

# 第 1 章

## 绪 论

### 引 例

#### 2 元钱的经济纠纷

2017 年 9 月 26 日下午,宁波海曙江厦派出所的廖警察出了一次警——一笔“涉案金额”为 2 元的小纠纷。根据当事双方的要求,他很快处理好了这次纠纷。回来后,他细想下来,总觉得哪里不对劲:似乎这道“数学题”不是这么算的!

这是怎么回事?到底是什么数学题这么费脑筋呢?

报警电话是一名小卖部老板打来的,说是“有人买东西不给钱”。廖警察到场后,了解了情况,大概是这么回事:有名小朋友在小卖部里买了一支雪糕,花了 2 元钱。后来,小朋友的外婆觉得雪糕有点儿变形,怀疑是变质的,于是到店里来要求退钱。小卖部老板当然不同意退。这时,一名女孩过来,买了一瓶水,也是 2 元。外婆就抢过她递给店老板的 5 元钱,然后找了对方 3 元钱。女孩也没说什么,拿了水就走了。对小朋友的外婆来说,这笔账似乎就这么了了。可在店老板看来,他没收到矿泉水的钱啊。两人就为这 2 元钱起了纠纷,于是报了警。

“那你们双方什么意见?”廖警察问道。“她把那 5 元钱还给我!”“那他得把买雪糕的 2 元钱退给我!”双方争先恐后地说。民警在场,两人倒是很快协调好了,各自拿了钱。当事双方没有多想,对这一协调结果表示满意。于是,大家都散了。但廖警察在返回所里的路上隐隐觉得:这道数学题不是这么算的吧?回去后,他特意把每次交易行为都一一列了下来,得出了结论——有人亏钱了。

资料来源:马涛,夏婧.案值 2 元的小纠纷很“烧脑”[N]. 宁波晚报,2017-09-28(A9).

### 1.1 统计与统计学

#### 1.1.1 统计学的产生与发展

统计作为一种人类认识自然、改造自然的实践活动,在原始社会就已经产生。统计学是伴随着资本主义古典哲学、古典政治经济学和空想社会主义的产生而发展起来的。一般认为,统计学从 17 世纪 60 年代算起,到现在已有 300 多年的历史。统计学的发展过程大体可分为以下四个阶段。

第一个阶段称为“城邦政情”阶段。“城邦政情”阶段始于古希腊的亚里士多德写的

“城邦政情”或“城邦纪要”。他一共撰写了 150 余条纪要,其内容包括城邦的历史、行政、科学、艺术、人口、资源和财富等社会及经济情况的比较与分析,具有社会科学特征。“城邦政情”式的统计研究延续了一两千年,直到 17 世纪中叶才逐渐被“政治算术”这个名词所替代,并且很快被演化为“统计学”(statistics)。统计学依然保留了“城邦”(state)这个词根。

第二个阶段称为“政治算术”阶段。“政治算术”阶段与“城邦政情”阶段并没有明显的分界点,两者也没有根本的区别。“政治算术”的特点是统计方法与数学计算和推理方法相结合,分析社会经济问题的方式更加注重运用定量分析方法。可以用 1690 年英国古典政治经济学的奠基人威廉·配第(William Petty, 1623—1687)出版《政治算术》一书作为这个阶段的起始标志。威廉·配第在书中强调了自己的立论的方法,“用数字、重量和尺度来表达自己的问题”。用数字、重量和尺度量化社会经济现象的方法是近代统计学的主要特征。因此,威廉·配第的《政治算术》被后来的学者评价为近代统计学的来源,威廉·配第也被评价为“近代统计学之父”。构成“政治算术”方法的核心是“数字”。配第在书中使用的数字有三类:第一类是对社会经济现象通过统计调查和经验观察得到的数字,第二类是运用某种数学方法推算出来的数字,第三类是为了进行理论性推理而采用的示例性数字。配第把这种运算数字和符号进行推理称为“代数的算法”。

另外,法国的布莱士·帕斯卡(Blaise Pascal, 1623—1662)和皮埃尔·德·费马(Pierre de Fermat, 1601—1665)将赌博中出现的各种具体问题归结为一般的概率原理,为概率论和统计学的发展奠定了重要的基础。特别值得提出的是比利时统计学家朗伯·阿道夫·雅克·凯特勒(Lambert Adolphe Jacques Quetelet, 1796—1874),他对统计学理论最大的影响是将概率与统计学相结合,从而提出了关于统计学的新概念,并努力使人们把统计学理解为一种可用于任何学科的一般研究方法。在“政治算术”阶段出现的统计学与数学的结合趋势逐渐发展形成了“统计分析科学”。

第三个阶段称为“统计分析科学”阶段。19 世纪末,欧洲大学里开设的“国情纪要”或“政治算术”等课程名称逐渐消失,取而代之的是“统计分析科学”课程。当时的“统计分析科学”课程的内容仍然是分析研究社会经济问题。“统计分析科学”课程的出现是现代统计学发展阶段的开端。

