

第 1 章

导 论



- 1.1 统计的初步认知
- 1.2 统计的概念界定
- 1.3 统计学的分科
- 1.4 统计学的基本术语
- 1.5 统计在商务和经济领域中的应用

本章学习目标

1. 了解和明确统计情境的界定。
2. 理解和掌握统计的内涵。
3. 掌握和理解统计学的几个基本术语。

趣味引例

希腊神话典故与统计学

故事 1：西壬(Syrens)。西壬是古希腊传说中专以美妙的歌声诱惑水手们的女妖。她们的上身是美丽的少女，胸部以下则完全是鸟的形状，长着双翅、鸟尾和鸟脚。这几个女妖遇有船只经过，便用美妙的歌声来诱惑水手们。受了诱惑的水手或被沉入海底或留下来被她们害死，因此在她们居住的海岛上到处都有死人的白骨。希腊英雄奥德赛在大海上漂流时，女巫瑟西事先警告了他，因此当他的船快靠近西壬的海岛时，奥德赛就用蜡把同伴们的耳朵都封起来，并叫同伴把他紧紧地绑在船桅上。当他听到女妖们美妙的歌声时，由于受不了歌声的诱惑，几次要求他的同伴把他从船桅上放下来，但同伴们反而把他绑得更紧，并加快划桨前进，就这样驶过了海妖的岛屿。传说这几个女妖以为她们的歌声失去了作用，就跳入大海化作了岩石。

故事 2：阿喀琉斯(Achilles)。阿喀琉斯是古希腊国王珀琉斯和爱琴海的海神涅柔斯之女西蒂斯所生的。他 9 岁时，母亲不愿意他在战争中牺牲，便把他装扮成一个少女，藏到斯库洛斯岛国王吕科墨得斯的宫廷里。希腊人远征特洛伊时，奥德赛打扮成商贩，带了一些首饰和武器到岛上来找他。阿喀琉斯没有挑选首饰，却拿起了武器来挥舞，因此暴露了真相，只好随着奥德赛前往特洛伊参加战争。

引例思考

从这两个希腊神话典故可以得出什么共同点？与统计学存在什么关联？

统计是一种具有悠久历史的社会实践活动。可以说，自从有了国家，就有了统计工作。最初的统计活动是为统治者管理国家的需要而进行的资料搜集工作，涉及计算国家的人口、军队、领土、居民职业以及资源财产等重要事项。到了今天，统计已经发展成为各行各业开展活动时必不可少的一项基础工作。第 64 届联合国大会第 90 次会议于 2010 年 6 月 3 日通过决议，确定 2010 年 10 月 20 日为第一个“世界统计日”。

1.1 统计的初步认知

在我们大多数人的初步印象里，统计是大量数字的复杂收集与计算分析过程，可以说是枯燥、无趣的事情。无论是在中国，还是在外国，与其他行当相比，统计成为人们调侃对象的机会似乎更多一些。大仲马的作品多曲折感人，而大仲马又多私生子，所以取笑讥讽他的人，往往把他的作品比作他的私生子。最使他头痛的是巴黎统计学会的秘书长李昂纳，这人是他的朋友，每次举统计数字的例子时，总是说大仲马的情妇和私生子有多少。有一年该统计学会开年会，大仲马估计，李昂纳又要大放厥词，说他的坏话了。于是他请求参加年会，获得了批准。果然不出大仲马所料，李昂纳又举他的情妇和私生子的例子。李昂纳报告完毕，请大仲马致辞。一向不愿在大庭广众之下发表演讲的大仲马，这次却破例登台说：“所有统计数字都是撒谎的，包括有关本人的数字在内。”听众哄堂大笑。

在国内，一说到统计数字我们总要保持一定的保留态度。在国外，曾经有这么个说法，“谎话有三种：一般的谎话、糟糕透顶的谎话和统计”，居然和糟糕透顶的东西并列，可见统计在某些人眼中的地位。当然在国外，当你说起是统计学者时，别人多半会特别地看你一眼，那是在看你的脑袋瓜，觉得还够用，不过其“眼外之意”或许还有一层，那就是统计很难很枯燥，你居然能守得住，可见是个阿呆。所以，国外的统计还需要去讨好人，比如有人写了一本书，起名字叫《没有眼泪的统计学》，以吸引别人来读。

