



21世纪高等院校教材

文科通用统计学

许涤龙 主编



科学出版社

013926182

C8-43
193

21 世纪高等院校教材

文科通用统计学

许涤龙 主编



科学出版社

北京



北航

C1633394

C8-43
193

S81330810 ..

内 容 简 介

本书是为文科类专业编写的统计学教材,在编写中力求贯彻“通俗、实用、有趣”的指导思想:用简明通俗的语言介绍统计学的思想与方法;按在文科领域实用的原则取舍内容和组织结构,不追求统计方法的推导证明,而讲求统计方法的实际应用;将文科领域的典型案例和有趣事件大量穿插于各种统计方法的应用之中,挖掘统计方法的趣味性,使得对统计学的学习更加贴近实践、贴近生活,并且易于理解和掌握。

本书内容围绕统计学数据及其搜集处理、分析应用的主线而展开,主要内容包括统计简介、统计数据及数据来源、统计数据的描述、机会的度量、抽样推断、假设检验、非参数统计、相关与回归、多元统计分析和综合评价。

本书可作为大学文科类各专业的统计学教材,也可作为人文、政治、法律、行政、新闻等领域的实际工作者学习、使用统计方法与工具的参考书。

图书在版编目(CIP)数据

文科通用统计学 / 许涤龙主编. —北京: 科学出版社, 2013

ISBN 978-7-03-036751-8

I. ①文… II. ①许… III. ①统计学—高等学校—教材 IV. ①C8

中国版本图书馆 CIP 数据核字 (2013) 第 034787 号

责任编辑: 兰 鹏 林 建 / 责任校对: 黄江霞

责任印制: 徐晓晨 / 封面设计: 蓝正设计

科 学 出 版 社 出 版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

新科印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2013 年 3 月第 一 版 开本: 787×1092 1/16

2013 年 3 月第一次印刷 印张: 18

字数: 430 000

定价: 39.00 元

(如有印装质量问题, 我社负责调换)

前言

国内已出版的大学统计学入门教材，基本上有两类：一类是主要面向理工类专业的“统计学”，具有比较浓厚的概率论与数理统计色彩，讲求数学推导和论证；另一类是主要面向经济和工商管理类专业的“统计学”，比较注重统计方法的应用，但主要针对的是统计方法在经济管理领域的应用。

随着社会的进步和教育的发展，现在越来越多的“纯”文科类专业（如法学、政治学、行政管理、新闻学、传播学、历史学甚至文学艺术等专业）开设了统计学课程，但国内目前很少见到面向文科类专业的统计学教材，一些开设了统计学课程的文科类专业只好使用为理工类或经管类专业编写的统计学教材，老师组织教学比较难，学生学起来也不对路。

本书是尝试为文科类专业编写的统计学教材，在编写中力求贯彻“通俗、实用、有趣”的指导思想：用简明通俗的语言介绍统计学的思想与方法；按在文科领域实用的原则取舍内容和组织结构，不追求统计方法的推导证明，而讲求统计方法的实际应用；将文科领域的大量典型案例和有趣事件穿插于各种统计方法的应用之中，挖掘统计方法的趣味性，使对统计学的学习更加贴近实践、贴近生活，并且易于理解和掌握。

本书由许涤龙教授主编并设计编写思路、基本框架和体例要求，由编写组共同讨论编写大纲和具体内容。各章的编写者分别为：第一章许涤龙教授，第二章张运讲师，第三章李正辉教授，第四章任英华副教授，第五章谭朵朵讲师，第六章李庭辉副教授，第七章倪青山副教授，第八章陈黎明副教授，第九章马守荣讲师，第十章张芳副教授。博士生欧阳胜银协助主编搜集了部分资料并编写了部分初稿。书稿完成后，主编对全书进行了统稿和总纂。

在本书编写中，编者参考了大量的国内外相关论著，其中有的已在书中列出，有的由于篇幅或体例所限未能在书中一一列出，谨向未列出论著的作者致歉，并向所有参考论著的作者诚恳致谢！同时，对科学出版社在本书编写和出版过程中所给予的大力支持表示衷心的感谢！

编 者

2013年1月

目 录

前言

第一章 什么是统计	1
第一节 统计的含义	1
第二节 现实中的随机性和规律性	3
第三节 统计学的基本范畴	5
第四节 统计学的应用	8
第二章 统计数据及数据来源	12
第一节 统计数据的分类与特点	13
第二节 统计数据的来源	17
第三节 统计数据的质量	31
第三章 统计数据的描述	34
第一节 描述数据的统计指标	34
第二节 数据的图表显示	40
第三节 数据的概括性度量	51
第四章 机会的度量	61
第一节 机会与概率	61
第二节 离散型随机变量	71
第三节 连续型概率分布	82
第四节 正态分布在社会科学中的应用	93
第五章 抽样推断	100
第一节 抽样推断概述	100
第二节 抽样分布	105
第三节 参数估计	110
第四节 抽样组织设计	122
第六章 假设检验	131
第一节 假设检验的基本原理	131
第二节 单个总体参数的检验	136
第三节 两个总体参数的检验	141
第四节 方差分析	147