经过历代统计学家的努力和发展,到 19 世纪末建成了古典统计学的基本框架。1908 年,英国的威廉·希利·戈塞特(William Sealy Gosset, 1876—1937)用笔名 Student 发表了关于  $t$  分布的论文,这是一篇在统计发展史上划时代的文章。他创立了小样本代替大样本的方法,开创了统计学的新纪元。后来统计学家罗纳德·艾尔默·费希尔(Ronald Aylmer Fisher, 1890—1962)给出了  $F$  统计量、极大似然估计、方差分析等;耶日·内曼(Jerzy Neyman, 1894—1981)和埃贡·S. 皮尔逊(Egon S. Pearson, 1895—1980)提出了置信区间估计和假设检验;亚伯拉罕·沃尔德(Abraham Wald, 1902—1950)发明了序贯抽样和统计决策函数;到 20 世纪 50 年代,构筑了现代统计学的基本框架。20 世纪 50 年代后,统计学进入一个全新的发展阶段,新的研究领域层出不穷,像多元统计分析、非参数统计、现代时间序列分析和线性统计模型等。

第四个阶段称为“统计学帝国主义”阶段。从世界范围看,自 20 世纪 60 年代以后,统计学的发展有三个明显的趋势:第一,随着数学的发展,统计学依赖和吸收数学方法越来越

越多;第二,向其他学科领域渗透,或者说以统计学为基础的边缘学科不断形成;第三,随着统计学应用的日益广泛和深入,以及受计算机和新兴科学的影响,统计学越来越依赖计算技术,成为数量分析的方法论科学。这一时期,统计学的研究和运用范围越来越广。美国的 *Science* 有一篇文章列出近百年来最有用的科学,统计学位居前 10 名。

### 1.1.2 统计的含义

“统计”一词,在不同的场合,所代表的含义也有所不同。一般来讲,“统计”有三种含义:统计工作、统计资料和统计学。

#### 1. 统计工作

统计工作是指对社会现象或自然现象的总体数量方面进行收集、整理和分析的实践活动,它是一种社会调查研究活动。统计工作在人类历史上出现得比较早。随着历史的发展,统计工作逐渐发展和完善起来,并成为国家、企业和个人及科研单位认识与改造客观世界和主观世界的一种有力工具。统计工作一般包括统计设计、统计调查、统计整理、统计分析、统计资料的提供和管理等环节。

#### 2. 统计资料

统计资料是指可以用以推导出某项结论的一些事实或数字。统计资料一般由元素、变量和观测值三部分组成。

(1) 元素。我们所研究的对象由各元素组成,统计资料就是关于各元素特征的信息。如统计某公司员工基本情况时,各员工就是元素。

(2) 变量。变量是关于元素的一种属性和特征。年龄是各员工的一个特征,这些特征对于不同元素取不同的结果,所以称为变量。其中结果用数字表示的,称为定量变量;不可用数字表示的,称为定性变量。例如,年龄、身高、体重、公司服务年限和受教育年限是定量变量,性别、民族、籍贯、学历是定性变量。

(3) 观测值。观测值是指通过测量或测定所得到的样本值。如统计每小时通过一个高速公路收费站的汽车数量就是一个观测值。

#### 3. 统计学

统计学是以收集、整理、分析和研究等统计技术为手段,对所研究对象的总体数量关系和数据资料去伪存真、去粗取精,从而达到显示、描述和推断所研究对象的特征、趋势和规律性的目的。简言之,统计学是以少量的数据(称为样本)所提供的信息来推断欲研究对象(称为总体)特征的一门科学。

统计工作与统计资料和统计学的相互关系如图 1-1 所示。

### 1.1.3 统计学的研究对象及特点

统计学的研究对象是客观事物的总体数量特征、数量关系和数据资料,以反映其发展过程及规律性。统计学的特点可以归纳为以下五个方面。

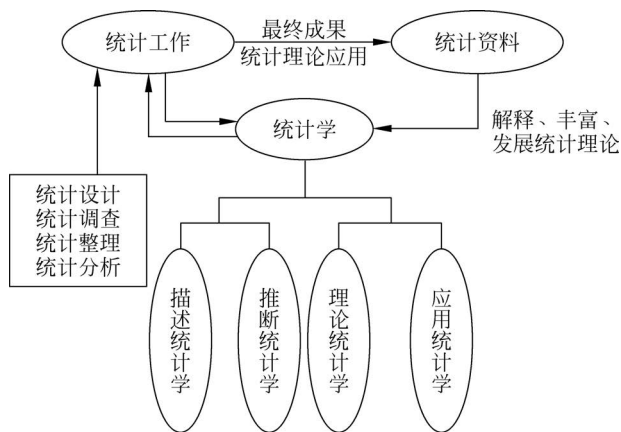


图 1-1 统计工作与统计资料和统计学的相互关系

### 1. 数量性

统计学的特点首先表现为以准确的事实为基础；同时，这些事实用数字加以表现，具有简短性和明显性。数量性是统计学研究对象的重要特点，这一特点也可把其他实质性的社会科学（如政治经济学）区别开来，如各种统计年鉴就是用大量数字资料说明事物的规模、水平、结构、比例关系、差别程度、普遍程度、发展速度等。应当注意，统计学不是单纯地研究社会现象的数量方面，而是在质与量的密切联系中研究社会现象的数量方面。