人是感性的，对奇异事物的印象往往深刻，受到强烈刺激当然反应也强烈。反过来，对平淡无奇的事物，人们往往兴趣索然，也不容易留下什么印象。而统计注重平均，正是要剔除异端影响，以得出对一般趋势的认识。许多人在电视里看过歌手大奖赛，对给最高分和最低分的评委，观众会有很深的印象，有时过了好多年还会说起。但现场公布成绩时，人们也总能听到这样的话：“去掉一个最高分……，去掉一个最低分……”这里，统计硬把人家的偏好给砍没了，人家偏好感性，统计却硬要求人家理性。

美国统计学家 David Freedman 认为统计是对令人困惑的问题做出数字设想的艺术。把统计学称为艺术显然有些夸张，但这一说法也提示了统计可以创作性地提出和解决问题，而不拘泥于条条框框。我们来看一个实际生活中发生的例子。在一个水库养了许多鱼，管理人员希望了解鱼的大致数量，这是一个实践中的统计学问题。由于鱼是不听指挥，会在各处自由游动的，所以解决方案可以是先从水库的不同位置一共捕上来 1000 条鱼，在每条鱼的尾部做上一个标记，并且保证标记不会影响鱼的自由游动，然后，将鱼全部放回水库；几天后，从水库中再捕上来 2000 条鱼，检查其中尾巴上有标记的鱼的数量。假定在第二次

捕上来的 2000 条鱼中,有 20 条鱼尾巴上做了标记,则可以推断,水库中鱼的总数大致为 $1000/(20/2000) = 10$ 万条。

上述这个案例在实践中是经常见到的,但对于一个统计人员来说,做出一个 10 万条鱼的估计是不够的,还应该对这一估计的精度做出判断。这种收集统计数据并进行分析的方法,更多地可以说是一种艺术。著名统计学家弗朗西斯·高尔登说:“一些人厌烦统计数字,甚至听到这个字眼就皱眉头,而我却发现它们妙趣横生。当人们不是将这些数字胡乱堆放,而是用精明手段去处置它们,小心翼翼地做出解释时,它们就显示出应付复杂现象的非凡能力。”



趣味小贴士

啤酒与尿布

全球最大的零售商沃尔玛通过分析顾客购物的数据后发现,很多周末购买尿布的顾客同时也购买啤酒。经过深入观察和研究发现,美国家庭买尿布的多是爸爸。年轻的父亲们下班后要到超市买尿布,同时“顺手牵羊”带走啤酒,好在周末看棒球赛的同时过把酒瘾。后来,沃尔玛就把尿布和啤酒摆放得很近,从而双双促进了尿布与啤酒的销量。“啤酒与尿布”的故事,很容易理解统计数据的作用和意义。统计是用数字来说话的,它通过数据来解释事物的本质特征及其规律性。“国立”台湾大学生物统计学博士、世界著名统计学家谢邦昌教授曾将统计比喻为开罐头的器具,而未来的趋势就像是罐头里的食物,资料就是罐头的铁皮。

1.2 统计的概念界定

统计是一种社会调查活动,不论是宏观社会的整体调查研究,还是微观事物的观察活动,都需要统计。根据理解的角度不同,在日常生活中“统计”一词可以有以下三种含义。

含义 1: 统计工作,是指收集、整理和分析数字资料的工作活动。例如,开学时,辅导员要统计到校的学生人数;篮球比赛中教练员要统计每个队员的投篮命中率和犯规的次数;农民在农作物收获后统计其产量等。

含义 2: 统计资料或统计数据,是统计工作的结果,表现为一系列的数字资料。例如,一个班的学生数为 60 人,女生占 30%,男、女生的比例为 2.33:1 等。国家统计局每年出版统计年鉴,反映国家的经济、文化教育以及科技发展等情况,这些都是这个意义上的统计。

含义 3: 统计学。关于什么是统计学有多种解释,在第七次国际统计会议上讨论关于统计学的含义时,据说竟有 180 余种之多。《不列颠百科全书》中指出“统计学是收集、分析、表述和解释数据的科学和艺术”。也有学者提出“统计学是在面对不确定情况时,向人们提供做出聪明决策的科学方法,这些方法包括资料的收集、整理、陈列、解释和分析,运用统计方法可进行推论,从而获得合理的判断与有效的结论”。本书认为,统计学是一门搜集、整理、分析统计数据的方法的科学,其目的是探索数据的内在规律性,认识客观事物的

