

统计学

袁 卫 赵彦云 易丹辉 编著
邹伟卫 汪叔夜 高敏雪

中国统计出版社

统 计 学

袁 卫 赵彦云 易丹辉
邹伟东 汪叔夜 高敏雪 编著

中国统计出版社

(京) 新登字 041 号

统 计 学

TONG JI XUE

袁 卫 赵彦云 易丹辉 编著
邹伟东 汪叔夜 高敏雪

*

中国统计出版社出版

(北京三里河月坛南街 38 号 100826)

北京通县永泰印刷厂印刷

*

850×1168 毫米 32 开本 10.5 印张 25.5 万字

1992 年 9 月第 1 版 1994 年 3 月北京第 2 次印刷

印数: 4 001—9 000

ISBN7—5037—0894—8/C·577

定价: 8.50 元

前 言

《统计学》是我国高等院校财经类专业通开的核心课程。本教材就是为了适应这一教学需要而编写的，可以作为高等院校财经各专业统计学课程的必读教材及社会学、人口学、新闻学、法律、档案学等专业学习统计学的选读教材，也可以作为广大经济管理人士的参考书。

本教材是在国家教委所制订《统计学》核心课程教学大纲的基础上，结合高等院校开设统计学的特定教学要求，从我国社会主义经济建设和改革开放的实际需要出发，按照理论联系实际和学以致用原则，总结了我国统计学课程建设的经验，参阅了国外同类统计学教材的一些有益作法，并兼顾宏观与微观经济管理的需要而编写完成的。

本教材包括：一、描述统计；二、推论统计；三、经济与管理中常用的统计方法；四、国民经济统计指标体系等四个部分，共十四章。财经各专业，一般学习第一、二、四部分，第三部分内容将视教学需要和学员的水平酌情选用。对于社会学、人口学、新闻学、法律、档案学等专业，则视需要重点选学第一至第三部分。

本教材的第一、二部分由袁卫教授、汪叔夜博士编写，第三部分由易丹辉副教授、邹伟东副教授编写，第四部分由赵彦云副教授、高敏雪硕士编写，袁卫、赵彦云、易丹辉对全书进行了审定。在编写中，**韩嘉骏**教授、**徐前**教授、袁寿庄教授、倪加勋教授对全书的编写提出了许多建设性的意见；刘文卿硕士对书中部分章节提出了许多宝贵意见。此外，黄强、孟昭红、张利双同志在本书排版、印刷过程中给予了大力协助。在此一并表示由衷的感谢。

由于我们对编写核心教材缺乏经验，以及编者水平的局限，书中不妥之处在所难免，欢迎读者批评指正。

编 者 1992年6月 于中国人民大学

目次

第一部分 描述统计	(1)
第一章 绪论	(1)
§ 1. 1 统计学的概念	(1)
§ 1. 2 统计数据的来源	(4)
§ 1. 3 数据的计量尺度	(5)
第二章 统计数据的描述	(10)
§ 2. 1 统计调整方案和方式	(10)
§ 2. 2 统计数据的整理	(14)
§ 2. 3 集中趋势的测度	(23)
§ 2. 4 离散程度的测度	(33)
§ 2. 5 茎叶图	(38)
习 题	(45)
第二部分 推论统计	(47)
第三章 概率与概率分布	(47)
§ 3. 1 概率基础	(47)
§ 3. 2 概率分布	(57)
习 题	(72)
第四章 参数估计	(76)
§ 4. 1 统计量与抽样分布	(76)
§ 4. 2 点估计	(81)
§ 4. 3 区间估计	(87)
习 题	(96)
第五章 假设检验	(98)
§ 5. 1 假设检验的基本思想与步骤	(98)
§ 5. 2 总体均值的假设检验及两个总体均值之差的假设检验	(102)
§ 5. 3 总体比例的假设检验	(106)
§ 5. 4 统计质量控制	(109)
习 题	(115)
第六章 方差分析	(117)

