

第一章 导论



学习目的

- 了解统计理论和实践活动的产生与发展。
- 领会统计工作、统计资料和统计学的含义,描述统计学和推断统计学的区别与联系。



重点难点提示

- 本章介绍统计、统计学及相关概念,为以后章节的学习打下基础,勾勒了本课程的框架结构——描述统计学和推断统计学。
- 本章重点是统计的含义、统计学的基本术语及统计学的分科。

在日常工作和生活中,人们经常接触各种各样有用的宏观、微观经济信息,从一国的国内生产总值、经济增长速度、人口规模与结构、居民收入支出情况、物价指数到企业的产值利润、商品库存、产品市场占有率再到高校的招生就业信息等,所有这些数据信息都离不开统计实践活动,而我们在自然科学研究及社会经济预测和管理决策中,也必须借助相关的统计分析方法,建立统计模型进行定量研究。统计作为一门数据处理技术的方法论的科学,在自然科学和社会科学领域发挥着日益重要的作用。本章主要介绍统计及统计学的含义、统计学的分科、统计学的基本术语及本教材的篇章结构。

第一节 统计学的产生和发展

一、统计的含义

英语中“统计”(statistics)一词,源于古希腊语 state(城邦)、拉丁语 status(状态)以及意大利语 stato(国家)和 statisti(政治家)。可见,历史上“统计”一词的含义与国家的政治经济活动密切相关,它是用来描述和计量一国重要政治经济事项的方法。在汉语中,从朴素直观的角度理解,“统计”是对大量事物的数量进行的汇总和计数,即“统而计之”的意思。随着社会经济的发展,人们对统计的功能提出了更高的要求,“统计”一词的内涵和外延都发生了巨大的变化。从统计实践角度看,统计是对客观事物的数量表现、数量关系和数量变化进行描述和推断的计量活动,这种计量活动的直接成果,就是反映客观事物数量特征和发展规律的各种数据资料,即统计资料。而对统计实践过程中积累的丰富经验加以科学总结和理论概括,就是统计科学。所以,从不同的角度理解,“统计”一词主要有三层含义:统计工作、统计资料及统计学。

统计工作即统计实践活动,是人们对客观事物的数据资料进行收集、整理、分析的工作活动的总称。

统计工作是“统计”一词最基本的含义,没有统计工作就不会生成统计资料,而没有丰富的统计实践活动,也不会产生统计科学。

统计资料是统计工作的成果,包括各种统计报表、统计图形及文字资料等。

如国家统计局及地方统计局定期发布的国民经济和社会发展公报、各种统计年鉴等,都属于统计资料。

统计学是一门收集、整理、描述、显示和分析统计数据的方法论的科学,其目的是探索数据的内在数量规律性,以达到对客观事物的科学认识。

统计工作、统计资料与统计学三者之间既有所区别,又紧密联系,是一个事物的三个方面。统计工作和统计资料是过程与结果的关系,统计工作过程的好坏直接影响统计资料的质量。统计学与统计工作是理论与实践的关系,理论来源于实践,统计学是统计工作实践与经验的科学总结,它的基础是统计工作实践,但反过来指导实践的发展,为统计工作提供理论和方法指导。

举一反三

举出生活中的统计实践和统计资料的实例。

二、统计实践活动的产生与发展

统计实践活动至今已有四五千年的历史,是随着社会经济的发展和治国管理的需要而产生和发展起来的。

(一)统计实践活动的起源

早在原始社会时期,就有结绳记事、刻契记事的方法,即所谓“事大,大结其绳;事小,小结其绳;结之多少,随物众寡”。这可以说是统计的萌芽。据晋朝皇甫谧(公元215~282年)在《帝王世纪》中记载,早在4 000多年前的夏朝,为了治国治水的需要,就进行过初步的国情统计,查明当时全国人口为13 553 923人,土地为24 308 024公顷,并依山川土质、人口物产及贡赋多寡等,将全国分为九州。在埃及,早在建造金字塔时,也为征集所需财物和劳力而对全国人口、劳力和财力进行过调查。古希腊在公元前600年就进行过人口普查,古罗马在公元前400年就建立了人口的出生和死亡登记制度。这些都是人类早期的统计活动。

(二)统计实践活动的发展

在整个奴隶社会和封建社会时期,由于生产力水平低下、发展缓慢,统计实践活动局限于为适应奴隶主和封建王朝实现税赋、扩张、征兵等需要而进行人口、土地、财产等的登记、汇总工作。17世纪后,资本主义在欧洲萌芽并迅速发展,简单的汇总计数的统计活动已经不能适应资本主义国家的管理需要,必须辅之以统计调查、整理、计算分析等各种手段,为经济和社会发展提供综合数据资料。统计实践随着社会生产力的发展得到推动和发展。统计工作也由人口、财产统计等领域,逐步扩展到工业、贸易、运输业、保险业等其他领域。

统计在治国与管理中的重要作用,引起各国政府的高度重视,许多国家都建立了统计调查

和报告制度,相继成立专门的统计机构,开展专项统计工作。统计实践的发展,催生了一批统计学术团体和刊物,它们将统计实践经验加以总结和升华,形成系统的统计理论和方法,又反过来指导统计实践活动。

1853年,在比利时首都布鲁塞尔召开了第一届国际统计会议,意味着统计逐渐成为一项国际性的事业。1995年8月,在北京召开了第50届国际统计年会,标志着我国的统计理论和实践已发展到较高的水平。统计在我国社会经济生活中的地位日益重要,已成为认识客观世界的有力武器和国家宏观管理的重要工具,也成为企业、个人等微观经济主体经营与决策的得力助手。

三、统计科学的发展

随着统计实践活动的深入发展和统计实践经验的日益丰富,加之数学及其他科学发展的推动,人们开始系统总结统计工作的经验,逐渐形成了比较科学系统的统计理论与方法,这便是统计科学。在统计学300多年的发展史上,经历了古典统计学、近代统计学和现代统计学三个阶段,每个阶段都出现过一些重要的学派。

(一)古典统计学时期

这一阶段是指17世纪中后期至18世纪中后期的统计学萌芽时期,比较有代表性的是政治算术学派和记述学派。

1. 政治算术学派

政治算术学派发源于英国伦敦,产生于17世纪中叶,代表人物是英国的威廉·配第(1623~1687年)和约翰·格朗特(1620~1674年)。该学派以威廉·配第所著的《政治算术》一书而命名。约翰·格朗特在1662年发表的《关于死亡表的自然和政治的考察》一文中,采用大量观察法,以当时伦敦人口的出生率、死亡率和性别比例等人口现象为依据,揭示了人口现象中的某些规律性事实,这是政治算术学派的开篇论文。

1676年,威廉·配第的《政治算术》一书问世,标志着社会经济统计学的一个分支学派的初步创立。在书中,配第用算术方法和大量的统计资料,对英国、法国和荷兰三国的经济、军事和政治实力从数量上进行了系统的比较和分析。威廉·配第“用数字、重量和尺度来说话”,将自己首创的数量对比和分析方法运用于宏观政治经济的分析和说明,因此被马克思尊称为“政治经济学之父”。虽然威廉·配第没有使用“统计学”这一名词,但他使用的数量对比和分析方法奠定了统计学的方法论基础,因此后人将政治算术学派看作“有实无名”的统计学派。