第七章 非参数统计	156
第一节 χ^2 检验	158
第二节 秩和检验	162
第三节 游程检验	168
第八章 相关与回归	172
第一节 相关分析与回归分析概述	173
第二节 两变量的相关分析	175
第三节 一元线性回归分析	179
第四节 一元线性回归分析的扩展	186
第五节 离散因变量模型	191
第九章 多元统计分析	199
第一节 多元统计分析概述	199
第二节 多元线性回归分析	200
第三节 主成分分析	209
第四节 因子分析	219
第五节 聚类分析	233
第十章 综合评价	252
第一节 综合评价概述	252
第二节 评价指标选取与数据预处理	257
第三节 权数的确定与综合评价模型	265
参考文献	282

第一章

什么是统计

如果你是一个汉语、影视传媒等文科专业的学生，学统计对你来说或许会很难；如果你与某国有企业的一位老员工交谈，他可能会告诉你统计就是数“数”，根本没什么用途；如果你与一个学统计的“海归”聊天，那么就很有可能听到这样一句话，“你不懂计算机，根本学不懂统计学”；但如果你找一个学统计的国内大学生，问其“什么是统计学，统计学到底是干什么的”，或许这位同学也会有些困惑，答曰“我也不知道”。怎么样，有点犯糊涂了吧！

那么统计学究竟是什么呢？统计学真的就是说不清、道不明的“四不像”吗？下面我们将带大家一起来了解、学习统计学。

第一节 统计的含义

一、统计的由来

最早的统计是作为国家或地区对生活、自然规律等重要事项记录的一种方式而存在的，这样的统计方式存在于远古文明时期。例如，中国自夏王朝开始就有记录重大历史事件，并被统治阶级作为治国的手段；古埃及农民通过记录尼罗河水位的涨势，来确定最佳的畜牧时期，促进农业发展，并根据月圆情况与潮汐涨落确定日常作息时间。这些都可以视为早期的统计工作。

随着时间的推移，统计实践逐步得到发展。从产生和发展的过程来看，其主要分为三个时期，即统计学的萌芽期、统计学的近代期和统计学的现代期。统计学的萌芽期为 17 世纪中叶至 18 世纪中下叶，主要包括以康令(H. Conring)、阿亨瓦尔(G. Achenwall)和斯廖采尔等为代表的国势学派，他们记载了关于国家组织、人口、军队、领土、居民职业以及资源财产等的统计资料，并为这门学科起了一个至今仍为世界公认的名词“统计学”，同时也提出了“统计数字资料”、“数字对比”和“显著事项”等重要的术语和概念；在统计学的萌芽期还包括以威廉·配第(W. Petty)和约翰·格朗特(J. Graunt)为代表的政治算术学派。统计学的近代期为 18 世纪末至 19 世纪末，主要包括以法国天文学家、数学家、统计学家拉普拉斯(P. S. Laplace)和比利时统计学家、数学家、天文学家凯特勒(A. Quetelet)为代表的数理统计学派，以德国大学教授尼斯(K. G. A. Knies)、恩格尔(C. L. E. Engel)、

梅尔(G. V. Mayr)等为代表的社会统计学派。统计学的现代期为 20 世纪初至现在的数理统计时期, 主要代表人物有费希尔等。

专栏 1.1 统计的由来

德国是最早将“统计”当做学名使用的国家。1660 年, 德国在大学开设“国势学”课程, 并对各国情况进行比较介绍。当这门课传到哥丁根大学政治学教授阿痕瓦尔手中后, 阿痕瓦尔于 1749 年为这门课“国势学”定了一个新名称“statistik”(统计)。在英国, 虽然早在 17 世纪就用数字来说明个中社会经济现象, 但使用的是另一个完全不同的名称“政治算术”(political arithmetic), 当时还不知道“统计”这一名称。直到 18 世纪末, 《英格兰统计记事》的首任主编辛克莱才将英语的 statistic 作为德语的 statistik 的译文传入英国, 并赋予“统计”以新的内容, 即用数字作为语言来表述事物。“统计”一词很快就流传到美国, 直至今日, 其已广泛传播到各国。这个词约在 20 世纪初才由日本传入我国。

二、统计工作与统计资料

在我国, “统计”一词是由英文单词 statistics 翻译过来的, 具有统计学、统计工作和统计资料等含义。

统计工作是一个工作的过程, 具体来说是对社会、经济以及自然现象等进行搜集、整理和分析的工作过程。我们日常生活中无时无刻不在接触着统计工作, 如对全班考试成绩进行加总并求平均分, 对大学期间不同月份的生活开支进行统计, 对每天进出图书馆的学生数量进行统计, 这些都可以看做简单的统计工作。

