

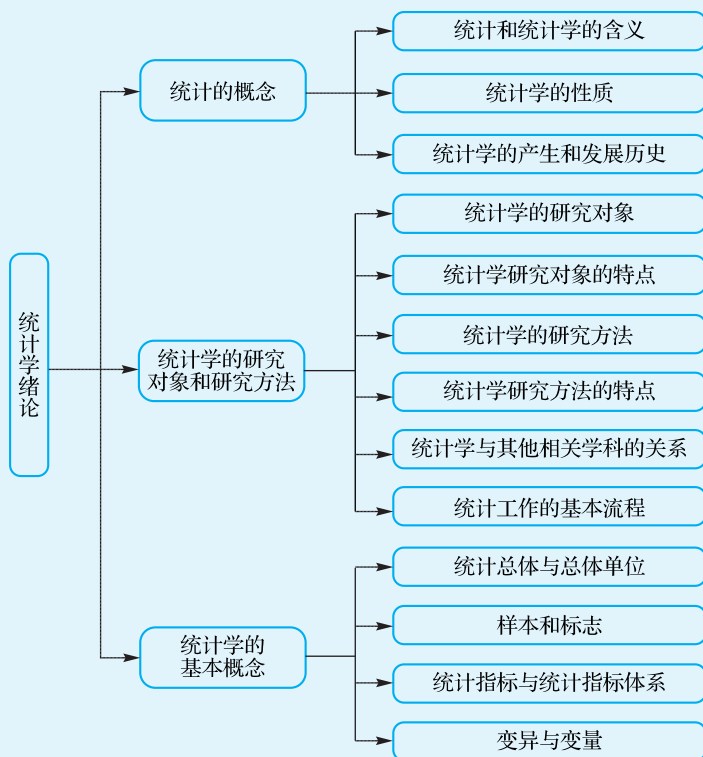
# 项目一

## 统计学绪论

### ● 学习目标

- 理解统计和统计学的含义；
- 熟悉统计学的研究对象和研究方法；
- 了解统计工作的基本流程和工作方法；
- 掌握统计学的基本概念和相关指标。

### ● 项目导航





## 案例导入

## 妙用统计学预测身高

19世纪后期,英国著名生物学家兼统计学家弗朗西斯·高尔顿(Francis Galton)在研究父母身高与子女身高的关系时,通过对1 078个家庭中父母身高和成年男孩身高的观察,得出了一个很有意思的结论:如果将父亲与母亲身高的平均值记作 $x$ ,其中一个成年男孩的身高记作 $y$ ,那么成年男孩的身高与其父母身高的平均值就满足一个方程: $y=33.73+0.516x$ 。

通过这个方程,可以估算一个家庭中幼年男孩成年之后的身高状况。但是,时隔100多年后的今天,人类的物质生活和精神生活都发生了巨大的变化,父母身高与子女身高之间将呈现出一种什么样的关系呢?在现实生活中,人们都知道父母身高对子女身高是有影响的,但是父亲与母亲的影响分别有多大?对男孩和对女孩的影响程度是否一样?我们能否以定量的形式回答这个问题呢?

资料来源:豆丁网,<https://www.docin.com/p-1603838301.html>,有删改。

学习统计学理论与方法,可以帮助我们进一步揭示父母身高与子女身高之间量化关系的秘密,帮助那些关注自己后代身高的年轻父母进行早期的预测,同时,为那些未婚青年男女选择理想配偶提供科学的参考依据。而且,通过对统计学内容的深入学习,我们还可以从现实生活中所存在的大量数据中找出事物之间的内在联系,通过量化分析方法来把握事物发展的规律性,为我们未来的生产和生活提供理论依据与实践方法。

## 模块一 统计的概念

统计学是一门研究数据的科学,是人们认识客观世界总体数量变动关系和变动规律的有力工具,是用以收集数据、筛选数据、分析数据、处理数据并由数据得出结论的一组概念、原则和方法。在科学知识和信息技术快速发展的今天,统计学已经被越来越广泛地应用于国家管理、社会经济管理、自然科学以及我们日常工作、学习和生活的各个方面。社会发展问题、经济可持续发展问题、国际竞争力问题、金融风险管理问题、保险精算问题、人口与社会保障问题、环境保护问题等,都亟待我们去深入地研究并解决。要解决这些问题,只进行定性分析是不够的,还必须进行定量分析。

统计学所研究的对象是一个总体现象,并且是从数量方面对其加以分析研究以达到对其本质的认识。本模块主要介绍统计与统计学的含义、统计学的性质、统计学的产生和发展历史,为后续各部分的学习奠定一定的基础。

### 一、统计与统计学的含义

什么是“统计”?什么是“统计学”?这是学习本课程首先应了解的问题。

统计的英文单词为 statistics,其语源最早出自中世纪拉丁语的 status,原意是指各种现



象的状态和状况。由这一词根组成的意大利语 state,表示国家的概念及关于国家结构和国情方面知识的总称。

统计的产生和发展具有悠久的历史,最早将“统计”一词作为学科名称使用的是18世纪德国哥廷根大学政治学教授戈特弗里德·阿亨瓦尔(Gottfried Achenwall),他把国势学称为statistik,即统计学。在英国,早在17世纪就出现用数字来说明社会现象的科学,但使用的是另一个完全不同的名称——政治算术(political arithmetic)。直到18世纪末,statistics才作为德语statistik的译文传入英国,即用数字表示事实。

所谓统计,就是人们认识客观世界总体现象数量特征、数量关系和数量变动规律的一种调查研究方法。这种方法是对总体现象数量方面进行收集、整理和分析研究的总称,是人们认识客观世界的一种最有效的工具。随着社会生产实践的快速发展,统计的含义也在不断地丰富。在日常的工作和生活中,我们经常提到的“统计”一词一般包含统计工作、统计资料和统计学三种含义。

统计工作是指收集、整理、分析、展示和编制统计数据资料等实践工作的全过程。

统计资料是统计工作的结果,是通过统计工作所取得的各项数字和有关情况的资料,用以反映客观事物的规模、水平、发展速度等多方面特征,它是我们进行社会经济管理和科学研究的重要依据。

随着统计方法和统计工具在各个社会领域中的应用,统计学已成为具有多个分支的综合学科。国内外统计学家曾经从不同的角度给统计学下了不同的定义,主要有:统计学是作为指导统计工作科学运行的理论研究的学科;统计学是对事物从数量的规律性进行分析的方法论科学;统计学是研究社会现象特别是经济现象数量方面的科学。我们认为,统计学是一门关于收集、整理、分析统计数据的科学。

总的来说,统计工作是统计实践活动,统计资料是统计工作的成果,统计学是统计实践经验的理论概括和深化,是研究如何测定、收集、整理、归纳和分析反映客观现象总体数量的数据,以便给出正确认识的方法论科学。统计学与统计实践活动的关系是理论与实践的关系,理论源于实践,高于实践,理论反过来又指导实践。统计学作为一门独立的学科形成以后,又反过来指导统计工作实践和统计资料表述,它们是理论与实践的关系。

统计工作的开展离不开统计数据,统计是对数据的操作过程,从另外一个角度来说,统计也研究变量之间的关系,实际上,研究变量之间的关系也是通过研究数据来实现的。统计数据的收集是获取数据的过程,它是进行统计分析的前提和基础,离开了统计数据,统计资料的整理和统计分析就无从谈起。因此,如何取得所需的数据是统计学的研究内容之一。统计数据的整理是对统计数据加工处理的过程,目的是使统计数据系统化、规范化和条理化,符合统计分析的需要。统计数据的整理是介于数据收集与数据分析之间的一个必要的环节。统计数据的分析是统计学的核心内容,它是通过统计方法认识和分析客观事物的规模、水平、内部比例关系、发展变化规律等特征的过程。可见,统计学是一门关于统计数据的科学。



## 知识拓展

## 中国古代的计数方法

我们现在可以数数,用笔、计算器等进行计算,那么古代的人们是怎么计算的呢?从文献资料中发现,我国古代用以计数的方法大致可以分为四种。

## 一、结绳计数

结绳计数是原始社会创始的以绳结形式反映客观经济活动及其数量关系的记录方式。结绳计数是被原始先民广泛使用的记录方式之一。

《周易·系辞下》中记载:“上古结绳而治,后世圣人易之以书契,百官以治,万民以察。”意思是上古时期,人们使用结绳记事的方式管理政务,后来帝王改变了这种方式,使用文书据以替代,众多官员凭借文书来治理政务,百姓也通过此来知晓世情。

直到20世纪中期,云南的少数民族地区仍在沿用结绳计数。而且不只是中国,世界各地的不同民族都有类似的计数方法。据说,古秘鲁印加人(南美洲印第安人)用来打结的绳子名为“魁普”(quipus),表示的数目清楚、完备,用来登录账目、人口数及税收数。

## 二、书契计数

书契计数是结绳计数之后出现的计数方法,当时主要用于剩余粮食数量的计算。“书契”指的就是文字。

由于这些刻有文字的竹木简经常被用作订立契约关系的凭证,因此“契”和“书契”也有“契约”的意思,我们今天常用的“地契”“房契”等词的意思正源于此。“契”字在“默契”等词语中表示的情义相投的意思,则是后来发展出来的。

《管子·揆度篇》中记述:“上古结绳,后易之以书契。”即是用刻刀将数刻在兽骨、竹木、龟甲、土石崖上,以便长久保存,不易损坏。

## 三、算筹计数

根据史书的记载和考古材料的发现,古代的算筹实际上是用竹子、木头、兽骨等材料制成一些长短、粗细差不多的小棍子以计算数目,不用时则把它们放在小袋子里面保存或携带。

算筹是我国古代广泛应用的一种计算工具,它的出现年代现在难以考证,但据史料推测,至迟在春秋晚期战国初年时已经出现。算筹制作规范、体积小、便于携带,更利于精确计算,作为一种计数方式,显然要比结绳计数成熟得多。

事实也的确如此,一直到算盘发明推广之前,算筹都是我国古代最重要的计算工具。算筹计数法遵循十进制,在世界数学史上是一个伟大的创造,跟世界上其他古老民族的计数法相比,具有显而易见的优越性。

## 四、画“正”字计数

在古代,民间还有画“正”字的计数方法,这种计数方法起源于戏院里面记的流水账。例如,每天戏院里要迎来很多的观众,当时还没有门票这个概念,所以就安排店小二在门口招揽生意,领满5位入座。还有专门的人在大水牌上写“正”字表明来了多少客人看戏,然后由账房先生负责计算收费,等等。后来,戏院都实行了门票制度,画“正”字的计数方法也就被废除了,不过很多中国人还是有采用“正”字计数的习惯。



## 二、统计学的性质

统计学最初是作为一门实质性科学而建立起来的,主要用来记述和反映关于国家的重要事项。但是,随着统计学研究范围的不断扩大以及统计方法在社会经济领域和自然科学领域的广泛应用,再加上统计方法体系本身的不断完善和发展,统计学的研究对象也发生了深刻变化。统计学已从实质性学科中分离出来,转而研究统计方法,成为一门方法论学科。本书所讨论的统计学,是关于统计学的基本理论、基本原则和基本统计方法,这些方法既可以用于社会经济现象方面的研究,也可以用于自然现象数量方面的研究,它是应用统计学的基础。

### (一) 统计学的研究对象是客观现象的数量

在统计学学科发展初期,所研究的问题是关于人口调查、出生与死亡的登记等现象。随着社会的发展,研究对象又扩大到社会经济和生物实验等方面。目前,无论是社会的、自然的还是实验的,凡是有大量数据出现的地方,都要用到统计学。凡是能以数量来表现的均可作为统计学的研究对象。统计方法已渗透到其他科学领域,成为当前最活跃的学科之一。

### (二) 统计学研究的是总体现象的数量特征与规律性

统计学所研究的是总体现象的数量特征及其分布的规律性。总体是由许多个体组成的,各个个体在数量特征上受到必然和偶然两种因素的支配,必然因素反映了该总体的特征,但由于受偶然因素的影响,个体是有差异的,通过这些个体的差异来描述或推断总体的特征,就产生了统计学。

### (三) 统计学是一门方法论学科

在统计学界对应用统计学的性质有实质性学科和方法论学科之争。方法论学科不以现实客体为研究对象,所研究的只是方法,提出通用性的方法工具,为其他学科的研究服务。与方法论相对,实质性学科以现实客体为研究对象,研究客体的特征、存在形式和变动规律。统计学到底属于实质性学科还是方法论学科,统计学界至今尚未形成统一的认识。基于目前统计学在国内的发展和应用实践,本书认为统计学是实用性很强的方法论学科。就统计工作来说,它总是研究实际问题的,应用统计学的方法也是从现实问题中产生的。然而统计学的发展有一个过程,早期的国势学派和政治算术学派虽然也利用一些统计方法来记述和分析现实问题,但那时还没有形成独立的统计学。随着统计方法的应用日益广泛,其内容也不断发展和充实,尤其是概率论的发展为统计方法提供了理论基础,使统计方法相对独立地形成了自己的科学体系,即统计学。其内容包括收集资料,对收集的资料加以整理、概括和表示以及对取得的数据进行分析和推断等一系列方法。这些方法和原理构成了统计学的基本内容。目前,统计方法已成为科学研究和各种管理的重要工具,它是一门年轻而引人入胜的科学,并且还在不断地发展。

## 三、统计学的产生和发展历史

统计活动的产生和发展至今已经有几千年的历史,但是统计学在学术方面作为一门独立的学科的历史没有这么长。学术界普遍认为,统计学产生于17世纪中叶的欧洲,距现在已有300多年。其发展主要可分为三个阶段。



### （一）古典统计学时期

这个时期大致是从 17 世纪中叶至 19 世纪初,其代表学派是政治算术学派和国势学派。政治算术学派产生于英国,后人称之为统计学中的政治算术学派。其主要代表人物是英国的威廉·配第(William Petty)和约翰·格朗特(John Graunt)。

配第于 1676 年出版了《政治算术》一书,此书以一系列分析和大量计算手段清晰地描述了英格兰、荷兰、法兰西和爱尔兰等地的经济、军事、政治等方面的情况,为英国称霸世界提供了各种有说服力的实证分析资料。《政治算术》的意义主要表现在研究问题的方法方面。配第在该书的序言里写道:“我进行这种工作所使用的方法在目前还不是常见的,因为我不采用比较级或最高级的词语进行思辨式的议论,相反地采用了这样的方法(作为我很久以来就想建立的政治算术的一个范例),即用数字、重量和尺度来表达自己想说的的问题,只进行诉诸人们感觉的议论,借以考察在自然界中有可见的根据的原因。”

配第用“数字、重量和尺度”研究现象的方法为统计学的产生奠定了基础。自配第之后的 200 年间,以用数量方法研究社会经济问题为基本特征的“政治算术”模式,成为统计学发展的主流。《政治算术》的出版,标志着统计学的诞生。马克思对配第及其《政治算术》评价很高,他说配第创造的政治算术,即一般所说的统计,还说配第是“政治经济学之父”,在某种程度上也可以说是“统计学的创始人”。

英国人约翰·格朗特早在 1662 年就出版了《关于死亡表的自然观察与政治观察》。他根据伦敦市发表的人口自然变动公报,通过大量观察,对人口的出生率和死亡率做了许多分类、计算和研究,发现了人口与社会现象中重要的数量规律。例如,新生儿的性别比例稳定在 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数量规律,而且对伦敦市总人口数量做出了较科学的估计。如果说配第是政府统计的创始人,那格朗特则可被认为是人口统计的创始人。

国势学派又称记述学派或国情学派,产生于 17 世纪封建制的德国,其主要代表人物是海尔曼·康令(H. Conring)和戈特弗里德·阿亨瓦尔。这一学派最早提出了“统计学”的名称。在康令之前,欧洲各国已出版有记述各国情况的著作。从 1660 年开始,康令在西尔姆斯特大学以“国势学”为题讲授一门课程,内容是各个国家的显著事项,方法则是文字叙述,目的是满足政治家的知识需求。阿亨瓦尔是国势学派的主要继承人和最有名的代表人物,一生在大学任教。他在 1749 年出版的《近代欧洲各国国势学概论》中,首次使用“统计学”(statistik)这个名称代替了“国势学”(阿亨瓦尔说过,statistik 的语源是拉丁语 status 和意大利语 state,前者是“现状”或“现势”的意思,后者是“国家”的意思)。阿亨瓦尔对统计学的性质做了解释,他认为统计学是关于各国基本制度的学问,其研究对象是一个国家显著事项的整体。这里的“国家显著事项”是指一个国家的领土、人口、财政、军事、政治和法律制度等,用这些来说明和比较国家的形势,就称为国势学。

#### 知识拓展

#### 中国古代的统计制度与方法

中国古代统计活动主要由官方组织开展,大多采用全面调查的方法,通过政府派



员调查和被调查者报告的方式进行,有时也采用重点调查、典型调查和统计估算等方法。此外,由于中国古代统计活动主要由各部门组织进行,很多统计数据产生于行政登记,各部门之间相互利用行政记录数据也较为普遍。

全面调查的方法运用于政府主持的人口、土地、赋税、仓储等多项统计活动中,通过基层填报、逐级上报、逐级审核、逐级汇总的方式收集统计资料。如中国古代历史最悠久的上计制度,就是以全面调查为基础而进行的。全面调查的具体方式有实地调查法和报告法两种。

(1) 实地调查法。实地调查法即由政府主管部门制定调查制度方法,按照由下至上的程序,通过基层官员组织填报,形成各类户籍、土地、税赋等的簿册,然后层层上报、审核、汇总,最终形成全国总簿册。唐代的计簿、宋代的丁产簿、明代的黄册和鱼鳞图册、清代的赋役全书等,均是通过这种方法编制而成的。

(2) 报告法。报告法即由被调查者自行申报或填报的一种统计调查方法。这种方法始创于秦朝,在人口、土地等统计调查活动中使用。东晋时期,报告法有所创新。句容县县令刘超在进行户籍调查时,将调查内容以信件方式发至各村,由百姓自报家产数目,填好后报送县府。百姓大多据实填报,税收收入反而超过往年。这种方法类似当今邮寄调查表的方法,不仅实施效果较好,而且在统计调查史上具有创造性。

在唐代,报告法进一步发展为全国性的“手实法”,官府规定,每年年终,各户须把全家人的姓名、年龄及所有田产数量填报至乡、里;乡、里将其汇编为册,称为“乡账”;然后由县、州、户部逐级汇编,最终编成全国的人口与土地资料,作为课税征夫的根据。每份手实调查表的最后都印有一段话,要求百姓保证如实填报。

元代的土地调查也采用过报告法,也称“经理法”。官府将土地上报的内容予以公示,限百姓 40 天内将其田产和田赋情况向官府呈报。报告过程中,允许他人检举,如有瞒报或欺骗,予以相应处罚;官吏若因工作不力导致差错,也要量事受罚。

中国古代统计活动还运用重点调查、典型调查和统计估算等方法。

(1) 重点调查。战国时期,商鞅提出“强国知十三数”,意为想要强国,就必须掌握能说明国情的 13 个数据,其可被视为较早关于重点调查的思想。中国古代重点调查主要用于矿冶、仓储等与国计民生密切相关的活动中。

(2) 典型调查。最早的典型调查是战国时期李悝所进行的。他通过对一个五口之家农户的收支情况进行调查,得出“农夫常困”的结论。此后,通过典型调查来进行社会经济问题的分析研究活动多有所见,如司马迁通过对资金周转和盈利的调查,得出当时社会的合理年利润率为 20%;林则徐通过调查一个烟民吸食鸦片一年的耗费,推断全国白银流失严重。

(3) 统计估算。春秋时期,管仲根据食盐消费量的调查结果,用平均数法估算提高盐价后的全国财政收入。战国时期,李悝用比例数估算耕地总数,又用平均每亩增减数估算耕地的粮食增减总量。此后估算法应用较多,如唐代杜佑、元代卢世荣都曾运用过统计估算进行财经分析。

资料来源:中国信息报, <http://www.zgxxb.com.cn/tjdk/201903120013.shtml>, 有删改。



## （二）近代统计学时期

这个时期大致是从 18 世纪末到 19 世纪末。著名的大数定律、最小二乘法、相关与回归分析、指数分析法、时间序列分析以及正态分布等理论都是在这个时期建立和发展起来的。

这一时期的代表学派主要有数理统计学派和社会统计学派。

数理统计学派产生于 19 世纪中叶,创始人是比利时学者阿道夫·凯特勒(Adolphe Quetelet)。他在统计理论上的主要贡献是把概率论引进了统计学,从而提出了关于统计学的新概念。凯特勒根据大数定律的原理提出了大量观察法,利用统计观察资料计算和研究社会现象和自然现象的数量规律性,并用于预测未来的情况。他创立了大数定律,认为统计学就是数理统计学。

凯特勒开创了统计理论和实际应用的一个新领域,即应用概率论认识随机现象数量规律性的理论和方法。这个新领域起初没有确定的名称,1867 年,德国数学家维特根斯坦(L. Wittgenstein)在其题为“关于数理统计学及其在政治经济学和保险学中的应用”的论文中提出“数理统计学”这个词,随即成为该学科的正式名称。数理统计学产生较晚,但发展很快。后经过葛尔顿、皮尔逊、鲍莱、戈赛特、费雪等人的研究和实践,发展成为一门完整系统的新学科。

社会统计学派产生于 19 世纪末期,首创者是德国人卡尔·古斯塔夫·阿道夫·克尼斯(Karl Gustav Adolf Knies),主要代表人物有格奥尔格·冯·梅尔(Georg Von Mayr)和恩斯特·恩格尔(Ernst Engel),他们认为统计学的研究对象是社会现象,研究方法是大量观察法,并提出统计学是一门实质性的社会科学。

### 知识拓展

#### 恩格尔系数

恩格尔系数(Engel's coefficient)是食品支出总额占个人消费支出总额的比重。19 世纪德国统计学家恩格尔根据统计资料,对消费结构的变化研究得出一个规律:一个家庭收入越少,家庭收入中(或总支出中)用来购买食物的支出所占的比例就越大,随着家庭收入的增加,家庭收入中(或总支出中)用来购买食物的支出所占的比例则会减少。推而广之,一个国家越穷,每个国民的平均收入中(或平均支出中)用于购买食物的支出所占的比例就越大;相反,一个国家越富裕,这个比例就越小。