### 2. 总体性

统计学研究社会现象的数量方面指的是总体的数量方面。从总体上研究社会现象的数量方面，是统计学区别于其他社会科学的一个主要特点。社会现象是各种社会规律交错作用的结果，它呈现出一种复杂多变的情景。统计学对社会现象总体数量方面的调查研究，用的是综合研究方法，而不是对单个事物的研究，但其研究过程是从个体到总体，即必须对足够大量的个体进行登记、整理和综合，使它过渡到总体的数量方面，从而把握社会现象的总体规模、总水平及其变化发展的总趋势。

### 3. 具体性

统计研究对象是自然和社会经济领域中具体现象的数量方面。它不是单纯数量的研究，而是具有明确的现实含义，这一特点是统计学与数学的分水岭。数学是研究事物的抽象空间和抽象数量的科学，而统计学研究的是客观存在的、具体实在的数量表现。正因为统计的数量是客观存在和具体实在的数量表现，它才能独立于客观世界，不以人们的主观意志为转移。统计资料作为主观对客观的反映，必然是存在第一性、意识第二性，存在决定意识，如实反映具体的已经发生的客观事实，才能为我们进行统计分析研究提供可靠的基础，才能分析、探索和掌握事物的统计的规律性。统计研究对象的这一特点，也正是统计工作必须遵循的基本原则。

#### 4. 社会性

统计学研究社会现象,这一特点与自然技术统计学有所区别。自然技术统计学研究自然技术现象(如天文、物理、生物、水文等),自然现象的变化发展有其固有的规律,在其变化过程中,通常表现为随机现象可能出现也可能不出现的现象。而统计学的研究是人类社会活动的过程和结果,人类的社会活动都是人们有意识、有目的的活动,各种活动都贯穿着人与人之间的关系,除了随机现象而外,还存在确定性的现象,即必然要出现的现象。所以统计学在研究社会现象时,还必须注意正确处理好这些涉及人与人之间关系的社会矛盾。

#### 5. 广泛性

统计学研究的数量方面非常广泛,指全部社会现象的数量方面。广泛性这一特点,可区别于研究某一特定领域的其他社会科学(如政治学、经济学、社会学、法学等)。统计学研究的领域包括整个社会,它既研究生产关系,也研究生产力以及生产关系和生产力之间的关系;它既研究经济基础,也研究上层建筑以及经济基础和上层建筑之间的关系。此外,它还研究生产、流通、分配、消费等社会再生产的全过程以及社会、政治、法律、军事、文化、教育等全部社会现象的数量方面。

### 1.1.4 统计学的分科

从统计方法的构成看,统计学可分为描述统计学和推断统计学;从统计方法研究和应用角度看,统计学可分为理论统计学和应用统计学。

#### 1. 描述统计学和推断统计学

描述统计学,研究任何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处理,进而通过综合、概括与分析得出反映客观现象的规律性的数量特征的方法。其内容包括统计数据的收集方法、数据的加工整理方法、数据分布特征的概括与分析方法等。它是整个统计学的基础。

推断统计学,研究根据样本数据去推断总体数量特征的方法,它是在对样本数据进行描述的基础上,对统计总体的未知数量特征作出以概率形式表述的推断。推断统计学是现代统计学的主要内容。

描述统计学和推断统计学的划分,既反映了统计方法发展的前后两个阶段,也反映了统计方法研究和探索客观现象内在数量规律性的先后两个过程。从描述统计学发展到推断统计学,既反映了统计学发展的巨大成就,也是统计学发展成熟的重要标志。

#### 2. 理论统计学和应用统计学

理论统计学主要是阐明统计学的一般原理和一般方法,主要内容包括概率论、随机化原则理论、各种估计原理、假设检验原理和一般预测与决策原理,其侧重点在于统计学原理的推导与证明。

应用统计学主要研究如何应用统计方法去解决实际问题。统计方法的应用几乎扩展

到所有的科学研究领域。统计方法在经济领域的应用就形成了经济统计学及其若干分支,在管理领域的应用就形成了管理统计学,在商务领域的应用就形成了商务统计学,在生物学中的应用就形成了生物统计学,在医学中的应用就形成了医疗卫生统计学,等等。这些应用统计学的不同分支,其所应用的基本统计方法是相同的,即都是描述统计和推断统计的主要方法,只是由于各应用领域的特殊性形成了各分支间一些不同的特点。

## 1.2 统计的几个基本概念

### 1.2.1 总体与总体单位

总体是在某种共性的基础上由许多个别事物结合起来的整体。构成总体的个别事物称为总体单位,总体单位也称个体。总体和总体单位都是客观存在的事物,是统计研究的客体。例如,在工业普查中,“工业企业”就是一个总体,它是由所有从事工业生产活动的企业所组成的,其中的每一个工业企业就是一个总体单位。总体必须同时具备三个性质:同质性、大量性和差异性。

#### 1. 同质性

同质性是形成统计总体的必要条件,是统计总体各个单位在某一点上具有共同的属性和特征,否则就不能形成一个统计总体。同质性是确定统计总体的基本标准,它是根据统计的研究目的而定的。研究目的不同,则所确定的总体不同,其同质性的意义也随之变化。例如,研究城镇居民的生活状况,城镇居民就构成了统计总体,凡是城镇居民都是同质的。