统计资料是统计工作的成果, 是统计工作过程所取得的各项数据和有关情况的资料, 是实实在在的、可以确切地知道的东西。如对全班同学成绩进行加总得出的总分和平均分是个确切的数据, 大学期间不同月份的生活开支和每天进出图书馆的学生数量也都是一个确切的数据。运用这些统计资料, 人们便可以及时、准确、便捷地了解 and 掌握所需要的信息, 进而进行情景分析与预测。

统计资料来源于统计工作, 因此, 统计工作的优劣直接关系到统计资料的数量和质量。如何计划好统计工作, 是统计实践取得理想效果的前提。

三、统计科学

统计科学简称统计学, 是在统计工作经验积累到一定程度时自然产生的, 它是搜集、整理、描述、分析统计数据的方法和技术, 为我们的决策提供“量”方面的依据。《大不列颠百科全书》(1993 年第 15 版第 28 卷, 第 217 页)对统计学的定义是: 搜集、分析数据, 并从中做出推断的科学和艺术。显然, 这里所说的艺术, 不是传统意义上的音乐和美术等艺术, 并不是说统计学属于这样的艺术范畴, 而是指统计方法应用的技巧性、技术性, 它展现数据的过程和结论往往能让我们享受到美感和实用性。总之, 统计学阐明了关于数据的基本概念、原则和方法, 是关于数据的一门科学。

统计学来源于统计工作,是统计工作经验的理论概括和科学升华。它来源于统计实践,反过来又服务于统计实践,即通过提供统计理论和方法指导统计工作,并推动统计工作的不断进步。

由于统计工作、统计资料、统计科学三者之间具有如此紧密的联系,所以习惯上将这三者通称为统计。

第二节 现实中的随机性和规律性

一、哲学本源

自然科学中的随机性是反映在一定条件下某一事物可能发生也可能不发生的偶然关系中的。随机性涉及的是事物发生可能性的概念,只要这种约定条件的相同试验大量重复进行,该事件出现的频率必定稳定于一个具体的数值——概率。唯物主义认为,概率是事物发生可能性的客观度量,而绝不是其他什么主观的信念程度,随机性的概念就是建立在这样的基础之上的。基于随机性概念建立的科学命题称为随机性定律,显而易见,随机性定律是从偶然性这一侧面出发来反映事物之间的联系。

在哲学中,随机性和规律性的矛盾对应着偶然性和必然性的矛盾。偶然性起着为必然性开辟道路的作用,偶然性是必然性的表现和补充。具体到人类社会,历史的必然性是通过人的有意识的活动(偶然性)来开辟道路的。更进一步,从原始社会,经奴隶社会、封建社会,再到资本主义社会,最终到共产主义社会,是历史的必然趋势,无论中间有多少曲折、多少反复,都只不过是必然性开辟道路的偶然性而已。

自然科学中的随机性定律,集中反映了大量同类偶然现象所蕴含的必然规律,是客观自然规律的反映。我们既要看到偶然性与必然性的对立,又要承认偶然性与必然性的统一。如果只看到对立的一面,不承认统一的一面,那是形而上学的观点。长期以来,自然科学领域内形而上学观点的影响很大,恩格斯曾经尖锐地指出:“常识和具有常识的大多数自然科学家,都把必然性和偶然性看做永远互相排斥的两个范畴。”《自然辩证法》在持形而上学观点的人的头脑里,一个事物、一个关系、一个过程不是偶然的,就是必然的,但不能既是偶然的,又是必然的。基于这种形而上学的观点,就会导致认识论上的错误。

关于偶然性与必然性的一种错误认识,是把反映偶然性的随机性定律看做是根本的、绝对的东西,从而陷入了不可知论的泥潭。不可知论者把偶然性视为解释不了的东西,恩格斯曾经尖锐地批判这种不可知论的唯心主义本质,指出他们实质上是相信上帝。不可知论者把随机性定律当做是不可逾越的信条,竭力反对探索偶然性中蕴含的必然性,不承认更深层次上的因果律的存在。和决定论一样,不可知论也对自然科学的发展起着阻碍作用。伴随着科学的发展,不可知论者总是以不同的形式唱着反调。20世纪的量子力学纯概率解释派正是持这种观点的典型代表,尽管他们用概率解释有效地说明了当今量子力学领域的各种现象,但是纯概率解释不应该是也不可能是最终的结论。我们无法想象反映原子层次运动规律的随机性定律已经到了自然规律的尽头,不再隐含新层次的因果律了。虽然目前还没有实验结果推翻量子力学纯概率解释派的这一观点,但是量子力学在解释基本