2. 记述学派

记述学派又称国势学派,创始人是德国赫姆斯特大学教授赫尔曼·康令(1601~1681年)和哥丁根大学教授哥特弗里德·阿亨瓦尔(1719~1772年)。这一学派认为,“统计学是研究一国或多数国家的显著事项之学”。阿亨瓦尔在其主要代表作《近代欧洲各国国势学概论》中强调,国势学是以现实的国家的领土、人口、财产、贸易、货币、阶级、政治制度等领域为研究对象,采用记述的方法,通过文字描述方式,罗列出各国的显著事项。由于德文中“国势”与“统计”一词词源相通,故这一学派以统计学命名。但由于它缺乏数量分析的方法和结论,缺乏现代统计学所研究的实质内容,因此后人称之为“有名无实”的统计学派。

3. 图表学派

图表学派的代表人物是丹麦的安彻逊(1700~1765年)和德国的克罗姆(1753~1833年)。

1741年,安彻逊第一次编纂了欧洲15个国家的比较统计表,用数字对欧洲各国的土地、人口、财政、教育、军队状况等进行了系统的比较。1782年,克罗姆第一次用几何图形来表现统计资料,以统计图表来表现和保存统计资料的图表学派从此诞生。

(二)近代统计学时期

近代统计学时期是指18世纪末到19世纪末这100多年的统计学发展阶段,其中代表性学派是数理统计学派和社会统计学派。

1. 数理统计学派

数理统计学派的主要创始人是比利时的统计学家阿道夫·凯特勒(1796~1874年)。凯特勒首先将概率论原理引入社会经济现象的研究中,在《社会物理学》、《统计学的研究》、《论人类》等著作中研究人口统计、道德与犯罪率统计等问题。他最先运用大数定律论证了社会生活现象并非偶然,而是有其发展规律可循的。此外,他还利用概率论原理提出了“平均人”的概念,即具有平均身高、平均体重、平均智力和平均道德水准的典型人或抽象人,并认为若社会上现实人与平均人的差异越小,社会矛盾就越缓和。

凯特勒在统计学发展中的主要功绩是将概率论引入社会现象的研究中,初步完成了统计学与概率论的结合。1867年,这门兼有数学和统计学双重性质的学科被命名为“数理统计学”,凯特勒因此被欧美统计学界誉为“近代统计学之父”。

2. 社会统计学派

社会统计学派于19世纪产生于德国,主要代表人物有乔治·蓬·梅尔(1841~1925年)和厄·恩格尔(1821~1896年)。他们认为,统计学是一门社会科学,是研究社会现象变动原因和规律的实质性科学。社会统计学派最大的贡献在于厄·恩格尔提出的衡量家庭生活质量高低和富裕程度的恩格尔系数。该学派认为,统计研究的是社会现象总体而不是个别现象,主张采用大量观察法等统计研究方法,对社会经济现象的数量方面及其发展规律进行研究,揭示社会经济现象发展变化的因果关系。

社会统计学派对统计学发展的贡献在于:从研究对象上看,它更关注社会现象总体;从研究方法看,它主要采用大量观察法,这构成了“实质性”科学的两大特点。

社会经济在发展,社会科学本身也在不断向专业化和定量化的方向发展,这都要求统计能提供更多、更有效的调查、整理和分析资料的方法和手段,社会统计学派在上述力量的推动下,逐步从实质性科学向方法论科学转化。该学派的理论和方法经过不断发展和完善,逐渐形成目前的社会经济统计学。

(三)现代统计学时期

现代统计学时期是指自20世纪初至今的数理统计学时期。这一时期发展的主要成果是数理统计在随机抽样基础上建立起来的推断统计学,它是一种建立在随机抽样基础上的推断总体特征的方法,起源于英国数学家哥塞特(1876~1936年)的小样本 t 分布理论。其后得到费雪(1890~1962年)的充实及尼曼、毕尔生等人的发展,逐步建立了统计假设理论。

20世纪60年代后,数理统计学的发展有三个明显的特点:一是随着数学的发展,数理统计学越来越广泛地应用数学方法;二是数理统计学的分支(如抽样理论、非参数统计、多变量分析等)或以数理统计学为基础的边缘学科(如计量经济学、天文统计学、统计力学等)不断形成;三是借助于电子计算机等现代化手段,数理统计学的应用日益广泛而深入。

第二节 统计学的分科

从学科性质分析,统计学是一门研究数据收集、整理、计算、分析及推断的方法论的科学,其目的在于通过对客观现象内在数量规律性的探索,达到对客观事物的科学认识。随着统计实践活动的发展,各种统计方法广泛应用于自然科学和社会科学领域,统计学也已发展成为由若干分支学科组成的学科体系。从统计方法的构成角度看,统计学可以分为描述统计学和推断统计学;从统计方法的研究和应用角度看,统计学又可以分为理论统计学和应用统计学。

一、从统计方法的构成角度分

(一)描述统计学

描述统计学(descriptive statistics)是研究如何取得反映客观现象的数据资料,对所收集的数据进行加工整理并通过统计图表等形式显示出来,进而通过综合、概括与分析得出反映客观现象的规律性数量特征的方法。具体内容包括统计数据的收集与整理、数据的显示方法、数据分布特征的描述与分析方法等。

(二)推断统计学

推断统计学(inferential statistics)是研究如何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分析的基础上,对统计总体未知的数量特征做出以概率论为基础的推断和估计。具体内容包括抽样估计、假设检验、方差分析及相关和回归分析等。

(三)描述统计学与推断统计学的关系

描述统计学和推断统计学的划分,一方面是统计方法发展过程的反映,同时也反映了应用统计方法研究客观事物数量规律性的不同过程。从一个完整的统计研究过程看,起始于统计数据的收集与调查,结束于探索出客观现象总体内在的数量规律性。如果在统计数据的收集与调查中得到的是总体数据(如根据普查形式得到的数据),则经过描述统计过程之后就可直接得出总体的数量特征和规律性;如果收集与调查到的仅是部分数据(如通过抽样调查获取的数据),则经过描述统计只能得出样本总体的数量规律性,还必须借助于概率论等数理统计知识对总体数量特征做出科学的推断和估计,或先对总体的某一数量特征做假设,再利用样本数据构造样本统计量,对假设的正确与否做出推断。

描述统计学和推断统计学是现代统计学的两个组成部分。其中,描述统计学是现代统计学的基础和统计研究工作的第一步,推断统计学则是现代统计学的核心和统计工作的关键。从描述统计发展到推断统计,既是统计学发展的巨大成就,也是统计学向纵深发展和成熟的标志。由于在统计研究中获取的一般都是样本数据,仅靠描述统计无法得知总体的数量特征,因此,推断统计在现代统计中的地位和作用日益重要。但另一方面,若没有客观、准确、可靠的数据资料,再科学的推断方法也难以得出可靠的结论。可见,描述统计和推断统计相辅相成、缺一不可。