恩格尔系数达 59% 以上为贫困,50%~59% 为温饱,40%~50% 为小康,30%~40% 为富裕,低于 30% 为最富裕。

## （三）现代统计学时期

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做推断。因为对大量产品做全面的检验,既费时费钱,又费人力,加之有些产品的质量检验为破坏性检验,所以全部检验已不可能。1908 年,戈赛特(W. S. Gosset)发表了关于  $t$  分布的论文,创立了小样本代替大样本的理论,即利用  $t$  统计量就可以从大量的产品中只抽取较小的样本完成对全部产品质量的检验和推断。费雪(R. A. Fisher)又对小样本理论进行进一步研究,给出了  $F$  统计量、最大似然估计、方差分析等方法和思想,





标志着现代统计学的开端。1930年,尼曼(J. Neyman)与小皮尔逊(E. S. Pearson)共同对假设检验理论做了系统的研究,创立了“尼曼-皮尔逊”理论,同时,尼曼又创立了区间估计理论。美国统计学家瓦尔德把统计学中的估计和假设理论予以归纳,创立了“决策理论”。这些研究和发现大大充实了现代统计学的内容。

20世纪50年代,统计理论、方法和应用进入一个全面发展的新阶段。一方面,统计学受计算机科学、信息论、混沌理论和人工智能等现代科学技术的影响,新的研究领域层出不穷,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、探索性数据分析、数据挖掘等;另一方面,统计方法的应用领域不断扩展,几乎所有的科学研究都离不开统计方法。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,现在连纯文科领域的法律、历史、语言、新闻等都越来越重视对统计数据进行分析,国外的人文与社会学科普遍开设统计学的课程,可以说统计方法与数学、哲学一样成了所有学科的基础。

### 知识拓展

#### 中华人民共和国统计法

《中华人民共和国统计法》作为我国唯一的一部统计法律,于1983年12月8日由第六届全国人民代表大会常务委员会第三次会议通过,1996年5月15日经第八届全国人民代表大会常务委员第十九次会议修正,2009年6月27日再次经第十一届全国人民代表大会常务委员会第九次会议修订通过,于2010年1月1日起施行。全文共七章五十条。

第一章总则中的前四条如下:

**第一条** 为了科学、有效地组织统计工作,保障统计资料的真实性、准确性、完整性和及时性,发挥统计在了解国情国力、服务经济社会发展中的重要作用,促进社会主义现代化建设事业发展,制定本法。

**第二条** 本法适用于各级人民政府、县级以上人民政府统计机构和有关部门组织实施的统计活动。

统计的基本任务是对经济社会发展情况进行统计调查、统计分析,提供统计资料和统计咨询意见,实行统计监督。

**第三条** 国家建立集中统一的统计系统,实行统一领导、分级负责的统计管理体制。

**第四条** 国务院和地方各级人民政府、各有关部门应当加强对统计工作的组织领导,为统计工作提供必要的保障。

## 模块二 统计学的研究对象和研究方法

### 一、统计学的研究对象

统计工作是对客观现象的数量资料进行收集、整理、分析研究的工作过程。统计学与统



计工作之间存在理论与实践的关系。

## 二、统计学研究对象的特点

本书中的数量方面是指社会经济现象的规模、水平、结构、速度、比例关系、差别程度、普遍程度、普及程度、发展速度、平均规模和水平、平均发展速度等。由于事物的质和量是密切联系的,因此,统计学在研究社会现象时,首先从定性研究开始,其次进行定量分析,最后达到认识社会现象的本质、特征或规律,这就是质—量—质的统计研究过程和方法。归纳起来,统计学研究对象的特点可以分为以下五个方面。

### (一) 数量性

统计学研究的是大量社会经济现象总体的数量方面的特征,主要包括社会经济现象的数量表现、各种现象之间的数量关系、质量互变的数量界限及其规律性三个方面。与其他经济学科(如政治经济学)相比,统计学最根本的特点就在于数量性,数字是统计的语言,它所研究的事物的量从社会现象的定性认识开始,以质的规定性为基础。

例如,我们观察一定时期人民的物质消费水平,就要统计一定范围、一定时期的社会商品零售额,并与相应的人口数进行比较。统计社会商品零售额,首先要明确这个指标的含义,然后再统计它的“量”。而确定社会商品零售额的含义,需要以政治经济学的理论为指导,与实践相联系,并解决什么是商品流转、什么是零售、社会商品零售额的统计范围包括哪些,以及怎样收集、整理和汇总社会商品零售额统计资料等问题。在确定这一统计指标含义的同时,必须考虑指标的可操作性。当我们统计了社会商品零售额以后,再把它与相应的人口联系起来对比分析,就可以观察到相应的物质消费水平。我们还可以把社会商品零售额与社会商品购买力进行对比,观察购买力的实现程度,分析两者之间的比例关系。至于这两者之间应该维持一个什么样的比例关系,才能稳定市场、满足人民的物质文化需要,还应进一步分析它们之间的数量界限。由此可见,统计虽然是研究社会经济现象数量方面的问题,但任何时候都不能离开社会经济现象的质。

## 知识拓展

### 社会消费品零售总额

社会消费品零售总额是指企业(单位、个体户)通过交易直接售给个人、社会集团的非生产、非经营用的实物商品金额以及提供餐饮服务所取得的收入金额。社会消费品零售总额指标涉及的商品包括售给个人作为生活消费用的商品和修建房屋用的建筑材料,也包括售给社会集团用作非生产、非经营的商品等;不包括企业(单位、个体户)用于生产经营和固定资产投资所使用的原材料、燃料和其他消耗品的价值量,也不包括城市居民用于购买商品房的支出和农民用于购买农业生产资料的支出费用。

社会消费品零售总额反映了国内消费品市场的总规模和地域分布情况,也反映了居民和社会集团对实物商品消费需求的总量和变化趋势。

在商业统计中,社会消费品零售总额可做如下分类:



- (1) 按行业类型可分为商业零售额、饮食业零售额、工业零售额等。
- (2) 按经济类型可分为全民所有制经济零售额、集体所有制经济零售额、个体经济零售额、中外合资和外资独资经济零售额等。
- (3) 按销售对象可分为对城乡居民的消费品零售额、对社会集团的消费品零售额、对农村的农业生产资料零售额等。
- (4) 按商品类别可分为食品类零售额、日用品类零售额、文化娱乐品类零售额、衣着类零售额、医药类零售额、燃料类零售额、农业生产资料类零售额等。

## (二) 总体性

统计学研究社会现象的数量方面不是指个别现象的数量特征,而是指由许多个别现象构成的总体的数量特征,是通过对许多性质相同的个别现象所组成的总体进行大量观察和综合分析,来反映现象总体的数量特征,揭示社会经济现象的一般状况。

统计学研究对象的总体性特点,是由社会经济现象的特点和统计学的研究目的来决定的。由于社会经济现象错综复杂,个别现象所处的时间、地点和条件不同,表现出明显的偶然性和不确定性,难以说明社会经济现象总体的本质和规律。只有以社会经济现象的总体为研究对象,即以构成总体的全部或足够多的单位作为研究对象,才能消除偶然性因素的影响,防止“只见树木,不见森林”的片面性,从而正确地揭示出社会经济现象的本质和规律性。当然,任何一个总体都是由个体构成的,要认识社会经济现象总体的数量特征,必须从调查个体的表现入手,从个体到总体。

例如,要研究中国农民的生活水平,就需要把全国各地的所有农民组成一个总体来统计,不论是哪一个地区或哪一个民族,还是高收入或低收入,只要是农民,都要包括在内。这样就可以消除地理环境、民族特征、收入高低等方面的差异,反映出中国农民生活水平的一般情况。当然,为了深入分析农民的生活水平,还可以就高收入、中收入、低收入农民的典型进行调查分析,探究不同典型的差异和形成原因,从而全面、客观地说明被研究对象的情况。

## (三) 具体性

统计学研究的对象是社会经济现象中具体事物的数量方面,而不是抽象的数量及其相互关系,这是统计学与数学的重要区别。因为社会经济现象中的事物都是具体的,现象都是在一定的地点、时间、条件下发生的,所以,其量的表现就必然带有特定场合和特定历史的痕迹,离开具体的地点、时间和条件,是无法说明社会经济现象的本质及其运行规律的。

例如,中国农民的生活水平,既存在地区上的差异,也存在时间上的不同。因此,在研究时,除考虑地区因素以外,还必须明确是哪一年的生活水平,并结合我国农业的生产条件、科学技术在农业生产中的作用,结合农村消费品价格等因素进行分析,来说明中国农民的生活水平是高还是低,是好还是差。单凭一个孤立的统计数字是难以说明问题的。

## (四) 社会性

统计学属于社会科学,这一点与自然科学不同。社会科学研究的是社会经济问题,而自然科学研究的是自然现象。除了研究对象不同以外,由于社会科学的认识主体所立场、所持观点、所用方法不同,会得出差别很大甚至完全不同的结论。

统计学通过对社会经济现象总体数量的调查研究来认识人类社会活动的条件、过程和



结果,反映物质资料的占有关系、分配关系、交换关系以及其他的社会关系。统计学研究的社会经济现象与各种利益关系是密切联系的。其定量研究是以定性分析为提前的,而定性分析使统计学在客观上就有了社会关系的内涵。所以,统计学在研究社会经济现象时,就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。

例如,在研究劳动者的收入时,可以根据劳动者在社会再生产过程中的地位和作用,先将劳动者分为经营者、管理者、技术员和一般工人等,然后再统计不同类型劳动者的收入,分析他们之间的相互关系,研究社会分配的合理性,从而修改或编制劳动报酬分配政策,以调整不同类型劳动者之间的相互关系,达到稳定社会秩序、调动广大劳动者积极性的目的。如果不加区别地把经营者、管理者、技术员和一般工人的收入混为一谈,势必得出错误的结论,以致制定出错误的政策,从而引起各种不必要的矛盾和社会问题。所以,社会性是统计学区别于其他自然科学的主要特征之一。

### (五) 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。这个特点,是统计学区别于研究某一领域的其他社会科学(如政治学、经济学、社会学、法学等)的特征之一。

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,统计学还研究生产、流通、分配、消费等社会再生产的全过程以及社会、政治、经济、军事、法律、文化、教育等全部社会现象的数量方面。

## 知识拓展

### 人均可支配收入

人均可支配收入在实际生活中,常用来代指人均居民可支配收入,严格来说这么使用不够准确。居民可支配收入是居民可用于最终消费支出和储蓄的总和,即居民可用于自由支配的收入。居民可支配收入既包括现金收入,也包括实物收入。按照收入的来源,可支配收入包含四项,分别为:工资性收入、经营性净收入、财产性净收入和转移性净收入。

2019年,全国居民人均可支配收入30733元,比上年名义增长8.9%,扣除价格因素,实际增长5.8%。其中,城镇居民人均可支配收入42359元,增长(以下如无特别说明,均为同比名义增长)7.9%,扣除价格因素,实际增长5.0%;农村居民人均可支配收入16021元,增长9.6%,扣除价格因素,实际增长6.2%。

2019年,全国居民人均可支配收入中位数为26523元,增长9.0%,中位数是平均数的86.3%。其中,城镇居民人均可支配收入中位数为39244元,增长7.8%,是平均数的92.6%;农村居民人均可支配收入中位数为14389元,增长10.1%,是平均数的89.8%。

资料来源:新浪财经, <http://finance.sina.com.cn/china/gncj/2020-01-17/doc-iihnzahk4668531.shtml>,有删改。



### 三、统计学的研究方法

研究方法在科学研究活动中是一个非常重要的问题,统计学在研究大量社会经济现象总体数量特征的过程中,要使用多种统计方法,包括大量观察法、统计分组法和综合指标法、统计推断法等。

#### (一) 大量观察法

大量观察法是统计学所特有的研究方法。所谓大量观察法,就是指对所研究事物的全部单位或足够数量的个体进行观察以取得数据的方法。在现象总体中,个别单位往往受偶然因素的影响,如果只对其中之一或很少个体进行观察,其结果不足以代表总体的特征,只有观察全部单位或足够的单位并加以综合,影响个别单位的偶然因素才会相互抵消,使现象的一般特征显现出来。例如,要了解一个地区某一时期出生婴儿的性别比例,如果只对一两户或很少几户家庭进行观察,结果可能是女婴占很大比例或男婴占很大比例,这不能反映该地区的客观实际。只有对该地区的所有居民户或足够多的居民户进行调查,才能消除偶然性的影响,比较真实地反映出生婴儿的性别比例。再如,在掷硬币实验中,每掷一次硬币只有两种结果:正面朝上和反面朝上。实验次数越多,正面朝上(或反面朝上)的概率就越接近于50%。通过大量观察,一方面,可以掌握认识事物所必需的总体的各种数量;另一方面,还可以通过个体离差的相互抵消,在一定范围内排除某些个别现象或偶然因素的影响,从数量上反映总体的本质特征。在我国的统计实践中,大量观察法在多种统计调查中被广泛运用,如全面统计调查、普查、重点调查和抽样调查等。当然,在统计观察和分析中,也常常对个别典型单位进行深入细致的研究,但是,它的最终目的仍然是说明总体的本质特征。

大量观察法的数学依据是大数定律,大数定律认为,在观察过程中,每次取得的结果不同,这是由偶然性造成的,但大量、重复观察结果的平均值几乎接近确定的数值。大数定律的本质意义在于,经过大量观察,个别偶然的差异相互抵消,而必然的、集体的规律性便显示出来。

#### (二) 统计分组法

统计分组法是研究内部差异的重要方法。通过分组可以研究总体中不同类型的性质以及它们的分布情况,如产业的经济类型及其行业分布情况;可以研究总体中的构成和比例关系,如三次产业的构成、生产要素的比例等;可以研究总体中各种现象之间的依存关系,如企业的经营规模和利润率之间的关系。

社会经济现象是十分复杂的,具有多种多样的类型。从数量方面认识事物不能离开对事物的分析,将所研究的现象总体区分为不同性质的组成部分是统计进行加工整理和深入分析的前提。例如,要研究工业行业结构及其对国民经济的影响,就必须首先把工业区分为冶金、电力、煤炭、石油、化工、机械、建材、食品、纺织、造纸等若干部门,然后分别调查和分析各个部门的产量、固定资产、能源消耗、资金占用、利润及职工工资总额等方面的情况;要研究改革开放以来我国经济结构特别是非公有制经济的发展情况,就应选择所有制作为分组标志进行分组。

统计分组法在整个统计工作研究过程中具有重要意义,贯穿于统计工作的全过程。统计调查离不开分组,在对统计资料进行加工整理的过程中,统计分组更是关键的环节,统计指标和指标体系是统计分析的基本工具,在统计分析中,综合指标的应用更是建立在统计分



组的基础之上。

### （三）综合指标法

综合指标法是指运用各种统计指标来反映和研究客观现象的一般数量特点和现象之间数量关系的方法。综合指标可以说明现象的规模、水平、集中趋势和离中趋势,说明现象发展变化的趋势和规律。综合指标法是应用十分广泛的一种统计分析方法。

在统计实践中,总量指标、相对指标、平均指标等综合指标被广泛应用,这些指标分别从静态和动态上综合反映与分析现象总体的规模、水平、结构、比例和依存关系等数量特征与数量关系。综合指标和统计分组是密切联系、相互依存的。统计分组如果没有相应的综合指标来反映现象的规模水平,就不能揭示现象总体的数量特征;而综合指标如果没有科学的统计分组,就无法划分事物变化的数量界限、掩盖现象的矛盾,成为笼统的指标。所以在研究社会经济现象的数量关系时,必须科学地进行分组,合理地设置统计指标,统计指标体系和统计分组体系应该相互适应。综合指标法和统计分组法是结合起来应用的。

### （四）统计推断法

统计推断法就是以一定的置信水平,根据样本的数据,推断总体数量特征的归纳推理的方法。在实际中,由于时间、经费、人员等因素的限制,我们只可能对研究对象中的一部分单位或个体进行观察,根据取得的数据,用局部的观察结果对总体的特征做出估计。例如,要了解一批灯管的使用寿命是否达到规定的照明时数,我们只能从中抽出一部分检查,以检查的结果推断该批灯管的情况。统计推断法既可以用样本的数据对总体参数进行估计,也可以研究某一现象与另一些现象的依存关系,还可以用作对总体的某些分布特征的假设检验。从某种意义上说,统计所观察的资料都是一种样本资料,因而统计推断法也就广泛地应用于统计研究的许多领域,如建立统计模型存在模型参数的估计和检验问题,根据时间序列进行预测存在原序列的估计和检验问题。因此,可以说统计推断法是现代统计学的一类基本方法。

我们常将以上几种统计研究方法分为两大类,即描述统计和推断统计。描述统计是指对实验或调查得到的数据进行登记、审核、整理、归类,计算出各种能反映总体数量特征的综合指标并加以分析,从中抽出有用的信息,用表格或图像表示出来。描述统计通过对分散无序的原始资料进行整理归纳,运用统计分组法和综合指标法得到现象总体的数量特征,揭露客观事物内在数量规律性,达到认识的目的。大量观察法、统计分组法和综合指标法都属于描述统计法。推断统计则是以一定的置信标准要求,根据样本数据来判断总体数量特征的归纳推理的方法。推断统计是逻辑归纳法在统计推理中的应用,所以又称为归纳推理的方法,具体包括参数估计、假设检验、方差分析和相关回归分析方法。

## 四、统计学研究方法的特点

### （一）定性认识与定量认识相结合

虽然统计属于定量认识的范畴,但统计是研究大量社会经济现象的总体数量特征,并不是从定量认识开始的,而是从定性认识开始的。要从数量方面认识现象总体的特征,就必须确定总体的内涵和范围,进行统计分组,设置统计指标和指标体系。这就首先要对统计分组的标准和方法以及统计指标的性质和口径有明确而具体的认识。所以,离开定性认识,定量认识就失去了可以依据的基础和方向,没有定性认识,就不会有真正的定量认识。可以说,



对事物仅仅停留在定性认识阶段还远远不够,也可以说,统计的定量认识是定性认识的深化和具体化。

### (二) 从个体认识到总体认识

统计的最终目的是要认识现象的总体特征,然而它是从认识个体特征开始的,对个体特征的调查、了解和反映是统计研究的基础,但它又不停留在个体特征认识上,而是通过归纳个体特征综合概括出总体特征,最后达到对现象总体规律性的认识。

### (三) 从已知量的描述到未知量的推断

统计总是对已经存在的事实进行观察和调查,并描述现象在具体时间、地点、条件下的数量表现。但统计的目的通常是要根据已知的数据去推断所关心的未知数量或情况。例如,根据已知的样本资料推断未知的总体数量特征,根据已知的资料推断未来的发展趋势,根据已知的一方面的资料推断另一方面的相关情况等等。从描述统计到推断统计,是统计认识的延伸和拓展。

在运用统计研究方法时,还必须根据实际情况,按照需要与可能,分别采用不同的统计方法,多种统计方法结合运用、相互补充。

## 五、统计学与其他相关学科的关系

### (一) 统计学与数学

统计学与数学都是研究数量规律的,都要利用各种公式进行运算。数学中的概率论为统计学提供了数量分析的理论基础。在统计学解决实际问题的步骤中,涉及数据的特征描述环节会用到数学的一些公式和结论,但是基本上不需要数学推导和证明。统计学中的理论以抽象的数量为研究对象,其大部分内容也可以看作数学的分支。

统计学与数学的主要区别在于:首先,从研究对象看,数学以最一般的形式研究数量的联系和空间形式,统计学特别是应用统计学则总是与客观的对象联系在一起。其次,从研究方法看,数学主要是逻辑推理和演绎论证的方法,而统计的方法,本质上是归纳的方法,统计学家特别是应用统计学家需要深入实际,进行调查或实验去取得数据,研究时不仅要运用统计的方法,而且要掌握某一专门领域的知识。

### (二) 统计学与经济学

统计学与经济学有密切的联系。经济学涉及大量经济数据,而统计学是开展经济学研究不可或缺的重要工具。经济学等实质性学科对经济统计学起着重要的指导作用。统计学家未必是经济学家,经济学家也未必是统计学家。但经济统计学家应当既是统计学家又是经济学家。统计学与经济学的区别可从定义和特点等方面进行比较。

### (三) 统计学与相关实质性学科

实质性学科研究该领域现象的本质关系并对有关规律做出合理的解释和论证。而统计学只是为实质性学科研究和认识数量规律提供专门的方法和工具,并不直接对规律产生的原因和机理做进一步的分析。在统计学解决实际问题的步骤中,需要与相关实质性学科的知识相结合。统计的实证研究,可以帮助人们认识相关实质性学科的数据规律,同时检验相关实质性学科理论的真实性和完善程度。相关实质性学科对统计学应用起着重要的指导作



用。统计指标体系的设定离不开实质性学科的问题,应用统计方法也在很大程度上受所研究对象性质的影响。

## 六、统计工作的基本流程

人类认识世界的活动是一个由现象到本质、由矛盾的特殊性到普遍性、由感性认识到理性认识的不断深化的过程。统计活动也不例外,从具体的统计认识活动来看,统计工作主要由统计设计、统计调查、收集数据、统计整理、统计分析和统计资料的开发与应用六个环节组成。

### (一) 统计设计

统计设计是指根据统计研究对象的性质和研究目的,对统计工作各方面和各环节所做的通盘考虑与安排,它的结果表现为各种标准、规定、制度、方案和办法,如统计分类标准、目录、统计指标体系、统计报表制度、统计调查方案、普查办法、统计整理和汇总方案等。

统计工作是一项高度集中统一和科学性很强的工作。在具体工作中,根据所要研究问题的性质,在有关学科理论的指导下,制定统计指标、指标体系和统计分类,给出统一的定义和标准,同时提出收集、整理和分析数据的方案和工作进度等。要做好统计设计,不仅要以统计学的一般理论和方法为指导,而且要求设计者对所研究的问题本身具有深刻的认识和相关的学科知识。统计设计的主要内容有统计指标和指标体系的设计、统计分类和统计分组的设计、统计表的设计、统计资料收集方法的设计、统计工作各个部门和各个阶段的协调与联系以及统计力量的组织与安排。

### (二) 统计调查

统计调查是根据统计方案的要求,采用各种调查组织形式和调查方法,有组织、有计划地对所研究总体的各个单位进行观察和登记,准确、及时、系统、完整地收集统计原始资料的过程。统计调查是统计认识活动由定性认识过渡到定量认识的阶段,这个阶段所收集的资料是否客观、全面、系统、及时,直接影响到统计整理的好坏,关系到统计分析结论的正确性,决定着整个统计工作的质量。所以,统计调查是整个统计工作的基础。