粒子时所产生的谬误,已经预示着量子力学在新的认识层次上将被其他规律所取代,我们相信这一天终将到来。

偶然性与必然性的矛盾存在于人类认识的每一层次,这标志着人们对自然界认识的深化,新定律将在新的近似程度上更精确地反映客观世界的自然过程。当然,这种取代丝毫不意味着否定了原来的定律在原层次上的成立,而只是指出了它们的局限。我们要以辩证的观点承认并且正确对待这种交替式的更新,绝不能像决定论者那样,认为依凭着因果连锁就能通向绝对真理。

二、随机性与规律性的统一

从中学开始,我们就知道自然科学中有许多定律,如物理中的牛顿三大定律、物质不灭定律以及化学中的各种定律等。但是在许多领域,很难用确定的公式或论述来描述一些现象。例如,人的寿命是很难预先确定的,一个吸烟、喝酒、不锻炼且喜好油荤食物的人可能比一个很少得病、生活习惯良好的人活得长。因此可以说,活得长短有一定的随机性。这种随机性可能和人的经历、基因、习惯等无数不易说清的因素都有关系。但从总体来说,我国公民的预期寿命却是非常稳定的,而且由于生活水平的提高,其在逐步增长,如1996年的平均预期寿命是70.80岁,而2000年为71.40岁,2010年为71.90岁,这就是规律性。一个人可能活到这个平均年龄,也可能活不到这个平均年龄,这是随机的。但是总体来说,预期寿命的稳定性说明了随机之中又有规律性,这种规律性就是统计规律。

玩过硬币的人都知道,往天空抛出的硬币掉落在地上以后,有时候有字的一面会朝上,有时候有花纹的一面会朝上,但较少有人去研究其内在关系。我们简单总结和比较各国统计学家、数学家进行抛硬币试验所得的数据,如表1.1所示。

表 1.1 各国统计学家、数学家进行抛硬币试验所得的数据

实验者	试验次数	正面朝上的次数	正面朝上的几率
布丰	4 040	2 048	0.506 9
德·摩根	4 092	2 048	0.500 5
费勤	10 000	4 979	0.497 9
皮尔逊	24 000	12 012	0.500 5
罗曼洛夫斯基	80 640	39 699	0.492 3

可以发现,各国统计学家、数学家通过进行随机性的大量抛硬币的试验得出了较为一致的结论:硬币落地后正面和反面朝上的几率几乎是相等的,均为0.5。可见,在我们随机性的试验当中,潜在的规律是可以被发掘的,自然现象的随机性和规律性是统一的。

当然值得注意的是,硬币落地后正面和反面朝上的几率是一样的,并不等于抛两次硬币,其中必然有一次正面朝上,有一次反面朝上,也不意味着硬币落地后正面朝上和反面朝上的硬币次数一定相等,有可能在10次试验当中,10次都出现正面朝上或者反面朝上,这并不能由此否定前述结论的正确性。在统计中,讲究的是概率,也就是机会,只有进行大量的随机性、重复性试验,才有可能得出正面和反面朝上均为0.5的概率。

专栏 1.2 抛硬币与掷骰子中的随机性和偶然性

我们都做过抛硬币或者掷骰子的游戏，都知道随机地抛一次硬币或者骰子不能事先确定出现正面或反面或者某个点数，这也就是说个别游戏或试验中充满了不确定性或偶然性。机会游戏或者赌博正是利用了这种不确定性和偶然性才能够吸引人。但当我们进行大量观察，即不断做重复试验时，就会发现抛一枚均匀硬币出现正面和反面的概率会大体相同，即 $0.5:0.5$ 。试验次数越多，出现正面和反面的可能性就越接近 $1/2$ 这一稳定的数值。同时，在掷一枚均匀骰子时，出现 $1\sim 6$ 点中任何一点数的可能性也都接近 $1/6$ 。这里的 $1/2$ 和 $1/6$ 就是抛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随机性与偶然性之间的关系。

三、随机事件

日常生活中存在着大量的随机事件(也称随机性事件)。比如前面谈及的抛硬币，落地后正、反面朝上就具有随机性。又比如掷骰子，只要没有人在骰子上做手脚，随便掷出的骰子均有可能得到 $1\sim 6$ 点中的任何一点，并且假如第一次得到的是 2 点，我们完全不能确定第二次投掷的骰子在停止转动之前得到的点数会是多少，这就是典型的随机事件。

第三节 统计学的基本范畴

一、总体(个体)

总体是总量、全部的学术性代称，是统计问题当中最基本的概念之一，是统计研究和分析中最基本的要素。所谓总体，就是我们要调查或统计的某一现象的全部数据的集合。例如，大学中新生总数、池塘中所有的鱼、卡车中的全部麦子等，都可以称为总体。