二、从统计方法的研究和应用角度分

(一)理论统计学

理论统计学(theoretical statistics)是利用数学原理研究统计学的一般理论和方法的统计学。由于数学知识在现代统计学中广泛而深入的应用,特别是概率论在统计推断中的理论基

基础地位,要求从事统计理论和方法研究的人员具有扎实的数学基础。理论指导实践,理论统计学为统计方法提供理论基础,没有理论统计学的发展,统计学是不可能发展成为目前如此完善的学科体系的。但在统计研究领域,从事理论统计研究的人员相对很少,绝大部分人员致力于应用统计学的研究。

(二)应用统计学

应用统计学(applied statistics)是研究如何应用统计方法解决实际问题,即如何根据收集的数据资料分析和解决问题的统计学。应用统计学大多是以数理统计为基础形成的一些边缘学科,在自然科学和经济社会科学领域都有广泛的应用。例如:在自然科学领域,统计方法在生物学中的应用形成了生物统计学,在天文学中的应用形成了天文统计学,在物理学中的应用形成了统计力学,在农业科学实验、育种方面形成了农业统计学;在经济和社会科学领域,运用数学方法、数学模型来研究经济现象数量变化规律形成了计量经济学,统计学在管理领域的应用形成了管理统计学,在人口学中的应用形成了人口统计学,在犯罪学领域的应用形成了犯罪统计学等。可以说,统计学已经渗透到自然科学和社会经济生活的各个方面,形成了不同的统计学分支,尽管应用领域各异导致这些应用统计学的特殊性,但其基本原理相同,都是数学和概率理论,基本方法都是描述统计学和推断统计学。

本教材的编写目的在于为高等院校经济管理类学生提供一本统计学的入门读物,以应用统计学为主,侧重于介绍统计方法的应用条件,重点突出统计的数据处理功能、方法和技巧,使学生在今后的学习和工作中解决一些实际问题。

第三节 统计学与其他学科的关系

一、统计学与哲学的关系

哲学是关于世界观和方法论的科学,对人们如何认识世界具有普遍的指导意义。统计学作为一门收集、整理、描述和分析统计数据的方法论的科学,其目的在于探索数据的内在数量规律性,以达到对客观事物的科学认识,它的一切原理、原则和方法都离不开哲学的指导。在众多的哲学理论中,首要当推辩证唯物主义和历史唯物主义的哲学思想对统计科学和统计工作的指导意义。例如,统计调查要取得客观、准确、及时、全面的数据资料,就必须以物质第一、意识第二和实践是检验真理的唯一标准等哲学思想为指导,深入实际调研,实事求是地反映客观情况。又如,统计活动要反映客观事物发展变化的过程、结果及数量规律性,必须遵循矛盾的对立统一法则,在质与量的密切联系中研究数量表现、数量关系和数量变化规律,并寻求量变引起质变的数量界限。在统计实践活动中,必须坚持“实践—感性认识—理性认识—再实践—再认识”的认识规律,经过从个别到一般、从具体到抽象、从感性到理性的反复实践和认识的过程,深刻而全面地总结统计工作经验,不断推动统计科学的发展。

二、统计学与数学的关系

由于统计学运用到了大量的数学知识,无论是研究理论统计学还是应用统计学均需要深厚的数学基础,这给人一种错觉,认为统计学只是数学的一个分支。事实上,统计学与数学之间有本质的区别,主要表现在:

第一,统计学研究客观事物在具体时间、地点、条件下的数量表现、数量关系及内在的数量

规律性,研究的数据有指向的对象和具体的计量单位,如实物量、劳动量或价值量等,数学则撇开了事物的具体内容而抽象地研究数量之间的空间形式,这些抽象的数字没有量纲或计量单位。

第二,统计学与数学研究的逻辑方法也不同。数学研究采用的是纯粹的演绎方法;统计学运用的是演绎和归纳的结合,必须深入客观实际收集数据,并与研究对象相结合,经过归纳、推断才能得出有用的结论。

数学与统计学的联系在于,数学为统计研究提供公式、模型和分析方法。具体表现为:

第一,数学为统计数据的整理、汇总、计算、分析提供计算分析方法和依据。

第二,数学为统计抽样推断、假设检验、趋势预测等提供可供选择的理论依据和计算模型。统计学应根据客观现象的具体特点和内在规律性,选择恰当的数学方法,而不能为统计而统计,为数字而数字,沦为流于形式的数字游戏。

三、统计学在管理学中的应用

统计学是一门应用性很强的学科,又由于几乎所有学科的研究都离不开数据的收集、整理和分析,因此,统计学与几乎所有的学科领域都存在或多或少的联系。这在应用统计学中已有表述。此处简要介绍统计学在管理学中的应用。

统计学以其科学的研究方法、完整的指标体系和先进的技术手段,在政府宏观调控和企业微观管理中有着广泛和深入的应用。宏观、微观经济主体利用获取的丰富的数据信息,深入开展综合分析和专题研究,为科学决策提供依据。统计实践活动提供的数据信息,还可对宏观、微观经济的运行状态进行监督、检查和预警,它既是决策不断修订和调整的重要依据,也是判断和检验决策方案正确与否的尺度。如从宏观领域看,政府可通过统计报表制度及普查、抽样调查等方式获取国民经济和社会发展的数据资料,核算国民经济和社会发展指标,在此基础上制订和完善经济与社会发展规划。在微观经济领域,统计理论和方法也渗透到企业的存货控制、产品质量控制、市场前景预测、人力资源开发与培训、投资决策、业绩考核等经营过程和环节。

统计学与管理学相结合,形成了管理统计学。管理统计学是一门关于社会经济学和企业经济管理问题研究中数据的收集、整理和分析技术的科学,是专门用以研究社会经济、政府宏观决策和企业微观管理等方面的应用统计学。

第四节 统计学中的基本术语

统计学中涉及的基本术语主要有四个,分别是总体、样本、参数和统计量。无论在描述统计学还是推断统计学中,这些术语都有广泛的应用。

一、总体和样本

(一)总体

总体(population)也称统计总体,是在特定研究目的下需要调查或统计的某一现象全部个体的集合。构成总体的个体称为总体单位。

1. 总体的特征

总体必须具备同质性、大量性和变异性三个特性。

同质性是指构成总体的个体必须具有某些共同属性。物以类聚、人以群分,“类”和“群”即为这里的“同质”。同质性是构成总体的前提和基础。

大量性是指构成总体的个体单位要足够多,即总体容量足够大。总体容量在统计学中以 N 表示。统计工作主要通过抽样调查收集数据,进而分析、描述和推断现象总体的数量特征和发展趋势,若总体容量本身过小,抽取的样本单位必然受限,则无法反映现象的总体特征和发展规律。因此,大量性是构成总体的必要条件。