### (三) 收集数据

统计数据的收集有实验法和调查观察两种基本方法。对于大多数自然科学和工程技术研究来说,如果有可能通过有控制的科学实验去取得数据,则可以采用实验法。对于社会经济现象来说,一般无法进行重复实验,要取得有关数据,就必须进行调查观察。

### (四) 统计整理

统计整理是根据统计研究的目的和任务,对统计调查阶段所取得的原始资料进行审核、分组和汇总,将分散的、零星的、反映总体单位特征的资料转化为反映各组总体数量特征的综合资料的过程。统计整理是将对总体单位特征的认识过渡到对总体数量特征的认识的桥梁和纽带,它既是统计调查的继续,又是统计分析的必要前提,在统计工作中,处于中间环节,起着承上启下的作用。

### (五) 统计分析

统计分析是指在统计调查和统计整理的基础上,用科学的分析方法,对所研究的现象总





体进行全面、系统的数量分析,认识和揭示事物的本质和规律性,进而向有关单位和部门提出咨询建议,以及进行必要的分析、预测的统计工作过程。统计分析是统计工作的最后阶段,也是统计发挥信息咨询和监督职能的关键阶段。

#### (六) 统计资料的开发与应用

对于已经公布的统计资料,需要加以积累,也可以进行进一步的加工,结合相关的实质性学科的理论知识去分析和利用。如何更好地将统计数据 and 统计方法应用于各自的研究领域是应用统计学研究的一个重要方面。

## 模块三 统计学的基本概念

在社会经济统计工作活动中,对事物的定量描述一般是通过定性认识开始的,并且在定性认识的基础上,运用一些专门的概念,通过一系列桥梁和纽带,从定性认识逐步过渡到定量的认识和描述,这些起到桥梁和纽带作用的统计量便是统计指标和统计指标体系等。在论述统计学的理论与方法的过程中,熟悉这些概念对于掌握统计学的学科体系具有重要意义。

### 一、统计总体与总体单位

统计指标是社会经济统计活动和社会经济统计学中最重要的基本概念。统计正是用统计指标来反映总体的实际情况,并用统计指标来研究认识总体的发展变化情况、总体内部以及它与外部的数量关系。在社会经济统计中,统计指标占有中心地位,许多统计方法都是围绕统计指标而产生的。在统计工作中,统计指标大体上可以分为两类:一类是说明总体数量特征的指标名称,另一类是说明总体数量特征的指标数值。

#### (一) 统计总体

统计总体,简称总体,是指客观存在的在同一性质基础上结合起来的许多个别事物的整体。凡是客观存在的,至少在某一性质相同的基础上结合起来的许多个别事物所组成的整体,都可以称为统计总体。例如,要研究全国农村居民的可支配收入情况,就需要将全国的农村居民作为一个总体,所有的统计活动都要围绕着全国农村居民这一特定的对象来进行。再如,要研究我国的人口分布状况,就应将全国人口看作一个总体,从普查方案的设计到普查登记、资料汇总,再到最后公布普查数据等环节,都要围绕这一对象来进行。统计总体应具有以下三个特点。

##### 1. 同质性

同质性是指构成统计总体的各个单位必须在某些方面而且至少在某一个方面具备共同的性质。同质性是构成总体的前提。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,所确定的总体也不同,其同质性的意义也随之变化。例如,在国家实施的精准扶贫活动中,研究城镇居民贫困户的生活状况,贫困线下的城镇居民户则构成了统计总体,贫困线下的城镇居民户是同质的,而贫困线上的城镇居民户是非同质的。



## 2. 大量性

大量性是指统计总体是由许多总体单位组成的,只有一个单位的总体是不存在的。当然,研究目的不同,统计总体就不同,总体中所包含的总体单位的数量也就不同,一个总体究竟包含多少总体单位,最终取决于统计研究的目的。

## 3. 变异性

变异就是事物之间的差异或不同。从统计研究的角度来说,变异性是指构成总体的各个单位之间存在的差别。例如,我国目前有 34 个省级行政区,有 56 个民族,一个特定的省份包含 18 个地级市等。

### (二) 总体单位

总体单位,简称单位,是组成总体的各个个体。根据研究目的的不同,单位可以是人、物、机构等实物单位,也可以是一种现象或活动等非实物单位。例如,某市高等院校中的每一所高校、该市所有高校在校生的每一名学生也是总体单位。

### (三) 总体与总体单位的关系

总体与总体单位的概念是相对而言的,二者之间的关系是整体与个体、集合与元素的关系。总体与总体单位的具体形式随着统计研究目的的不同而不同,可以是人,也可以是物,还可以是组织(企业或家庭)或时间、空间、行为等。总体与总体单位的关系不是一成不变的,二者可以相互转化。在一定的研究目的下,一个事物可以作为总体而存在,然而当研究目的发生变化后,这个事物就可能成为总体单位。例如,当研究某省高等院校的办学状况时,该省所有的高等院校就是一个总体,其中的每一所高校就是一个总体单位;当研究某一所高校的办学状况时,这所高校就成了总体;当研究全国的高等院校办学状况时,全国的高等院校便组成总体,而该省的高等院校又变成总体单位。

根据总体所包含的单位数量,可以将总体分为有限总体和无限总体两类。有限总体是指总体中的总体单位数可以计数或穷尽的总体,如一个企业的全体职工、一个城市的全部人口等都是有限总体。如果总体中的单位数是一个无穷大量,或者准确地度量它的单位数是不经济或没有必要的,这样的总体称为无限总体,如在连续生产的生产线上产出的全部零件数、一片草原上的小草数量、海洋里的各类鱼的数量,等等。对于有限总体,既可以进行全面调查,也可以进行非全面调查。但对于无限总体,只能抽取一部分单位进行非全面调查,据以推断总体。

## 二、样本和标志

### (一) 样本

当总体单位数量很多甚至无限时,不必要或不可能对构成总体的所有单位都进行调查。这时,需要采用一定的方式,从由作为研究对象的事物全体构成的总体(又称母体)中,抽取一部分单位,作为总体的代表加以研究。这种由总体的部分单位组成的集合称为样本(又称子样)。样本也由一定数量的单位构成,样本所包含的总体单位数称为样本容量。

### (二) 标志

统计是从对个体的观察开始,逐步过渡到对总体数量特征的认识的。总体中各单位普



遍具有的属性或特征称为标志。例如,每个工人都具有性别、工种、文化程度、技术等级、年龄、工龄、工资等属性和特征,这些就是工人作为总体单位的标志。标志分为品质标志和数量标志两种。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、语言来描述,如工人的性别。数量标志表明单位数量方面的特征,可以用数值来表现,如年龄、身高、体重等。

如果一个总体中各单位有关标志的具体表现都相同,则称之为不变标志。例如,在工人这一总体中,职业是不变标志。在一个总体中,当一个标志在各单位的具体表现有可能不同时,这个标志便称为变异标志。例如,各工人的工龄可能表现不同,因而是变异标志。至少要有一个不变标志,才能够使各单位结合成一个总体。不变标志是总体同质性的基础,作为总体,还必须存在变异标志,这表示所研究的现象在各单位之间存在差异,才需要进行统计研究。

### 三、统计指标与统计指标体系

#### (一) 统计指标

##### 1. 统计指标的概念

统计指标简称“指标”,是反映社会经济总体现象数量特征的概念和数值。一个完整的统计指标包括指标名称和指标数值两个部分,如 2019 年我国国内生产总值为 990 865 亿元<sup>①</sup>。指标名称反映一定的社会经济范畴,指标数值是根据指标名称的内容所计算的统计数字,同一名称的指标在不同时间和地点可以表现为不同的指标数值。在实际工作中,人们有时只把指标名称称作指标,而不包括指标数值。

##### 2. 统计指标与标志的关系

(1) 统计指标和标志的区别。统计指标和标志的区别主要在于两点:首先,指标和标志的概念明显不同,指标是说明总体的综合数量特征的,具有综合的性质;标志是说明单位属性的,一般不具有综合的特征。其次,统计指标分为数量指标和质量指标,它们都是可以用数量来表示的。标志分为数量标志和品质标志,它们并不是都可以用数量来表示,品质标志只能用文字表示。

(2) 统计指标和标志的联系。统计指标数值是由各单位的标志值汇总或计算得来的。数量标志可以综合为数量指标和质量指标,品质标志只有对它的标志表现所对应的单位加以总计才能形成统计指标。

根据研究目的不同,指标与标志之间可以互相转化。二者体现这样的关系:指标在标志的基础上形成,指标又是确定标志的依据。例如,如果要研究的是全国工业企业的情况,则各企业的职工人数、固定资产、工业增加值等都是总体单位(即各个企业)的标志;如果变成研究某一企业的职工状况,则该企业就变成一个总体,企业职工人数则变成统计指标,每个职工的文化程度、技术等级、性别等就成为标志。

##### 3. 统计指标的分类

指标按其所反映总体现象的内容和特点,可分为数量指标和质量指标。数量指标是反映总体规模大小的各种总量指标,一般是把总体单位个数加总或把总体单位的某个标志值

<sup>①</sup> 数据来自国家统计局网站。



加总计算出来的,如人口总数、工业企业数、职工人数、工农业总产值、商品进出口总额等。这些指标反映现象或过程的总规模和水平,所以也称为总量指标,用绝对数来表示。质量指标是说明总体内部或总体之间数量关系的指标,通常是由两个有联系的指标对比计算出来的,如职工平均工资、人口密度、工人出勤率、劳动生产率、单位产品成本、产品合格率、工时利用率、单位面积产量等。质量指标是总量指标的派生指标,用相对数或平均数来表示,以反映现象之间的内在联系和对比关系。

## (二) 统计指标体系

社会经济现象是一个复杂的总体,各类现象之间存在相互依存、相互影响的关系。一个统计指标往往只能反映复杂现象总体某一方面的特征,要了解客观现象在各个方面及其发展变化的全过程,仅靠单个的统计指标是不行的,必须要建立和运用统计指标体系。

所谓统计指标体系,就是指若干个反映社会经济现象数量特征的相对独立又相互联系的统计指标所组成的整体。例如,一家民营企业把产品产量、净产值、劳动生产率、产品质量、消耗、成本、销售收入等统计指标联系起来就组成了指标体系。这便于人们全面、准确地评价该企业的生产经营情况。

统计指标体系按其作用不同,可分为描述性指标体系、评价性指标体系和决策性指标体系。描述性指标体系主要反映社会经济现象的现状、运行过程和结果;评价性指标体系主要是比较、判断社会经济现象的运行过程及结果是否正常;决策性指标体系是为了保证社会、经济、科技等方面的有序、协调发展。

## 知识拓展

### 国内生产总值与国民生产总值

国内生产总值(gross domestic product,GDP)是指一个国家或地区的所有常住单位在一定时期内生产活动的最终成果,常被公认为是衡量国家经济状况的最佳指标。GDP是核算体系中一个重要的综合性统计指标,也是我国新国民经济核算体系中的核心指标,它反映了一国或地区的经济总体规模和经济结构。

2020年1月17日,国家统计局发布数据,经初步核算,2019年中国国内生产总值为990865亿元,按可比价格计算,比上年增长6.1%,符合6%~6.5%的预期目标。分季度看,一季度同比增长6.4%,二季度同比增长6.2%,三季度同比增长6.0%,四季度同比增长6.0%。

国民生产总值(gross national product,GNP)是一个国民概念,是指一个国家或地区所有常住单位在一定时期内所获得的初次分配收入总额。举例说明:一个在日本工作的美国公民所创造的财富计入美国的GNP,但不计入美国的GDP,而是计入日本的GDP。在1991年之前,美国均是以GNP作为经济总产出的基本测量指标,后来因为大多数国家都采用GDP,加之国外净收入数据不足,相对于GNP,GDP是衡量国内就业潜力更好的指标,易于测量,所以美国才改用GDP。

20世纪80年代初,中国开始研究联合国国民经济体系的GDP指标。1985年,中国开始建立GDP核算制度。1993年,中国正式取消了国民收入核算,GDP成为国民经济核算的核心指标。



## 四、变异与变量

### (一) 变异的概念

当各个总体单位在某一品质标志上的具体表现都相同时,则为不变标志。如某市的民营企业,这些企业的性质都是民营,因此,对该市的每一家民营企业来说,其企业性质这个品质标志的具体表现都是相同的。

当各个总体单位在某一品质标志上的具体表现不相同或不完全相同时,则为可变标志,这个可变标志就称为变异。因此可以说,变异是指可变的品质标志。例如,把所有城市的民营企业作为一个统计总体,那么,“厂址”和“企业类型”等就是变异。

### (二) 变量的概念

说明现象的某一数量特征的概念也被称为变量,变量的具体取值是变量值,统计数据就是统计变量的具体表现。例如,固定资产是一个变量,各企业固定资产的具体数值是变量值。为了区别,在本书中,凡是变量均用大写的英文字母表示,而变量值则用小写英文字母表示。

### (三) 变量的分类

#### 1. 按其取值是否连续分类

按其取值是否连续,变量可分为连续型变量和离散型变量。

连续型变量是指变量的取值在数轴上连续不断,无法一一列举,即在一个区间内可以取任意实数值,如气象上的温度与湿度、零件的尺寸等。

离散型变量是指变量的取值是整数值,可以一一列举,如某省内的企业数、职工人数等。

#### 2. 按其所受因素影响的不同分类

按其所受因素影响的不同,变量可分为确定性变量和随机变量。

确定性变量是受确定性因素影响的变量,即影响变量值变化的因素是明确的,是可解释和可控制的。

随机变量则是受许多微小的不确定因素(又称随机因素)影响的变量。变量的取值无法事先确定。社会经济现象既有确定性变量也有随机变量。统计学研究的主要是随机变量。

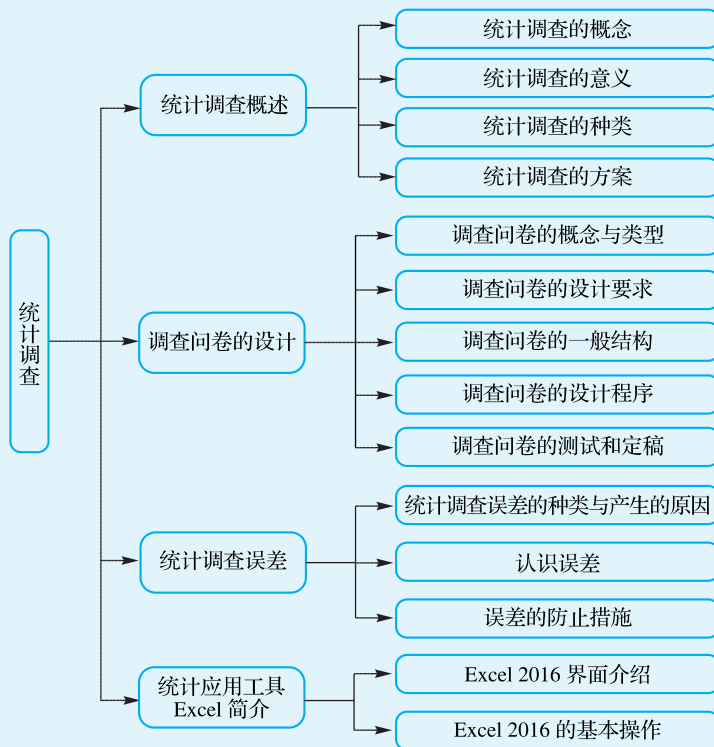
# 项目二

## 统计调查

### ● 学习目标

- 了解统计调查的概念、意义和种类；
- 掌握统计调查方案的内容和方法；
- 了解调查问卷的概念和类型；
- 掌握调查问卷的基本结构和设计程序；
- 了解统计调查误差的概念、种类以及防止登记性误差的措施；
- 了解 Excel 工作界面；
- 掌握 Excel 的基本操作。

### ● 项目导航





## 案例导入

### 您好，我是第四次全国经济普查调查员

“普查员能较快完成数据普查，一方面，与被调查对象对经济普查认识到位、积极配合有关；另一方面，更得益于扎实、充分的普查准备工作。”武汉市江岸区统计局工作人员李某说，经济总量大，单位数量多，涉及领域广，经济业态丰富，跨行业、异地经营情况复杂。自接受第四次全国经济普查国家综合试点任务以来，江岸区抓紧推进各项准备工作。

此次经济普查试点工作，是为了检验普查方案的科学性和可行性，为2019年1月正式开展的全国经济普查入户登记提供经验，做好准备。

在此次试点前期清查摸底环节，江岸区共对单位名录底册中的5万多条法人、产业活动单位、个体经营户信息进行实地核查。一张调查表最少涉及7个指标，最多有177个指标，绝大多数指标在70个左右，光指标解释就有80多页；行业分类共计20多个门类、97个大类、473个种类、1380个小类；试点工作必须在一个多月的时间内全面完成，时间紧迫、任务艰巨。

自2018年5月4日起，江岸区1000多名普查员开始走街串巷，按建筑物对企业法人、产业单位逐一摸清；对个体户、专业市场逐一登记，对夜市摊位进行拉网式登记，确保不漏巷、不漏楼、不漏门、不漏户。

在普查登记阶段，报表多、指标复杂、内容综合，对数据质量要求更高，直接关系到整个普查工作的成败。为确保普查登记工作顺利推进，江岸区进一步加强了培训指导，对普查登记所涉及的重点调查表式进行了详细讲解。而在普查过程中，有的普查对象对经济普查认识不到位，特别是财务资料不健全的小规模企业和个体户，不认真回答普查员的询问，普查员只能多次上门宣传沟通，努力获得企业的理解与支持。

数据质量是普查工作的生命，在保持工作进度的同时，江岸区严把数据质量关，不仅邀请相关专家加强培训指导，还派出专业人员陪同普查员入户，帮助普查员审核报表，协助做好登记工作。此外，在报表审核中，采取区、街、社区分级审核机制，对普查表中指标间逻辑关系、指标漏填情况及主要经济指标填写情况层层审核把关，并重点关注重要指标、新增指标和易错易漏指标，做到不漏项、不缺项、不存在明显逻辑错误，确保上传到平台的数据真实、完整、准确。

国家统计局有关负责人表示，经济普查是对中国经济家底的一次大盘点。只有尽可能获取更为翔实的信息，提高普查信息质量，准确摸清查实各类单位的基本情况和主要产品产量、服务活动等，才能为加强和改善宏观调控、深化供给侧结构性改革、科学制定中长期发展规划、推进国家治理体系和治理能力现代化，提供科学准确的统计信息支持。

资料来源：搜狐网，[http://www.sohu.com/a/234161194\\_118392](http://www.sohu.com/a/234161194_118392)，有删改。

为了完成统计工作的任务，实现统计研究的目的，必须根据统计设计的规划，深入研究社会经济现象总体的每个单位，收集数据，摸清情况，对研究对象进行统计调查。



## 模块一 统计调查概述

### 一、统计调查的概念

统计调查是统计工作过程的第二个阶段,它是根据统计任务和统计设计的要求,采用科学的调查方法,有计划、有组织地向调查单位收集统计资料的工作过程。统计调查收集的资料有两类:一类是未做任何加工整理的原始资料,又称初级资料;另一类是次级资料,即已经经过加工整理过的,可以在一定程度上说明总体情况的统计资料。

### 二、统计调查的意义

当今社会是信息社会。哪个企业信息掌握得迅速、准确、可靠,生产计划安排得当,产品更新换代快、适销对路,哪个企业的经济效益就好。人们常说,信息就是金钱,就是生命。信息从哪里来?只能从统计调查中来。

世界上许多国家和企业都很重视调查与研究,还成立了不少专门的组织和机构。我国不少企业也越来越重视调查和研究,有的还有专门的组织和机构。例如,某电冰箱厂在市场竞争非常激烈、商品供大于求、销售越来越困难的情况下,通过调查了解到农村消费者对电冰箱冷冻的要求比较高,而且有夏季购买集中的特点,便提高电冰箱的冷冻效果,加大冷冻空间,并派销售人员到较富裕的农村进行现场介绍,免费提供上门安装和维修服务,使产品销售量迅速增加。但是,有一些企业至今还不重视调查,甚至盲目生产,最后导致偏离市场,缺乏竞争能力,从而陷入困境。

客观现象是错综复杂的,人们要系统、准确地认识事物的发展变化规律,必须客观实际地收集资料并进行统计调查。统计调查在统计工作中占有重要地位,它是统计整理和统计分析的基础,影响着后两个阶段工作的进行,直接关系到整个统计工作的质量。只有做好统计调查,才能保证统计工作达到对于客观事物性的认识,从而预测未来。因此,统计调查活动要求收集来的统计资料必须是准确、及时和全面的。可以想象,如果调查得来的资料不准确、不及时、不全面,就不能客观反映情况和正确认识现象总体的本质特征,最终也就难以对所研究现象总体的规律性做出明确判断,甚至还会导致错误的结论。

### 三、统计调查的种类

社会经济现象错综复杂,根据不同的调查对象和调查目的,需要采用不同的统计调查方式和方法。统计调查按不同的分类标准,可分成多种相应的种类。不同分类标准下所形成的各种统计调查,具有不同的特点和作用。

#### 1. 按调查对象的范围不同分类

按调查对象的范围不同,统计调查可分为全面调查和非全面调查。

全面调查是对被调查对象中的全部单位进行调查,目的是取得总体的全面、系统和完整的总量资料。例如,要了解全国人口状况,就要对全国所有人口都进行调查。全面调查要耗费大量的人力、物力、财力和时间。





非全面调查是对调查对象总体的一部分单位进行调查。例如,为了了解乡镇企业的经济效益状况以及经营的新情况、新问题,可不必对全部乡镇企业一一进行调查,只需选择部分乡镇企业进行调查即可。重点调查、抽样调查、典型调查都属于非全面调查的范畴。

### 2. 按调查的组织形式不同分类

按调查的组织形式不同,统计调查可分为统计报表和专门调查。

统计报表是按一定的表式和要求,自上而下统一布置,自下而上提供统计资料的一种统计调查方式。我国的有关统计报表制度规定,我国所有企事业单位和机关单位都应按照规定的表式、项目、程序、时间等,向上级领导机构提供相关报表。统计报表是我国统计工作中的一项重要内容。

专门调查是为了某一特定目的而专门组织的统计调查,如普查、抽样调查、重点调查、典型调查等。

### 3. 按调查登记时间连续与否分类

按调查登记时间连续与否,统计调查可分为经常性调查和一次性调查。

经常性调查是指随着调查对象的不断发展变化,随时进行连续不断的登记或观察,其目的是取得事物在一定时期内全部发展过程的总资料。例如,企业生产所需要的各种原材料、燃料、动力的消耗量,企业产品的产量,商品零售企业的存货等,这些指标的数值变动很频繁,必须进行经常登记,才能满足企业进行适时管理控制的要求。经常性调查能够反映事物在一定时期内发展过程的全貌。

