭反馈的市场信息科学地推断出全市 200 万户家庭对该化妆品的购买意愿这一数量规律是统计工作者的任务。

为什么统计方法能够通过对数据的大量观察和处理而研究和探索出其内在的数量规律性呢？这是由客观事物本身的特点和统计方法的特性共同决定的。从客观事物方面来说，根据辩证法的基本原理，任何客观事物都是必然性与偶然性的对立统一。同样，任何一个数据，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。如果客观事物只有必然性一个方面的特征，事物的表现形式就会比较简单，就可以比较容易地把握它的规律性。正是由于偶然性的存在，造成了事物的表现形式与必然性和规律性发生偏移，从而形成了表面形式的千姿百态，形成数据表现形式的千差万别。这样，必然性的数量规律性就被掩盖在表面的差异之中了。我们在上面举出的几个例子，本身都存在必然性的数量规律，但就每个新生儿的性别、每次掷硬币的结果、每次施肥带来的增产、每个家庭的购买意愿看都是不同的、有差异的，表现形式是充满偶然的。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表面的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（或推断）出事物总体的数量规律性。当然，如果事物本身的规律比较简单，所用的统计方法也就相对容易；如果事物本身的规律错综复杂，所用的统计方法也就相对复杂。那么显然，用什么统计方法，取决于我们有什么样的数据和要解决什么问题。

下面我们给出统计方法探索客观现象数量规律性过程的框图（见图 1）：

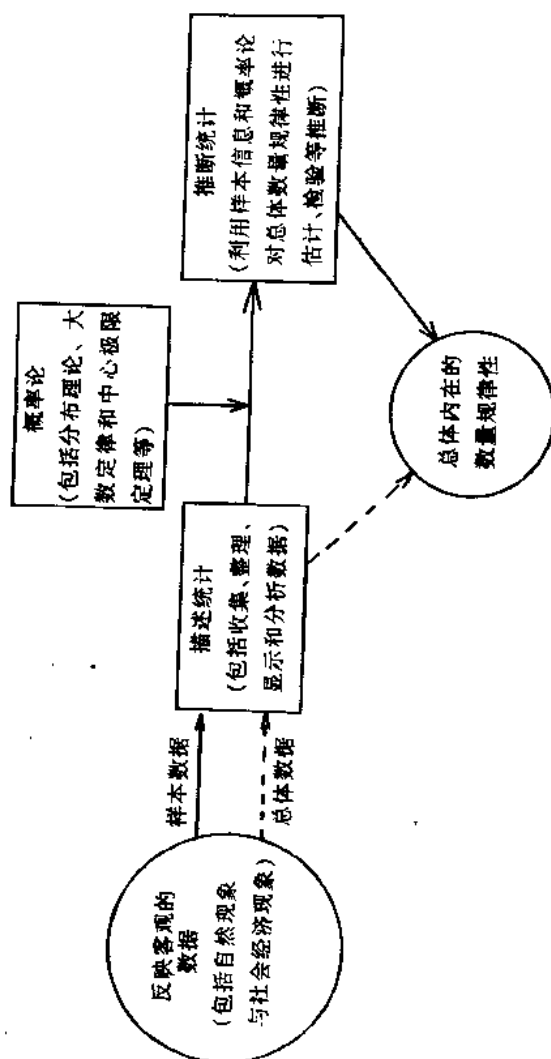


图 1 统计学探索客观现象数量规律过程的框图

由图 1 我们看到，统计研究过程的起点是数据，终点是探索到客观事物总体内在的数量规律性。要达到统计研究的目的，如果我们收集到的是总体数据（如普查），则经过描述统计之后就可以达到探索内在数量规律性的目的了；但如果我们所获得的数据只是研究总体的一部分数据，要探索到总体的数量规律性，必须应用概率论的理论并根据样本整理出的信息对总体做出科学的推断。显然，描述统计是整个统计学的基础和统计研究工作的第一步。它包括对客观现象的度量、调查方案的设计、及时、快速、经济地收集数据、整理数据、用图表显示数据、分析和提取数据中的有用信息以最终推断总体；推断统计是现代统计学的核心和统计研究工作的关键环节，因为统计最终能否科学准确地探索到总体内在的数量规律性与选用何种统计量、选用什么推断方法、如何进行推断有着直接的联系。