§ 6. 1	单因素方差分析	(117)
§ 6. 2	双因素方差分析	(124)
§ 6. 3	试验设计初步	(129)
习 题		(136)

第三部分 经济与管理中常用的统计方法

.....	(138)
第七章 相关与回归	(139)
§ 7. 1 相关分析	(139)
§ 7. 2 一元线性回归	(148)
§ 7. 3 多元线性回归	(162)
习 题	(168)
第八章 时间数列	(172)
§ 8. 1 时间数列的概念和分类	(172)
§ 8. 2 时间数列分析	(174)
§ 8. 3 时间数列的构成分析	(182)
习 题	(193)
第九章 指 数	(196)
§ 9. 1 指数的概念和分类	(196)
§ 9. 2 简单指数	(199)
§ 9. 3 综合指数	(202)
§ 9. 4 加权平均数指数	(209)
§ 9. 5 指数体系	(214)
习 题	(224)
第十章 统计决策	(227)
§ 10. 1 统计决策的基本概念	(227)
§ 10. 2 决策准则	(229)
§ 10. 3 决策树	(242)
§ 10. 4 贝叶斯决策	(250)
习 题	(255)

第四部分 国民经济统计指标体系	(258)
第十一章 国民经济统计指标体系的基本问题	(258)
§ 11. 1 国民经济统计基本概念	(258)
§ 11. 2 国民经济活动基本分类	(261)
§ 11. 3 国民经济统计指标体系的基本内容	(264)
第十二章 国民经济生产与使用统计指标	(266)
§ 12. 1 国民经济生产总量统计指标	(266)
§ 12. 2 国民经济生产结构统计指标	(270)
§ 12. 3 国民经济使用总量及结构统计指标	(273)
§ 12. 4 国民经济生产和使用统计的价格	(275)
§ 12. 5 国民经济生产和使用分析统计指标	(277)
第十三章 国民经济流通与分配统计指标	(283)
§ 13. 1 国民经济流通统计指标	(283)
§ 13. 2 财政统计指标	(285)
§ 13. 3 金融统计指标	(288)
§ 13. 4 国际收支统计指标	(293)
§ 13. 5 居民收支统计指标	(295)
§ 13. 6 国民经济分配总量统计指标	(297)
第十四章 国民经济生产条件统计指标	(301)
§ 14. 1 人口和劳动力统计指标	(301)
§ 14. 2 国民财产和自然资源统计指标	(305)
习 题	(308)
参考书目	(315)
附表:	(316)
1. 二项分布表	(316)
2. 泊松分布表	(322)
3. 正态分布表	(324)
4. 七分布表	(325)
5. T 分布表	(326)
6. X^2 分布表	(328)
7. D. W. 检验表	(329)

第一部分 描述统计

第一章 绪 论

§ 1.1 统计学的概念

统计学是一门关于大量数据资料的收集、整理、描述和分析的学科，其目的是探索数据内在的数量规律性。

统计学的英文是“Statistics”。这个英文名词有两个含义：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“统计资料”或“统计数据”。显然，由其英文名词复数的特征我们知道“统计数据”不是指个别的单个数据，而是指同类的较为大量的数据。因为单个数据在没有其它数据进行比较时，是说明不了什么问题的。例如某次统计学考试某个考生得 85 分，如果仅仅这么一个数字我们很难对这个考生的成绩做出分析和评价，因为我们不知道其他考生的考试成绩如何。当我们知道了其他考生的考试成绩时，就可以对该班考生的成绩做些整理和分析。可能这个 85 分是班上最高分，可能是较高的分数，可能处于中等水平，也可能是较低的分数。这时，我们就可以对这个考生的统计学知识做出评价了。