一次性调查是不连续登记的调查,它是指对事物每隔一段时期后在一定时点上的状态进行登记。例如,土地面积、生产设备的数量调查、人口普查等,这些指标的数值或者是在一定时期内变动不大,或者是经常调查难度较大,往往要用一次性调查的方式收集资料。一次性调查的目的是获得事物在一定时点上的水平状态资料,既可以是定期的,也可以是不定期的。

### 4. 按收集资料的方式不同分类

按收集资料的方式不同,统计调查可分为直接观察法、报告法、采访法和问卷调查法。

(1) 直接观察法。直接观察法是指调查人员亲自到现场对调查对象直接进行观察、检验、测定、计量等,以取得统计资料的一种调查方法。直接观察法可以获得大量真实的第一手资料,是获得感性认识和发现问题的重要途径,能够保证所收集资料的准确性。例如,为了及时了解农作物产量而进行农产品抽样调查时,调查人员亲自参加抽选样本、实割实测、脱粒、晾晒、保管和过秤计量。直接观察法的局限性有:涉及私人的一些问题不宜采用;只能在有限的范围内使用;花费的人力、物力和时间较多。

(2) 报告法。报告法是由被调查单位利用各种记录和核算资料,按照统一的要求和表格形式,向有关部门提供统计资料的方法。现在我国各企事业单位、各机关向上级填报统计报表的方法就是报告法。这种方法由于有原始启示录、统计台账或有关核算资料为依据,因此只要记录完整可靠,符合制度规定,所获得的资料就有较高的准确性。

(3) 采访法。采访法是由调查人员向被调查者提问,根据被调查者的答复取得调查资料的方法。采访法又分为个别访问和开调查会两种。个别访问是指通过调查人员向被调查者逐一访问收集资料的方法,这种方法可以了解被调查者的真实想法,不受其他人意见的影响。开调查会是由调查人员召集相关的被调查者,以座谈会的形式,按一定的调查提纲进行



调查并收集资料的一种方法。由于参加单位较多,会议时间有限,因此开调查会前一定要认真做好准备,要明确会议的主题,否则就不能取得好的效果。

(4) 问卷调查法。问卷调查法是为特定目的,以问卷形式提问,发给被调查者,由被调查者根据实际情况自愿、自由回答的一种采集资料的方法。要科学地进行问卷调查,必须精心设计调查问卷,问题要简明扼要,被调查者填写答案不需要花费多少时间,保证做到为被调查者保密。在实施中,要尽量防止提出回答率或答案质量不高的问题。

随着科学技术的不断进步以及卫星遥感技术和互联网技术的广泛传播与应用,统计调查的手段也越来越先进,与此相适应,这些方法在统计调查中也越来越被重视。取得统计调查资料的方法取决于统计研究对象的特点。根据统计调查的内容和被调查者的具体条件,可以选择最适宜的取得资料的方法,有时根据需要,还可以同时应用几种方法。

#### 四、统计调查的方案

统计调查工作量大、涉及面广、调查项目多,是一项复杂、细致的工作。因此,无论采用什么样的方式收集资料,都要事先根据需求和可能对被调查对象进行定性分析。制订一个正确的调查方案是保证统计调查有计划、有组织地进行的前提,是准确、及时、系统、完整地取得调查资料的重要条件,也是统计设计调查阶段的具体化。一个完整的统计调查方案,应该包括以下几个方面。

##### 1. 确定调查的目的和任务

不同的研究目的和任务,决定着不同的调查内容和范围,目的不明、任务不清就无法确定向谁调查、调查什么及采用什么方式进行调查,整个调查工作就会陷入盲目混乱,造成人力、物力和财力的浪费。统计调查的目的和任务可以根据党的方针政策、党政领导提出的任务要求来确定,也可以根据企事业单位经营、管理及其他工作需要来确定。

##### 2. 确定调查对象和调查单位

调查对象是在统计调查中所要进行统计研究的总体范围,它是由许多性质相同的调查单位组成的。例如,2010年人口普查规定,人口普查对象是具有中华人民共和国国籍并在中华人民共和国境内常住的人。再如,要对某地区中学学生情况进行普查,则调查对象就是该地区各中学的全部学生。

调查单位是指在被调查总体中被记录其标志的单位。调查单位的确定取决于调查目的和调查对象。例如,工业调查的目的是了解工业企业的生产经营情况,调查单位就是每一个工业企业。这里就需要把工业企业和农业企业、建筑业企业、商业企业、运输业企业等其他企业区别开来,即划清总体界限,然后明确工业企业的含义。再如,如果调查目的在于取得国有大中型企业的产品产量、产值、生产成本等资料,那么,所有国有大中型企业就是调查对象,而构成国有大中型企业这个总体的每个国有企业则是调查单位。在明确调查单位的同时,要注意将调查单位和填报单位区别开来。填报单位是指负责向上级汇报调查内容、提交统计资料的单位,与调查单位有时一致,有时不一致。例如,在工业企业普查中,每个工业企业既是调查单位又是填报单位;在工业企业设备调查中,每一台设备是调查单位,而每一个工业企业是填报单位。

##### 3. 确定调查项目,制作调查表

调查项目就是调查中所要登记的调查单位的特征,这些特征在统计上又称为标志。确



定调查项目就是明确向调查单位调查什么问题。反映调查单位特征的标志是多种多样的,在调查中确定调查项目,应根据调查目的和调查单位的特点而定。确定调查项目时,应注意取得资料的必要性与可能性、项目的统一性、项目间的衔接性和时间上的可比性。

调查表就是把调查项目按一定的逻辑关系顺序以表格形式列示出来,所要解决的问题是明确向调查单位调查的内容。调查表是整个调查方案的核心部分,是收集原始资料的基本工具,利用调查表进行调查非常便利,它不仅能够条理清晰地登记所需资料,而且便于以后的整理和汇总。

调查表一般由表头、表体和表脚组成。表头是用来表明调查表的名称以及填写调查单位的名称、性质、隶属关系等;表体是调查表的主要部分,包括统计调查所要说明的社会经济现象的项目和这些项目的具体表现,也就是数字、计算单位等;表脚包括调查者的签名和调查日期等,以便明确责任,若发现问题,便于查询。

调查表的形式一般有一览表和单一表两种。一览表是指在一张表上登记若干个单位的调查资料,它的调查项目不宜过多,使用这种表不仅节省人力、物力,而且一目了然;单一表是指在一张表上只登记一个调查单位的调查资料,它可以容纳较多的调查项目,适用于较详细的统计调查。

#### 4. 确定调查时间和调查期限

调查时间是指调查资料所属的时间,在统计调查中,如果所调查的是时期现象,就要明确规定调查资料所反映的起止日期,所登记的资料就应包括该时期第一天到最后一天的累计数字。例如,调查2012年第一季度的钢铁产量,则调查时间是1月1日—3月31日共3个月。如果所调查的是时点现象,调查时间就是规定的统一标准时点。例如,我国第六次人口普查的标准时间是2010年11月1日零时。

调查期限是指整个调查工作的起始至结束的时间,包括收集资料和报送资料的整个工作所需时间。例如,某管理局要求所属企业在2015年1月底上报2014年工业总产值资料,则调查时间是2014年1月1日至2014年12月31日,上报的总产值为这一年的总产值,调查期限是一年。

#### 5. 确定调查工作的组织实施计划

为了保证整个统计调查工作的顺利进行,在调查方案中需要一个考虑周密的组织实施计划,它的主要内容包括:调查工作的组织领导机构和调查人员的组织;调查的方式和方法;调查前的准备工作,包括宣传教育、干部培训和文件印发等;调查资料的报送办法;调查经费的预算和开支办法;提供或公布调查成果的时间即完成工作后的总结等。对一些大规模的调查,在方案制订好后,还要进行试点,以便根据试点情况,修正、补充方案的不足,使得调查方案更科学、更符合实际,便于以后的操作。

#### 6. 编写填表说明

各调查人员对调查表上许多项目的理解可能不完全相同,为了取得数据一致的资料,对这些项目的含义、计算方法以及在调查时需要注意的问题都必须予以简要的解释和说明。

填表说明的内容广泛,但是不能太过冗长复杂,内容过多会导致主次不分,影响对问题的注意力。填表说明中,文字应力求准确、简单易懂、清楚了。



## 模块二 调查问卷的设计

在数据整理工作中,一旦数据资料收集方法确定以后,实际的调查问卷的设计过程也就开始了。调查者事先根据统计调查的目的和要求设计由一系列问题、说明及备选答案组成的调查问卷。

问卷调查是一种特殊的且在目前调查业中所广泛采用的调查形式。统计调查的组织者根据调查的目的和要求设计各类调查问卷,采取抽样的方式(随机抽样或整群抽样)确定调查样本,通过调查员对样本的访问完成事先设计的调查项目,然后由统计分析得出调查结果。调查问卷是用来收集调查数据的一种工具,调查问卷的设计在调查中十分重要,它是调查方案设计的核心内容,是将调查目的转化为一些被调查对象可以回答的问题的重要环节。

### 一、调查问卷的概念与类型

调查问卷是以书面形式了解被调查对象的反应和看法,并以此获得资料 and 信息的载体,是收集统计资料的基本工具之一。标准化的问卷不仅有利于准确、迅速地收集资料和信息,而且便于高速、高效地对这些数据进行处理和分析。

调查问卷必须与调查目的和要求、调查主题、调查对象以及调查方式相适应,因而不同的研究问题、对象和方式应采用不同的问卷类型。

(1) 按调查方式的不同,调查问卷可分为派员访问调查问卷、电话调查问卷、邮寄调查问卷、网上调查问卷和座谈会调查问卷等。

(2) 按问卷的填答方式不同,调查问卷可分为自填式问卷和代填式问卷。自填式问卷即由被调查者自己填答的问卷,主要适用于邮寄调查、宣传媒体发放的问卷调查、网上调查以及派员送发的问卷调查等;代填式问卷是由访问员根据被调查者的口头回答来填写的问卷,主要适用于派员访问调查、座谈会调查以及电话调查等。

### 二、调查问卷的设计要求

合格的调查问卷首先应满足两个条件:其一,调查者可以通过调查问卷获得统计研究中所需要的、完整的、准确的、适合统计数据处理的信息;其二,调查问卷所列的问题能够使绝大多数被调查者理解并愿意作答。因此,调查问卷的设计应满足以下要求:

(1) 有明确的主题。应根据调查主题,从实际出发拟定调查问卷的题目,符合调查的总体信息要求,题目的明确,重点突出,没有可有可无的问题。

(2) 表述准确、通俗易懂。调查问卷中各问题的表述要清楚,措辞准确,使作答者一目了然并愿意如实回答。调查问卷中语气要亲切,符合作答者的理解能力和认识能力,避免使用专业术语及不规范用语。对敏感性问题,要采取一定的技巧,使调查问卷具有合理性和可答性,避免主观性和暗示性,以免答案失真。

(3) 结构合理、逻辑性强。问题的排列应有一定的逻辑顺序,既符合调查对象的特点,也符合被调查者的思维程序,做到结构合理。一般是先易后难、先简后繁、先具体后抽象。

(4) 保持中立。调查问卷中所提出的问题应客观,不能带有任何倾向性,只有保持中



立,才能使得到的资料反映出被调查者的真实想法,切合实际。如果在调查问卷中出现有强烈暗示性答案的问题,容易诱导被调查者给出并非自己真实想法的答案。

(5) 简明扼要。无论是设计问题还是设计答案,所用的语言应该简单,问题的陈述应尽可能简短。此外,调查问卷的长度也应该进行控制,尽量将被调查者回答问卷的时间控制在20分钟左右,问卷中既不浪费一个问题,也不遗漏一个问题,否则容易引起反感。

### 三、调查问卷的一般结构

#### 1. 调查问卷的题目

调查问卷的题目体现了调查的主题,使被调查者对要回答的问题有一个大致的了解,是调查内容的高度浓缩。调查问卷的题目应该设计得既简明扼要,又准确、醒目和突出,如“大学生消费状况调查”“汽车消费状况调查”“城市家庭收入分配情况调查”等,不能只写“问卷调查”,这样容易造成被调查者因不必要的怀疑而拒答。

#### 2. 调查问卷的说明

调查问卷的说明一般放置在调查问卷的开头,通过它可以使被调查者消除顾虑,积极合作,正确填写问卷调查表。应该以简明扼要的语言向被调查者说明项目调查的目的、意义、涉及的主要内容等;同时,要说明调查者的身份、调查结果的使用和依法保密的措施等。例如,以下是某课题研究小组关于“某地区企业财务报告的现状调查问卷”的说明。

各有关企业及会计人员:

为调查了解我区企业财务报告的现状、问题及改进意见,为会计标准建设提供实务依据,我们组织开展了“财务报告的列报”课题研究。现有课题组起草的《关于企业财务报告列报问题的调查问卷》需向贵企业及有关会计人员进行调查,请结合本企业及自身工作,填写调查问卷,我们将遵照《中华人民共和国统计法》,对您的回答严格保密。对您的合作与支持,我们表示衷心的感谢!

#### 3. 指导语

指导语就是告诉被调查者如何填写问卷的说明和解释,一般位于封面信之后,并标有“填写说明”或“注”的标题,是对问卷填写的要求、方法、注意事项等做的一个总的说明,指出应注意的事项等。只有按照要求进行填写的调查问卷才能被视为有效,无效的调查问卷不能作为统计资料被收集。

#### 4. 调查事项的问题和答案

问题和答案是调查问卷的主体,调查资料的收集要通过这部分才能完成,问题和答案设计的好坏直接影响到调查结果的使用价值。既要提出问题,又要明确问题的回答方式。根据问题的形式不同,可分为开放式问题和封闭式问题。

开放式问题指调查问卷中只提出问题,但是对所做的命题不设置任何备选答案或可供参考的提示,而是要求调查对象给出自己的回答。一般要求在问题之后要留有足够大的空白供被调查者填写答案。例如:你喜欢看什么题材的电影?你最爱的娱乐方式是什么?如果你最近决定旅游,会选择哪个城市?这类问题都属于开放式问题。开放式问题的优点是可以让作答者自由发挥,能收集到生动的资料,也可能反映出被调查者之间的一些较细微的差异,甚至得到意外的发现。其缺点是被调查者的答案各异,复杂多样,从而给统计资料的



整理、分类以及汇总带来一定的困难。

封闭式问题在提问的同时提供若干答案,由被调查者根据自己的实际情况选择问题答案。根据问题的性质、要求和回答方法的不同,封闭式问题可以分为单项选择、多项选择、排序回答、等级评定等多种类型。封闭式问题的优点是问题清楚、具体,被调查者容易作答,节约时间;答案是标准化的,填写方便,容易进行资料整理和分析。其缺点是增加了调查问卷设计的难度,尤其对复杂问题,答案的列示是否全面、合理会直接影响调查结果的准确程度。

#### 四、调查问卷的设计程序

调查问卷的设计是由一系列相关工作过程构成的,为使调查问卷具有科学性和可行性,需要按照一定的程序进行。

##### (一) 调查问卷的准备阶段

###### 1. 确定调研主题和调查项目的信息

问卷设计的第一步就是界定调研主题,根据调研主题确定所需要的信息。准确地界定调研主题,一项调研就成功了一半。在调查问卷的设计过程中,设计人员必须不断地提醒自己要准确地反映调研主题,任何不能体现调研主题的问题必须坚决摒弃。调查问卷设计者要明确需要哪些资料以及如何获取资料,哪些问题必须包含在调查问卷中,也就是要找出变量或变量群,再建立变量群间的关系,然后根据这种关系绘制调查问卷设计流程图,根据流程图设计问题。这样,在设计问题时,就会有一定的思路,不易产生疏漏,同时也可以避免产生不必要的问题。另外,调查者可以随时验证其调查问卷结构是否合乎逻辑。

###### 2. 分析调查对象的各种特征

分析调查对象的各种特征即分析各被调查者的社会阶层、行为规范、社会环境等社会特征,文化程度、知识水平、理解能力等文化特征,行为、需求动机等心理特征,以此作为拟定调查问卷的基础。

##### (二) 调查问卷的初步设计

这一步主要是确定调查问卷的结构,拟定并编排问题。

###### 1. 确定调查方式和调查问卷类型

在明确了所需要的信息后,就要确定调查方式和调查问卷类型。不同的调查方式采用不同的调查问卷类型,调研者应从研究目的、研究对象、资料分析和解释的种类出发,确定调查问卷类型。

###### 2. 确定调查问卷的内容

确定调查问卷的内容具体指问题的设计和选择,根据调查问卷设计流程图,根据目的分析你需要的内容,确定调查问卷中具体包括的问题以及应询问的内容。在确定所包含的问题及问题的内容时,要确保能将所要调查的问题明确地传达给被调查者,并取得真实、准确的答案。这一阶段的重点就是要以被调查者为中心考虑问题,应检查以下内容:

(1) 检查问题是否必要。与研究目的无关的问题,不仅会增加调查者和被调查者的负担,还会浪费时间和费用,因此应该剔除。

(2) 检查被调查者是否能回答问题的内容。如果被调查者都不能回答或臆测式回答,会



影响调查问卷的准确程度。

(3) 检查被调查者是否愿意回答问题的内容。如果被调查者了解问题的含义,也具备回答的能力,但拒答,也会影响调查的准确性。

(4) 检查问题是否具有很强的专业性。如果问题具有很强的专业性,被调查者可能需要收集资料才能回答,被调查者就可能不愿意接受调查。

### 3. 确定问题的形式

问题的形式是指调查者为让被调查者表示其意见或圈选其答案而设计的问题的呈现方式,主要有:开放性问题 and 封闭性问题;直接性问题、间接性问题和假设性问题;事实性问题、行为性问题、动机性问题和态度性问题。不同类型的问题有不同的优缺点和适用范围。

### 4. 确定问题的顺序

调查问卷必须具备逻辑的连续性,一般应注意以下几点:

- (1) 时间上由远到近(或由近到远)。
- (2) 先易后难,先熟悉后生疏。
- (3) 先封闭性问题后开放性问题。
- (4) 先一般性问题后敏感性问题。
- (5) 先行为资料和态度资料,后基本资料。
- (6) 专业性强的具体而细致的问题应尽量放后面。

### 5. 确定问题的措辞

问题措辞的基本原则是简短、明确、通俗、易懂。在调查问卷设计中,对问题的语言表达和提问方式应注意以下几点:

- (1) 避免使用不确切的、含混模糊的词和语言。
- (2) 避免过于抽象的问题和专业化术语。
- (3) 避免提具有引导性和断定性的问题。
- (4) 避免提令被调查者难堪及其忌讳和敏感的问题。
- (5) 避免双重提问和使用复杂句。
- (6) 拟定问题要考虑时间性和明确的界限。

### 6. 确定问卷的版面格式

问卷的版面格式将影响调查者询问时或被调查者答题时的意愿及情绪,从而影响调查的质量。对于版面格式的确定,应注意以下几点:

(1) 调查问卷要整齐、清楚,注重纸和印刷质量,低质量的纸张和印刷会影响调查问卷的量和效果。

(2) 在调查问卷中要有足够的空间供被调查者填写答案,同时要考虑数据处理的方便性。

- (3) 适当的图案或图表可以调动被调查者的积极性。
- (4) 重要的地方要进行强调,引起被调查者的注意。
- (5) 避免将一个题目印到两页,各题目之间应有一定的间隔,以增加可读性。
- (6) 调查问卷应装订在一起,防止数据丢失。



## 五、调查问卷的测试和定稿

在调查问卷的主要内容基本确定以后,一般需要对调查问卷进行测试,主要包括两个方面的内容:第一,站在调查者的立场进行提问,主要考查问题是否清楚明白,是否便于调查中对资料进行记录和整理;第二,站在应答者的立场上进行回答,主要考查被调查者对所有问题的回答能力和态度。根据试答情况对调查问卷不合理的地方进行修改,直到完全合格之后才能最后定稿。

## 模块三 统计调查误差

统计调查误差是客观存在的,不论采取什么途径或调查方法,收集到的数据总会由于各种原因产生一定程度的误差。统计调查误差是指通过调查或实验所获得的统计数据与调查对象的实际数量的差别。例如,对某镇居民年人均收入调查的结果为3 500元,而事实上的数据为3 560元,出现了比较小的误差。这种现象非常普遍,几乎所有的统计调查都会产生误差。为了获得准确的统计调查资料,必须把统计调查误差缩小到最低限度或控制在允许的范围之内。

### 一、统计调查误差的种类与产生的原因

统计调查误差分为登记性误差和代表性误差。

#### 1. 登记性误差

登记性误差是指在调查过程中的测量、记录、填报、计算等活动中产生的误差,主要包括:由计量手段的局限性所带来的难以绝对符合实际而出现的误差;由登录、计算、抄报、汇总错误及被调查者所报不实或调查者有意虚报等所带来的误差;被调查者因人为因素干扰形成的有意虚报或瞒报调查数据所带来的误差。

#### 2. 代表性误差

代表性误差是指在非全面调查中,由于该部分单位对总体缺乏充分的代表性,而使推断的总体数值同总体的实际数值之间发生的误差。其产生的原因主要有:抽取样本时没有遵循随机原则,样本容量不足,总体结构存在差异等。比如在某乡调查居民家庭拥有电冰箱的比例,通常对这种统计不需要进行全面调查,只抽取其中一部分家庭即可,假如在所抽中的家庭中没有电冰箱的恰好较多,使总的比例偏低,则形成代表性误差。大量的统计活动都是非全面调查,难免会存在一定程度的“以偏概全”。

### 二、认识误差

(1) 误差是难以避免的。“误差”一词本身就是理想化的,因为我们通常不知道“实际值”到底是多少。我们一般所说的“实际值”是通过大量观察所得到的平均值,因为误差多是正负互现的,通过大量观察,误差可以正负抵消并趋于零,大量观测值的平均数就被我们视为“实际值”,这显然不完全可靠。

(2) 误差降低了统计数据的质量。统计数据的准确性是数据质量的核心问题,它是人





们进行研究和分析的基础,基于错误的数据,不可能形成正确的认识。因此,从这个意义上讲,误差越小越好。

(3) 人们对精确性的要求是相对的。例如,没有人说自己的体重是 50.034 56 kg;质量检验机构的衡器精度标准与体重秤的标准不同。这些情况说明,我们在不同的场合对准确程度的要求不是绝对的,而是相对的,只要达到我们所能接受的精度即可。