二、统计学的产生和发展

统计学产生于 17 世纪中叶，是从几个不同的领域开始的。

统计学的一个源头是来自英国威廉·配第（William Petty）的《政治算术》（1676 年）。配第在书中用大量的数字对英、法、荷三国的经济实力进行比较，用数字、重量、尺度等定量的方法进行分析 and 比较，表达他的思想和观点。马克思在《资本论》中评价配第“是政治经济学之父，在某种程度上也可以说是统计学的创造人。”

统计学的另一个创始人是英国的约翰·格朗特（John Grannt）。格朗特在 1662 年出版了《关于死亡表的自然观察与政治观察》。他通过大量观察的方法，研究并发现了人口与社会现象中重要的数量规律性，如新生儿的性比例稳定在 14：13；男性在各年龄组中死亡率高于女性；新生儿和大城市的死亡率较高；一般疾病与事故的死亡率较稳定而传染病的死亡率波动较大等等。在研究中，格朗特不但探索了人口变化和发展的一些数量规律，而且还对伦敦市总人口

数做出了较科学的估计。如果说配第是政府统计的创始人，则格朗特可被认为是人口统计的创始人。

统计学的第三个源头是古典概率论，奠基人包括法国的帕斯卡尔（B·Pascal）和费马特（Pierre de Fermat）。早在他们之前，已有一些数学家在研究赌博中的数量规律性了。意大利诗人但丁早在15世纪就讨论过掷三颗骰子可能出现的各种点数。16世纪中，意大利科学家伽利略讨论了掷三颗骰子出现10点次数多于9点次数的原因。在数学家们对机会游戏研究的基础上，帕斯卡尔和费马特通过通信的方式，将赌博中出现的各种具体问题，归纳为一般的概率原理，为后来概率论和统计学的发展奠定了重要的基础。

自17世纪中叶上述几位统计学家从不同角度开始了统计学研究后，经过数代统计学家的努力，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。

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。因为大批量产品生产要做全面检验，既费时、费钱又费人力，加之有些产品质量的检验要做破坏性检验，全部检验已不可能。1907年，英国的戈赛特（W. S. Gosset）提出了小样本 t 统计量理论，使统计学进入了现代统计学（主要是推断统计学）的阶段。以后经过著名统计学家费喧（R. A. Fisher）给出了F统计量、极大似然估计、方差分析等方法和思想；内曼（J. Neyman）和皮尔逊（E. S. Pearson）的置信区间估计和假设检验；沃尔德（A. Wald）的序贯抽样和统计决策函数到本世纪中基本构筑了现代统计学的基本框架。

从本世纪50年代以后，统计理论、方法和应用进入了一个全面发展的新阶段。统计学受计算机科学、信息论、混沌理论等现代科学技术的影响，新的研究领域层出不穷，例如多元统计分析、现代时间序列分析、贝叶斯统计、非参数统计、线性统计模型等。另一方面，统计方法的应用领域不断扩展，几乎所有科学研究都离不开

统计方法。因为不论是自然科学、工程技术、农学、医学、军事科学还是社会科学都离不开数据，要对数据进行研究和分析就必然要用到统计方法。因而可以说统计方法与数学、哲学一样成为所有学科的基础。