在英文中，作为单数的“统计学”和作为复数的“统计资料”间有着密切的联系。“统计学”是由收集、整理、描述和分析“统计资料”的统计方法组成的，这些方法来源于对“统计资料”的研究，目的在于对“统计资料”进行研究。显然，离开了统计资料，统计学就失去了它存在的意义。这正如俗语所说“巧媳妇难为无米之炊”。这里的“巧媳妇”就是掌握了统计方法的统计学家或统计工作人员，“米”就是统计数据或统计资料。没有统计资料，再好的统计学家也难有所作为。这里的“炊”就是我们在本

节第一句话所说的“探索数据内在的数量规律性。”

什么是数据内在的数量规律呢？下面我们用几个例子来加以说明。

【例 1.1】 就每一个家庭中一个新生儿的性别来说，是不确定的。生出来的可能是男孩，也可能是女孩。在过去没有实行计划生育的时候，某个家庭可能连生几个男孩，另一个家庭也可能连生几个女孩。因而，新生儿的性别从表面上看或从个别家庭看，似乎是没有什么规律的。但如果对新生儿的性别大量观察，即观察几百个、几千个或者更多的新生儿，就会发现男孩比女孩稍多一点，大致为每生 107 个男孩，就会有 100 个女孩。这个性别比例 107 : 100 就是新生儿性别比例的稳定的数值，也就是新生儿性别比例的数量规律性。之所以将这个 107 : 100 称为数量规律性，是因为古今中外新生儿性别比例都大致如此。它是由人类遗传和人类社会发展需要所决定的。我们知道，人类社会要保持稳定的发展，要求男女总人数大致相同。那么有人会问，新生儿中男多于女，不是性别不平衡了吗？是的，新生儿时男多于女是不平衡的。但男孩的死亡率高于女孩，到了中年时，该年龄组的男女人数就大致相同了。同时，又由于男性的平均预期寿命短于女性，这样，在老年组中，男性就比女性少了。从一个国家乃至全人类看，婴幼儿时男多于女，中年时大致相同，而老年时女多于男。总体上还是大致平衡的，中青年结婚生育时，也是大致平衡的。这是人类经过长期进化的结果，是一种和谐的平衡。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。

【例 1.2】 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是不能事先确定会出现哪一面的，随机地掷一次骰子也不能事先确定掷出的点数，也就是说在单独的游戏和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。然而，当我们进行大量观察时，即进行不断的重复试验时，我们会发现掷一枚均匀硬币出现正面和反面的次数都接近 0.5。

试验的次数越多，就越接近 0.5 这一稳定的数值。同样，在掷骰子中，出现 1—6 点的次数也逐渐接近 $\frac{1}{6}$ 。这中间的 0.5 和 $\frac{1}{6}$ 就是掷硬币和掷骰子中出现某一种特定结果的概率，也就是我们所要探索的数量规律性。

【例 1.3】 在进行农业试验时，发现某种粮食作物的产量会随某种施肥量的增加而增加。当开始追加施肥量时，产量增加较快，以后同样的追加施肥量所带来的粮食产量增量逐渐减少。当施肥量增加到一定数值时，产量不再增加了。此时如再增加施肥量，产量反而会减少。这一施肥量与产量之间的数量关系就是我们要探索的数量规律性。当我们从大量资料中把握了这一数量规律性，即用方程式描绘出了施肥量与产量之间的数量关系，就可以考虑肥料的费用而选择最佳的施肥效果了。

【例 1.4】 党的十一届三中全会制定了改革开放的方针政策以来，广大城乡人民群众的收入到底提高了多少？现在比十年前的分配状况是更平均化了还是出现明显的分配不公、两极分化了呢？这也需要我们探索城乡人民群众十年来收入分配变化的数量规律性。这一数量规律性可通过分别列出十年前和现在城乡人民群众收入分配曲线并对两条曲线进行比较而得出。

为什么统计方法能够通过通过对数据的大量观察和处理进而研究和探索到其内在数量规律性呢？这是由客观事物本身的特点和统计方法的特性所共同决定的。从客观事物来说，根据唯物辩证法的基本原理，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联系，是比较稳定的，因而它决定了事物的表现形式是有规律可循的。如果客观事物只有必然性一个方面，那么事物的表现形式就会比较简单、单一，恐怕不需要统计方法，直观地看一看就可以把握住它的规律性了。也正是由于偶然性的存在，它影响事物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，形成数据表面的千