现实状态,认识随机现象的统计规律,减少由于人们对系统的“无知”而导致的对它在认识上的不确定性,从而实现对客观事物的科学认识。

统计学已经成为很多学科的通用语言。作为其潜在的使用者,既要知晓正确使用统计方法所需的“科学”,又要掌握其中的“艺术”。对统计方法的小心使用,仔细界定情境、收集数据、精确地描述数据、推断并提出有意义的结论,才能够获取数据的真实信息,做出正确的应对决策。希腊神话典故中,西壬引吭高歌也好,阿喀琉斯装扮少女也罢,其本来面目虽一时难辨,但并非无法推测。只要使用适当的方法,总会从待查事物对某种方法的反应中,推测出其本来面目,而作为方法的统计学的作用就在于此。

“统计”一词虽然有不同的表述,但它们之间又具有密切联系。统计资料是统计工作的成果,统计学是统计实践活动的经验总结和理论概括,统计工作是在统计理论的指导下进行和完成的。



深入辨析:统计学的内涵

由统计学的概念可见其核心是处理信息,并用数字和视觉图形加以描述和分析解释。当然,不同的人对统计学有不一样的理解。有人认为,统计学是人们企图用不正确的信息和结论去说服他人的“把戏”;另一些人觉得它是收集和展示信息的手段;还有人把它视为“在面对不确定性时进行决策”的方法。确切地说,上述观点都是对的。



趣味小贴士

数据的误导

对于统计数据的分析也要谨慎。运用统计学工具要以相关学科的知识为基础,以定性分析为前提,当你听到一种统计关系时,切勿轻率地对其因果关系做出结论,否则将会误入歧途。例如,统计资料表明,大多数汽车事故出在中等速度的行驶中,极少事故出现在大于150km/h的行驶速度,但这并不说明高速行驶比较安全。由于大多数人是以前中等速度开车,因此数据的统计关系并不能够完全代表事件的因果关系。

1.3 统计学的分科

统计学的内容十分丰富,研究与应用的领域非常广泛。从统计教育的角度,统计学大致有以下两种分类。

1.3.1 描述统计和推断统计

描述统计(Descriptive Statistics)是用图形、表格和概括性的数字对数据进行描述的统计方法。图1-1的框图清楚地表明描述统计是对数据进行处理的第一阶段,即用直观的图形、汇总的表格和概括性的数字(如平均数)表示数据的分布、形状等特征,并为进一步的统计推断提供根据。

推断统计(Inferential Statistics)是根据样本信息对总体进行估计、假设检验、预测或其他推断的统计方法。例如,根据 100 张样本发票的统计结果,审计人员可以对所有 55 400 张发票中有错误的发票数做出估计,然后公司可据此决定是否调整其营业额。又如,美国盖洛普(Gallup)调查公司在美国总统大选前通常会从全美国的选民中随机抽取 1500 人左右,对大选结果进行调查和预测,并会给出不超过 3% 的预测误差。以上这两个例子都是利用样本信息和概率论原理进行统计推断的过程。

统计学分为描述统计和推断统计,一方面反映了统计发展的前后两个阶段,另一方面也反映了统计方法研究和探索客观事物内在数量规律性的先后两个过程。图 1-1 即为统计方法探索客观现象数量规律性过程的框图。

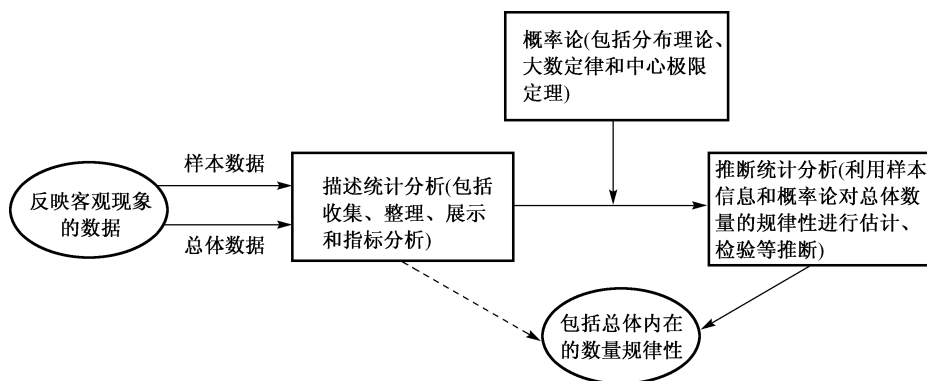


图 1-1 统计方法探索客观现象数量规律性过程的框图

由图 1-1 可以得知,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,就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,科学、及时、快速、经济地收集与整理数据,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。一个出色的统计工作者的能力和技巧在推断统计中将得到充分的体现和检验。但如果没有描述统计收集可靠的数据并提供有效的样本信息,即使很高明的统计学家利用很科学的推断方法也难以得出准确的结论。因而,推断统计对描述统计又有很强的依赖性。