#### 2. 大量性

大量性是指总体由许多单位组成,仅个别和少数单位不能构成总体。这一点是由统计研究的目的所决定的。统计研究的目的是要揭示自然和社会经济现象的规律性,我们只有通过大量事物的观察、分析和研究,才能发现从其普遍联系中表现出来的规律性。例如,要研究某市职工的工资水平,我们只观察少数几个职工的工资是达不到目的的,因为这少数几个职工的工资不能代表该市全体职工的工资水平,它可能偏高或偏低。如果我们观察许多职工的工资,就可降低或抵消偶然性的偏差,计算他们的平均工资,就可以反映该市职工工资的一般水平。每个职工工资差别越大,需要观察的职工人数就越多。如果要提高观察职工工资的代表性,就需要增加观察的职工人数。可见,统计总体的大量性是一个相对概念。总体的大量性与各单位标志表现的差异性密切联系在一起。同时,总体的大量性也与研究目的的要求有关,精确度要求越高,总体单位数也越大。

#### 3. 差异性

差异性是指构成统计总体的各个总体单位除了某一方面是同质的以外,在其他方面又要有差异,否则就没有必要进行统计分析了。以某企业职工总体为例,这里每位职工除了工作单位相同外,在姓名、性别、年龄、学历、职称和工龄等许多方面都存在差异。因此,

事物存在差异性是进行统计研究的前提。

总体可分为有限总体和无限总体。如果一个总体中包含的总体单位数是有限的,称其为有限总体;如果一个总体中包含的总体单位数是无限的,称其为无限总体。例如,我们要研究海洋中的鱼类、某地区的空气污染情况,则海洋中的鱼类、某地区的空气污染情况就是无限总体。但社会经济现象总体大多是有限的,如全国工业企业总体、某企业职工总体、某商场库存商品总体以及某地交通事故总体等。

总体和总体单位具有相对性,它们随着研究目的不同是可以变换的。例如,如果要研究某地区工业企业的生产经营情况,则该地区全部工业企业构成总体,而每一个工业企业是总体单位;如果要研究该地区某一个企业的生产经营情况,那么该企业就成了总体,该企业下属的各个职能部门就是总体单位。由此可见,一个工业企业由于研究目的不同,既可以作为一个总体单位来研究,也可以作为一个总体来研究。

### 1.2.2 样本

样本是指从统计总体中抽取出来作为代表这一总体的部分单位组成的集合体。与样本相对应的统计总体称为全及总体。由于样本是从总体中抽取出来并代表总体的,全及总体也可称为母体,而样本则称为子体。样本也有自己的几个特点。

#### 1. 代表性

抽取样本的目的是用来推断总体,这就必然要求样本能够代表总体。样本代表总体的程度越高,样本计算的抽样指标与总体指标的误差就越小。因此,抽样推断时,总是要求样本具有较高的代表性。

#### 2. 客观性

从总体中抽取样本时,必须排除主观因素的影响。样本可以用随机抽样的方法抽取,也可以用非随机抽样的方法抽取,但都必须保证取样的客观性,保证样本的选择不受调查者和被调查者的主观影响。

#### 3. 随机性

一个总体可以抽取许多不同的样本,至于到底抽取的样本是哪一个,完全取决于样本的随机性。随机性是用样本数据资料推断总体特征性质的基本原则。

#### 4. 排他性

样本单位必须取自总体内部,而不能抽取总体外部的单位。因为统计推断是利用样本作为总体代表,用样本的数据资料来推断总体的数据资料。

### 1.2.3 指标与标志

#### 1. 指标

指标也称统计指标,其含义一般有两种理解:一种是反映总体现象数量特征的概念。

例如,国内生产总值(GDP)、国民总收入(GNI)、商品销售额、人口出生率等。这种理解的统计指标包括三个构成要素:指标名称、计算方法及计量单位。这种含义一般在进行统计设计或理论研究时所使用,它是仅有数量概念而没有具体数字的统计指标。另一种是由反映总体现象数量特征的概念(指标名称)和具体数值构成的统计指标。例如,《湖南统计年鉴 2023》显示,2022 年长沙市地区生产总值 486 707.37 亿元。从完整的意义上讲,指标由六个要素构成:时间限制、空间限制、指标名称、指标数值、计量单位和计算方法。这种含义的统计指标是统计工作中经常使用的。上述例子对应的六要素依次是:时间限制是 2022 年,空间限制是长沙市,指标名称是地区生产总值,指标数值是 486 707.37,计量单位是亿元,计算方法是支出法。

指标通常分为数量指标和质量指标。数量指标是反映现象总规模、总水平和工作总量的统计指标,如人口总数、企业总数、工资总额、国内生产总值等。由于数量指标反映现象的总量,所以也称总量指标,并且由于用绝对数表示,也称统计绝对数。质量指标是反映现象相对水平或工作质量的统计指标,如人口密度、出生率、出勤率、劳动生产率、职工平均工资等。质量指标通常是由两个总量指标相比而派生的指标,用相对指标或平均指标来表示,反映现象之间的内在联系和对比关系。

## 2. 标志

标志是说明总体单位属性和特征的名称。例如,某企业全体职工作为一个总体,每一位职工是总体单位,职工的性别、年龄、籍贯、民族等是说明每一位职工的特征的名称,都是标志。