差万别。这样，必然性和数量规律性就被掩盖在千差万别的表面差异之中了。统计学提供了一系列的方法，可以对表面上杂乱无章的数据加以整理和分析，可以根据需要和客观数据的特点探索出隐藏在数据之中的数量规律性来。然后，再进一步对这些数量规律进行分析和解释。

§ 1.2 统计数据的来源

我们在第一节中提到统计数据是利用统计方法进行分析的基础，是统计科学研究的原料。那么，统计数据从哪里获得呢？这一节我们简单介绍统计数据的来源。

1.2.1 来源于公开的出版物

对于中国的国民经济统计数据，国家统计局和各级统计部门都定期地向社会公开发表，如统计公报、经济形势的报告、中国统计信息报等等。同时，中国统计出版社和其它一些出版单位还定期出版《中国统计年鉴》、《中国统计摘要》、《中国农村统计年鉴》、《中国分县农村经济统计摘要》、《中国工业经济统计年鉴》、《中国建筑业统计资料》、《中国社会统计资料》、《中国商业和对外经济统计资料》、《中国科技统计资料》、《沿海经济开发区经济研究和统计资料》、《国际经济和社会统计摘要》等。各省、市、自治区的统计部门都向社会提供各地区的统计数据，如《北京市社会经济统计年鉴》、《山东省物价调查统计年鉴》等。如要了解世界各国的统计数据，还可以在各地图书馆查阅《联合国统计年鉴》、《美国统计摘要》、《日本统计月报》等公开出版物。以上这些公开出版物为我们进行统计研究提供了大量的数据，这是统计数据的一个主要来源。

1.2.2 来源于国家统计部门和地方各级统计部门的统计报告制度

我国是社会主义国家，现在实行的是有计划的商品经济。建国 40 年来，我国已形成了一套以各级统计部门为主的统计报告制度。例如，工业企业有定期的日报、旬报、月报、季报和年报，向上级主管业务部门和统计部门呈报生产产量、产值、物耗等情况。上级主管业务部门和统计部门在汇总后又向上一级呈报，最后汇总到国家统计局和国务院各主管部门。农业上也有定期的产量报告制度。为搞清城乡居民生活的变化和农产量的数据，经国务院批准，由国家统计局代管两支抽样调查队，一支是城市居民生活调查队，另一支是农村经济调查队，各省、市、自治区均有专职调查队搜集城乡居民生活和农产量的数据。这些数据中的一小部分以年鉴和公报的形式供社会使用，大量数据作为统计部门内部研究分析使用，因而这是统计数据的又一主要来源。

1.2.3 来源于专门组织的调查

对于统计研究来说，以上两方面的资料和数据有时还满足不了要求。这时，就需要针对所要研究的问题，专门组织一些统计调查，来获取所需要的数据。从全国范围的专门性调查来说，1982 年 7 月 1 日零时我国进行了第三次人口普查工作；1985 年进行了全国工业普查工作；1990 年 7 月 1 日零时我国又进行了第四次人口普查工作。从各省、市、自治区及更小的范围来说，这种专门的调查（如小型普查和抽样调查等）就更多了。例如在北京市进行家用电器购买愿望的抽样调查，对某一种商品市场需求的抽样调查等。通过这些小型的调查搜集必要的资料，为我们的专题研究提供数据。

§ 1.3 数据的计量尺度

要对数据进行统计分析，首先要对数据进行计量。要对数据进行计量，就要确定数据的计量尺度。数据尺度要根据数据所衡

量的客观事物的不同特征来加以确定。同时，不同的数据计量尺度也就决定了所要应用的不同的统计方法。从对客观事物加以精确计量的角度分类，由低到高、由粗略到精确可以分为如下四种计量尺度。