尽管描述统计可以在获得总体数据时直接探索出总体数量的规律性,但这种情况在实际工作中却很少见到。自然现象的总体多数是无限的。例如,统计物理研究中要弄清楚分子运动的规律,而分子又是无穷多的,不可能全部观察和试验。社会经济现象的总体虽然多数是有限的,但要考虑获得数据以及推断总体的时效性、经济性和准确性,抽样调查往往比普查更有效,因而应用也更普遍。例如,全国的人口数量和变化、耕地面积、企业数量

和经营状况等,虽然可以通过普查得到全部数据,但普查要投入大量的人力、财力和物力,而且要很长时间才能收集、整理出所要的数据,不是每年都能做到的,因而我国确定每10年进行一次人口普查,每5年进行一次经济普查(分别在逢3、逢8的年份进行,由于“非典”影响,原定于2003实施的第一次全国经济普查延至2004年),其他各年均以抽样调查数据进行推断。此外,大量的管理和研究工作不可能组织普查。例如,城市居民每月的家庭收入支出调查、某种商品的市场调查、某个事件的民意调查等都只能通过抽样调查方式,对总体数量规律性进行科学的推断。因而,在描述统计中收集、整理和分析的多是样本数据。这样,科学地整理样本数据、显示样本数据的特征和规律、提取样本数据中的有用信息就显得格外重要了。

1.3.2 理论统计和应用统计

理论统计是指统计学的数学原理。由于现代统计科学几乎用到了所有的数学知识,要成为优秀的统计工作者就必须经过严格的统计训练,特别是从事统计理论和方法研究的人员就必须有很好的数学基础。从广义来讲,统计学是应该包括概率论的,因为概率论是统计推断的数学基础,而概率论是数学的一个分支,则理论统计是应该包括概率论在内的对统计方法数学原理的研究。

在统计工作者中,从事理论统计研究的人只是很少一部分,大部分是应用统计方法去解决实际问题的应用统计工作者。统计学是一门数据学科,由于在自然科学、社会科学的所有研究和实际工作中都要通过数据来分析和解决问题,统计方法的应用就自然而然地扩展到几乎所有的研究领域。例如,统计方法在物理研究中的应用形成了统计物理,统计方法在生物学中的应用形成了生物统计,统计方法在医学中的应用形成了医疗卫生统计,统计方法在风险管理与保险中的应用形成了保险精算学,统计方法在微观企业管理中的应用形成了管理统计,等等。以上这些应用统计学的不同分支所应用的基本方法都是一样的,即都是描述统计和推断统计的主要方法。但由于各应用领域都有其特殊性,统计方法在应用中就具有了不同的特点。例如,在经济应用中要测量和探索物价变动的数量规律性,就在比率和平均数的基础上形成了物价指数法。又如,正态分布在教育学中得到了广泛应用,在教育测量和分数转化问题研究中得到了发展。

作为一名优秀的应用统计工作者,不但要熟练地掌握和应用各种统计方法,而且必须具备所研究和应用领域的专业知识。因而,专业知识和统计方法是做好统计应用的两个基本功。例如,要做好统计工作,其基本条件是要具备理论和统计方法的知识和能力,当然还需要有计算机操作能力、写作能力和其他能力。这就要求统计应用人才是一名复合型人才,这也是人文、社科、理、工、农、医所有院校都开设应用统计方法这门课程的缘由。