标志按其性质不同可分为品质标志和数量标志。品质标志是表明总体单位品质属性或特征的名称,它不能用数值表示,只能用文字说明。例如,企业职工性别、籍贯、民族、文化程度就是品质标志。数量标志是表明总体单位数量方面的特征,它用各种不同的数值表示。例如,企业职工的年龄、工龄就是数量标志。数量标志的具体表现为标志值。例如,A 公司某职工年龄为 28 岁,工资为 6 500 元/月。

## 3. 指标和标志的区别与联系

指标和标志,两者既有区别又有联系。

指标和标志的区别有以下四点。

第一,标志是说明总体单位特征的,而指标是说明总体特征的。

第二,标志有能用数值表示的数量标志和不能用数值表示的品质标志,而指标不论是数量指标还是质量指标,都是用数值表示的。

第三,指标数值是经过一定的汇总取得的,而标志中的数量标志不一定经过汇总,可直接取得。

第四,标志一般不具备时间、地点等条件,但作为一个完整的统计指标,一定要包含时间、地点、范围等信息。

指标和标志的联系有以下两点。

第一,统计指标的数值是以总体单位数量标志的值进行直接汇总或间接计算而来的。

例如,某企业职工的月工资总额是该企业所属职工月工资额汇总而来的,而职工的月平均工资则是通过进一步计算得到的。

第二,指标与数量标志之间存在着变换关系。由于研究目的的不同,原来的统计总体变成了总体单位,则相应的统计指标也就变成数量标志了;反之亦然。例如,在研究某公司职工情况时,该公司的全部职工是总体,该公司的工资总额是统计指标。而在研究该公司所属的集团公司职工工资时,该公司就是总体单位,则该公司的工资总额为数量标志,具体的工资总额数值为标志值。于是,该公司的工资总额由统计指标相应变为数量标志了。

#### 1.2.4 变异与变量

统计中的变异是普遍存在的,一般意义上的变异是指标志在总体单位之间的不同具体表现,但严格地说,变异仅指品质标志的不同具体表现,如性别表现为男、女,民族表现为汉族、满族、蒙古族、回族、苗族等。而数量标志的不同具体表现则称为变量值或标志值,如某职工的年龄为35岁、工龄为14年、月工资为8500元等。可变的数量标志称为变量。品质标志的变异最后表现为综合性的数量时,如按职工的性别,汇总计算出男、女各多少人,才构成统计研究的对象。观察登记总体各单位的品质标志和数量标志的变异与变量,是统计研究的起点。

变量按其取值是否连续,可分为离散变量和连续变量。凡变量值只能以整数出现的变量,称为离散变量;凡变量值可做无限分割的变量,称为连续变量。例如,一个地区的人口数、工厂数、一个工厂的机器台数等,都只能是整数而不可能带小数,这些就是离散变量。但人的身高、体重、工厂的产值、利润等却是可以带小数,而且其数值是可以无限分割的,在两个数字之间还可以有不断变化的其他数字,这种变量就是连续变量。

变量按其所受因素影响的不同,可分为确定性变量和随机性变量。确定性变量是具有某种(或某些)起决定性作用的因素致使其沿着一定的方向呈上升、下降或水平变动的变量。例如,随着人们生活水平的提高以及医疗卫生条件的完善,这些确定性因素的影响,使得人们的期望寿命这个变量的值不断提高。期望寿命就是确定性变量。随机性变量是指变量值的变化受不确定性因素的影响,变量值的变化没有一个确定的方向,有很大的偶然性。例如,在同一台机器设备上加工某种机械零件,其尺寸大小总是存在差异的。造成这些差异的因素可能有原材料质量、电压的不稳定,气温和环境的变化以及操作工人的情绪波动等。这些影响该种机械零件尺寸变动的因素都是随机发生的,是不确定的。这里的机械零件尺寸就是一个随机性变量。

### 1.3 商业世界中的统计学

统计学作为一门基础学科,其方法适用于所有学科领域。统计分析方法是一种通用数据分析方法。可以说,只要有数据的地方就会用到统计方法。随着人们对定量研究的日益重视,统计方法已被应用到自然科学和社会科学的众多领域。几乎所有的研究领域都要用到统计方法,如政府部门、学术研究领域、日常生活、公司或企业的生产经营管理

等。下面简要介绍统计在商业中的一些应用。

### 1. 企业发展战略

发展战略的重要性对一家企业而言是不言而喻的。发展战略的制定需要把握宏观经济和市场的状况及发展变化趋势。同时,还要对本企业进行合理的市场定位,以把握企业自身的优势和劣势。所有这些都离不开统计。它不仅需要统计提供可靠的数据,还需要利用统计方法对数据进行科学的分析和预测。

### 2. 市场研究

市场竞争是激烈的。企业要在激烈的市场竞争中获胜,首先必须了解市场。为此,企业需要做广泛的市场调查,取得所需的信息,并对这些信息进行科学的分析,以便作为生产和营销的依据,这些都需要统计的支持。

### 3. 产品质量管理

质量是企业的生命,是企业持续发展的基础。质量管理离不开统计的应用。质量控制已成为统计学在生产领域中的一项重要应用。各种统计质量控制图被广泛应用于监测生产过程。当今,6 $\sigma$  准则已成为一种重要的产品质量管理理念。