(4) 追求过高的准确性往往得不偿失。误差程度的缩小或精确程度的提高通常是与工作量或工作费用紧密相连的。误差越小,意味着工作量越大,所花费的时间和费用也就越多。我们的目标是使用尽可能少的费用,取得准确性合乎要求的数据,而不是一味地提高准确程度。

### 三、误差的防止措施

#### (一) 登记性误差的防止措施

##### 1. 制订科学的调查方案

统计调查方案的设计要合理、科学,调查方案所包括的各项内容都要详细、具体、明确地加以规定和说明,使调查人员和填报单位能够正确理解与执行。

##### 2. 加强统计基础工作

加强统计基础工作,包括建立相应的统计机构,配备必要的人员,建立健全计量工作、原始记录、统计台账等制度和相关的责任制,保证数据填报的正确性。

##### 3. 全面加强统计法制建设

健全统计法规、严明统计纪律、增强法制观念、坚持实事求是,是从根本上杜绝弄虚作假、虚报瞒报等违法行为的重要途径。

#### (二) 代表性误差的防止措施

##### 1. 要正确选定抽样方法

选择正确的抽样方法,有利于使抽取的样本真正代表总体,减少误差。要根据调查目的和要求以及调查所面临的主客观、内外部条件进行权衡,以选择合适的抽样方法。一般条件下,随机抽样法具有更大的适应性。

##### 2. 要正确确定样本数目

一般来说,样本数与抽样误差呈反比关系,即样本越大,抽样误差越小,反之亦然。总体中差异越大,在同样样本数的条件下,误差越大;总体中的差异越小,在同样样本数的条件下,误差越小。也就是说,在确保同样的差异误差的前提下,如果总体中的差异大,则需抽取的样本数应该大一些,反之亦然。

## 模块四 统计应用工具 Excel 简介

在大数据时代,数据收集是一个复杂的过程,很多情况下,数据收集是“多合一”的,其中,“多”指的是数据的提供方很多,我们应该对这些数据进行挖掘,收集有价值的东西,形成数据资产。Excel 2016 不仅具有强大的数据组织、计算、分析和统计功能,能够通过图表、图



形等多种形式形象地显示处理结果,而且可以方便地与 Excel 2016 其他组件相互调用数据,实现资源共享。

## 一、Excel 2016 界面介绍

Excel 2016 沿用了前一版本的功能区(Ribbon)界面风格,将 Excel 2007 版本的传统风格菜单和工具栏以多面选项卡功能面板代替。此外,在窗口界面中设置了一些便捷的工具栏和按钮。

Excel 2016 的界面如图 2-1 所示。

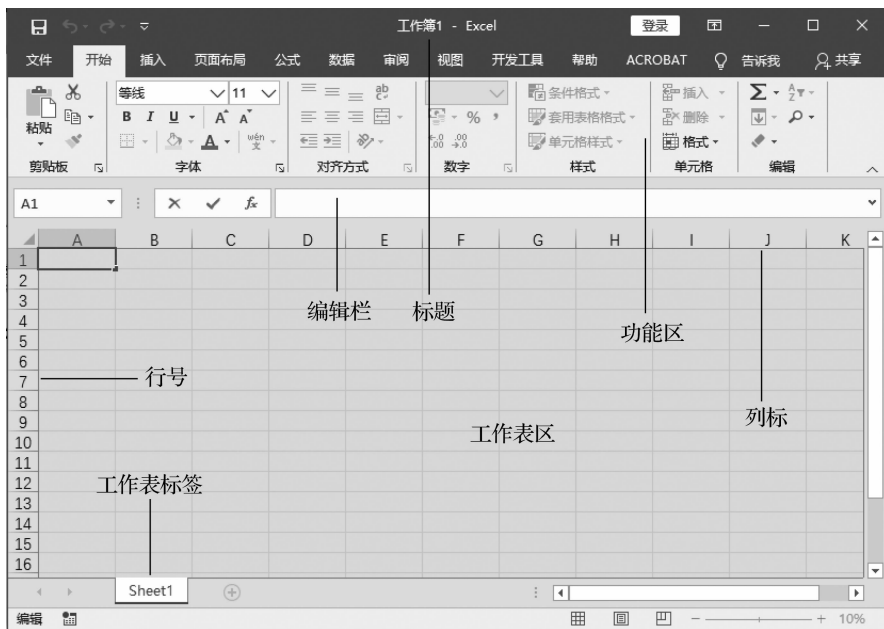


图 2-1 Excel 2016 的界面

## 二、Excel 2016 的基本操作

### (一) 工作簿的创建

用户可以通过以下两种方法创建新的工作簿。

#### 1. 在 Excel 工作窗口中创建工作簿

在 Excel 2016 的工作窗口中,可以通过以下两种等效操作方式创建新的工作簿:

- (1) 在功能区上执行“文件”→“新建”命令,再选择“空白工作簿”命令。
- (2) 按 Ctrl+N 组合键。

上述方法所创建的工作簿同样只存在于内存中,并会依照创建次序自动命名。

#### 2. 在 Windows 系统中创建工作簿

在 Windows 系统中,可以通过以下两种操作方式创建新的工作簿:

- (1) 双击桌面快捷方式,打开 Excel 程序窗体,在右侧单击“空白工作簿”命令。
- (2) 安装了 Excel 2016 的 Windows 系统,会在鼠标右键快捷菜单中自动添加“Mi-



icrosoft Excel 工作表”的快捷命令,通过这一快捷命令也可以创建新的 Excel 工作簿文件,并且所创建的工作簿是一个存在于磁盘空间内的真实文件,操作方法如下:

在 Windows 桌面或者文件窗口的空白处右击,在弹出的快捷菜单中选择“新建”→“Microsoft Excel 工作表”命令,如图 2-2 所示。完成操作后可在当前位置创建一个新的 Excel 工作簿文件,双击此新建的文件,即可在 Excel 工作窗口中打开此工作簿。

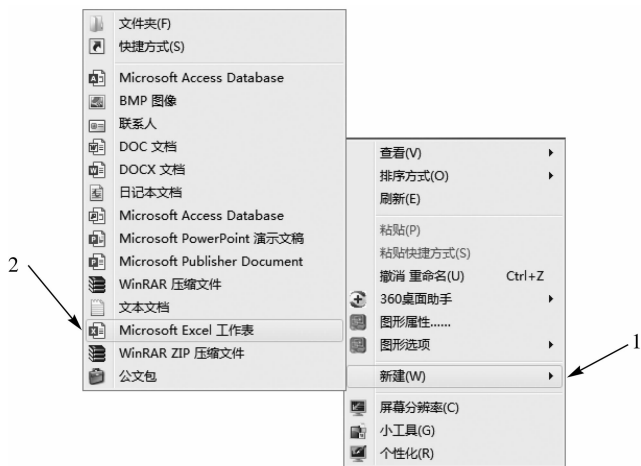


图 2-2 通过右键快捷菜单创建工作簿

## (二) 保存工作簿

事实证明,计算机最重要的功能是保存数据。在 Excel 中,用户能以多种方式保存自己的文件,可以使用“保存”“另存为”“发布”“保存工作区”“关闭”和“退出”命令,也可以使用最简单的保存方法——单击“快速访问工具栏”上的“保存”按钮。

在 Excel 2016 之前,保存工作簿的路径只能为本地,当用户使用 Excel 2016 保存工作簿时,执行“文件”→“另存为”命令,打开“另存为”窗口(见图 2-3)。

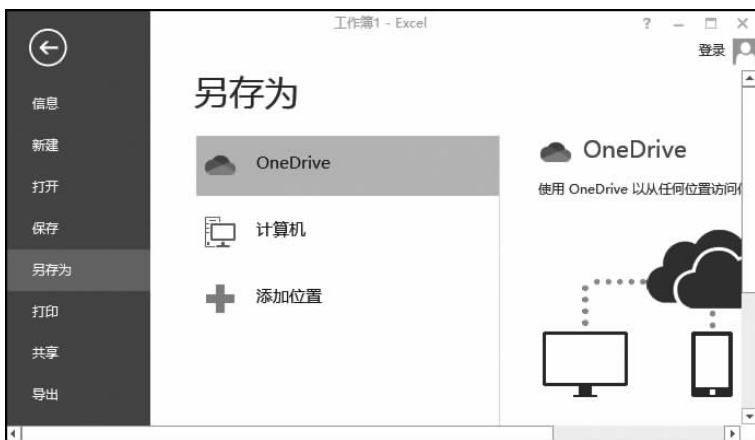


图 2-3 “另存为”窗口

(1) OneDrive:将工作簿保存到当前登录的 Office 2016 的 Microsoft 账户对应的个人 OneDrive 空间。



- (2) 计算机:将工作簿保存到本地。
- (3) 添加位置:添加保存的路径位置。

## 知识拓展

### 保存和另存为

需要注意的是,Excel 有两个和保存功能有关的菜单命令,分别是“保存”和“另存为”,它们的名字和实际作用都非常相似,但是实际上有一定的区别。

对于新建的工作簿,在第一次执行保存操作时,“保存”和“另存为”命令的功能完全相同,执行“保存”或“另存为”命令,都将打开“另存为”对话框,供用户进行路径定位、文件命名和格式选择等一系列设置。

对于之前已经被保存过的现有工作簿,用户再次执行保存操作时,这两个命令则有以下区别:

- (1) 执行“保存”命令,不会打开“另存为”对话框,而是直接将编辑修改后的内容保存到当前工作簿中。工作簿的文件名、存放路径不会发生任何改变。
- (2) 执行“另存为”命令,将会打开“另存为”对话框,允许用户重新设置存放路径、命令和其他保存选项,以得到工作簿的另一个副本。

### (三) 打开现有工作簿

经过保存的工作簿在计算机磁盘上形成文件,用户使用标准的计算机文件管理操作方法就可以对工作簿文件进行管理,诸如复制、剪切、移动、删除和重命名等。无论工作簿文件被保存在哪里,或者是复制到不同的计算机上,只要所在的计算机安装有 Excel 程序,工作簿文件就可以被再次打开并进行读取和编辑等操作。

打开现有工作簿的方法如下。

#### 1. 直接通过文件打开

如果用户知道工作簿文件所保存的确切位置,直接双击文件图标即可打开。

#### 2. 使用“打开”对话框打开

如果用户已经启动了 Excel 程序,那么可以通过执行“打开”命令在弹出的“打开”窗口(见图 2-4)中打开指定的工作簿。



图 2-4 “打开”窗口



执行“文件”→“打开”命令或按 Ctrl+O 组合键,用户可以通过以下三种方式打开已有的工作簿:

(1) 最近使用的工作簿:在“打开”界面的右侧会显示用户最近使用的工作簿列表,在列表中单击工作簿名称即可打开该工作簿。

(2) OneDrive:打开用户保存在 OneDrive 中的工作簿。

(3) 计算机:打开本地工作簿,右侧会显示“最近访问的文件夹”,用户单击指定文件夹,弹出“打开”对话框,文件夹路径为用户选择的列表文件夹路径。如果在“最近访问的文件夹”列表中没有指定打开工作簿的文件夹选项,用户还可以单击“浏览”按钮(见图 2-5),弹出“打开”对话框,如图 2-6 所示。



图 2-5 “打开本地工作簿”窗口



图 2-6 “打开”对话框

在“打开”对话框中,用户可以通过左侧的树形列表选择工作簿文件的存放路径,在目标路径下选中具体文件后,双击文件图标或者单击“打开”按钮即可打开文件。

### 3. 通过历史记录打开

用户近期打开过的工作簿文件,通常情况下都会在 Excel 程序中留有历史记录,如果用户需要打开最近操作过的工作簿文件,也可以通过历史记录来快速打开文件。



调用历史记录的方法为:在 Excel 功能区中执行“文件”→“打开”→“最近使用的工作簿”命令,就会列出最近打开过的文件记录(显示 25 条最近记录),如图 2-7 所示。用户单击文件名即可打开相应的工作簿文件。



图 2-7 最近打开过的文件记录

用户可以自行修改最近使用的工作簿的显示数量,其操作方法是在功能区中执行“文件”→“选项”命令,打开“Excel 选项”对话框,在左侧选中“高级”选项卡,然后在右侧的“显示”区域中,通过“显示此数目的‘最近使用的工作簿’”的微调按钮,调节需要显示的最近使用的工作簿个数(见图 2-8),最后单击“确定”按钮保存设置。

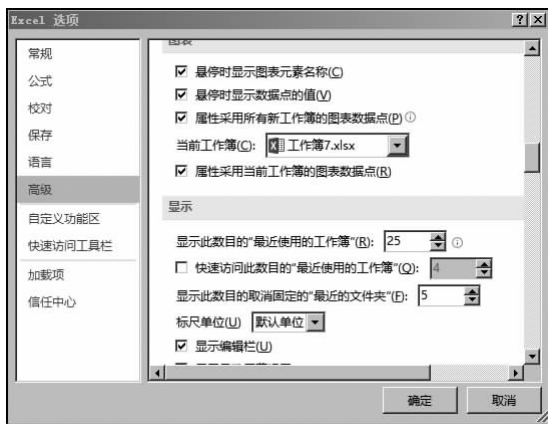


图 2-8 修改最近使用的工作簿的显示数量

如果选中“快速访问此数目的‘最近使用的工作簿’”复选框,同时通过右侧的微调按钮确定显示数目(工作簿数目默认为 4 个),可以在“文件”选项卡底部显示“快速访问工作簿”列表,单击列表中的文件名称,即可打开相应的工作簿文件。

#### (四) 关闭工作簿和退出 Excel 程序

##### 1. 关闭工作簿

用户结束工作后,可以关闭 Excel 工作簿以释放计算机内存。关闭当前工作簿的等效操作方式有以下几种:

(1) 在功能区执行“文件”→“关闭”命令。



- (2) 按 Ctrl+W 组合键。
- (3) 单击工作簿窗口右上方的“关闭”按钮。

## 2. 退出 Excel 程序

通过以上操作方式虽然可以关闭当前工作簿,但是并不能退出 Excel 程序。退出 Excel 程序的等效操作方式有以下几种:。

- (1) 在键盘上按 Alt+F4 组合键。
- (2) 单击 Excel 工作窗口右上方的“关闭”按钮。

如果当前的 Excel 工作窗口中同时打开了多个工作簿,也可以通过以上方法来关闭所有工作簿。

## (五) 工作表的基本操作

工作表包含在工作簿之中,是工作簿的必要组成部分,工作簿总是包含一个或者数个工作表,它们之间的关系就好比图书与书中的书页。下面将对工作表的创建、复制等基本操作进行介绍。

### 1. 工作表的创建

工作表的创建通常分为两种情况:一种是随着工作簿的创建而一同创建,另一种是从现有工作簿中创建新的工作表。

(1) 随着工作簿的创建而一同创建。默认情况下,Excel 在创建工作簿时,自动包含名为“Sheet1”的一张工作表。用户可以通过设置来改变新建工作簿时所包含的工作表数目。

打开“Excel 选项”对话框,用户可以在“常规”选项卡中“包含的工作表数”的文本框内设置新建工作簿时包含的工作表数,数值范围为 1~255,单击“确定”按钮保存设置并退出“Excel 选项”对话框,如图 2-9 所示。调整数值后,新建工作簿时,自动创建的内置工作表数目与设置的数目相同,并且自动命为 Sheet1~Sheetn。

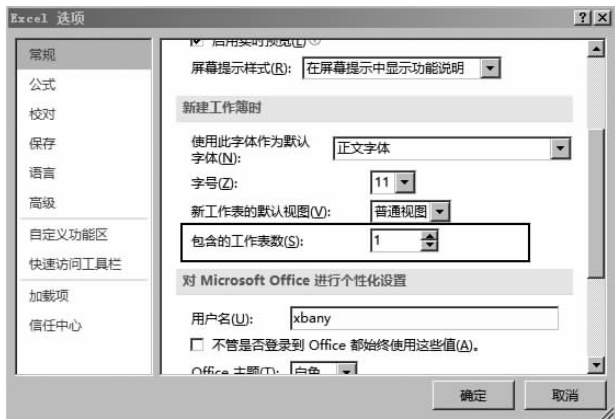


图 2-9 设置新建工作簿时包含的工作表数

**提示:**在大多数情况下,用户的工作簿中并没有包含太多工作表的必要,而且空白的工作表会增加工作簿文件的内存,造成不必要的存储容量占用。所以,建议用户将新工作簿内的工作表数设置得尽可能少,在需要的时候增加工作表比不需要的时候删除空白工作表更容易。



(2) 从现有的工作簿中创建新的工作表。用户可以通过以下几种等效操作方式在当前工作簿中创建一个新的工作表：

① 在“开始”选项卡中执行“插入”→“插入工作表”命令，则会在当前工作表左侧插入一张新工作表，如图 2-10 所示。

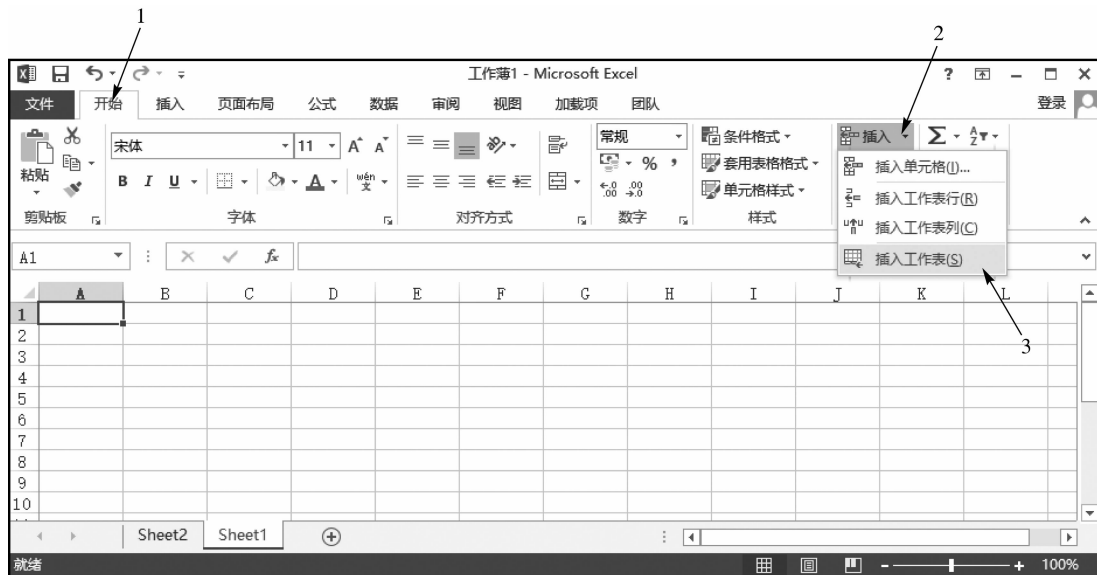



图 2-10 通过“插入工作表”创建新工作表

② 在当前工作表标签上右击，在弹出的快捷菜单中选择“插入”命令，在弹出的“插入”对话框中选中“工作表”，再单击“确定”按钮。

③ 单击工作表标签右侧的“新工作表”按钮, 则会在工作表的末尾快速插入新工作表。

**提示:** 新创建的工作表，依照现有工作表数目自动编号命名。

如果用户通过右键快捷菜单插入新工作表，需要批量增加多张工作表时，可以在第一次插入工作表操作完成后，按 F4 键重复操作；如果用户通过工作表标签右侧的按钮创建新工作表，则无法按 F4 键重复创建。也可以在同时选中多张工作表的情况下使用功能区按钮或使用工作表标签的右键菜单命令插入工作表，此时会一次性创建与选定的工作表数目相同的新工作表。

**注意:** 创建新工作表的操作无法通过“撤销”按钮进行撤销操作。

## 2. 设置当前工作表

在 Excel 的操作过程中，始终有一个“当前工作表”作为用户输入和编辑等操作的对象和目标，用户的大部分操作都是在“当前工作表”上得以体现的。在工作表标签栏上，“当前工作表”的标签背景为白色，如图 2-11 所示。要切换其他工作表为当前工作表，可以直接单击目标工作表标签。

如果工作簿内包含的工作表较多，标签栏上不一定能够全部显示所有工作表标签，可以通过单击标签栏左侧的工作表导航按钮来滚动显示工作表标签。

除此以外，通过拖动工作窗口上的水平滚动条边框，用户可以改变工作表标签的显示宽度，以显示更多的工作表标签。





图 2-11 当前工作表

如果工作簿中的工作表很多,需要滚动很久才能看到目标工作表,还可以在工作表导航栏上右击(见图 2-12)。此时会显示一个工作表标签列表,从中选择一个工作表名称,单击“确定”按钮,就可以显示相应的工作表,双击工作表名称,也可以显示相应的工作表。

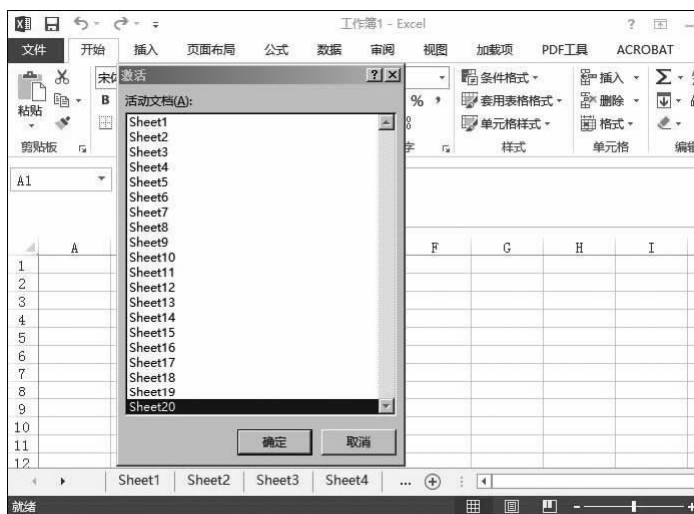


图 2-12 工作表标签列表

另外,常用于切换工作表的快捷键是  $\text{Ctrl} + \text{Page Up}$  组合键和  $\text{Ctrl} + \text{Page Down}$  组合键,它们的作用分别是切换到上一张工作表和切换到下一张工作表。

### 3. 同时选定多张工作表

除了选定某个工作表作为当前工作表,用户还可以同时选中多个工作表形成“工作组”。在工作组模式下,用户可以方便地同时对多个工作表对象进行复制、删除等操作,也可以进行多数据联动编辑等操作。