1.4 统计学的基本术语

统计基本术语的理解对于理论的学习与应用至关重要。在不同的统计领域进行分析研究,所侧重应用的统计术语是有所不同的。在描述统计领域,常用的基本术语包括统计总

体和总体单位、标志和变量、统计指标和指标体系；在推断统计领域，常用的基本术语包括总体和样本、参数和统计量。

1.4.1 统计总体和总体单位

统计总体(简称总体)，它是指客观存在的、在同一性质基础上结合起来的许多事物的整体。总体单位(简称单位)，就是组成总体的每一个个体，是调查项目的具体承担者。例如，要调查某市居民人均用水量，该市所有居民户的集合为总体，每一户居民为总体单位。

统计总体的形成必须具备一定的条件，作为统计研究具体对象的统计总体，其形成条件主要有以下三条，这也是统计总体的特点。

第一，同质性。组成统计总体的所有单位必须在某些性质上是相同的。例如，工业企业总体，必须是由进行工业生产经营的基层单位组成的。如果是国有工业企业总体，便又多了一个所有制性质上的相同标志，它的范围便小于工业企业总体了。

第二，大量性。统计总体是由许多总体单位构成的。小型总体(抽样总体)的单位数也要足够多。

第三，差异性，又称变异性。构成总体的各单位除了有同质性的一面外还必须有差异性的一面，否则便不需要进行统计调查研究了。例如，某校大学生总体中的每个学生，在性别、年龄、专业、兴趣爱好等方面都有差异。

一个统计总体所包含的总体单位数可以是无限的，此时称为无限总体；如果是有限的，则称为有限总体。例如，太平洋里所有的鱼构成的总体就是无限总体，因为这里鱼的数目是数不清的；又如，某所大学有 23 000 名学生，那么该校的学生总体就是一个有限总体，因为其总体单位数是有限的。在社会经济现象中统计总体大多是有限的。在统计调查中，对无限总体不能进行全面调查，只能调查其中一小部分单位，据以推断总体。对有限总体既可做全面调查，也可只调查其中的一小部分，如职工普查(全面)及职工抽查(小部分)。凡是调查总体的一小部分单位时，往往要根据局部资料来推算总体。这就涉及推断统计学的研究内容了。



深入辨析：总体和单位

(1)总体和总体单位不是固定不变的，随着研究目的不同，它们是可以变换的。例如，当研究某班级学生的学习状况时，则每个学生是这个总体的一个单位；如果把研究范围扩大到以班级为单位的全校学生，则班级为总体单位。

(2)描述统计学中，总体是研究对象的全体，其本身并不承载数据；具体数据的载体是单位。例如，当研究某班级学生的统计学成绩时，则该班所有的学生是研究的总体，每一位学生统计学成绩是总体上的一个单位。

1.4.2 标志和变量

标志是说明总体单位属性和特征的名称，其具体表现结果称为标志值。标志值(也就是标志表现)可能是数值，也可能是文字。当标志值是数值时，我们称这样的标志为数量标志；

当标志值是文字时,我们称这样的标志为品质标志(属性标志)。例如,某职工张三的性别为女、民族是汉族等。这里的“性别”和“民族”是品质标志,而“女”和“汉族”是这类品质标志的属性的具体表现,即其标志值。又如,该职工的年龄是35岁,工资是3096元,则“年龄”和“工资”是数量标志,而“35岁”和“3096元”则是它们的标志值。再如,在某市房地产企业组成的总体中,每个房地产企业为总体单位,房地产企业的各种特征为一个个标志。如公司的“已开发面积”、“房屋平均售价”为数量标志,而公司的“所有制性质”、“公司总部所在地”则称为品质标志。数量标志和品质标志这两类标志在统计研究中都很重要,是形成统计指标的基础。

此外,标志按其每个个体上的表现结果是否相同可以分为不变标志和可变标志两种。不变标志是指每个个体上的具体表现完全相同的标志。例如,高校学生总体中,其身份“大学生”是不变标志。很显然,不变标志是相关个体集成总体的基础,总体的同质性也正体现在要求具有至少一个不变标志;可变标志则是指每个个体上的具体表现不相同的标志。例如,高校学生总体中,出生地、年龄、身高、专业等都是可变标志,因为它在不同学生身上的体现是不相同的。这种个体标志表现不相同的现象,在统计学上称为变异。总体的差异性也正体现在至少要有一个可变标志。

可变标志的抽象化就是统计学中经常涉及的一个概念——变量。变动的标志值就是变量的具体取值,称为变量值。例如,我们可以把年龄、民族称为变量,而年龄和民族的具体表现则是变量值。