### 4. 财务分析

企业的财务分析不可避免地使用统计分析方法。一方面,企业自身的投资,离不开对财务数据的分析,其中要用到大量的统计方法;另一方面,投资者往往根据上市公司提供的财务报表和相关的统计数据进行分析,以此作为投资决策的参考依据。

### 5. 经济预测

企业经营者必须对未来的市场状况进行预测。在进行预测时,要使用各种统计信息和统计方法。例如,企业要对产品的市场潜力作出预测,以便及时调整生产计划,这就需要利用市场调查取得数据,并对数据进行统计分析。

### 6. 人力资源管理

企业人力资源管理人员利用统计方法对企业员工的年龄、性别、受教育程度、工作业绩、工资等进行分析,并作为企业制订工资计划、奖惩制度的依据。

当然,统计并不是仅仅为了管理才有用,它是为自然科学、社会科学的多个领域而发展起来的,为多个学科提供了一种通用的数据分析方法。从某种意义上说,统计仅仅是一种数据分析的方法。与数学一样,统计是一种工具,它是一种数据分析的工具。利用统计方法可以简化繁杂的数据,如用图表展示数据,建立数据模型。有人认为统计的全部目的就是让人看懂数据,其实这仅仅是统计的一个方面。统计更重要的功能是对数据进行分析。它提供了一套分析数据的方法和工具。统计不是万能的,它不能解决你所面临的所有问题。统计可以帮助分析数据,并从分析中得出某种结论,但对统计结论的进一步解

释,则需要你的专业知识。例如,吸烟可能使患肺癌的概率增加,这是一个统计结论,但要解释吸烟为什么能引起肺癌,这就不是统计学家所能解释的,而需要有更多的医学知识才行。

商务统计学应用的广泛性,在于它的分析方法的有效性,其中就包括统计模型。统计模型用统计学方法描述经济活动。根据所采用的数学方法不同,对经济活动揭示的程度不同,构成各类不同的统计模型,主要有相关回归分析模型、时间序列分析模型以及统计决策模型等,相关知识我们将在后续章节相应介绍。

## 1.4 统计数据的计量尺度与类型

### 1.4.1 统计数据的计量尺度

要对客观对象进行计量,就要弄清楚统计数据的计量尺度问题。根据对研究对象计量的不同精确程度,人们将计量尺度由粗略到精确分为四个层次:定类尺度、定序尺度、定距尺度和定比尺度。

#### 1. 定类尺度

定类尺度是最粗略、计量层次最低的计量尺度。它是按照客观现象的某种属性对其进行分类,不能进行加、减、乘、除等数学运算。这一场合所使用的数值只是作为各种分类的代码,并不反映各类的优劣、量的大小或顺序。例如,人口按性别分为男、女,用“1”表示男性,“0”表示女性。

#### 2. 定序尺度

定序尺度也称顺序尺度,是对现象之间的等级或顺序差别的一种度量。利用定序尺度不仅可以研究对象分为不同的类别,还可以反映各类的优劣、量的大小或顺序,但不能进行加、减、乘、除等数学运算。例如,学生成绩可分为优、良、中、及格和不及格五类,产品质量可分为一、二、三等,一个人对某件事情的态度可以分为非常赞同、赞同、不赞同、坚决反对等。

#### 3. 定距尺度

定距尺度是对现象类别或顺序之间距离的度量。定距尺度不仅可以用数表示现象各类别的不同和顺序大小的差异,而且可以用确切的数值反映现象之间在量方面的差异。定距尺度的计算结果表现为数值,可以进行加、减等数学运算。反映现象规模水平的数据必须以定距尺度计量,如产品数量、人口数、企业数、国内生产总值等。定距尺度在统计数据中占据重要地位,统计中的总量指标也是用定距尺度来计量的。

#### 4. 定比尺度

定比尺度是在定距尺度的基础上,确定相应的比较基数,然后将两种相关的数加以对

比而形成相对数(或平均数),用于反映现象的结构、比重、密度、速度等数量关系。定比尺度的计量结果也是数值,可以进行加、减、乘、除等数学运算。例如,将一个企业创造的增加值与该企业的职工人数对比,计算全员劳动生产率,以此反映该企业的生产效率。在统计的对比分析中,要广泛地运用定比尺度进行计量。

### 1.4.2 统计数据类型

统计数据是对现象进行计量的结果。例如,对经济活动总量的计量可以得到国内生产总值的数据,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。由此可见,统计数据不仅仅是数字,也可以是文字的。

按计量层次,可以将统计数据分为分类数据、顺序数据和数值型数据。统计数据还可以从其他角度进行分类,如按收集方法,可以将其分为观测数据和实验数据;按时间状况,可以将统计数据分为截面数据和时间序列数据。

#### 1. 分类数据、顺序数据和数值型数据

(1) 分类数据。只能归于某一类别的非数字型数据,称为分类数据。分类数据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。例如,人口按照性别分为男、女两类;学位按照等级分为博士、硕士和学士三类。

在分类数据中,各类别之间是平等的并列关系,无法区分优劣或大小,各类别之间的顺序是可以任意改变的。它仅具有等于或不等于的数学特性。虽然分类数据只是表现为某种类别,但为了便于统计处理,特别是为了便于计算机识别,我们可以对不同类别用不同的数字或编码来表示,如用“1”表示男性人口,“0”表示女性人口;用“1”表示国有企业,“2”表示集体企业,“3”表示私营企业;等等。这些数字只是给不同类别的一个代码,并不意味着这些数字可以区分大小或进行任何数学运算。