用户可以通过以下几种方式同时选定多张工作表,形成工作组:

- (1) 按住  $\text{Ctrl}$  键,同时依次单击需要选定的工作表标签,就可以同时选定多个工作表。
- (2) 如果用户需要选定的工作表为连续排列的工作表,则可以先单击其中的第一个工



作表标签,然后按住 Shift 键,再单击连续工作表中的最后一个工作表标签,即可同时选定工作表。

(3) 如果要选定当前工作簿中的所有工作表组成工作组,可以在任意工作表标签上右击,在弹出的快捷菜单上选择“选定全部工作表”命令。

(4) 多个工作表被同时选中后,会在 Excel 窗口标题栏上显示“工作组”字样,被选定的工作表标签都将以白色显示。

用户如果要取消工作组的操作模式,可以单击工作组以外的工作表标签(如果所有工作表都在工作组内,则单击任意工作表标签即可)或者是在工作表标签右击,在弹出的快捷菜单上选择“取消组合工作表”命令。

#### 4. 工作表的复制和移动

通过复制操作,可以在另一个工作簿或者不同工作簿中创建工作表的副本。还可以通过移动操作,在同一个工作簿中改变工作表的排列顺序,也可以在不同的工作簿间转移。下面介绍两种复制和移动工作表的方法:

(1) 通过菜单操作,有以下两种等效操作方式可以显示“移动或复制工作表”对话框:

① 在工作表标签上右击,在弹出的快捷菜单上选择“移动或复制”命令。

② 选中需要移动或复制的工作表,在“开始”选项卡中执行“格式”→“移动或复制工作表”命令。

在“移动或复制工作表”对话框中,单击“工作簿”的下拉按钮,可以选择将要移动或复制的目标工作簿,也可以选择新建工作簿。“下列选定工作表之前”文本框中显示了指定工作簿中所包含的全部工作表,用户可以选择要移动或复制的工作表在全部工作表中的排列位置。

“建立副本”复选框是一个操作类型开关,勾选此复选框则为“复制”方式,取消勾选则为“移动”方式。

在移动或复制的操作中,如果当前工作表与目标工作簿中的工作表名称相同,则会被自动重新命名,例如,“Sheet1”会被更名为“Sheet1(2)”,取消勾选则被移至相应的工作表前。

设置完毕后,单击“确定”按钮退出“移动或复制工作表”对话框,完成工作表移动或复制操作。

(2) 拖动工作表标签。拖动工作表标签来移动或者复制工作表的方法更为直接。将光标移至需要移动的工作表标签上,按下鼠标左键,鼠标指针显示出文档的图标,此时就可以拖动鼠标将此工作表移动至其他位置。如果在按住鼠标左键的同时,按住 Ctrl 键,则执行“复制”操作,此时鼠标指针下显示的文档图标上还会出现一个“+”号,以此来表示当前操作方式为“复制”。

#### 5. 删除工作表

用户也可以选择将当前工作簿中的一个或者多个工作表删除,有以下两种等效操作方式:

(1) 选中要删除的工作表,在“开始”选项卡中执行“删除”→“删除工作表”命令。

(2) 在工作表标签上右击,在弹出的快捷菜单中选择“删除”命令。

**注意:**在 Excel 程序中,删除工作表是无法进行撤销的,如果用户不慎误除了工作表,将





置,例如“此单元格两行之上的单元格”。

(2) 绝对引用。绝对引用是指该单元格在工作表中的固定位置,如“此单元格位于列 A 中和行 2 的交叉处”。

(3) 混合引用。混合引用则既包含相对引用又包含绝对引用,如“此单元格位于列 A 中活动单元格两行之上”。

当用户将公式从工作表中的某个位置复制到另一个位置上时,绝对引用和混合引用就特别有帮助。当用户进行复制和粘贴时,相对引用会自动调整,而绝对引用则不会自动调整。例如,对单元格 A1 的相对引用,看上去应该是这样的: =A1。而对于单元格 A1 的绝对引用,看上去应该是这样的: =\$A\$1。可以将对单元格 A1 的相对引用和绝对引用组合起来创建混合引用: =\$A1 或 =A\$1。

#### 4. 函数的编制与操作

简单的函数可以根据计算要求自己编制。例如,求和函数为: SUM(number1, number2, …), 其中“number1, number2, …”是要对其求和的各个参数。例如,在编辑栏输入函数“=SUM(A3:H4)”,表示 A3 至 H4 区域的数据加总。

#### 5. 常见的统计分析函数

Excel 2016 有 60 个内置的数学和三角函数,下面介绍一些常见的统计分析函数。

(1) 使用内置统计函数。

① AVERAGE 函数: AVERAGE 函数通过将一系列数值相加后的结果除以数值的个数来计算算术平均值。AVERAGE 函数使用(number1, number2, …)作为参数,包括多达 255 个参数,它忽略了包含逻辑值、文本或者空白的单元格。例如,要计算 B4—B15 的单元格中数值的平均值,可以用公式“=(B4+B5+B6+B7+B8+B9+B10+B11+B12+B13+B14+B15)/12”,但是很明显使用“=AVERAGE(B4:B15)”效率更高。

② MEDIAN、MODE、MAX、MIN 以及 COUNT 函数: MEDIAN 计算一组数值的中数; MODE 返回一组数值中出现最频繁的数值; MAX 返回一组值中的最大值; MIN 返回一组值中的最小值; COUNT 返回给定区域中包含数值的单元格的数目。

③ SUMIF、SUMIFS 和 COUNTIF 函数: SUMIF 函数类似于 SUM 函数,但是它首先使用定制的条件检验每个单元格,然后将其回到总和之中; SUMIFS 函数与 SUMIF 函数的功能相似,不同之处在于前者可以指定多达 127 个不同的求和区域,每个区域都具有各自的判别标准; COUNTIF 函数计算区域中满足给定条件的单元格的个数,它使用参数(range, criteria)。

(2) 使用分析排位和百分比的函数。PERCENTRANK 函数返回数据集中任意成员的百分比排位; PERCENTILE 函数可以确定区域中数值位于某个百分比的成员; QUARTILE 函数使用参数(array, quart),工作原理类似于 PERCENTILE 函数,不同之处在于它返回代表最低百分点的值或者输入集中任一四分位数值; SMALL 和 LARGE 函数分别是返回输入区域中的第  $k$  个最小值和最大值; RANK 函数返回某个特定的数在数据集中的排位。

(3) 使用样本和总体统计函数。方差和标准偏差是一组数据或者是总体离散度的统计测量值。

VAR 函数和 STDEV 函数计算平方差和标准偏差,假定它们的参数仅仅代表总体的一



份样本。如果所研究的数值代表整个总体而不是某个样本,就可以使用 VARP 函数和 STDEVP 函数来计算平方差和标准偏差。

(4) 计算线性回归的函数。等式  $y=mx+b$  用代数方法描述了单个变量数据集的直线,其中  $x$  是自变量, $y$  是因变量, $m$  代表直线的斜率, $b$  代表  $y$  轴截距。如果一条直线代表对预期结果多次回归分析中的许多自变量,则回归直线的等式应采用这样的形式: $y=m_1x_1+m_2x_2+\cdots+m_nx_n+b$ ,其中  $y$  是因变量, $x_1\sim x_n$  是  $n$  个自变量, $m_1\sim m_n$  是每个自变量的系数, $b$  是常数。

给定  $y$  的一组已知值以及每个自变量的一组已知值,LINEST 函数使用上述更一般的等式返回  $m_1\sim m_n$  以及  $b$  的值,返回已知数据最佳拟合直线的数学描述。它的参数形式为:  
=LINEST(known\_y's, known\_x's,const,stats)。

TREND 函数找出位于直线上的点以及落入未知位置的点,可以根据 TREND 函数所返回的数值画出趋势线,帮助理解实际数据意义的直线,也可以使用 TREND 做插值计算,或者根据已知数据所展示的趋势对将来的数据做出合理的猜测。它的参数形式为:  
=TREND(known\_y's, known\_x's,new\_x's,const)。

FORECAST 函数类似于 TREND 函数,不同之处在于它返回直线上的一个点而不是用来定义直线的数组。它的参数形式为:  
=FORECAST(x,known\_y's, known\_x's)。

SLOPE 函数返回线性回归直线的斜率,它的值与 LINEST 函数所返回中的第一个数值相同。它的参数形式为:  
=SLOPE(known\_y's, known\_x's)。

STEYX 函数计算回归的标准误差。它的参数形式为:  
=STEYX(known\_y's, known\_x's)。

(5) 计算指数回归。指数回归沿直线画出数值,与此不同的是,指数回归通过计算所需数值来描述曲线,描述指数回归曲线的等式如下:

$$y=b * m_1^{x_1} * m_2^{x_2} * \cdots * m_n^{x_n} \quad (2-1)$$

如果只有一个自变量,则等式如下:

$$y=b * m^x \quad (2-2)$$

LOGEST 函数类似于 LINEST 函数,不同之处在于它用来分析非线性的数据,并返回指数曲线(而不是直线)的坐标。LOGEST 函数返回每个自变量的系数取值以及常量  $b$  的值。它的参数形式为:  
=LOGEST(known\_y's, known\_x's,const,stats)。

GROWTH 函数返回最佳拟合已知数据的指数回归曲线的数学描述,并找出位于曲线上的点。GROWTH 函数类似于纯属回归分析中的同类函数 TREND,它的参数形式为:  
=GROWTH(known\_y's, known\_x's,new\_x's,const)。

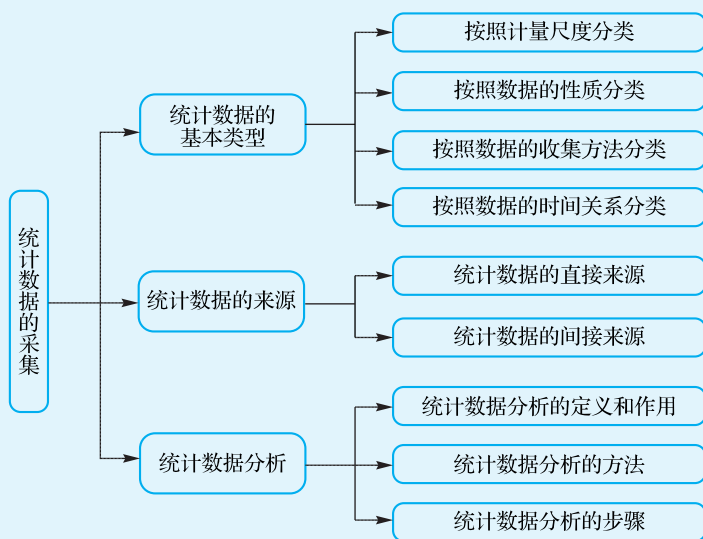
# 项目三

## 统计数据的采集

### ● 学习目标

- 了解各类统计数据的结构和特点；
- 了解各种数据分类的标准；
- 熟悉各种统计数据的采集方法；
- 掌握统计分析的方法和步骤。

### ● 项目导航





## 案例导入

### 我国人口结构的统计

2019年末,中国大陆总人口140 005万人,比上年末增加467万人。全年出生人口1 465万人,人口出生率为10.48‰;死亡人口998万人,人口死亡率为7.14‰;人口自然增长率为3.34‰。从性别结构看,男性人口71 527万人,女性人口68 478万人,总人口性别比为104.45(以女性为100)。从年龄构成看,16~59周岁的劳动年龄人口89 640万人,占总人口的比重为64.0%;60周岁及以上人口25 388万人,占总人口的18.1%,其中65周岁及以上人口17 603万人,占总人口的12.6%。从城乡结构看,城镇常住人口84 843万人,比上年末增加了1 706万人;乡村常住人口55 162万人,减少了1 239万人;城镇人口占总人口的比重(城镇化率)为60.60%,比上年末提高了1.02个百分点。全国人户分离人口(即居住地和户口登记地不在同一个乡镇街道且离开户口登记地半年以上的人口)28 000万人,比上年末减少了613万人;其中流动人口23 600万人,比上年末减少了515万人。

资料来源:扬子新闻网,<http://www.yzpz.com/view-28765-1.html>,有删改。

#### 思考与讨论:

- (1) 在年末的人口统计工作中,工作人员是如何采集到这样的海量数据的?
- (2) 人口普查中的相关数据有哪些具体特征?
- (3) 在具体工作中,各类数据的表示方法具体有哪些?

统计学的研究对象实际上就是统计数据。我们的实际生活中充斥着各种统计数据。例如,对某商品在各店的价格进行调查可以得到价格数据,对某班学生的成绩进行观测可以得到成绩数据,等等。本项目就来研究采集这些数据的方法。

## 模块一 统计数据的基本类型

统计数据是统计工作活动过程中所取得的反映国民经济和社会现象的数字资料以及与之相联系的其他资料的总称。统计数据是对现象进行测量的结果,比如:对国家经济活动总量进行测量,可以得到国内生产总值(GDP)的数据;对股票价格变动水平进行测量,可以得到股票价格指数的数据;对人口性别进行测量,可以得到男性人口数和女性人口数的数据。本模块将从不同角度说明统计数据的分类。

### 一、按照计量尺度分类

统计数据作为统计工作的成果,要“符合标准”,要准确地反映客观现实。根据对研究对象计量的不同精确程度,可以将计量尺度由低到高、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。



### （一）定类尺度

定类尺度是按照客观现象的某种属性对其进行分类。这一尺度所使用的数值只是作为各种分类的代码,并不反映各类现象的优劣、量的大小或顺序。例如,按照性别可以将人口分为男性人口和女性人口两类;按照经济性质可以将企业分为国有企业、集体企业、私营企业、混合制企业等。这里的“性别”和“经济性质”就是两种定类尺度。定类尺度是最粗略、计量层次最低的计量尺度,利用它只可测度事物之间的类别差,而不能了解各类之间的其他差别。定类尺度计量的结果表现为某种类别,但为了便于统计处理,例如,为了计算和识别,也可以用不同数字或编码表示不同类别。例如,用1表示男,用0表示女;用1表示国有企业,用2表示集体企业,用3表示私营企业,等等。这些数字只是不同类别的代码,绝不意味着它们有大小的区分,更不能进行任何数学运算。定类尺度能对事物做最基本的测度,是其他计量尺度的基础。

### （二）定序尺度

定序尺度是对客观现象各类之间的等级差或顺序差的一种测度。利用定序尺度不仅可以研究将对象分成不同的类别,而且可以反映各类的优劣或顺序。例如,“产品等级”就是一种测度产品质量好坏的定序尺度,它可以将产品分为一等品、二等品、三等品和次品等;“考试成绩”是一种定序尺度,它可以将成绩分为优、良、中、及格和不及格等;“对某一事物的态度”也是一种定序尺度,它可以将人们的态度分为非常同意、同意、保持中立、不同意和坚决不同意,等等。显然,定序尺度对事物的计量要比定类尺度精确些,但它至多测度了类别之间的顺序,而未测量出类别之间的准确差值。因此,定序尺度的计量结果只能比较大小,不能进行加减乘除等数学运算。

### （三）定距尺度

定距尺度是对现象类别或次序之间间距的测度。定距尺度不但可以用数字表示各类别现象的不同和顺序的差异,而且可以用确切的数值反映现象之间在量方面的差异。定距尺度通常使用自然或物理单位作为度量单位,如收入用人民币“元”度量,考试成绩用“分”度量,气温用“ $^{\circ}\text{C}$ ”或“ $^{\circ}\text{F}$ ”来度量,重量用“ $\text{kg}$ ”度量,长度用“ $\text{m}$ ”度量等。区间尺度的计量结果表现为数值。区间尺度的数值可做加减法运算,例如,考试成绩80分与90分之间相差10分,一个地区的气温 $20^{\circ}\text{C}$ 与另一个地区的气温 $25^{\circ}\text{C}$ 相差 $5^{\circ}\text{C}$ ,等等;但不能做乘除法运算。

### （四）定比尺度

定比尺度是指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、速度、密度等数量关系。定比尺度不仅能测度各类别的大小和多少,还有一个绝对零点(absolute zero)作为起点。这个绝对零点是它跟定距尺度的明显差别,也就是说,定距尺度中没有绝对零点,即使其计量值为“0”,这个“0”也是有客观内容的数值,即“0”水平,而不表示“没有”或“不存在”。例如,某个学生统计学的考试成绩为“0分”,这个“0分”是他的统计学的客观成绩,并不表示他没有考试成绩或没有任何统计学知识;一个地区的气温为 $0^{\circ}\text{C}$ ,这表示一种气温的水平,并不是说没有气温。而定比尺度中绝对零点的“0”,表示“没有”或“不存在”。例如,一个人的身高为“0米”,表示这个人不存在;一个人的收入为“0元”,表示这个人没有收入;一个产品的产量为“0千克”,表示没有这种产品;等等。在现实中的大多数场合,人们使用的都是定比尺度。





## 二、按照数据的性质分类

在统计学中,根据以上四种计量尺度计量的结果及数据的性质,又可以将统计数据分为定性数据(qualitative data)和定量数据(quantitative data)两种类型。

### (一) 定性数据

定性数据又称为品质数据或字符型数据,是说明事物的品质特征表现的具体类别或者观察或实验结果,每个观察结果或实验结果之间没有量的大小区别,表现为互不相容的类别或属性。定性数据不可以用数值大小表示,只能用文字描述,如反映“职业”“教育程度”等现象的属性特点的数据,不能说明具体量的大小和差异,一般不带有度量衡单位。根据观察结果的等级或顺序,可以将定性数据分为定类数据和定序数据两类。

#### 1. 定类数据

定类数据又称名义数据或计数数据,是对事物按照其属性进行分类或分组的计量结果,没有量的特征,只有分类特征。这类数据表现为文字型的无序类别,可以对每一类别出现频数进行计算,没有等级之分,也不能进行加减乘除的数学运算。例如,一个学校的学生按性别分为男生和女生两类,按民族分为汉族和少数民族两类。这种分类把所考察的个体(学生)划分为不同的类型,但各类型之间是平行的关系,没有优劣或大小之分。

#### 2. 定序数据

定序数据又称有序数据或等级数据,同样是按事物属性进行分类或分组,但可排序。也就是说,定序测度不仅可以划分类型,而且可以确定这些类别的优劣或顺序,其数据表现为有序类别,可以进行类别的频数计算和排序,同样不能进行加减乘除的数学运算。例如,通过定序测度可以把企业的信誉分为好、较好、一般和较差四个等级,可以把某种药物的疗效分为轻度疗效、中等疗效和强效等。显然,这些类别具有等级差异,但不能具体测定等级之间差异的大小,这类变量的具体表现就是定序数据。

### (二) 定量数据

定量数据又称数量数据或数值型数据,是说明现象数量特征表现的,能够甚至必须用数值来表现。因这类数据由定距尺度和定比尺度计量形成,故又可细分为区间数据和比率数据。对不同类型的数据,可采用不同的统计方法来处理和分析。例如,对定性数据一般只采用分组法计算和分析各组的频数或频率,而对定量数据则可用更多的统计方法去处理,计算和分析更多的统计指标或统计量,这类数据一般带有度量衡单位。

例如,居民个人收入就以实际的0元作为起点,0元表示没有收入。假如甲的月收入为5 500元,乙的月收入为5 000元,可以说甲的月收入比乙多500元,也可以说甲的月收入是乙的1.1倍。因此,定量数据的每个观察值或实验值之间有量的大小的区别,既可以进行频数计算和排序,又可以进行加减乘除的数学运算。

根据数据的分布情况,可以将定量数据分为连续型数据和离散型数据两类。

#### 1. 连续型数据

连续型数据的取值不是间断的,表现为非单独的整十整百的数字,包含若干位小数且取值密集,相邻两个数值可做无限分割,即可取无限个数值。例如,生产零件的规格尺寸与人



体的身高、体重和胸围等为连续数据,其数值可以用测量或计量的方法获取。这类数据通常是非整数,虽然有时候记载的是整数,如身高的厘米数,但是当提高精确度后,总会出现小数。

## 2. 离散型数据

离散型数据是指其数值只能用自然数或整数单位计算的数据,如企业个数、职工人数、设备台数等,这些数据的数量可能是有限的,也可能是无穷的,无论是有限的离散型数据还是无穷的离散型数据,都可以依据一定的规则和顺序按照计量单位数进行计数。

## 三、按照数据的收集方法分类

按照数据的收集方法不同,数据可分为观测数据和实验数据。观测数据指的是通过直接调查或测量而收集到的数据,可能是全面数据,也可能是样本数据(局部);实验数据指的是通过实验中控制实验对象以及所处的实验环境收集到的数据,观测数据、实验数据一般都是样本数据。

### (一) 观测数据

观测数据是对客观现象进行实地观测所取得的数据,这类数据是在没有对事物进行人为控制的条件下而得到的,社会经济现象的统计数据几乎都是观测数据,在数据取得的过程中,一般没有人为的控制和条件约束。在社会经济问题研究中,观测是取得数据的最主要的方法。

### (二) 实验数据

实验数据一般是指在科学实验环境下取得的数据。在科学实验中,实验环境是受到严格控制的,数据的产生一定是某一约束条件下的结果。在自然科学研究中,实验的方法应用非常普遍,如医药研究实验数据、动植物杂交品种实验数据等。自然科学领域的大多数数据都是实验数据。

## 四、按照数据的时间关系分类

按照数据的时间关系,数据可分为时间序列数据(time series data)、横截面数据和面板数据(panel data)。

### (一) 时间序列数据

时间序列数据又称动态数据,是指在不同时间对同一总体的数量表现进行观察而获得的数据,用于所描述现象随时间变化的情况。这类数据反映了某一事物和现象等随时间的变化状态或程度。例如,我国自改革开放以来到2009年的国内生产总值就是时间序列数据。时间序列数据可做季度数据、月度数据等细分,其中很有代表性的季度时间序列模型就是因为其数据具有四季的变化规律,虽然变化周期不尽相同,但是,整体的变化趋势都是按照周期变化的。

### (二) 横截面数据

横截面数据又称静态数据,是指在同一时间对同一总体内不同单位的数量进行观察而获得的数据。与时间序列数据相比较,其区别在于数据的排列标准不同,时间序列数据是按



时间顺序排列的,横截面数据是按照统计单位排列的。因此,横截面数据不要求统计对象及其范围相同,但要求统计的时间相同,也就是说,必须是同一时间截面上的数据。与时间序列数据完全一样,横截面数据的统计口径和计算方法(包括价值量的计算方法)也应当是可比的。