变异性是指构成总体的个体间存在各种属性或数量差异。个体之间的变异是广泛存在的,以某市所有工业企业总体为例,尽管每家工业企业的经济职能相同,都是从事工业活动的基本单位,但在行业隶属、企业规模、职工人数、经营范围、销售收入、产值利润等方面千差万别。统计工作的任务就是在对个体的差异研究中探寻共性规律。因此,变异性是统计研究的目的所在。

2. 总体的分类

根据总体中包含的个体(也称为总体单位)不同,可将总体分为有限总体和无限总体。总体中包含的单位总数称为总体容量,以 N 表示。如果总体中包含的个体数量有限或可分,称为有限总体,如 2017 年 12 月 31 日中国内地的人口总体,尽管人数众多,但数量有限。若总体中包含的个体数量无限或不可分,则称为无限总体,如连续生产的流水线上生产的产品总体、海洋生物群体、原始森林植被总体等。

区分有限总体和无限总体的统计学意义在于:对于有限总体,既可以采用全面调查,也可以采用非全面调查方式收集数据信息;对于无限总体,只能采用非全面调查方式收集数据,以样本数据推断总体的数量特征。

(二)样本

样本(sample)也称抽样总体,是从总体中选取出来进行调查或统计的部分个体的集合。

样本中包含的总体单位数称为样本容量,以 n 表示。根据样本容量的大小,可将样本分为大样本和小样本。样本容量大于 30 时称为大样本,否则称为小样本。根据总体的大量性要求,绝大多数抽样调查中采用大样本。样本容量与总体容量之比 n/N 称为抽样比,用 f 表示。

二、参数和统计量

(一)参数

参数(parameter)是用来描述总体数量特征的概括性度量。

统计学中,常用的描述总体数量特征的度量有均值、成数、方差、标准差等。对于某一特定总体而言,参数是确定的常数,但往往是未知的,要借助于统计调查获取的数据资料去推断。

(二)统计量

统计量(statistic)是根据样本数据计算出来反映样本数量特征并据以推断总体数量特征的概括性度量。

与总体参数类似,常用的描述样本数量特征的统计量也有均值、成数、方差、标准差等,由于样本统计量的数值随着样本选取的不同而变化,因此,样本统计量是随机变量。

第五节 本教材的篇章结构

本教材总体上按照现代统计学的两大分科——描述统计学和推断统计学——的框架来编排,分为四篇、十章。其中,统计基础篇介绍统计的概念、学科沿革、学科构成及统计数据的收集与调查方法;描述统计学作为整个统计学科的基础和统计研究工作的第一步,是本教材的主体,主要介绍了表格与图形法、静态数据描述法和动态数据描述法;推断统计学是现代统计学的核心和关键,也是统计学课程的重点、难点所在。鉴于教材使用者大多为经济管理专业学生,本教材最后常识性地介绍了我国国民经济常用统计指标,以帮助学生拓宽视野,增强对国民经济和社会发展的感性认识,利用所学指标对经济运行状况进行分析。在教材内容的编排上,充分考虑读者的专业方向和接受能力,兼顾需要与可能,侧重介绍统计学基本理论和基本方法,做到难易有别、详略得当。

具体安排如图 1-1 所示。

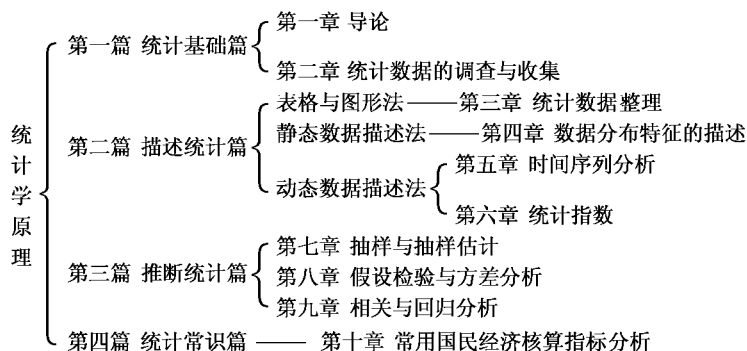


图 1-1 本教材的基本结构

本章小结

本章共五节,主要介绍了统计的含义、统计理论和实践的产生发展过程、统计学的分科、统计学与其他学科的关系、本教材的内容框架等。

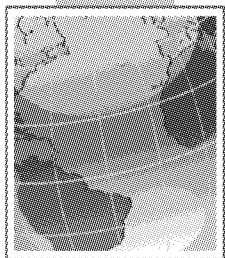
1. 统计的三层含义:统计工作、统计资料和统计学。三者相互联系,其中最基本的含义是指统计工作。
2. 从统计方法的构成角度看,统计学可分为描述统计学和推断统计学。描述统计学是现代统计学的基础,而推断统计学是现代统计学的核心。从统计方法的研究和应用角度看,统计学又可分为理论统计学和应用统计学。
3. 由于几乎所有学科的研究都离不开数据的收集、整理和分析,因此,统计学与几乎所有的学科都存在一定的联系。本章重点介绍统计学与哲学、数学及管理学的关系。
4. 统计学中的基本术语主要包括总体、样本、参数和统计量。
5. 本教材按描述统计和推断统计的框架来编排,共分四篇,分别是统计基础篇、描述统计篇、推断统计篇和统计常识篇。

常用术语

- | | |
|---------------------------------|---------------------------------|
| 1. 统计学 statistics | 2. 描述统计学 descriptive statistics |
| 3. 推断统计学 inferential statistics | 4. 总体 population |
| 5. 样本 sample | 6. 参数 parameter |
| 7. 统计量 statistic | |

复习思考题

1. 从不同的角度理解,“统计”一词有几种含义?它们之间的关系如何?
2. 在日常生活中,你接触过哪些与统计有关的问题?
3. 描述统计学和推断统计学的研究方法、内容有什么不同?两者之间有什么联系?
4. 举例说明统计在企业管理中的应用。
5. 举例说明统计在政府宏观经济管理中的应用。
6. 统计总体有哪些特征?如何理解?
7. 参数和统计量有什么区别?



第二章 统计数据的调查与收集



学习目的

- 了解统计调查的概念和分类。
- 领会统计数据的计量尺度、数据和变量的类型及它们之间的关系。
- 明确统计调查方案的设计内容。
- 掌握统计数据的来源、统计调查的方法和组织方式。



重点难点提示

- 重点:数据的计量尺度、数据和变量的类型及数据的直接来源,普查、抽样调查等统计调查组织形式。
- 难点:如何区分不同类型的计量尺度、数据和变量及其相互之间的关系。

统计数据是对客观现象进行计量的结果,也是统计汇总、加工和分析的直接对象。大到一国的经济总量、人口的平均预期寿命,小到个人的收入支出、职业学历等都属于统计数据。获取统计数据是统计工作的初始环节和首要问题。从统计数据的最初来源看,主要是统计调查和科学试验。统计调查是获取统计数据最重要的直接来源,但对于大多数信息使用者来讲,通过直接的统计调查或科学试验获取第一手数据的可能性不大;更多的是利用他人调查或试验的数据,即获取第二手数据。信息时代为我们提供了众多的间接数据来源:公开的统计出版物,各种报纸、杂志、图书、广播、电视传媒及计算机网络等。本章主要介绍数据的计量尺度与类型、统计数据的调查和收集方法。