变量按表现形式不同可分为数量型变量和属性变量。例如,住房面积、收入为数量型变量,而文化程度、性别为属性变量。

数量型变量又可以分为离散型变量和连续型变量。如果变量的取值可以一一列举,在相邻的两个数值之间不再有其他数值,这样的变量称为离散型变量,如在校学生人数、全国上市公司企业数。如果变量的取值不能一一列举,在任意两个数值之间都可以再取无限多的数值,这样的变量称为连续型变量,如人的年龄、股票的价格。

1.4.3 统计指标和指标体系

统计指标(简称指标)是反映统计总体综合数量特征的概念和数值。例如,通过对某市每户用水量、家庭人口数的登记,汇总计算得出全市的总用水量及人均用水量。我们将全市的总用水量、人均用水量称为指标,因为它们反映了总体的特征,而每户的用水量则称为标志,因为它是总体单位的特征。

统计指标按所反映的总体特征的不同性质,分为数量指标和质量指标。数量指标能够直接或间接地说明总体的规模大小,数量多少。例如,某省总人口、粮食总产量,前者说明了某省人口总体的总规模;后者说明了该省粮食生产成果总体的总规模。质量指标是说明总体内部数量关系或总体单位某种数量标志的标志值水平的指标。例如,产品合格率、市场占有率,这两个指标反映了总体内部的数量关系。产品合格率是就全部产品这个总体而言,合格品数量与总数量相比所占的份额;市场占有率则是某一市场上成交的所有的同类产品这一统计总体中某一品牌的产品销售量占同类产品全部销售量的百分比。再比如某班学生平均成绩,就这个例子而言,统计总体是这个班级的全体学生,总体单位是每一名学

生的期末数学成绩。每名学生有这样或那样的特征,期末数学成绩也是其中的一项特征(标志),具体的分数为标志值。我们用全班的总分数(总体各单位的标志值之和)除以全班总人数,自然就得到了平均成绩这个指标。我们把平均成绩称为质量指标,因为它代表了各单位标志值的一般水平。

指标体系是由一系列相互联系的统计指标所构成的整体。一个统计指标构不成体系,彼此相互独立的统计指标放在一起也构不成指标体系。统计指标体系是一个整体,这个整体由许多统计指标构成,其中的各个统计指标之间是相互联系的。单个的统计指标说明了事物的一个侧面,把它们结合在一起说明事物的全貌。如就工业企业的固定资产指标,它只能反映企业固定资产的拥有数量,但如果把企业的固定资产、流动资金的占用周转以及利润税金等指标放在一起,则能反映企业的财务状况,这几项指标之间存在着密不可分的联系。



深入辨析:指标和标志

正确理解统计指标,还要注意不要把它同统计标志混为一谈。

(1) 指标与标志的区别

其一,指标是对总体而言,说明总体数量特征,具有综合性质;而标志是对总体单位而言,说明单位的属性或特征的,一般不具有综合性质。例如,在研究企业职工收入状况时,职工个人的工资是标志,企业全体职工的工资总额或平均工资就是指标。区别指标与标志,看它是否经过了数量上的综合。

其二,指标一定是可量的;而标志有数量标志和品质标志之分,不一定可量,其中品质标志不是直接可量的。

(2) 指标与标志的联系

其一,指标是由标志值汇总而来的。例如,工业企业职工总数是由各工业企业职工人数汇总得到的。品质标志虽然本身不具有数值,但有些指标是按照品质标志分组计算得出来的。例如,人数总数中,男生、女生的人数各有多少;工业企业总数中,国有、集体、私营等性质的企业各有多少等。

其二,由于承担指标的总体和承担标志的单位,在不同研究目的下可以转换,因此,指标和标志之间也有变换关系。例如,调查了解一个工业局的职工状况,则整个工业局为总体,局内的各企业、各单位为统计单位,企业的职工人数为标志,全局的职工人数为指标;而当研究企业内部职工状况时,企业内部各部门的人数为标志,而企业职工人数则成了指标。



趣味小贴士

企业经济效益评价指标体系

按照建立现代企业制度的要求,为了综合评价和反映企业经济效益的状况,财政部在反复研究的基础上,制定了一套企业经济效益评价指标体系。这套体系包括销售利润率、总资产报酬率、资本收益率、资本保值增值率、资产负债率、流动比率(或速动比率)、应收账款周转率、存货周转率、社会贡献率、社会积累率10项指标。