三、统计学与其他学科的关系

一、统计学与数学的关系

统计学与数学有密切的联系，但又有本质的区别，将统计学说成是应用数学的一个分支是不妥当的。我们说统计学与数学的关系密切，一方面是因为现代统计方法用到了几乎所有的现代数学知识，国内外的数理统计专业都要学较深的数学，一些杰出的统计学家是从数学专业转过来的。这就造成了一种错觉，似乎统计学是数学的一个分支。但实际上，数学对于统计学而言，仅仅是提供了统计理论和方法的数学基础，而统计学的主要特征是研究数据；另一方面是统计方法与数学方法一样，并不能独立地直接研究和探索自然现象和社会现象的规律，而是给各学科提供了一种研究和探索客观规律的数量方法。但统计学与数学又有本质的差别，主要表现在：第一，虽然表面上看来统计学与数学都是研究数量规律性，都是与数字打交道的，但实际上却有着明显的差别。数学研究的是抽象的数量规律性，而统计学研究的是具体的、实实在在的数量规律性。数学研究的是没有量纲或单位的抽象的数，统计学研究的是有具体实物或计量单位的数据；第二，数学研究所使用的是纯粹的演绎，而统计学是演绎与归纳相结合，占主导地位的是归纳。数学家可以坐在屋里，凭借着笔、纸和聪明的大脑从假设命题出发而推导出漂亮的结果。统计学家则必须深入实际收集数据，才能有所作为。因为没有对大量数据的归纳，统计学家就得出任何有益的结论。

二、统计学与其他学科的关系

由于统计学是一门应用性质很强的学科，而又由于几乎所有学科都要研究和分析数据，因而统计学与几乎所有学科都有或多或少的联系。这种联系表现为，统计方法可以帮助其他学科探索各学科内在的数量规律性，而对这种数量规律性的解释和由数量规律性进而研究各学科内在的规律只能由各学科的研究完成。例如，古老的大量观察法已经发现了新生儿性比例是 $107:100$ （或 $14:13$ ），但为什么会是这个性比例？这是由什么原因造成的则是人类遗传学或医学的任务而非统计方法而所能了。又例如，统计方法通过对吸烟和不吸烟者患肺癌的数据进行分析，得出吸烟是肺癌的病因之一。但为什么吸烟能导致肺癌就需要医学作出解释了。因而我们可以这样说，统计方法仅仅是一种有用的工具，是一种定量分析的工具。但能否真正解决各学科的问题，第一要看使用这种工具的人是否能选择正确的工具；第二要在定量分析的同时与定性分析相结合，也即应用各学科的专业知识对统计的结果作出合理的解释和分析。当我们应用统计方法解决社会经济现象和问题时，就更要小心，因为社会经济现象比自然现象更加错综复杂。自然现象的影响因素尽管有时也很多，但多数情况下我们可以在实验室中控制一些影响因素而观察另一些影响因素的变化规律。社会经济现象与时间因素交织在一起，加之人为因素的影响，所有的因素都同时发生作用。给我们探索必然性的数量规律性带来了较大的困难。但只要认真地进行分析研究，社会经济现象的统计分析同样也能得到令人满意的结果。

第一章 统计数据的搜集与整理

统计数据是我们利用统计方法进行分析的基础，离开了统计数据，统计方法就成了“无米之炊”，失去了用武之地。那么，统计数据从哪里取得呢？在取得统计数据之后，怎样才能使这些数据适合于我们分析的需要呢？本章主要介绍统计数据的搜集与整理方法，为以后各章介绍的统计分析方法打下基础。

第一节 数据的计量与类型

一、数据的计量尺度

统计数据是对客观现象进行计量的结果，在搜集数据之前，我们总是要先对现象进行计量或测度。但不同事物我们能够予以计量或测度的程度是不同的，比如有些事物只能对它的属性进行分类，有些则可以用比较精确的数字加以计量。按照我们对客观事物测度的程度或精确水平来看，可将所采用的计量尺度由低级到高级、由粗略到精确分为四个层次，即定类尺度、定序尺度、定距尺度和定比尺度；采用不同的计量尺度可以得到不同类型的统计数据。

1. 定类尺度

定类尺度也称列名尺度，它是最粗略、测度层次最低的计量尺度。这种计量尺度只能按照某种属性对客观事物进行平行的分类或分组。例如，按照性别将人口分为男、女两类；按所有制性质将企业分为国有、集体、私营、合资企业等等。定类尺度只是测度了事