(2) 顺序数据。只能归于某一有序类别的非数字型数据,称为顺序数据。顺序数据也是对事物进行分类的结果,但这些类别是有顺序的。例如,学生考试成绩可以分为优、良、中、及格、不及格,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。

顺序尺度的计量结果虽然也表现为类别,但这些类别之间是可以比较顺序的。很显然,顺序数据要比分类数据精确一些,它除了具有等于或不等于的数学特性外,还具有大于或小于的数学特性。

(3) 数值型数据。按数字尺度测量的观察值,称为数值型数据。数值型数据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。例如,考试成绩用百分制度量,重量用克度量,长度用米度量,等等,其结果都表现为具体的数值。因此,数值型数据可以进行加、减、乘、除运算。

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品质数据;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。我们所处理的数据大多为

数值型数据。

## 2. 观测数据和实验数据

观测数据是通过调查或观测收集到的数据,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。实验数据是在实验中控制实验对象而收集到的数据。例如,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。

## 3. 截面数据和时间序列数据

不同对象在同一时期或时点发生的数据,称为截面数据。截面数据所描述的是现象在某一时刻的变化情况。例如,2023 年我国 31 个省区市的生产总值数据就是截面数据。

同一对象在不同时期或时点产生的数据,称为时间序列数据。时间序列数据所描述的是现象随时间而变化的情况,如 2001—2023 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。

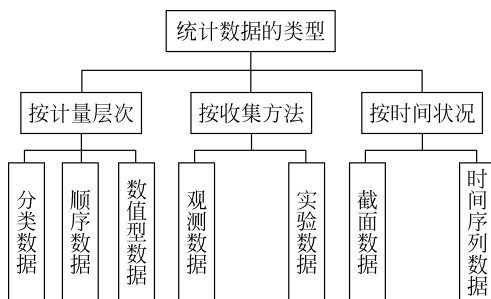


图 1-2 统计数据的分类

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。适用于低层次测量数据的统计方法,也适用于较高层次的测量数据,因为后者具有前者的数学特性。例如,在描述数据的集中趋势时,对分类数据,我们通常是计算众数,对顺序数据通常是计算中位数,但对数值型数据同样也可以计算众数和中位数;相反,适用于高层次测量数据的统计方法,则不能用于较低层次的测量数据,因为低层次测量数据不具有高层次测量数据的数学特性。例如,对于数值型数据我们可以计算平均数,但对于分类数据和顺序数据则不能计算平均数。理解这一点,对于选择统计分析方法是十分有用的。值得注意的是,对于 SPSS 等统计软件,必须定义变量的类型,否则无法进行统计运算。

# 1.5 统计研究方法过程

## 1.5.1 统计研究方法

统计学研究客观现象总体的数量特征与规律性所涉及的方法多种多样,这些统计研

究方法构成了统计研究方法体系。在统计研究方法体系中,最重要、最基本的研究方法有以下几种。

### 1. 大量观察法

大量观察法是根据总体中足够多数的单位进行统计研究的方法。由于在研究的总体中,总体单位受各种因素的影响往往具有差异性,总体单位不能反映总体的一般特征和规律性。这就要求统计必须对足够多的总体单位进行综合研究,以清除偶然因素的影响,反映出总体的必然性、数量特征和规律性。大量观察法可以是对总体所有单位进行研究,如普查;亦可以是对能表现现象本质特征和规律性的部分单位进行非全面统计,如抽样调查。

### 2. 统计分组法

统计分组法是对所研究的客观现象按照一定的分类标准或标志,把研究的全部总体单位划分为不同的组别,用以区别现象的各种不同类型,揭示总体的内部结构及其分别特征,反映现象之间的相互关系,达到认识客观现象的本质特征和规律性的目的的方法。统计分组法既是统计资料整理的基本方法,也是一种重要的统计分析方法。

### 3. 综合指标法

综合指标法是统计研究中不可缺少的基本方法。综合指标是表明客观现象总体数量特征的各种数量化的概念及其数值。常用的综合指标有总量指标、相对指标、平均指标等。综合指标法是统计整理、统计分析的基本方法,其他各种统计分析方法均以它作为基础,如时间序列法、指数法、抽样法、相关法等都离不开综合指标的对比研究。

### 4. 归纳推断法

归纳法是从个别到一般的推理方法。综合指标法是对个体的数值综合汇总成总体的数值,以概括反映总体的一般数量特征,就是归纳法的具体体现。统计分组法是将总体各个个体划分为不同的组别,以研究总体的内部结构和分别特征,也是一种归纳法。推断法是根据样本数据来推断总体特征的方法,常用的推断法有抽样推断、参数估计、假设检验、产品质量检测、统计预测决策法等。

## 1.5.2 统计研究过程

统计研究是一个复杂的过程,主要包括以下几个步骤。

### 1. 统计设计

统计设计是根据统计研究的目的和要求,对统计研究的对象、内容、方法及程序所做的通盘考虑和安排。统计设计的结果表现为各种设计方案,如统计调查方案、统计指标体系、分类目录、统计报表制度、统计整理与分析方案等。它是统计工作在实践中展开前所做的计划性安排,既包括统计方案的全面设计,又包括人、财、物等各种资源的安排。一项统计工作能否顺利进行与统计设计是否全面、周密有很大关系。