1.4.4 总体和样本

所谓统计推断,就是根据概率论所揭示的随机变量的一般规律性,利用抽样调查所获得的样本数据信息,对总体的某些性质和数量特征进行数理推断。这种推断必定是在量化数据的基础上进行的,其研究的对象应该是数据集合。因此,推断统计总体与描述统计总体的口径界定有所差别。当前我们要测定总体某个指标(或变量)时,该总体所有单位的标志值(或变量值)所组成的整体可视为总体。可见,推断统计总体是描述统计总体所有单位标志值(或变量值)的集合体,是全部数据的承载体。从随机抽样的角度来看,总体中部分数据的集合体称为样本。例如,从一批灯泡中随机抽取 100 只分析灯泡的使用寿命,则总体是这批灯泡中所有灯泡的使用寿命期的集合,这 100 只灯泡使用寿命期的集合就构成了总体的一个样本,然后根据 100 只灯泡的平均使用寿命去推断这批灯泡的平均使用寿命。



深入辨析:描述总体和推断总体

虽然在统计学中,在描述统计和推断统计领域,研究对象的全体都称为总体。但具体的口径界定应区别开来。描述总体仅仅是研究对象的整体,不承载数据;而推断总体是研究对象全部数据的整体,其本身就是全部标志值(或变量值)的数据集合体。

1.4.5 参数和统计量

根据总体数据计算的指标称为参数,以反映总体的某种特征。由于总体数据往往不知道,因此参数是一个未知常数,本书中用希腊字母表示。例如,总体平均数用 μ 表示,总体标准差用 σ 表示,总体比例用 π 表示,等等。

根据样本数据计算的指标称为统计量,以反映样本的某种特征。由于样本的抽取是随机的,因此统计量是样本的函数,本书中用大写英文字母表示。例如,样本平均数用 $\bar{X}$ 表示,样本标准差用 S 表示,样本比例用 P 表示,等等。

由于样本是抽样得出的,因此统计量总是可以计算得出的。抽样的目的就是要根据样本统计量去估计总体参数。例如,用样本平均数 $\bar{X}$ 估计总体平均数 μ ,用样本标准差 S 去估计总体标准差 σ ,用样本比例 P 去估计总体比例 π ,等等。除了样本均值、样本比例、样本方差这类统计量,还有一些是为统计分析的需要而构造出来的统计量,例如,用于统计检验的 Z 统计量、 t 统计量、 χ^2 统计量、 F 统计量等,它们的含义将在后面相关的章节中再做介绍。

1.5 统计在商务和经济领域中的应用

在今天全球化的经济和商务环境背景下,每个人都会接触到大量的统计信息。统计方法被普遍地应用于商业、经济的各个职能领域,如会计、金融、营销、生产、经济。

1.5.1 会计

会计师事务所在对客户进行审计时常常采用统计抽样的方法。例如，假如一家会计师事务所想确定客户的资产负债表中所显示的应收账款的数目，是否真实地反映了实际应收账款的数目。通常，应收账款的账目数量会很大，逐个审查和验证将会耗费大量的时间，成本巨大。在这种情况下的通常做法是：审计人员会从全部账目中抽取一部分，称为样本。通过审查所抽取的账目的准确性，审计人员就可以做出客户的资产负债表中所显示的应收账款数目是否准确的判断。

1.5.2 金融

财务分析师们使用各种统计信息对投资进行指导，在各种投资方案中做出选择或追踪趋势。对于股票投资来说，财务分析师们审查各种财务数据，包括市盈率、股息收益率等。通过把个股信息与股票市场的总体平均状况加以比较，财务分析师们就可以得出某一支具体股票的价格是被高估还是被低估的判断。例如，《巴伦周刊》(Barron's)在2008年2月18日曾经报道过30种道琼斯工业股票的平均股息收益率为2.45%，同期，奥驰亚集团(Altria Group)股票的股息收益率则为3.05%。在本例中，股息收益率方面的统计信息显示，与道琼斯股票的平均股息收益率相比，奥驰亚集团的股息收益率比较高。因此，一个金融顾问很可能得出以下结论：奥驰亚集团的现行价格被低估了。这一点以及其他与有关奥驰亚集团的信息有助于财务分析师做出对该股票是买入、卖出还是继续持有的建议。