1.3.1 列名尺度 (Nominal scale)

这是一种最粗略、计量精度最低的计量尺度。这种计量尺度的特点是只能对数据进行平行的分类或分组，各组各类之间关系平等。例如我国省、市、自治区行政区划分类可将大陆土地面积、人口等分成 30 个组，即 30 个行政区划。每个省、市、自治区可有一个代号，如北京为 01，河北为 02……。大陆的人口、企业、县等都可按这一标准分在不同的组内。每个组内分配的是单位的个数，如 1982 年北京市有 9 230 687 人，河北省有 53 005 875 人……。又如 1982 年北京有 3 900 个工业企业，工业总产值为 228.69 亿元（按 1980 年不变价计算），河北省 1982 年有 15 600 个工业企业，工业总产值为 229.94 亿元（按 1980 年不变价计算）……。这些分组都是按列名尺度的分组。所谓列名尺度，是指分组标志（即分组的标准）是对一些客观事物的名称进行排列，如北京、河北、……。这些排列不分大小，都是并列的关系。该种计量尺度在社会经济现象中有着广泛的应用，对其进行分析的统计量主要是计算各组次数占总体（即全部）的频率（即各组比重）。

1.3.2 顺序尺度 (Ordinal scale)

这种计量尺度比列名尺度稍精确一些，表现为它除具有列名尺度所具有的性质和特征外，各组之间还可以比较大小。例如产品的等级可以分为特等品、一等品、二等品等等，考试成绩也可以分为优、良、中、及格、不及格等等，对国家企业进行综合评价后可以分为一级企业、二级企业等等。显然，在产品的等级中“特等品”优于“一等品”，“一等品”优于“二等品”。考试成绩

“优”优于“良”，“良”优于“中”等等。国家企业“一级企业”优于“二级企业”，“二级企业”优于“一般企业”。由于各组之间能够进行排序和比较，在分析中就可以应用较多一些的统计方法。顺序尺度的统计量不仅可以用频率进行分析，而且可以大致计算出总体的众数和中位数，以及四分位数、十分位数等等，分析和处理的办法就更多了一些。

1.3.3 间隔尺度 (Interval scale)

这是一种较列名尺度和顺序尺度更高级更精确的计量尺度，它以某一种客观存在的计量单位作为其分组标志。例如考试的百分制成绩、温度等等。在考试成绩中，总分为100分，每分与每分之间的间隔是相等的。也就是说，80分与90分的差距和90分与100分之间的差距是相同的，即都差10分。温度也是一样，以摄氏温度为例，温度每提高一度其温度差也是相同的。显然，由顺序尺度的五级分制到间隔尺度的百分制，其计量尺度是更精确了。也就是在五级分制的每个等级中又细分出若干个等级，而且其等级之差是间隔相等的。间隔尺度具有顺序尺度所具有的特征，在顺序尺度中适用的统计量如众数、中位数等对间隔尺度完全适用。这是因为间隔尺度可以换算成顺序尺度，反之，顺序尺度则不能转换成间隔尺度。例如可将百分制考试成绩折合成五级分制成绩：

百分制成绩		五级分制成绩
90—100分	→	优
80--89分	→	良
70—79分	→	中
60—69分	→	及格
60分以下	→	不及格

由于间隔尺度中每一间隔是相等的，因而各数值之间不但可以比较大小，而且可以用加减法准确地给出差异大小的数值。作为其适用的统计量，除频率、众数、中位数等外，还可以计算数

据的均值和标准差，这样对其数据的分析就精确得多了。

1.3.4 比率尺度 (Ratio scale)