第一节 数据的计量尺度与类型

一、数据的计量尺度

在统计调查和收集数据之前,首先要对客观现象进行计量或测度,这就涉及计量尺度的问题。不同性质的事物,其测定尺度是不同的。有的事物只能按其品质属性进行分类,如人口的性别和职业,企业的所有制性质和行业隶属等;有的则可以用精确的数字加以衡量,如职工的工资,学生的成绩,企业的产值、利润等。显然,以数字计量要比用文字分类描述更加精确。根据计量学的分类方法,按照对事物计量的精确程度,可将计量尺度从低级到高级、从粗略到精确分为定类尺度、定序尺度、定距尺度和定比尺度四个层次。采用不同的计量尺度可得到不同的统计数据,不同的统计数据适用不同的统计方法。

(一) 定类尺度

定类尺度(nominal scale)又称定名尺度或类别尺度,是按照研究对象的某种属性将其划分为若干组或若干类的一种测度。

定类尺度只能测定事物之间的类别差异,但不对类别之间的关系做任何假定。定类尺度在现实生活中应用最广。如按照性别将班级学生分为男生和女生,按照民族将人口分为汉族、满族、回族等。为便于统计处理,特别是便于计算机识别,往往将不同类别的事物用不同的数字或编码表示,如“1”表示男生,“0”表示女生;“1”表示汉族,“2”表示满族,“3”表示回族等。需要指出的是,这些数字仅仅是不同类别的代码,并不意味着这些数字可以进行数学运算或可以比较大小。定类尺度是最粗略,也是最基本的计量尺度,是其他计量尺度的基础。它具有以下特征:

(1)只能区分事物的类别,无法比较事物的优劣或大小,也就是说,定类尺度只具有“=”或“≠”的数学性质。因此,使用该尺度对事物所做的分类中,各类别之间是并列关系,类别之间的顺序是可以改变的。

(2)对事物的分类必须遵循穷尽和互斥的原则。穷尽是指在所做的全部分类中,必须保证每一个体都能归属于某一类别,不致出现遗漏;互斥是指每一个体只能归属于一个类别,不能在其他类别中重复出现。简言之,就是每一总体单位有且只有一个类别归属。

(3)对定类尺度计量的数据进行分析的统计量主要是频数或频率。如对班级学生按性别分类后,可分别计算男生人数和女生人数(统计学中称为频数),也可计算男生、女生占班级总人数的比重(统计学中称为频率)。

举一反三

举例说明生活中还有哪些常见的定类尺度。

(二) 定序尺度

定序尺度(ordinal scale)又称顺序尺度,是对事物之间等级差别或顺序差别的一种测度。

教师的职称、学历,商品的质量等级,人们对某一事件的态度等都可以采用这种尺度测定。其主要特征有:

(1)对事物不仅可以分类,而且能够比较各类别之间的优劣和大小,借以对事物排序,但不能进行加减乘除等数学运算。从数学角度看,它不仅有“=”或“≠”的数学性质,还具有“>”或“<”的数学性质。例如,高校教师按职称分为助教、讲师、副教授、教授等类别,职称是从低到高排列的。产品按质量高低分为一等品、二等品、三等品等,事物之间的顺序是不能改变的。

(2)对事物的分类同样要求穷尽和互斥。

(3)对定序尺度计量的数据进行分析的统计量除了频数和频率外,还有累计频数和累计频率。我们同样要对定序尺度进行编码,以方便计算机的读取和统计运算,如在上述高校教师职称分类和统计中,可以用“1”、“2”、“3”、“4”分别表示助教、讲师、副教授和教授。但与定类尺度不同的是,定序尺度的计算结果不仅可以得到不同职称的教师人数及比重(分别称为频数和频率),还可以得到在某一职称级别之上或之下的教师人数,如具有副教授以上职称的教师人数

及比重(分别称为累计频数和累计频率)。

举一反三

举例说明生活中还有哪些常见的定序尺度。

(三)定距尺度

定距尺度(interval scale)又称间隔尺度,是对事物类别或次序之间间隔的一种测度。

定距尺度一般以自然单位来衡量,如考试成绩用“分”计量、温度用“℃”计量等。定距尺度的主要特征有:

(1)不仅能区分事物的类别、进行排序、比较大小,还可以精确地计量出两个数字之间的差距。定距尺度的计量结果表现为数值。由于这种尺度的每一间隔是相等的,既可以计算两个数值之间的差值,也可以进行加减数学运算。例如:某班级学生考试成绩最高分为98分,最低分为32分,两者之间的差距为66分;某日某市最高温度为38℃,最低温度为23℃,当日的温差为15℃等。

(2)没有绝对零点,即定距尺度中的“0”表示“0”水平,而不表示“没有”或“不存在”。例如,某学生考试成绩为“0”分不代表没有分数,某日某市最低温度为0℃不表示没有温度。

举一反三

举例说明生活中还有哪些常见的定距尺度。

(四)定比尺度

定比尺度(ratio scale)也称比率尺度,是对事物之间比值的一种测度。

常见的定比尺度如人的收入支出、企业的产值利润、某地区的人口总数、失业人数等。它与定距尺度属于同一层次,计量结果也表现为数值。但其功能强于定距尺度,其主要特征表现为:

(1)除能对事物分类、排序、比较大小、求出差异大小外,还可以计算出两个数值之间的比率,具有数学上加减乘除四则运算功能。例如,对某市工业企业普查的数据显示,甲、乙两家企业的年销售收入总额分别为5 000万元和500万元,则不仅可比较甲、乙企业的销售收入高低,计算销售收入之差,还可以计算出甲企业的销售收入是乙企业的10倍。

(2)具有绝对零点,即数字“0”表示“没有”或“不存在”。例如,某人本月收入为“0”表示本月没有收入,某售楼处6月份商品房成交量为“0”表明当月无销售,某市某月重大交通事故为“0”则说明无重大交通事故。

上述四种计量尺度对事物的测量层次是由低级到高级、由粗略到精确逐步递进的。高层次计量尺度具有低层次计量尺度的全部特征,我们可以方便地将高层次计量尺度的计量结果转化为低层次计量尺度的计量结果。如将考试分数转化为优、良、中、差四个等级,是将定距尺度转化为定序尺度,但若将考试成绩等级转化为分数则比较困难。

在统计分析中,一般要求测量的层次越高越好,因为高层次的计量尺度包含了更多的数学特征,对数据的处理方法就越多,处理结果提供的信息量也越大。在实际工作中,应尽量使用高层次的计量尺度来对事物进行测度。现将四种尺度的测定层次和数学特征归纳如表 2-1 所示。

表 2-1 四种计量尺度的数学特征比较

数学特征	定类尺度	定序尺度	定距尺度	定比尺度
分类(=, ≠)	可以	可以	可以	可以
排序(>, <)		可以	可以	可以
间距(+, -)			可以	可以
比值(×, ÷)				可以