总体具有三个基本标准，分别为同质性、大量性和变异性。同质性的含义是，研究的总体具有相同的属性、相同的性质，如对森林当中所有松鼠的统计，就应该只统计松鼠数量，兔子、鸟类等其他动物的数量则不用统计。大量性作为总体的另一个重要标准，顾名思义，就是要求总体数目足够大，因为统计对总体数量特征进行研究，其目的是为了探索、揭示现象中的规律，而现象中的规律只有通过大量观察才能较为准确地显示出来。因此，统计总体应该具有大量的同质性单位。变异性作为第三个标准，就是要求总体内部之间还必须存在差异性，这种差异性才是统计研究的主要内容。因为如果总体单位不存在差异性，说明总体内部不存在明显的差异，自然就没有进行统计和调查研究的必要了。

根据总体所包含的单位数量，总体可以分为有限总体和无限总体两类。有限总体是由有限量的个体构成的总体。例如，全国人口普查，尽管其包含的单位数量很大，但仍然是有限的，所以是有限总体。而当总体中的个体数目难以确定，其数量有可能是无限时，便构成了无限总体。例如，在检验某种新工艺是否真正能够改善产品的性能的问题中，由于该新工艺有可能一直延续下去，利用该工艺制造的产品包括已经生产和将要生产的产品，

其数量难以具体确定,因此属于无限总体。

样本是组成总体的各个个体。根据研究目的的不同,个体可以是人、物、机构等实物单位,也可以是一种现象或活动过程等非实物单位。

总体和个体的概念是相对而言的,随着研究目的不同、总体范围不同而变化。同一个研究对象,在一种情况下为总体,但在另一种情况下又可能变成个体。例如,要研究全国各省的人口情况时,全国为总体,各省为个体;而当要研究某省各县人口状况时,则该省变成了总体,而各县变成了个体。

二、样本

通常而言,总体都是数量比较庞大的,对于实际分析和研究缺乏可操作性,因此,统计工作人员可以考虑从总体中选取某一部分作为研究对象,而后以这一部分面积与总体面积的比例同比例地扩大,以此近似得到海量总体的具体规模,这一部分面积就是样本。

例如,当需要了解某一森林中松鼠的数量时,由于森林面积很大,工作人员无法从整个森林出发确切地统计所有松鼠的数量。这时,工作人员可以将森林平均分成若干个较小规模的区域,那么从单个较小规模区域中捕捉松鼠进行统计就轻松方便得多。然后依据某一较小规模区域中统计出的松鼠数量乘以较小规模区域个数,就可以近似得到整个森林中的松鼠数量。

在这个例子当中,整个森林的松鼠数量是一个庞大的总体,而平均分成的各个较小规模的区域就是各个样本。总体是难以简易、准确地获取数据的,通常通过设置有效样本,可以方便、快速地达到近似效果。

但要注意的是,对于样本的设置,通常要求是独立的,即对样本的划分要抛弃主观想法,进行随机划分,这样才能保证最后结果的可信性。

三、标志

标志又称标识,是指总体当中个体普遍具有的属性或特征。总体中每个个体从不同方面考虑都具有许多属性和特征,因而具有很多标志。例如,每个工人都具有性别、工种、文化程度、技术等级、年龄、工资、工龄等属性和特征,这些都是工人作为总体当中个体的标志。统计研究是从登记标志状况开始的,并通过对标志的综合反映出总体的数量特征。

标志按性质不同也可以分为品质标志和数量标志两类。品质标志表明单位属性方面的特征,品质标志的表现只能用文字、语言等来描述。例如,工人的性别是品质标志,其标志只能具体表现为男或女,而不能表现为其他形式。数量标志表明单位数量方面的特征,可以用数字来表现。例如,职工的工龄是数量标志,其标志具体表现为年数,是一个具体的数字。

标志按同一标志在各单位上的具体表现是否相同,可以分为不变标志和可变标志。不变标志就是所有总体单位同一标志在各单位上的表现是相同的,可变标志则是不相同的。例如,把某班的每位同学作为总体单位考察时,每位同学所属的班级就是一个不变标志,它是形成总体的前提,这就是总体的同质性;而在这个总体当中,每位同学的身高是不完