比率尺度与间隔尺度属于同一等级的尺度，因而在有些教科书中不对它们做出区别。这两种尺度的细微差别是间隔尺度没有绝对零点，而比率尺度有绝对零点。换句话说，间隔尺度中的“0”只表示某一个数值，即0数值，不表示没有。而比率尺度中的“0”则表示没有。应用比率尺度计量的客观事物很多，如人的身高、体重、产量、产值、物价等等。如果一个人的身高为0米或者体重为0公斤，此人则不存在。因为没有一个人身高和体重为0。如果一个企业某月产量和产值均为0，也表示该月份没有生产活动。然而，间隔尺度中的0含义就不同了，它表示的是某个0水平，不是没有或不存在。例如某个考生在研究生入学考试中数学得了0分，它表示在该种难度的考试中他的成绩是0分，但不等于他完全没有数学知识。又例如温度中的 0°C ，它表示冰点这一特定温度，不表示没有温度。由于绝大多数的客观事物都是按比率尺度计量的，间隔尺度只是其中的一些特例，因而某些教科书中不对它们进行区别也是有道理的。

在比率尺度中，由于“0”表示不存在，因而其数值的比较不仅可以比较大小、计算其差异的具体数值，而且可以计算数值之间的倍数。换句话说，不仅可以应用加、减法进行计算，而且可以用乘、除法四则运算处理。我们可以说体重100公斤的人其重量是体重50公斤的人的两倍，但不能说研究生数学成绩80分的考生其数学知识是40分考生知识的两倍，其道理在于是否存在绝对零点这一细小差别。

我们将数据的计量水准分为这四种尺度是为了后面学习统计方法时针对不同数据选择不同方法做准备的。下面，我们将四种计量尺度及其特征和适用的统计量列在表1-1中，当我们学完整部教材后再返回来理解，印象也许会更深。

表 1—1 数据的四种计量尺度及其相应的统计量

数据尺度	特 征	统计量	统计推断
列名尺度	组内等价 组间并列	频 率 列联系数	非参数 统计推断
顺序尺度	组内等价、组间可比较大小， 用 $>$ 、 $<$ 表示	众数、中位数 四分位数等	非参数 统计推断
间隔尺度	具有列名、顺序尺度特征， 不但可比较大小，而且可求出 大小差异数值，可用 $+$ 、 $-$ 法。	均值、标准差 等	参数与非参数 统计推断
比率尺度	具有前三种一切特征，在比 较中，有绝对 0 点，可用 $+$ 、 $-$ 、 $\times$ 、 $\div$ 等方法运算	所有统计计量	参数与非参数 统计推断

第二章 统计数据的描述

§ 2.1 统计调查方案和方式

2.1.1 统计调查方案

不论是全国性的大规模调查，还是小范围的专题调查，在调查之前都要求制定出一个较周密、完整的调查方案，以指导整个调查工作。调查方案应包括以下主要内容：

(1) 调查目的。例如国家统计局 1987 年 9 月颁布的城市住户调查方案的目的是：“系统地搜集、整理城镇各阶层非农业居民家庭的人口、就业、现金收支及主要商品消费量等资料，以便研究城镇居民的收入在消费与储蓄之间的分配和消费结构，为各级党政领导研究人民生活、市场物价、货币流通、劳动工资等政策提供信息。”在明确调查目的的前提下，调查方案的其他项目就容易设计了。

(2) 调查对象和单位。这是进一步明确向谁调查的问题。调查对象就是所要进行调查的社会经济现象的全部，也称为调查总体。这个调查总体是由组成这一总体的若干调查单位构成的，例如在城市住户调查中，调查对象确定为“在城市和县城关镇中选择部分非农业居民家庭及其成员作为调查对象”。尽管我们所要了解的是城市和县城关镇所有非农业居民家庭的情况，但调查总体不是居民家庭的全部，只是其中的一部分，因而是一种非全面的调查。作为构成调查对象的调查单位，该项调查确定为“以户及个人为统计单位。”

(3) 调查表。即用表格的形式列出调查项目，供调查单位填报或由调查员来负责填报。这是一项调查的最主要内容，因为调查表是根据目的和调查对象设计的，是搜集统计数据的工具。例如城市住户调查的调查表共有 8 种，其中“居民家庭成员基本情况一览表”表式如表 2—1。