考考你

以下所列出的分别属于哪一层次的计量尺度?
人的身高、体重、学历、政治面貌、籍贯、收入、支出。

二、数据的类型

(一)数据的类型和表现形式

1. 数据的类型

统计数据是计量尺度对事物进行计量的结果,采用不同的计量尺度会得到不同类型的统计数据。从上述四种计量尺度计量的结果来看,可将统计数据分为两种类型:定性数据和定量数据。

定性数据也称品质数据,它说明事物的品质特征,一般不能以数值表示,只能以文字表述。这类数据是由定类尺度和定序尺度计量形成的。

如人的性别只能以男、女表示,则“男”和“女”就是定性数据;考试成绩用优、良、中、差表示,则“优”、“良”、“中”、“差”即为定性数据。

定量数据又称数量数据,用以说明现象的数量特征,具体表现为数值。这类数据是由定距尺度和定比尺度计量形成的。

如产品销售额为 300 万元、学生考试成绩为 80 分、某市人口总数为 260 万人,则 300 万元、80 分、260 万人均为定量数据。

对于不同类型的数据,可采用不同的统计分析和处理方法。如前所述,对定性数据,可计算各组的频数或频率,计算众数,进行 χ^2 检验等。对定量数据,除可以进行对定性数据的计算分析外,还可以计算均值、进行参数估计、方差分析和假设检验等。这说明,适用于较低层次测量数据的统计方法,同样适用于较高层次的测量数据;反之,则不一定。

考考你

区分以下命题中数据的类型,并说明它们是由哪一类计量尺度测量形成的:

1. 某人具有研究生学历;

2. 武汉市 2016 年实现财政总收入 2 423 亿元。

2. 数据的表现形式

定量数据通常有两种表现形式:总量数据和分析数据。

总量数据又称总量指标或绝对数,反映现象在一定时期或一定时点上达到的总规模、总水平和工作总量,是统计数据的基本表现形式,也是计算分析数据的基础。按反映现象的时间状况不同,可分为时期指标和时点指标。时期指标反映现象在一段时期内达到的总规模或总水平,如某一地区 2017 年的总产值、财政收入、出生人口数等;时点指标反映现象在某一时刻(瞬间)达到的状态量,如某一地区 2017 年年末的总人口数、城乡居民储蓄存款余额等。

分析数据是两个总量数据(绝对数)的比值,具体分为相对数和平均数两种。

根据对比的数据性质不同,相对数有比例和比率之分。比例是同一总体中各部分的数量占总体总量的比重或部分数量与部分数量之比,用以反映总体的内部结构,同一总体中各部分比重之和等于 1。比率是不同类别的数量之比,可以反映同一现象在不同时间、空间上的对比,也可以反映两个有联系但性质不同的总体数量之比,如经济发展速度,人口密度,计划完成程度,人口出生率、死亡率、自然增长率等。

平均数又称均值,是用总体(或样本)各单位的某一数据总量与总体(或样本)单位总数的比值,反映总体(或样本)单位在某一数量方面的一般水平。

(二)变量及其类型

说明现象某种属性特征的概念称为变量(variables),变量的具体表现称为变量值。
统计数据就是变量的具体表现。

与统计数据的分类相对应,变量也分为品质变量和数值变量。

品质变量又称品质标志,是用来反映事物品质特征的名称,具体表现为定性数据。如“职称”是一个品质变量,“助教”、“讲师”、“副教授”等是其具体表现。

数值变量又称数量标志,是用来反映事物数量特征的名称,具体表现为定量数据。如人的身高、体重、年龄,企业的产量、产值、销售收入、商品库存量等都是数值变量,因为它们具体表现为数字。

数值型变量按变量值的计数特征不同,可分为连续变量和离散变量。连续变量是指变量的取值是连续不断的,其数值可以无限分割,一般表现为小数形式。连续变量一般由测量得出的结果表示,如上述人的身高、体重、年龄,企业的产值、利润、销售收入等。离散变量是指变量的取值是间断的,其数值一般只能以整数形式表示。离散变量一般由计数得出的结果表示,如产品产量、商品库存量、设备台数、人口总数等。

举一反三

常见的连续变量和离散变量还有哪些?

变量是统计学中常用的一个概念,多数情况下我们所指的是数值型变量,大多数统计方法所处理的也都是数值型变量,通常将数值型变量简称为变量,本教材以后章节所指的都是数值型变量。

第二节 统计数据的直接来源——统计调查

在统计数据两个直接来源中,统计调查是获取社会经济数据的重要渠道,统计调查可以是统计部门专门组织的统计调查,如全国人口普查和抽样调查、工业普查、农业普查等,也可以是其他部门或机构为特定目的而进行的调查,如民意调查、市场调查等;科学实验是获取自然科学数据的主要渠道,本节不多作介绍。本节介绍社会经济数据的主要收集方式和方法——统计调查。

一、统计调查的概念和分类

(一)统计调查的概念

统计调查是根据统计的目的和任务,运用科学的调查方法,有计划、有组织地向客观实际收集和登记统计数据的过程。

统计调查是统计数据的主要来源,也是获取直接统计数据的重要手段。统计调查在统计工作中具有重要的意义,它是统计工作的基础环节和初始阶段,是认识事物的起点,是统计数据整理和分析的前提。所以,对统计调查数据的要求是准确、及时、系统、完整。

(二)统计调查的分类

1. 按调查对象包括的范围不同,可分为全面调查和非全面调查

全面调查是对调查对象的所有个体一一进行调查登记的调查方法,目的是要取得总体的全面、系统、完整的统计资料。如全国人口普查、工业普查、经济普查等。

非全面调查是只对调查对象的部分个体进行调查登记的调查方法。如要了解城镇居民的消费支出情况,只要抽取该市一部分家庭进行调查即可。抽样调查、典型调查、重点调查等均为非全面调查。

2. 按照调查组织方式不同,可分为统计报表和专门调查

统计报表制度是根据《中华人民共和国统计法》的规定,按照一定的表式和要求,自上而下统一布置、自下而上逐级提供统计资料的一种报表制度。统计报表制度是各级政府主管部门获取有关国民经济和社会发展基本数据资料的主要方法,统计报表提供的数据也是政策制定和修正的重要依据。

专门调查是为了研究一些特殊问题而专门组织的统计调查,多属于一次性调查。如为了配合艾滋病的宣传预防工作而在青少年中进行艾滋病知识的问卷调查,为掌握我国的人口规模和结构情况而组织的以2010年11月1日零时为标准时点的第六次全国人口普查等。本章将要介绍的重点调查、抽样调查等也属于专门调查的范畴。

3. 按照调查登记的时间是否连续,可分为经常性调查和一次性调查

经常性调查是连续性调查,它随着被调查对象在时间上的发展变化进行连续不断的调查登记。经常性调查适用于时期现象调查,其主要目的是获得事物整个发展过程及结果的统计数据。如工业企业中进行的产量、产值、销售量、销售额、原材料消耗等统计。