例如,为了研究某一行业各个企业的产出与投入的关系,我们需要收集同一时间截面上各个企业的产出  $Q$ 、劳动  $L$  和资本投入  $K$  的横截面数据。这些数据的统计对象显然是不同的,因为是不同企业的数据。但是,关于  $Q$ 、 $L$ 、 $K$  的解释、统计口径和计算方法仍然要求相同,即本企业的  $Q$ 、 $L$ 、 $K$  在统计上要可比。

### (三) 面板数据

面板数据,也称平行数据,是指在时间序列上取多个截面,在这些截面上同时选取样本观测值所构成的样本数据。也可以说它是一个  $m \times n$  的数据矩阵,记载的是  $n$  个时间节点上  $m$  个对象的某一数据指标。

面板数据有时间序列和截面两个维度,当这类数据按两个维度排列时,是排在一个平面上的,与只有一个维度的数据排在一条线上有着明显的不同,整个表格像是一个面板,所以通常把 panel data 译作“面板数据”。但是,如果从其内在含义上讲,把 panel data 译作“时间序列-截面数据”更能揭示这类数据的本质特点。也有译作“平行数据”或“TS-CS 数据”(time series and cross section data)的。全国(不含港澳台)各省(区、市)2010—2018 年的 GDP 面板数据如表 3-1 所示。

表 3-1 全国(不含港澳台)各省(区、市)2010—2018 年的 GDP 面板数据 单位:元

省 (区、市)	2018 年	2017 年	2016 年	2015 年	2014 年	2013 年	2012 年	2011 年	2010 年
北京	30 319.98	28 014.94	25 669.13	23 014.59	21 330.83	19 800.81	17 879.40	16 251.93	14 113.58
天津	18 809.64	18 549.19	17 885.39	16 538.19	15 726.93	14 442.01	12 893.88	11 307.28	9 224.46
河北	36 010.27	34 016.32	32 070.45	29 806.11	29 421.15	28 442.95	26 575.01	24 515.76	20 394.26
山西	16 818.11	15 528.42	13 050.41	12 766.49	12 761.49	12 665.25	12 112.83	11 237.55	9 200.86
内蒙古	17 289.22	16 096.21	18 128.10	17 831.51	17 770.19	16 916.50	15 880.58	14 359.88	11 672.00
辽宁	25 315.35	23 409.24	22 246.90	28 669.02	28 626.58	27 213.22	24 846.43	22 226.70	18 457.27
吉林	15 074.62	14 944.53	14 776.80	14 063.13	13 803.14	13 046.40	11 939.24	10 568.83	8 667.58
黑龙江	16 361.62	15 902.68	15 386.09	15 083.67	15 039.38	14 454.91	13 691.58	12 582.00	10 368.60
上海	32 679.87	30 632.99	28 178.65	25 123.45	23 567.70	21 818.15	20 181.72	19 195.69	17 165.98
江苏	92 595.40	85 869.76	77 388.28	70 116.38	65 088.32	59 753.37	54 058.22	49 110.27	41 425.48
浙江	56 197.15	51 768.26	47 251.36	42 886.49	40 173.03	37 756.59	34 665.33	32 318.85	27 722.31
安徽	30 006.82	27 018.00	24 407.62	22 005.63	20 848.75	19 229.34	17 212.05	15 300.65	12 359.33
福建	35 804.04	32 182.09	28 810.58	25 979.82	24 055.76	21 868.49	19 701.78	17 560.18	14 737.12
江西	21 984.78	20 006.31	18 499.00	16 723.78	15 714.63	14 410.19	12 948.88	11 702.82	9 451.26
山东	76 469.67	72 634.15	68 024.49	63 002.33	59 426.59	55 230.32	50 013.24	45 361.85	39 169.92
河南	48 055.86	44 552.83	40 471.79	37 002.16	34 938.24	32 191.30	29 599.31	26 931.03	23 092.36



续表

省 (区、市)	2018年	2017年	2016年	2015年	2014年	2013年	2012年	2011年	2010年
湖北	39 366.55	35 478.09	32 665.38	29 550.19	27 379.22	24 791.83	22 250.45	19 632.26	15 967.61
湖南	36 425.78	33 902.96	31 551.37	28 902.21	27 037.32	24 621.67	22 154.23	19 669.56	16 037.96
广东	97 277.77	89 705.23	80 854.91	72 812.55	67 809.85	62 474.79	57 067.92	53 210.28	46 013.06
广西	20 352.51	18 523.26	18 317.64	16 803.12	15 672.89	14 449.90	13 035.10	11 720.87	9 569.85
海南	4 832.05	4 462.54	4 053.20	3 702.76	3 500.72	3 177.56	2 855.54	2 522.66	2 064.50
重庆	20 363.19	19 424.73	17 740.59	15 717.27	14 262.60	12 783.26	11 409.60	10 011.37	7 925.58
四川	40 678.13	36 980.22	32 934.54	30 053.10	28 536.66	26 392.07	23 872.80	21 026.68	17 185.48
贵州	14 806.45	13 540.83	11 776.73	10 502.56	9 266.39	8 086.86	6 852.20	5 701.84	4 602.16
云南	17 881.12	16 376.34	14 788.42	13 619.17	12 814.59	11 832.31	10 309.47	8 893.12	7 224.18
西藏	1 477.63	1 310.92	1 151.41	1 026.39	920.83	815.67	701.03	605.83	507.46
陕西	24 438.32	21 898.81	19 399.59	18 021.86	17 689.94	16 205.45	14 453.68	12 512.30	10 123.48
甘肃	8 246.07	7 459.90	7 200.37	6 790.32	6 836.82	6 330.69	5 650.20	5 020.37	4 120.75
青海	2 865.23	2 624.83	2 572.49	2 417.05	2 303.32	2 122.06	1 893.54	1 670.44	1 350.43
宁夏	3 705.18	3 443.56	3 168.59	2 911.77	2 752.10	2 577.57	2 341.29	2 102.21	1 689.65
新疆	12 199.08	10 881.96	9 649.70	9 324.80	9 273.46	8 443.84	7 505.31	6 610.05	5 437.47

资料来源:国家统计局数据。

## 模块二 统计数据的来源

在社会生产和经济活动中,从使用者的角度看,统计数据主要来源于两个渠道:一个渠道是通过自己直接调查和科学实验进行数据的采集,这是统计数据的直接来源,也称为第一手或直接的统计数据;另一个渠道是通过其他渠道获取由他人调查或实验所采集到的数据,这是统计数据的间接来源,也称为第二手或间接的统计数据。

### 一、统计数据的直接来源

统计数据的直接来源通常有两个:一个是相关部门或个人专门组织的统计调查,另一个是科学实验。专门调查有统计部门进行的统计调查,也有其他部门或机构为特定目的而进行的调查,如新产品投放市场前所做的市场调查、某个特定地区消费者的收入调查等,这是取得社会经济数据的一个有效手段。科学实验是取得自然科学数据的主要方式。

统计调查是获取统计数据的主要来源,更是取得直接数据的重要手段。常用的统计调查方式主要有普查、抽样调查、重点调查、典型调查和统计报表等。

#### (一) 普查

普查是为了某一特定目的而由相关部门专门组织的一次性全面调查,如人口普查、经济



普查、工业普查和农业普查等。普查是适用于特定目的和特定对象的一种社会调查方式,主要用于收集一定范围内不容易或者不适合用全面统计报表收集的时点现象的统计资料,也可以用于收集时点现象的统计资料,以掌握调查对象的全貌,为有关政策或措施的制定提供依据。

普查是统计调查的组织形式之一,普查资料常被用来说明现象在一定时点上的全面情况。进行重大的全国性的国情和国力普查,有以下优点:领导重视,声势浩大,调查经费容易解决;有利于各部门之间相互配合,共同协作;能够较快地制定出各种分类标准和目录等;统一行动,限期完成;可为普查后做好经常性统计报表和开展抽样调查打下较好的基础。

### 1. 普查的组织方式

普查的组织方式基本上有两种:一种是建立专门的普查机构,配备大量的调查人员,对调查单位进行直接的登记,如人口普查等;另一种是利用基层单位的记录和核算资料,发放调查表,由登记单位填报,如物资库存普查等。

为了满足党和国家研究某些问题或开展某种工作的紧急需要,也可以组织进行一种特殊形式的普查,即“快速普查”,它一般是由组织领导普查工作的最高机关,越过一切中间环节,直接把普查任务下达到基层调查单位,再由登记单位将填报的表格越过中间一些环节直接报送到最高一级机构集中汇总。

### 2. 普查的主要特点

(1) 普查是非连续性调查。由于普查涉及的面较广、调查的单位较多,通常需要耗费大量的人力、物力和财力,而且需要间隔较长的时间才能进行一次。例如,我国每逢尾数为“0”的年份进行人口普查,每逢尾数为“3”的年份进行第三产业普查,每逢尾数为“5”的年份进行工业普查,每逢尾数为“7”的年份进行农业普查,每逢尾数为“1”或“6”的年份进行统计基本单位普查。

(2) 普查规定调查标准时点。为了避免调查数据的重复或遗漏,保证普查结果的准确性,普查通常规定调查资料所属的标准时点。例如,根据《中华人民共和国统计法》和《全国人口普查条例》规定,国务院定于2020年11月1日零时作为开展第七次全国人口普查的标准时间,标准时间一般定为调查对象比较集中、相对变动较小的时间。

(3) 普查的规范程度高。普查是全面调查,数据准确,规范化程度高。普查时,还必须按照统一规定的项目和指标进行登记,不准任意改变或增减,以免影响汇总和综合,降低资料质量。同一种普查,每次调查的项目和指标应力求一致,以便于对比分析历次调查的资料和观察社会经济现象发展变化情况。它可以为抽样调查或其他调查提供基本依据。

(4) 普查主要适用于调查国民经济和社会发展的的重要资料。普查有两种组织形式:一种是组织普查机构、配备普查人员,对调查对象进行直接登记,如我国的人口普查;另一种是调查机关向被调查单位发放普查报表,由被调查单位利用其活动记录和核算资料,结合清仓盘点,自行填报普查资料的调查方式,如我国库存物资普查。



## 知识拓展

## 全国人口普查

全国人口普查由国家来制定统一的时间节点、方法、项目、调查表,严格按照指令依法对全国现有人口普遍地、逐户逐人地进行一次全项调查登记和数据汇总分析报告,普查重点是了解各地人口发展变化、性别比例等。全国人口普查属于国情调查,是当今世界各国广泛采用的收集人口资料的一种最基本的科学方法,是提供全国基本人口数据的主要来源。

2019年11月,经李克强总理签批,国务院印发《国务院关于开展第七次全国人口普查的通知》。根据《中华人民共和国统计法》和《全国人口普查条例》规定,国务院决定于2020年11月1日零时开展第七次全国人口普查。

人口普查工作包括对人口普查资料的收集、数据汇总、资料评价、分析研究、编辑出版等全部过程,它的目的是全面掌握全国人口的基本情况,为研究制定人口政策和经济社会发展规划提供依据,为社会公众提供人口统计信息服务。现代意义的人口普查,是从中华人民共和国成立后才开始的。从1949年至今,中国分别在1953年、1964年、1982年、1990年、2000年和2010年进行过六次全国人口普查。

资料来源:北晚新视觉, <https://www.takefoto.cn/viewnews-1952762.html>,有删改。

## (二) 抽样调查

抽样调查是一种非全面调查,抽样调查是根据随机的原则从总体中抽取部分实际数据进行调查,并运用概率估计方法,根据样本数据推算总体相应的数量指标的一种统计分析方法。抽样调查从研究对象的总体中抽取一部分个体作为样本进行调查,据此推断有关总体的数字特征,经济性好,时效性强,适应面广,准确性高。抽样调查虽然是非全面调查,但它的目的在于取得反映总体情况的信息资料,因而,也可起到全面调查的作用。

## 1. 抽样调查与其他非全面调查的区别

抽样调查与其他非全面调查的区别在于:第一,调查样本是按随机原则抽取的,在总体中,每一个单位被抽取的概率是相等的,因此能够保证被抽中的单位在总体中的均匀分布,不会出现倾向性误差;第二,所抽选的调查样本数量是根据调查误差的要求经过科学计算确定的,调查样本的数量有可靠的保证;第三,抽样调查的误差,在调查前就可以根据调查样本数量和总体中各单位之间的差异程度进行计算并控制在允许的范围以内,调查结果的准确程度较高。

## 2. 抽样调查的分类

根据抽选样本的方法,抽样调查可以分为概率抽样和非概率抽样两类。概率抽样是按照概率论和数理统计的原理从调查研究的总体中根据随机原则来抽选样本,并从数量上对总体的某些特征做出估计和推断,可以从概率意义上对推断出的可能出现的误差加以控制。在我国,人们习惯将概率抽样称为抽样调查。

## 3. 抽样调查的应用

抽样调查的应用范围非常广泛,适用于那些无法进行全面调查的现象,如只有通过破坏



性检验才能确定产品质量的工业产品,或者总体上从理论上讲可以进行全面调查,但实际上不能进行全面调查的事物,就像了解某个森林有多少棵树,等等。对于可以进行全面调查的现象,有时也可以采用抽样调查方法进行,能起到全面调查的作用。由于抽样调查还可以用来检查全面调查的准确性,因此抽样调查被公认为是非全面调查方法中用来推算和代表总体的最完善、最有科学根据的调查方法。

#### 4. 抽样调查的特点

(1) 节省人力、物力、财力和时间。由于调查的样本单位通常是总体单位中很小的一部分,调查的工作量小,因而可以节省大量的人力、物力、财力和时间。

(2) 可以迅速、及时地获得所需要的信息。由于抽样调查的工作量小,调查准备、登记资料、数据处理用时较少,因而可以提高数据的时效性。此外,抽样调查可以频繁地进行,随着事物的发生和发展及时取得有关信息,以弥补普查等全面调查的不足。例如,两次人口普查之间各年份的人口数据都是通过抽样调查取得的。

(3) 适用范围广。抽样调查适用于对各个领域、各种问题的调查。它既适用于全面调查能够调查的对象,又适用于全面调查所不能调查的对象,特别是对一些特殊现象的调查,如产品质量状况调查(破坏性调查如灯泡使用寿命调查、炮弹射程调查等)、客户满意度调查和医药临床试验等。

(4) 调查结果的准确性高。抽样调查的工作量小,能使各环节的工作做得更细致,误差往往很小,其数据质量往往较高。

(5) 可以事先控制误差。抽样调查能够根据调查的目的要求、调查对象的特点及掌握的资料情况,通过选择调查方式和确定样本数目,事先对误差的大小加以控制。



### 在上海,哪类人群的垃圾分类意识最强

2019年6月28日,上海市质量协会公布了垃圾分类调查报告。本次调查通过上海市质协用户评价中心在全市收集有效问卷共计3 000份,调查对象覆盖全市16个区,市民年龄层覆盖18~70周岁。

报告显示,上海学生的垃圾分类意识突出,其中,小学生的垃圾分类意识最强。此外,湿垃圾破袋工作等还需加强,46.9%的市民能准确地做到湿垃圾破袋入桶,年龄越大的市民,这方面的践行率越高。

调查认为,上海的垃圾分类工作厚积薄发,通过整理上海1995年开始至今的垃圾分类事件历程可以发现,上海通过多年的试点、推广、调整和实施阶段,积累了丰富的施行经验,为《上海市生活垃圾管理条例》(以下简称《条例》)的正式实施打下了坚实的基础。

2014年《上海市促进生活垃圾分类减量办法》为地方政府规章,2016年《关于进一步加强本市垃圾综合治理的实施方案》为地方工作文件,2018年《关于建立完善本市生活垃圾全程分类体系的实施方案》为地方规范性文件,至2019年《条例》升级为省级地方性法规,这代表了上海市政府对该条例执行的决心和态度。

调查反映,上海市民对法规的支持率为96.6%,知晓率高达98.7%。其中:81.1%的受



访者会在投递湿垃圾时沥干水分;79.1%的受访者会在投放立体包装物时清空里面的东西,压扁后投放;82.7%的受访者在投放废荧光灯管或节能灯时连带包装或包裹投放。

与此同时,上海自2018年来兴起了面向全市少儿开展的垃圾分类专题培训和实践工作,学校已成为垃圾分类工作的另一个重要宣传阵地。

调查显示,在有孩子的家庭中,91.9%的家长反馈家里的孩子有生活垃圾分类投放的意识,尤其是4~12岁的孩子,垃圾分类意识最强。其中,幼儿园学生的垃圾分类意识高达84.4%,小学生达92.4%,初中生达84.8%。

同时,46.9%的市民能准确地做到湿垃圾拆袋分开投放,48.1%的市民是将装有湿垃圾的垃圾袋一同投放到湿垃圾收集桶。通过数据还可以看出,年龄越大的市民在这方面的践行率越高。

此外,一次性纸杯的使用率仍高,不使用一次性纸杯的市民仅占31.0%。

《条例》第二十一条明确规定:“党政机关、事业单位应当带头使用有利于保护环境的产品、设备和设施,提高再生纸的使用比例,减少使用一次性办公用品,内部办公场所不得使用一次性杯具。政府采购应当按照规定,优先采购可循环利用的产品。鼓励企业、社会团体节约使用和重复利用办公用品,减少使用一次性杯具。”但是调查显示,只有31.0%的受访者表示自己所在的单位不使用一次性纸杯,其中工厂或车间、医院和政府机关、行政办事场所位列前三。

资料来源:搜狐网,[https://www.sohu.com/a/323630581\\_260616](https://www.sohu.com/a/323630581_260616),有删改。

### (三) 重点调查

重点调查是一种在调查对象中选择一部分重点单位进行调查的方法。它是一种非全面调查,既可以用于经常性调查,也可以用于一次性调查。其特点是所选择的调查对象的标志值在所研究的标志总量中或占很大比重,或有较强的代表性,能反映总体的基本状况。例如,要想了解工业贷款的增长原因,只要选择变化较大的有代表性的重点企业进行调查就可以分析清楚。

与抽样调查不同的是,重点调查取得的数据只能反映总体的基本发展趋势,不能用以推断总体,因而也只是一种补充性的调查方法,主要在一些企业集团的调查中运用。例如,为了掌握“三废”排放情况,可以选择对冶金、电力、化工、石油、轻工和纺织等重点行业的工业企业进行调查。重点调查的优点是,花费的力气较小,能及时提供必要的资料,便于各级管理部门掌握基本情况,采取措施。例如,要了解全国铁路运输情况,选择北京、上海、沈阳、郑州、兰州、广州、武汉、成都等枢纽站进行调查,就可以掌握全国铁路运输的基本情况,因为全国铁路客货运输量大多集中在这些枢纽站。

通过重点调查,可以了解到调查对象的基本情况,可以根据所研究对象在标志总量中所占的比重来选择重点调查单位。尽管重点单位的标志值在总体标志总量中占有绝对比重,了解了它们的情况,就等于了解了总体的基本特征,但重点单位的这些情况毕竟不能完整地说明总体总量,不具备推断总体总量的条件。因此,重点调查只适用于获取反映总体基本情况的统计资料,不适合推断总体指标。

### (四) 典型调查

典型调查是指根据调查目的,在对研究对象总体进行全面分析的基础上,有意识地从中





选取若干个总体单位进行系统周密调查研究的一种非全面调查,借以认识事物发展变化的原因和规律,从而认识这一类对象的本质特征、发展规律,找出具有普遍意义和有价值的经验与值得借鉴的教训。例如,要研究工业企业的经济效益问题,可以在同行业中选择一个或几个有代表性的单位做深入细致的调查,以探寻企业经济效益形成的过程、原因和特点。典型调查主要有两种形式:一种是对个别典型单位进行研究性调查,即所谓的“解剖麻雀”式的调查,主要用于探寻事物发生及发展的原因和规律;另一种是先按有关标志将研究对象分类,然后再在各类中选择有代表性的单位进行调查,即所谓的“划类选典调查”。如果典型单位选择适当,不仅能够科学推算总体全面资料,而且能够深入研究事物的发生及发展过程。典型调查有以下几个特点:

(1) 典型调查主要是定性调查。典型调查主要依靠调查者深入基层进行调查,对调查对象进行直接剖析,取得第一手资料,能够透过事物的现象发现事物的本质和发展规律。它是一种定性研究,难以进行定量研究。

(2) 典型调查的关键在于被调查单位的选择。典型调查是根据调查者的主观判断,选择少数具有代表性的单位进行调查。因此,调查者对调查单位的了解情况、思想水平和判断能力对选择典型的代表性起着决定性作用。

(3) 典型调查的方式是面对面的直接调查。典型调查主要依靠调查者深入基层,对调查对象进行直接接触与剖析,因此,对现象的内部机制和变化过程往往了解得比较清楚,取得的资料比较全面、系统。

(4) 典型调查方便、灵活,可以节省时间、人力和经费。典型调查的对象少,调查时间短,反映情况的速度快,调查内容系统周密,了解问题深入,使用调查工具不多,运用起来灵活方便,可以节省很大的人力和财力。



### 新冠肺炎疫情对我国汽车行业造成的冲击

2020年初,一场突如其来的新冠肺炎疫情从湖北武汉席卷大江南北,致使大部分行业停工停产。2月以来,这场疫情对汽车行业的冲击开始显现,各大车企陆续发布销量快报,2月汽车销售情况可谓一片惨淡,不少车企遭遇断崖式下跌。

3月初,某新闻网站对我国汽车行业进行了深入的调查研究,研究报告显示:2020年2月,上海汽车集团股份有限公司(上汽集团,600104)2月的汽车销量为4.73万辆,去年同期为36.29万辆,同比下降86.97%,1—2月累计汽车销量为44.76万辆,同比下降54.07%。比亚迪汽车销量为5501辆,去年同期为2.68万辆,同比下降79.5%,1—2月累计汽车销量为30674辆,同比下降56.65%。

在新能源汽车方面,比亚迪新能源汽车2月的销量为2803辆,去年同期为14429辆。1—2月累计新能源汽车销量为9936辆,同比下降76.95%。北汽新能源汽车2月的销量为1002辆,去年同期为2867辆;1—2月累计汽车销量为3008辆,同比下降59.24%。

疫情冲击下,跨国车企也难逃销量下滑的命运。1月在中国销量同比还为正的本田汽车,2月销量仅为11288辆,同比大跌85.12%;丰田汽车在中国的销量约为23800辆,同比



下降 70.2%；马自达 2 月在中国更是仅售出 2 430 辆汽车，同比下降 78.95%；韩系车企现代、起亚 2 月在中国的销量仅为 3 000 辆，同比大跌 95%，现代汽车 2 月全球销量同比下降 13%，创下了近 10 年来的销量新低。