全相同的, 身高便是可变标志。

专栏 1.3 统计标志不同于商品标志

统计学中讲的标志不同于社会生活中常见的商品标志(如某著名品牌的图标等), 统计学中的标志也是我们每天都能看到并且大部分能识别的, 但几乎从来不会将其想当然地认为成标志。统计学里的标志又称标识, 是表征总体各单位所具有的属性或特征, 而商品标志一般是指一个企业的名字或者产品的名字, 也是一种精神文化的象征, 如三一重工企业的标志, 德国大众汽车的标志, 李宁、耐克、阿迪达斯的服装标志等。从不同的统计学角度来考察, 每个总体单位都可以有许多不同的内在的特征, 如性别、民族、年龄和文化程度等都是每个职工的特征, 都可以看做是这些职工的标志, 而商品标志表现为外在属性, 通常都是请人专门设计、需要支付报酬的。按照不同的性质, 标志可以分为品质标志和数量标志, 按照标志的变异情况其又可分为不变标志和可变标志等。某些统计标志随着时间的推移会发生根本性的变化, 如国籍的变动、学历的提升等, 而商品标志通常具有识别性、领导性、显著性、准确性和持久性等特征, 商品标志能轻易地被认识, 并且和其他商品标志显著区分开来, 如果企业长久生存, 那么该标志也会一直使用, 不会轻易更换。此外两者的差异还表现在, 统计学中的标志通常体现的是研究对象的自然属性, 而商品标志则体现出社会属性的特征, 并且多数商品标志会受到不同的法律保护等。

四、变量(指标)

变量是统计中一个常用的重要概念。可变的数量标志和所有的统计指标称为变量, 数量标志和指标的具体表现称为变量值。例如, 职工的工资作为一个标志时, 每人的工资不都一样, 因此工资这一数量标志就是一个变量, 各工资水平如 1500 元、1600 元、1700 元等则是变量值。

更多情况下, 我们使用统计指标来替代统计变量。统计指标是反映统计总体数量特征的概念和数值, 因此统计指标都是可以用数值来表示的, 是对总体单位某一特征进行调查、登记并加以汇总整理而得到的数据, 是构成总体全部单位的综合结果, 而不是说明个别总体单位的数量特征, 可以反映一定时间、地点、条件下的数量表现。通过统计指标, 我们可以对研究对象的某一方面具有确切的了解。

统计指标有定量指标和定性指标之分。定量指标是指可以搜集到具体数据的指标, 可以完成对该类指标大小高低的具体评判。例如, 由国家统计局、国家商务部和中国人民银行等部门公布的权威指标, 包括国内生产总值(GDP)、地区人口总数、进口(出口)汽车数量以及高校在校大学生数量等, 所有这些指标是不需要用户自己计算和评价的, 均可以根据实际来源较为容易地获得相关数据。

而定性指标则是较难通过公开渠道获得数据的一类指标, 需要用户自己通过对比得出该指标的具体数值。例如, 评价一个人长得漂不漂亮, 是很难通过官方渠道获得具体数据的, 需要评价者通过对这个人与其他人的对比才能评价其是否漂亮; 更进一步, 如果需要

对某个人的漂亮程度进行打分,则更需要通过和其他人的对比,并通过设置一个标准(如最低分是0分,满分是100分等)来得到评价。定性指标由于无法获得直接数据,因而在评价过程中存在很大的主观性,致使每个评价者在对定性指标进行具体打分时,也会出现不一样的结果。

统计指标和标志有明显的区别。就说明的对象来说,统计指标是说明总体数量特征的,具有综合的性质;而标志是说明总体单位的属性和特征的,一般不具有综合的性质。就表现形式而言,统计指标分为数量指标和质量指标,它们都是可以用数字来表示的;而标志根据其性质不同可以分为用文字、语言等表示的品质标志和用数字表示的数量标志。

统计指标和标志还存在密切的联系。不仅有许多统计指标的数值是由总体单位的数量标志值直接汇总而来的,如全国总人口指标是由全国各省、自治区、直辖市的人口数据汇总而来的,而且统计指标和数量标志之间存在着转换关系,如由于研究的目的不同,原来的统计总体变成了总体单位,则相应的统计指标就变成了数量标志,反过来也是这样。

第四节 统计学的应用

统计学是一门应用非常广泛的学科,几乎所有的社会科学、行为科学、生物学的大学,教育学、护理学、心理学、社会福利和社会服务学以及人类学的研究生都需要选修这门课程。在很多高校,还单独设置了统计学专业进行招生。2011年2月,国务院学位委员会第28次会议通过了新的《学位授予和人才培养学科目录(2011)》,统计学上升为一级学科,设在理学门类中(编号为0714)。这样,统计学就完全从数学和经济学中独立出来,成为了名副其实的一级学科。由于应用甚广,统计学也成为目前就业前景最好的专业之一。

作为应用性很强的一门学科,统计学的生命力和发展动力在于它与其他学科的密切联系,割断了这种联系,统计学就会成为无源之水、无本之木,产生不出任何有意义的问题和方法。因此,统计学与其他学科和领域所形成的边缘和交叉性质的学科也特别多,如工业统计学、农业统计学、物理统计学、生物统计学、医药统计学、人口统计学和空间统计学等。不仅如此,统计学研究中处理数据的方法,无论是在自然科学领域、社会科学领域,还是在工程技术领域当中,都发挥着重要的作用。如果没有统计,大量杂乱无章的数据将无法显示其内含的有用信息,金子将被掩埋在沙砾之中。