一次性调查是不连续登记的调查,它对被调查对象每间隔一段时间登记一次,以取得现象在一定时点上的状态资料。一次性调查适用于时点现象的调查,如我国的人口普查,一般间隔10年左右进行一次,以了解在某一时点上的人口总量结构及几个时点上状况的对比变化。

考考你

某饮料生产厂家为了解消费者的口味和消费习惯进行了一次市场调查,试判断下面描述是否正确:

1. 此类调查适合采用全面调查方式。
2. 此类调查适合采用连续调查方式。
3. 这是一种专门调查。

二、统计调查方法

统计调查方法是指收集统计数据的具体方法。常用的统计调查方法有直接观察法、报告法、采访法和通信法等。

(一)直接观察法

直接观察法是调查人员亲自到现场,对调查对象进行观察计量以取得统计数据的一种方法。例如,农业科研人员对试验田里的新品种农作物进行实割实测,以获取亩产量等相关数据;又如,通过对商品物资的期末盘点以掌握本期发出及期末结存物资的数据资料。这种调查方法能保证资料的准确性和完整性,但需耗费较多的人力、物力和时间,所以不宜过多采用。

(二)报告法

报告法是基层单位根据上级部门的要求,以各种原始记录和核算资料为基础和依据,填写调查表并逐级上报的一种调查方法。我国现行的统计报表制度就属于这种调查方法。

(三)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以取得统计资料的一种调查方法。具体分为个别口头询问、被调查者填表、开调查会等形式。

(四)通信法

通信法是由调查者把调查问卷邮寄给被调查者,由被调查者填写回答以取得统计数据的一种方法。这种方法要求所提问题简明扼要、通俗易懂,不能太多、太复杂,由于调查效果取决于被调查者的配合程度,因此,这种方法通常只作为辅助调查方法。

(五)电话访问

电话访问是调查人员通过电话向被调查者询问、交谈以获取所需资料。这种方法具有经济高效的优点,适用于对一些简明扼要、数量有限的问题进行调查。但与通信法类似,也存在受被调查者配合程度局限等问题,成功率较低。

(六)网上调查法

网上调查法是20世纪90年代开始兴起的一种调查方法。具体通过收发电子邮件、电脑辅助电话访谈、被调查者在网上填写调查问卷等方式进行。

知识链接**——人口抽样调查方法**

我国2015年组织的1%人口抽样调查,采用的是由统计调查员入户调查询问、现场填报的方式。调查员按照调查表中的项目,逐户、逐人询问,逐项填写被调查者的基本情况、迁移流动、人口素质、就业和社会保障、婚姻生育状况等信息。

考考你

政府相关职能部门是如何获取每年应届本科毕业生的就业率等数据资料的?依靠的是哪种数据收集方法?

三、对统计调查数据的基本要求

为了保证整个统计工作的质量,统计调查收集的数据必须满足准确、及时、系统、完整和经济的要求。

(一)准确性

准确性要求统计数据要如实反映客观实际,这是保证统计工作质量的首要环节和关键所在。如果统计数据失去了真实性,也就失去了利用价值。

(二)及时性

及时性又称时效性,就是要求在尽可能短的时限内,尽快完成数据的收集和登记工作。如我国第六次人口普查,要求在2010年11月1日至11月10日之间全部完成,以保证数据的时效性。

(三)系统性

系统性要求所收集的数据有条理,符合逻辑顺序,不杂乱无章,便于下一步的整理和汇总。

(四)完整性

完整性要求统计数据项目齐全,不重复、不遗漏。若统计数据残缺不全,就无法反映被研究现象的全貌,无法正确认识客观现象总体的特征,甚至会得出错误的结论。

(五)经济性

经济性是指在满足一定准确度和把握度的前提下,通过精心设计调查方案,尽量减少人力、物力和财力投入,提高统计调查工作的经济效益。简言之,就是尽量提高投入产出比。

上述五个方面的要求既相互依存,又对立统一。在实际统计调查过程中,要兼顾调查目的的需要和客观条件(如经费来源、时间限制等)的可能,分清轻重缓急,实现综合平衡。一般而言,以准确为前提,稳中求快,力争以较少的投入取得完整、系统的数据资料。

四、统计调查方案的设计

统计调查方案是事先制订的、用来组织和协调统计调查全面开展的工作计划书。

在统计调查中,首要问题是设计调查方案,调查方案将调查目的、任务、内容、方法、步骤等高度统一,使调查工作有条不紊地进行。调查方案一般涉及以下内容:

(一)确定调查的目的和任务

设计调查方案的首要问题就是要明确调查目的和任务,做到有的放矢。它所解决的是“为什么调查(why)”的问题。

(二)确定调查对象和调查单位

调查对象是需要进行调查的某一社会经济现象的总体,它由许多性质相同的调查单位组成。调查单位是调查对象中需要调查的具体单位。

确定调查对象和调查单位要解决的是“向谁调查(who)”、由谁来具体提供调查资料的问

题。只有科学合理地确定了调查对象,才能划清统计研究的总体界限,避免因总体界限不清影响所收集数据的准确性和有用性。例如,要全面调查某省高校教师的基本情况,则该省所有的高校教师构成的整体就是调查对象,每位高校教师就是一个调查单位。

调查对象与调查单位之间是总体与个体的关系。调查单位可以是调查对象中的全部单位,也可以是部分单位。在全面调查中,调查单位是调查对象中的全部单位;在非全面调查中,调查单位只是调查对象中的部分单位。

考考你

统计学老师组织了一次课堂测验以了解学生对知识的掌握情况,是全面调查还是非全面调查?说明调查对象和调查单位。

(三)确定调查项目

调查项目是调查的具体内容,解决的是“调查什么(what)”的问题。它可以是调查单位的品质特征,也可以是数量特征。统计数据整理及分析都建立在调查的具体内容之上,所以拟定调查项目是调查方案设计的关键所在。

在实际工作中,应根据需要与可能以及少而精的原则确定必要的调查项目,还要注意调查项目的内涵明确、调查项目之间具有内在联系和彼此衔接,以全面、完整、系统地反映调查对象的特征,提高调查资料的质量。

(四)设计调查表

调查表也称调查问卷,就是将调查项目按照一定顺序排列而成的一种表格形式,是统计工作中收集数据的基本工具。本章附录将具体介绍调查问卷的设计方法。

在大多数统计调查中,调查项目要通过一定的表格形式表现出来。调查表有两种形式:单一表和一览表。一份表中只登记一个调查单位的是单一表,它可以容纳较多的调查项目,适用于比较详细的调查,如人口普查、工业普查等。一览表是一份表中可登记若干个调查单位,在调查项目不多时,为节省人力、物力和财力,可采用一览表。

调查表一般由表头、表体和表外附加三部分构成。表头用来说明调查表的名称,填写调查单位的名称、性质、隶属关系等;表体是调查表的主体部分,内容是调查的具体项目;表外附加一般由填报人签名盖章、报表报出日期及附注填表说明和指标解释等组成。调查表的具体内容将在第三章详细介绍。

(五)确定调查时间

调查时间有两层含义:

一是调查资料所属的时间。如果调查的是时点现象,要规定调查资料所属的标准时点。如我国 2015 年的 1%人口抽样调查规定的标准时点是 2015 年 11 月 1 日零时。如果调查的是时期现象,则要明确规定调查现象的起止时间。如 2016 年我国的 GDP 统计中,起止时间是 2016 年 1 月 1 日至 12 月 31 日。规定调查资料所属的时间是为了保证统计资料的准确性。

二是调查时限,即调查工作进行的起讫时间,也即收集和报送数据资料所需的时间,以保证统计数据的时效。如 2010 年第六次人口普查规定,调查登记工作的起止时间为 2010 年 11 月 1 日至 11 月 10 日。

考考你

某市组织财税大检查,要求在 2016 年 12 月 15 日至 12 月 31 日之间全部完成,这一规定是调查时间还是调查时限?