有分析认为，受新冠肺炎疫情影响，各大车企线下门店的销售受冲击明显，零部件的供应短缺还拖慢了节后的复工节奏，这些都导致了 2 月汽车销量的暴跌。

不过，随着国内疫情逐渐得到控制，车企与经销商逐步复工，消费者的购车活动正慢慢恢复正常。而且，各地政府正在积极出台政策稳定汽车消费，这也都对国内车市形成一定利好。

乘用车市场信息联席会曾预计，2 月全国乘用车市场零售增速初步判断同比下滑达到 80%，估计是 2020 年车市的最大一次下滑，估计 2020 年 1—2 月汽车销量累计同比下降幅度达到 41%，这也是近 20 年来的最大下降。

在行业整体黯淡的形势下，也有好消息，上汽集团海外销量依然保持大幅增长。2 月，上汽集团实现整车出口及海外零售量超 21 000 辆，同比增长 32%，1—2 月累计零售突破 45 000 辆，同比增长约 20%，继续保持全国第一。其中，上汽集团旗下全球化品牌名爵在全球市场延续强劲增势，1—2 月，名爵海外零售超过 25 000 辆，同比猛增 150%。

资料来源：腾讯网，<https://new.qq.com/omn/20200308/20200308A04KJ00.html>，有删改。

## （五）统计报表

统计报表是指按照国家有关法规，自上而下统一布置，各级企事业单位和行政单位按规定的表格形式、内容及时间要求和报送程序，提供统计资料的一种统计调查方式。它是我国取得国民经济和社会发展情况基本统计资料的一种重要手段。统计报表要以一定的原始记录为基础，按照统一的表式、指标、报送时间和报送程序进行填报，因而具有统一性、时效性、系统性和相对准确性等特点。但统计报表的组织方式也有一定的局限性，主要表现为缺乏灵活性和广泛的适应性。

统计报表是我国进行国民经济管理的重要工具。在市场经济条件下，国家各职能部门对国民经济实行宏观调控的过程中，大量经济政策的制定都需要以有关统计报表提供的资料作为决策的依据。通过统计报表取得的统计资料不仅可以反映我国社会主义现代化的建设成绩，而且可以作为分析研究、总结经验、认识规律的依据，还是指导工作、改善经营、加强管理的重要工具。

### 1. 统计报表的分类

（1）统计报表按报送周期长短不同，分为日报、周报、旬报、月报、季报、半年报和年报。各种报表报送时间的长短与内容的详略有一定的关系。一般来说，报表报送的时间越短，报表的项目就应该越少、越简单；反之，报表的项目就应该越多、越详细。日报、周报和旬报称为进度报表，主要用来反映生产、工作的进展情况；月报、季报和半年报主要用来掌握国民经济发展的基本情况，检查各月、季、年的生产工作情况；年报是每年上报一次，主要用来全面总结全年经济活动的成果，检查年度国民经济计划的执行情况等。

（2）统计报表按报送内容和实施范围的不同，分为国民经济基本统计报表、专业统计报表和地方统计报表。国民经济基本统计报表是根据国家统计调查项目和统计调查计划制定



的相应的统计报表,用以收集全国性的经济和社会基本情况,包括农业、工业、基建、物资、商业、外贸、劳动工资和财政等方面最基本的统计资料;专业统计报表也称部门统计报表,是根据有关部门统计调查项目和调查计划制定的相应的报表;地方统计报表是根据有关地方统计调查项目制定的相应的统计报表,是为本地区的计划和管理服务的。

(3) 统计报表按调查范围不同,分为全面统计报表和非全面统计报表。全面统计报表要求调查对象中的每一个单位都要填报,非全面统计报表只要求调查对象的一部分单位填报。

(4) 统计报表按填报单位不同,分为基层统计报表和综合统计报表。基层统计报表是由基层企事业单位填报的报表,综合统计报表是由主管部门或部门根据基层报表逐级汇总填报的报表。综合统计报表主要用于收集全面的基本情况,也常为重点调查等非全面调查所采用。

## 2. 统计报表的内容

(1) 表式。统计报表的表式是统计报表的主体,统计调查资料是通过这些表式的填报而取得的。表式的主要内容是主栏项目、宾栏指标及补充资料项目等。此外,每张统计报表还列有表名、表号、填报单位、日期以及报送单位的负责人和填报人的签名等。

(2) 填表说明。填表说明的内容包括:填报范围,它要求明确规定每种统计报表的填报人(即填报单位),各级主管部门和统计部门的综合范围(即汇总时应包括哪些单位);指标解释,就是对列入表式的统计指标的概念、计算范围及其他有关问题的具体说明;分类目录,就是统计报表主栏中应进行填报的有关项目一览表,它是填报单位进行填报的重要依据;对其他有关事项的规定,如报表日期、受表机关和报送份数等。

## 3. 统计报表的资料来源

统计报表的资料来源于基层的原始记录,从原始记录到统计报表,中间还要经过统计台账和内部报表。因此,建立健全原始记录和统计台账,完善基层单位的基本统计核算是保证统计报表质量的重要工作,应落实做好。

原始记录是基层单位对各项业务活动所做的经常性记载,如工厂产品的质量记录、机械设备工作日记、原材料消耗记录、工作出勤和工时记录、粮食入库单等都是原始记录。

统计台账是基层单位根据编制报表和本单位经营工作的需要,按时间顺序设置的用来集中登记原始记录的表册,如劳动工时台账、设备台账和产品台账等。

原始记录是基层单位进行生产管理、日常工作所不可缺少的工具,统计台账则是积累资料、研究经营管理的重要依据和编制统计报表的可靠资料来源。做好原始记录和统计台账这两项基础工作,对提高统计数字的真实性和统计工作的质量有着重要的作用。

## 4. 统计报表的特点

(1) 有利于基层单位对生产、经营活动进行监督和管理。由于统计报表的基础是企事业单位的经济活动记录,因此,企事业单位必须建立健全原始记录,从而有利于基层单位对其生产及经营活动进行全面监督和管理。

(2) 便于领导部门掌握其管辖范围内的基本情况。由于统计报表是逐级汇总上报的,因此各级领导部门能够获得其管辖范围内的报表资料,从而了解本部门、本地区社会经济发展的全面情况和问题。

(3) 便于积累历史资料,有利于开展动态分析。由于统计报表一般是经常性调查,且调



查项目比较稳定,因此可以根据需要积累研究对象较长时期的发展资料,从而形成时间序列数据,以分析现象发展变化的速度和规律。

## 二、统计数据的间接来源

在统计数据的收集过程中,有时候难以通过直接调查或实验取得所需的第一手数据。此时可以通过一定渠道获取别人调查或科学实验所取得的统计数据,这便是第二手资料,也就是间接来源数据。

如果与研究内容有关的原信息已经存在,我们只是对这些原信息重新加工、整理,使之成为我们进行统计分析可以使用的数据,这种数据就是间接来源数据。从收集的范围看,这些数据可以取自系统外部,也可以取自系统内部。系统外部获取数据的主要渠道有:统计部门和各级政府部门公布的有关资料,如定期发布的统计公报、定期出版各类统计年鉴;各类经济信息中心、信息咨询机构、专业调查机构、各行业协会和联合会提供的市场信息与行业发展的数据情报;各类专业期刊、报纸、书籍所提供的文献资料;各种会议,如博览会、展销会、交易会及专业性、学术性研讨会上交流的有关资料;从互联网或图书馆查阅到的相关资料;等等。随着计算机网络技术的发展和普及,通过网络来获取所需的各种数据资料成为获取间接统计数据的一种重要渠道。使用间接统计数据时,应注意数据的含义、计算口径和计算方法,以免误用或滥用,同时要注明数据的来源,以尊重别人的劳动成果。

### 知识拓展

#### 中国统计年鉴

《中国统计年鉴》是国家统计局编印的一种资料性年刊,它全面反映了中华人民共和国经济和社会发展情况。正常年度每年一刊,刊名为《××××中国统计年鉴》,如《2019中国统计年鉴》。每年的统计年鉴收录上一年全国和各省、自治区、直辖市经济和社会各方面的统计数据以及历史重要年份和近20年的全国主要统计数据,由国家统计局每年出版发行,是我国最全面、最具权威性的综合统计年鉴。

年鉴正文内容一般分为20余个篇章,于不同年份根据经济社会发展的不同情况略有调整,如综合,人口,国民经济核算,就业和工资,价格,人民生活,财政,资源和环境,能源,固定资产投资,对外经济贸易,农业,工业,建筑业,批发的零售业,运输、邮电和软件业,住宿、餐饮业和旅游,金融业,房地产,科学技术,教育,卫生和社会服务,文化和体育等主要社会经济指标。附录为一个篇章,为国际主要社会经济指标。为方便读者使用,各篇章前设有“简要说明”,对该篇章的主要内容、资料来源、统计范围、统计方法及历史变动情况予以简要概述,篇末设有“主要统计指标解释”。

资料来源: <https://www.bilibili.com/read/cv5725140>,有删改。

## 模块三 统计分析

统计分析的数学理论基础在20世纪初期就已经确立,但直到第二次世界大战以



后,随着计算机的出现并广泛应用,其实际操作才成为可能,数据分析的方法才得以被推广。统计分析的方法和过程是数学与计算机科学相结合的产物。

## 一、统计分析的定义和作用

统计分析是指用适当的统计分析方法对收集到的大量数据进行分析,将它们加以汇总和理解并消化,以求最大化地开发数据的功能,发挥数据的作用。统计分析是为了提取有用信息和形成结论,而对数据加以详细研究和概括总结的过程。

如前所述,数据也叫观测值,是通过实验、测量、观察、调查等方法所得到的结果。统计分析中所处理的数据分为定性数据和定量数据,只能归入某一类而不能用数值进行测度的数据称为定性数据。定性数据中表现为一定类别但不区分顺序的,是定类数据,如性别、品牌等;定性数据中表现为一定类别但区分顺序的,是定序数据,如学历、商品的质量等级等。

统计分析的目的是把隐藏在一大批看来杂乱无章的数据中的信息集中和提炼出来,从而找出所研究对象的内在规律。在实际应用中,统计分析可帮助人们做出判断,以便采取适当行动。统计分析是有组织、有目的地收集数据、分析数据,使之成为信息的过程。这一过程是质量管理体系的支持过程。在产品的整个寿命周期,包括从市场调研到售后服务和最终处置的各个过程都需要适当运用数据分析过程,以提升有效性。例如,设计人员在开始一个新的设计以前,要通过广泛的设计调查分析所得的数据,以判定设计方向,因此,数据分析在工业设计中具有极其重要的地位。

## 二、统计分析的方法

在社会生活实践中,常用的统计分析方法主要有聚类分析(cluster analysis)、因子分析(factor analysis)、相关分析(correlation analysis)、对应分析(correspondence analysis)、回归分析(regression analysis)和方差分析 analysis of variance, ANOVA 等。

### 1. 聚类分析

聚类分析是指将物理或抽象对象的集合分组变成由类似的对象组成的多个类的分析过程。聚类是将数据分类到不同的类或者簇的过程,同一个簇中的对象有很大的相似性,而不同簇间的对象有很大的相异性。聚类分析是一种探索性的分析,在分类的过程中,人们不必事先给出一个分类的标准,聚类分析能够从样本数据出发,自动进行分类。因为聚类分析所使用方法的不同,所以常常会得到不同的结论。不同研究者对于同一组数据进行聚类分析,所得到的聚类数未必一致。

### 2. 因子分析

因子分析是指研究从变量群中提取共性因子的统计技术。因子分析就是从大量的统计数据中寻找内在的联系,减少决策的困难。因子分析的方法有十几种,如重心法、影像分析法、最大似然估计法和最小平方法等。这些方法本质上类似,都是以相关系数矩阵为基础的,所不同的是相关系数矩阵对角线上的值,采用不同的共同性估值。在社会学研究中,因子分析常采用以主成分分析为基础的反复法。

### 3. 相关分析

相关分析研究的是现象之间是否存在某种依存关系,并对具体有依存关系的现象探讨



其相关方向以及相关程度。相关关系是一种具有非确定性的关系,例如,以  $X$  和  $Y$  分别记一个人的身高和体重,或分别记每公顷<sup>①</sup>小麦的施肥量与每公顷小麦的产量,显然,  $X$  与  $Y$  有关系,但其中的一项无法精确地决定另一项,这就表明  $X$  与  $Y$  是相关关系。

#### 4. 对应分析

对应分析也称关联分析、R-Q 型因子分析,它通过分析由定性变量构成的交互汇总表来揭示变量间的联系。通过对应分析可以揭示同一变量的各个类别之间的差异以及不同变量各个类别之间的对应关系。对应分析的基本思想是将一个联列表的行和列中各元素的比例结构以点的形式在较低维的空间中表示出来。

#### 5. 回归分析

回归分析是研究一个随机变量  $Y$  对另一个变量  $X$  或一组变量  $(X_1, X_2, \dots, X_k)$  的相依关系的统计分析方法。回归分析是确定两种或两种以上变量间相互依赖的定量关系的一种统计分析方法,其运用十分广泛。回归分析按照涉及的自变量的多少,可分为一元回归分析和多元回归分析;按照自变量和因变量之间的关系类型,可分为线性回归分析和非线性回归分析。

#### 6. 方差分析

方差分析又称“变异数分析”或“F 检验”,是费雪发明的,用于两个及两个以上样本均数差别的显著性检验。由于各种因素的影响,研究所得的统计数据有波动。造成统计数据波动的原因可分成两类:一类是不可控的随机因素,另一类是研究中施加的对结果形成影响的可控因素。方差分析是从观测变量的方差入手,研究诸多控制变量中哪些变量是对观测变量有显著影响的变量。

### 三、统计数据分析的步骤

在科学研究和具体社会生产实践中,完整的统计数据分析主要包括六个步骤,依次为:分析设计、数据收集、数据处理、数据分析、数据展现、报告撰写。统计数据分析的六个步骤也被称作数据分析的六部曲。

#### 1. 分析设计

在进行统计数据分析工作时,首先需要明确统计数据分析目的。只有明确了分析的目的,统计数据分析才能够沿着正常的轨道进行,不至于偏离方向,否则,得出的统计数据分析结果不仅没有指导意义,而且可能会使人陷入分析的误区。

明确了统计数据分析目的之后,需要对思路进行梳理分析,并搭建统计数据分析的工作框架,需要对分析目的进行层层分解,分解为若干个不同的分析要点,进一步明确该如何具体开展数据分析,需要从哪几个角度进行分析,采用哪些分析指标,采用哪些逻辑思维,运用哪些理论依据,等等。

明确统计数据分析目的以及确定分析思路,是确保数据分析过程有效进行的先决条件,它可以为数据收集、处理以及分析提供清晰的指引方向。

<sup>①</sup> 1 公顷=10 000 平方米。



## 2. 数据收集

数据收集是指根据系统自身的需求和用户的需要,按照确定的数据分析框架,收集相关数据的过程,它为数据分析提供了素材和依据。这里的数据包括一手数据与二手数据:一手数据主要指可直接获取的数据,如公司内部数据库、市场调查取得的数据等;二手数据主要指经过加工整理后得到的数据,如统计局在互联网上发布的数据、公开出版物中的数据等。数据收集的方法主要有当面调查询问法、电话调查法、会议调查询问法、邮寄调查询问法和问卷调查询问法等。

## 3. 数据处理

数据处理是指从大量的原始数据中抽取出有价值的信息,即将数据转换成信息的过程。数据处理主要是对所输入的各种形式的数据进行加工整理,其包含对数据的收集、存储、加工、分类、归并、计算、排序、转换、检索和传播的演变与推导的全过程。数据处理贯穿社会生产和社会生活的各个领域。数据处理技术的发展及其应用的广度和深度,极大地影响了人类社会发展的进程。

数据处理的基本目的是从大量的、杂乱无章的、难以理解的数据中抽取并推导出对解决问题有价值和有意义的信息。如果数据本身存在错误,那么,即使采用最先进的数据分析方法,得到的结果也是错误的,不具备任何参考价值,甚至还会误导决策。数据处理的方法主要包括数据清洗、数据转化、数据抽取、数据合并、数据计算等。一般的数据都需要进行一定的处理才能用于后续的数据分析工作,即使再“干净”的原始数据,也需要先进行一定的处理才能使用。

数据处理对数据(包括数值的和非数值的)进行分析和加工的技术过程包括对各种原始数据的分析、整理、计算、编辑等的加工和处理。数据处理比数据分析含义广。随着计算机的日益普及,在计算机应用领域中,数值计算所占的比重很小,通过计算机数据处理进行信息管理已成为主要的应用,如测绘制图管理、仓库管理、财会管理、交通运输管理,技术情报管理、办公室自动化等。在地理数据方面,既有大量自然环境(土地、水、气候、生物等)数据,也有大量社会经济(人口、交通、工农业等)数据,常要求进行综合性数据处理,故须建立地理数据库,系统地整理和存储地理数据,减少冗余,发展数据处理软件,充分利用数据库技术进行数据管理和处理。数据处理流程可以概括为四步,分别是采集、导入和预处理、统计和分析以及挖掘。

## 4. 数据分析

数据分析是指用适当的分析方法及工具对收集来的数据进行分析,提取有价值的信息,形成有效结论的过程。在确定数据分析思路的阶段,数据分析师就应当为需要分析的内容确定合适的数据分析方法。到了数据分析阶段,就能够驾驭数据,从容地进行分析和研究。

一般的数据分析我们可以通过 Excel 完成,而高级的数据分析就要采用专业的分析软件才能完成,如数据分析工具 SPSS、SAS、Python、R 语言等。

## 5. 数据展现

数据展现也称数据可视化,是指通过数据分析,使隐藏在数据内部的关系和规律逐渐浮现出来。那么通过什么方式展现出这些关系和规律,才能让人一目了然呢?一般情况下,数据是通过表格和图形的方式来呈现的,即用图表说话。常用的数据图表包括饼图、柱形图、



条形图、折线图、散点图、雷达图等,当然也可以对这些图表进一步整理加工,使之变成我们所需要的图形,如金字塔图、矩阵图、瀑布图、漏斗图、帕雷托图等。常见的数据可视化方法详见本书项目五。

多数情况下,人们更愿意接受图形这种数据展现方式,因为它能更加有效、直观地传递分析师所要表达的观点。一般情况下,能用图说明问题的,就不用表格,能用表格说明问题的,就不用文字。

## 6. 报告撰写

数据分析报告其实是对整个数据分析过程的一个总结与呈现。通过报告,把数据分析的起因、过程、结果及建议完整地呈现出来,以供决策者参考。数据分析报告是通过全方位的科学分析来评估企业运营质量,为决策者提供科学、严谨的决策依据,以降低企业的运营风险,提高企业的核心竞争力。

一份好的分析报告,首先需要有一个好的分析框架,并且层次明晰、图文并茂,能够让读者一目了然。结构清晰、主次分明,可以使读者正确理解报告内容;图文并茂,可以令数据更加生动活泼,提高视觉冲击力,有助于读者更形象、直观地看清楚问题和结论,从而产生思考。其次,需要有明确的结论,没有明确结论的分析称不上分析,同时也失去了报告的意义,因为最初就是为寻找或者求证一个结论才进行分析的,所以千万不要舍本逐末。最后,一定要有建议或解决方案,作为决策者,不仅仅要找出问题,更重要的是获取建议或解决方案,以便他们在决策时参考。所以,数据分析师不光需要掌握数据分析方法,还要了解和熟悉业务,这样才能根据发现的业务问题,提出具有可行性的建议或解决方案。



### 啤酒与尿布

“啤酒与尿布”的故事产生于20世纪90年代的美国沃尔玛超市中,沃尔玛的超市管理人员在分析销售数据时发现了一个令人难以理解的现象:在某些特定的情况下,“啤酒”与“尿布”两件看上去毫无关系的商品会经常出现在同一个购物篮中,这种独特的销售现象引起了管理人员的注意,经过后续调查发现,这种现象常出现在年轻的父亲身上。

在美国有婴儿的家庭中,一般是母亲在家中照看婴儿,年轻的父亲前去超市购买尿布。父亲在购买尿布的同时,往往会顺便为自己购买啤酒,这样啤酒与尿布这两件看上去不相干的物品就会经常出现在同一个购物篮中。如果这个年轻的父亲在卖场只能买到两件商品之一,则他很有可能会放弃购物而到另一家商店,直到可以同时买到啤酒与尿布为止。沃尔玛的工作人员发现了这一独特的现象,便开始在卖场尝试将啤酒与尿布摆放在相同的区域,让年轻的父亲可以同时找到这两件商品并很快地完成购物;而沃尔玛超市也可以让这些客户一次购买两件商品而不是一件,从而获得很可观的商品销售收入,这就是“啤酒与尿布”故事的由来。

“啤酒与尿布”故事的依据是商品的相关性(也称关联性),商品的相关性是指商品在卖场中不是孤立的,不同商品在销售中会形成相互影响的关系(也称关联关系),比如“啤酒与尿布”故事中,尿布会影响啤酒的销量。在卖场中,商品之间的关联关系比比皆是,比如咖啡的销





量会影响到咖啡伴侣、方糖的销量,牛奶的销量会影响面包的销量,等等。

事物之间的相关性是指当一个事物变化时,另一个事物也会发生变化。当事物之间的变化相互抵消时,比如猪肉价格上涨,猪肉销量下降,我们称这种相关性是负相关;当事物之间的变化向同一个方向发展时,如气温上升,冷饮销量也上升,我们称这种相关性是正相关。有些事物的相关性显而易见,有些则不是那么明显。美国华尔街股票分析师将女性超短裙的长度与道琼斯股票指数建立了关联,超短裙的长度与道琼斯股票指数成反比,据说十分灵验,这就是相关性在生活中的种种体现。

商店中商品的关联性更是比比皆是,如烟酒销售的关联关系:当门店附近有建筑工地时,低档烟、酒的销量就会上升;当附近有高档社区时,中华烟、葡萄酒的销量就会上升。提到商品相关性,很多人认为就是数据分析的事,其实,对于商品相关性来说,更重要的是客户心理层面的因素,毕竟是在提着购物篮,而不是猴子。由于客户在购物时的心理行为是产生商品之间关联关系最基本的原因,因此,在找到购物篮规律时,必须要从客户消费心理层面解释这些关联关系,否则“啤酒与尿布”会永远停留在啤酒与尿布两件商品身上,而没有任何推广意义。要想详细了解由商品相关性形成的客户心理因素,还需进行大量的客户消费行为观察,构建客户购物篮场景。