一、经济管理

统计学最初产生于对经济现象的研究,至今经济领域仍然是统计学最重要的研究领域,统计学在经济领域的应用形成了经济统计学,无论是宏观经济学研究还是微观经济学分析,都需要用到大量的统计学方法,通过各种调查搜集实际的经济统计数据,并分析其数量规律性。统计学已经成为经济学研究不可或缺的一种基本手段。在国内,几乎所有的经济管理学科都要用到统计知识,都要开设统计学这门课程,如国民经济学、数量经济学、政治经济学、会计学、财政学、国际贸易、工商管理、管理科学与工程等。

众所周知,量变引起质变。数量关系的背后,牵扯着市场的稳定和发展。金融业的现

代化推动了统计与数理方法的应用研究,反过来,当今世界的金融管理特别是金融风险防范,也越来越注重量化研究。早在1995年,美国斯坦福大学经济学家刘遵义教授就通过实证比较数量分析和模糊评价等方法,预测出菲律宾、韩国、泰国、印度尼西亚和马来西亚有可能发生金融危机,后来的事实证明了这一点。这从一个侧面提醒我们,没有完整、科学的分析预测工具,就有可能在国际金融竞争中蒙受重大损失,只有加强对作为金融信息的各种变量的研究,才能对金融运行规律形成良好的认识,才能把握市场的发展动向。

投资顾问使用各种统计数据信息进行投资分析并用来指导他们的投资建议。对于股票投资来说,顾问们审查各种财务数据,包括市盈率、股息率等。通过把个股信息与股票市场的平均数加以比较,投资顾问就可以得出个股价格是高估还是低估了的结论。例如,2005年3月16日上证综合指数中30只股票的市盈率是21.08,东方电子集团有限公司的市盈率是17.92。这时,市盈率方面的统计信息显示,与上证综合指数股票的平均收入相比,东方电子集团有限公司的股票价格较低。因此,投资顾问可以得出结论:东方电子集团有限公司的现行价格低估了。这一点以及其他有关东方电子集团有限公司的信息有助于投资顾问提出买入、卖出或持股的建议。

二、医疗卫生

统计学在医疗卫生事业当中已取得了广泛的应用,在国内,很多高等院校都要求强化对学生统计能力的培养,并成立了专门的“卫生统计学”这一学科。卫生统计学有助于提高医疗卫生管理决策的科学性,增强计划的可行性。其主要体现在以下两个方面。

(1)提供准确、及时、有价值的原始资料。其包括:

第一,完善卫生指标体系,选择适当的指标,确定其含义、计算方法与资料来源。

第二,改革报表制度,结合各种调查方法,力争以经济高效的方法取得真实可靠的数据。

第三,利用高科技手段收集与储存资料,提高统计信息的实效性。

(2)用统计方法对“死的”数据进行分析研究并加以开发利用,使其成为“活的”统计信息,及时提供给领导部门。其包括:

第一,用各种描述统计指标(如平均数、相对数等)及统计图表来反映资料分布、服务情况的质量、人群健康状况等的现状与变化,以便找出工作重点。例如,分析地区卫生资源的分布情况,提出基层卫生结构技术条件差、利用不足,而大医院门诊住院压力大等问题。

第二,利用动态分析方法分析一段时间内某地区卫生资源、服务情况、取得效益的变化趋势,找出当前主要矛盾加以改进。例如,某医院对10年来门诊人次进行剖析,提出门诊量下滑的影响因素。

第三,利用管理图(控制图)法分析医疗卫生工作进行过程的稳定性,及时发现缓慢变异和异常现象,探索其原因,以便及时纠正,也可以反映医疗卫生设备、检验结构的实际精度。

第四,统计估算与预测。利用现有资料或数据对将来的趋势作估量,或从局部资料来推断总体,为做出计划与决策提供依据。例如,卫生计划中估算工作量、人财物投入量、各种效益指标(治愈率、病床使用率等)的变化水平等。估算是否能接近实际情况,取决于

三点：一是现有数据是否真实；二是是否充分考虑与了解对所估算指标有重大影响的内外因素；三是选择的估算方法是否恰当。估算与预测的方法很多，较为简单的是利用已有的按时间排列的数据，用移动平均法(或指数平滑法、季节变动分析法等)修匀原有数列，找出变化趋势，如预测门诊工作量等，也可选择某一种回归(regress)模型或用灰色系统模型预测。评价预测方法是否合理的基准是预测结果的精确度，不宜用模型的复杂程度作为评价预测方法好坏的标准。

第五，评价工作效益。不能用单一指标比较效益时，可用综合指数分析，先对每一指标按确定的标准值计算出指数，再将各指标的指数加权综合，通过比较综合指数的变化来判断效益，如利用工作效率、经济效益、医疗效果三方面的指标比较历年的医院经营管理效益。