## 2. 统计调查

统计调查是根据统计设计的要求收集统计数据的阶段,是定量认识的起点。统计调查与其他调查活动的不同点在于,精心设计的统计调查方法更能保证调查活动收到事半功倍的效果。统计调查的方法分为普查、统计报表制度、重点调查、典型调查和抽样调查,这五种方法各有特点。普查收集的资料最全面、详细、准确,但缺点也很明显,那就是费钱、费时且费力。重点调查和典型调查是我们短期迅速获得数据的好方法,但容易产生无法控制的误差,且随意性很大。抽样调查是最经常使用的一种方法,原因在于它具有时效性强、低成本、高质量和科学性的特点,虽然抽样调查的结果存在误差,但这个误差是我们可以接受和控制的。抽样调查有一整套科学方法,是统计学的重要分支。

## 3. 统计整理

统计整理是对通过统计调查收集到的原始数据、统计资料,根据研究的需要进行分类、汇总等。统计整理是一个对统计调查结果去粗取精、去伪存真的过程,就是要对统计调查收集到的原始数据进行把关。统计整理后的结果就是我们通常看到的各种统计资料。

## 4. 统计分析

统计分析是对统计整理的数据进行再加工和深加工的过程,主要是采用各种分析方法,计算各种分析指标,以揭示被研究对象的总体数量特征和规律性,从而达到统计研究的最终目的。这一阶段充分体现了统计学的精华,主要表现为对各种统计分析方法的应用,如统计指标、时间序列分析、统计指数和回归分析等。

统计研究过程框架如图 1-3 所示。

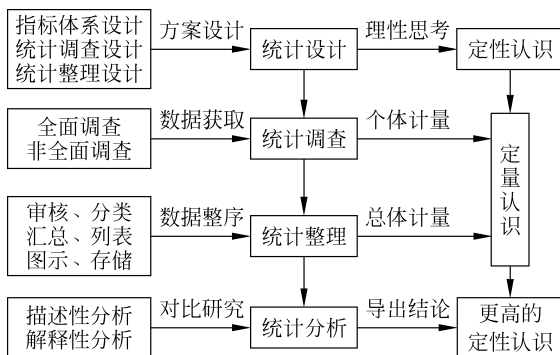


图 1-3 统计研究过程框架

## 1.6 案例：高考志愿

### 1.6.1 案例背景

湖南的赵大爷一家祖祖辈辈都是地地道道的农民,全家节衣缩食供唯一的孙女读书,

希望知识改变命运。孙女芳芳刻苦努力,不负众望,2017 年高考成绩 588 分(理科)。在填高考志愿之际,赵大爷问在中南大学工作的杨老师:能否填报中南大学?

## 1.6.2 案例分析

### 1. 高考志愿填报的基本步骤

考生该如何填报高考志愿呢?一般而言,应经过以下几个基本步骤。

(1) 熟悉规则。<sup>①</sup> 每一位考生可以填写 10 个平行志愿——10 所大学,每个志愿下可填写 6 个专业,高考录取遵循“分数优先、志愿优先”的原则。

(2) 收集资料。查询历年(尤其是近 3 年)国家最低控制分数线和各高校的抛档分数线。

(3) 数据分析。根据当年的最低控制分数线和历年的最低控制分数线,预测各高校当年的抛档分数线。

(4) 统计决策。根据考生当年的高考分数和预测的高校抛档分数线,选择学校。根据选择的学校,进一步查询该校历年各专业的录取线,结合当年实际分数和自己的兴趣,选择相关专业。

### 2. 芳芳的高考志愿填报

2017 年芳芳高考理科成绩 588 分,湖南省本科一批理科国家最低控制线是 505 分,在一本院校选择。而中南大学属于一本,可以进一步分析被中南大学录取的可能性。中南大学理工科 2014—2016 年最低控制分数线依次为 522 分、526 分、517 分,抛档分数线依次为 612 分、614 分、610 分。抛档分数线高出最低控制分数线 90 分、88 分和 93 分,平均高出约 90 分。预测 2017 年中南大学理工科抛档分数线至少 595 分,芳芳不宜填报中南大学。事实上,2017 年中南大学理工科抛档分数线为 601 分。

## 习 题 1

- 1.1 怎样理解统计学与统计数据的关系?
- 1.2 统计数据分为哪几种类型?比较不同类型数据的特点。
- 1.3 什么是变量?变量有哪些类型?
- 1.4 举例说明连续变量与离散变量。
- 1.5 举例说明总体、样本、参数、统计量等概念。
- 1.6 指标与标志的区别是什么?请举例加以说明。
- 1.7 举例区分不同的统计尺度。
- 1.8 结合实际问题论述统计研究过程。
- 1.9 常见的统计研究方法有哪些?请举例加以说明。
- 1.10 结合实际问题谈谈商务统计的应用。

<sup>①</sup> 说明:本案例基于 2017 年高考志愿填报规则分析,近年来高考志愿填报规则有变化。

## 即测即练