1.5.3 营销

各零售商的收银台使用的电子扫描仪收集到的数据可广泛用于各种市场营销调研。例如，像A. C. 尼尔森(A. C. Nielsen)公司和信息资源有限公司(Information Resources, Inc.)这样的数据提供商，都是从商场购买产品销售纪录的电子扫描数据，经过处理后，再把这些数据的统计汇总信息卖给制造商。制造商们为了获得此类数据，往往在每种产品上要花费成千上万美元。制造商们还购买诸如特价和店内展销等促销活动时相关产品的销售数据和统计汇总信息。相关产品管理人员可以分析电子扫描数据的统计结果与促销活动期间商品的相关销售统计结果，从而更好地把握产品促销和产品销量之间的关系。这种分析对制定各种产品未来的营销策略是大有好处的。

1.5.4 生产

当今社会高度重视产品的质量，因此在生产过程中对质量进行监控是统计学的一项重要应用。各种统计质量控制图被用来监控生产过程。特别地， $\bar{X} - R$ 质量控制图可用于监控平均产出。例如，假设利用机器自动向罐子中注入一种12盎司的软饮料，工人定期从产品中抽取一定数量灌装好的饮料作为样本，计算样本中这种软饮料的平均含量，并将这个平均值($\bar{x}$ 值)画在一张 $\bar{x}$ 质量控制图上。如果标出的值大于图中的控制上限值表示注入的饮料过量；如果标出的值小于图中的限制下限值，则表示注入的饮料不足。只要标出的 $\bar{x}$

值处于图中的控制上限与控制下限之间,则生产继续,这时生产过程处于受控状态。确切地说, $\bar{x}$ 质量控制图有助于确定何时必须调整和矫正生产过程。

1.5.5 经济

经济学家经常会对未来的经济形势或者某些经济问题做出预测。他们利用大量的统计信息来做这样的预测。例如,在预测通货膨胀率时,经济学家们会依据下列统计指标:生产价格指数、失业率、制造业开工率等。通常,将这些统计指标输入到预测通货膨胀率的计算模型中,就可以预测出相应的通货膨胀率。



趣味小贴士：统计资源网站

<http://www.lianshu.com/> 联数工作室(统计软件)
<http://www.sasor.com/SAS/> 爱好者
<http://www.statforum.com/> 数理统计论坛
<http://home.kimo.com.tw/g894730/> 统计生活馆(论坛、统计教学等丰富内容)
<http://estat.ncku.edu.tw/> 统计在线学习馆(论坛、统计教学等丰富内容)
<http://probbstat.nuk.edu.tw/> 概率统计在线学习馆(教学)
http://www.math.nsysu.edu.tw/NSC_prob/ 概率学习馆(教学)
<http://ibm9.math.nsysu.edu.tw/StatDemo/> 互动式统计教学(教学)
<http://140.128.104.155/wenwei/thesis/jj1.htm/> 多媒体互动式在线统计教学
<http://stat.ustc.edu.cn/> 中国科学技术大学统计与金融系(数理统计、软件下载)
<http://www.applstats.org.cn/> 中国应用统计网(人大应用统计科研所)
<http://stats.xmu.edu.cn/info/default.asp/> 中国经济统计(统计资源)
<http://www.china-actuary.com/default.asp/> 中国精算网
<http://www.fengsb.com/index.asp/> 研究人员之家——实验设计 & 数据分析
<http://www2.zzu.edu.cn/spq6/> 数据处理 & 实验设计(包括 PPT 教案、软件等)

本章小结

1. 通过趣味案例分析与应用、应用领域的简要介绍,加深印象、引起兴趣,确立统计有用、学好统计的信心。
2. 通过“统计”三种含义的分析,体会到统计学就是“数据的科学”,它的目的是“探索数据内在的数量规律性”。
3. 通过对统计学分科的阐述,明确不同统计领域的分析理念、分析思路、分析风格是有所差别的。
4. 不同统计领域所常用的基本术语是不相同的,描述统计常用总体、单位、标志、指标等,推断统计常用总体、样本、参数、统计量等。

思考与练习

1. 什么是统计？怎样理解统计学与统计数据的关系。
2. 试举出日常生活或工作中的统计应用实例。
3. 举例说明总体、单位、标志、变量、指标这几个概念。
4. 某研究部门准备抽取 2000 个职工家庭推断该城市所有职工家庭的年人均收入。请指出研究的总体、样本、参数、统计量。