三、教育测评

统计学在教育体系中也取得了广泛的应用，最典型的应用实例是对全国或全球高校的综合评估与排名。每年教育部都会公布中国大学排名情况，除了清华大学、北京大学等一流院校始终领跑在最前面外，其他高校排名都有不同程度的上升或下降。那么教育部公布的这些大学排名是如何确定的呢？为什么清华大学、北京大学等知名高校能基本保持稳定，而其他高校却有不同水平的波动呢？对于很多人而言，这是不太好理解的。但对于学习或接触过统计知识的人来说就会知道其中缘由。其实，这样的大学综合排名主要就是用到了统计当中的指数体系、赋权方法与综合评价等方法。

除了大学实力的综合评估，教育部也会对很多高校进行教学评估，教育工作者的升迁与否也是结合各方面的成绩，通过统计方法进行综合研究与打分后决定的。这样的统计评分方法在国内教育体系中正在广泛开展。

四、主要人文社科领域

统计学在法律中的应用非常广泛，但我国与发达市场经济国家的统计实践差别非常大，统计在法律业务中几乎是空白，因此统计在法律事务中具有广阔的应用空间。从美国看，当案件所面临的问题变得越来越复杂时，统计就愈显重要了。统计学家通过分析数据，帮助法官或陪审团决定某人是否有罪，或是否必须对所造成的伤害进行赔偿。律师需要聘请统计学家或具有统计专业知识的人进行咨询，或对大量事实进行验证。这样的统计学家不但要懂统计，还要能够向不熟悉统计的法官、陪审团和律师有效地解释自己的工作。法庭需要运用各种统计方法处理其业务，所以法庭统计是一个极具挑战性的应用统计的重要分支。

专栏 1.4 统计学在性别比例研究中的应用

一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，新生婴儿的性别比例似乎没什么规律可循。但如果对新生婴儿的性别进行大量观察，即观察成千上万个或者更多时，就会发现性别比例还是有规律的，即婴儿总数中男孩要多于女孩，大致为每生育 100 个女孩，就会生育 105 个左右的男孩。这个 105 : 100 的比例就是新生婴儿

男女性别的数量规律性，古今中外都大致相同，它是人类社会长期遗传与发展的结果。因为人类社会要延续、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，新生儿男多于女时会出现不平衡，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。进入中老年以后，男性的死亡率仍然高于女性，导致男性的平均寿命比女性短，长寿的男性要少于女性。从一个国家来看，如果没有B超检查后堕胎等干扰，其规律是婴幼儿时男性略多于女性，中青年时男女人数大致相同，老年时女性又略多于男性。这样既保证人类在中青年结婚生育时性别的大致平衡，又使得在人口总数上男女也大体相当，有利于人类社会的进化和发展。对人类性别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之一。

统计学在企业管理中也有广泛应用。企业管理是对企业生产经营活动进行策划、组织、营销、激励、调节、控制，以获得最大的经济效益。企业生产经营活动离不开统计，要运用统计方法开展市场调查、市场预测和 market 分析，调整企业经营策略，对产品质量进行监测与控制，对企业的经济效益进行评价，研究市场、认识市场、适应市场、驾驭市场，在激烈的市场竞争中提升竞争能力。因此，统计学是企业经营管理者的必备知识。

特别要指出的是，很多管理活动都要用到统计工具。将这个有用的工具与自身专业领域相结合，能提高管理的效率和拓展活动的空间。例如，零售付账柜台使用电子扫描仪收集数据，可供各种营销研究应用。市场调查公司或营销专家需要的各种数据，都可以从零售商店购买销售点的扫描纪录中获得，经过处理后，这些数据的统计汇总信息再卖给制造商。产品公司经理们可以检查并评论扫描得到的数据以及促销活动的相关数字，从而更好地理解销售与促销活动之间的关系。

此外，在发达国家，统计在社会学、新闻学、文献计量学、心理学、教育学、伦理学、宗教学、政治学、环境学等社会科学领域也有广泛和深入的应用，而在我国应用得相对较少。我们要向发达经济国家学习，开发新的统计应用领域，为我国经济、社会发展贡献力量。

► 习题

1. 什么是统计？统计的作用是什么？
2. 什么是统计总体和统计样本？总体与样本的概念是不是一成不变的？以一个实例说明总体与样本概念的相对性。
3. 什么是统计指标？统计指标和标志有什么联系和区别？
4. 举出几个在日常生活中可能存在的规律性和随机性统一的实例。
5. 小王家种了200亩（1亩=0.067公顷）水稻，今年风调雨顺，收成很好。小王想估算一下今年水稻的总产量，从中随机选取了10亩水稻进行产量计算。试指出其中什么是总体、样本和变量，并说明统计推断的大致过程。