(六)制订调查组织实施计划

周密细致的组织实施,是调查工作顺利开展的重要保证。它包括的内容有:调查工作完成的期限及工作进度;调查的方式与汇总方法;调查的组织领导;调查经费的预算和开支管理;调查前的其他准备工作等。

五、统计调查的组织方式**(一)普查**

普查是专门组织的、一次性的全面调查。

普查是一种重要的调查方式,主要用来收集那些不能够或不适宜采用定期的全面统计报表收集的统计资料,如我国历次的人口普查、经济普查等。通过普查可摸清一个国家或地区的人力、物力、财力等资源情况,还可以取得某些专门问题的详细资料,为政府进行宏观经济决策、制定长远发展规划等提供依据。所以对普查资料的要求是全面、详尽、系统,但普查的工作量大,耗时、耗资较多,一般不宜经常采用。

普查的组织方式一般有两种:一是组织专门的普查机构和人员,对调查单位直接进行登记,如我国历次的人口普查;二是利用调查单位的原始记录和核算资料,通过发放统一的调查表格,由调查单位自行填报,如我国的物资库存普查。

普查工作应遵循以下原则:

- (1)为避免数据资料的重复和遗漏,必须统一规定普查的标准时间。
- (2)在普查范围内,各调查单位和调查点要尽可能同时行动,并尽可能在最短时间内完成,以做到步调一致、报送及时。
- (3)统一规定普查的内容和间隔时间,以便于同一次普查中资料的汇总及不同时间点或时间段上资料的纵向对比。

知识链接——我国的普查制度

每逢年份的末尾数字为“0”的年份,进行人口普查;
每逢年份的末尾数字为“1”或“6”的年份,进行基本统计单位普查;
每逢年份的末尾数字为“3”的年份,进行第三产业普查;
每逢年份的末尾数字为“5”的年份,进行工业普查;
每逢年份的末尾数字为“7”的年份,进行农业普查。

(二)统计报表

统计报表制度是依照《中华人民共和国统计法》的规定,自上而下统一布置、自下而上逐级提供基本统计数据的一种调查方式。

我国的统计报表有多种类型。按报送范围不同,可分为全面报表和非全面报表。全面报表是一种全面调查方式,要求所有的调查单位都必须填报,如我国的工业企业统计报表;非全面报表只要求一部分单位填报。按报表所属的时期长短不同,可分为日报、月报、季报、半年报和年报等。

经过几十年的统计实践,在我国已形成比较完备的统计报表制度,是收集国民经济和社会发展状况统计数据的主要方式。但也有其固有的局限性:要花费大量的人力、物力和财力;内容和程序比较固定呆板;数据资料的取得时间较长;还有一些重要的统计数据无法利用统计报表制度取得,如居民消费数据、人口再生产数据等。所以,统计报表必须与其他统计调查方式相结合,才能更好地发挥作用。

知识链接 ——“企业一套表联网直报数据”封堵统计作假

地方 GDP 总和为何比全国的数据高? 不同部门公布的房价、菜价为何不一致? 湖北省在全国率先试点的“企业一套表联网直报数据”有望逐步解决这些难题。截至 2017 年 6 月,湖北省已有 4 万多家规模以上工业企业、限额以上批发零售住宿餐饮企业、资质以内的建筑企业和房地产开发经营企业告别纸质报表,改用电脑终端向国家统计局直接报送统计数据。任何人、任何环节的数据修改都将在操作平台上留下痕迹,任何弄虚作假行为都会暴露在众目睽睽之下。随着直报数据涵盖的范围越来越广,GDP、CPI 等公众高度关注的综合数据也将通过数据平台直接生成和公布。这无疑是统计生产方式的一次重大变革,也是实现统计数据客观真实、及时准确的重要保障。

资料来源:湖北省统计局官方网站。

(三)抽样调查

抽样调查是一种专门组织的非全面调查。它是按照随机原则从总体中抽取一部分单位组成样本进行观察分析,以样本数量特征去推断总体数量特征的一种调查方法。

抽样调查方式是实际工作中应用最为广泛的一种调查方式。抽样调查有概率抽样和非概率抽样两种,本书所指的是概率抽样。

1. 抽样调查的特点

第一,经济性。这是抽样调查最大的优点。由于调查的样本单位通常只是调查对象中很小的一部分,调查的工作量小,可节省大量的人力、物力和财力。

第二,时效性高。抽样调查可以迅速、及时地获取所需要的数据资料。由于工作量小,调查的准备时间、调查时间、数据处理时间等均可大大缩短,从而提高数据的时效性。与普查相比,抽样调查可以根据需要,较为频繁地进行,以弥补普查难以经常开展的不足。

第三,适应面广。抽样调查适用于各个领域、各种问题的调查,可以获取广泛的数据信息,特别适合一些特殊现象的调查。从调查的内容看,抽样调查可以设计更多的调查项目,获取更详细和深入的数据资料。

第四,准确度高。与普查相比,抽样调查的工作量小、环节少,可有效降低登记性误差,从而获取比普查更精确的数据资料,提高数据资料的质量和可靠程度。

2. 抽样调查的作用

(1)对于不可能进行全面调查的社会现象,为了测算全面数据,必须采取抽样调查的方法。

一种情况是对无限总体的调查,如环保部门对河水水质进行检测,只能采取抽样调查方式;另一种情况是某些产品的质量检测具有破坏性,如日光灯管的使用寿命、食品的保质期等,也不可能进行全面调查,只能采取抽样调查。

(2)对于某些没必要或很难进行全面调查的社会现象,也可采用抽样调查。如要调查某个乡镇的小麦亩产量,理论上可逐家、逐户进行全面调查,但要浪费大量的人力、财力、物力和时间,而抽样调查既可提高调查的实效性,又能达到全面调查的目的和效果。

(3)其他作用。如通过抽查产品合格率对生产过程进行质量控制,利用抽样调查对全面调查的结果进行检验修正等。

(四)重点调查

重点调查是在调查对象中选择一部分重点单位所进行的非全面调查。

所谓重点单位,是指在总体中虽然个数不多,但它们的数据总量在总体数据总量中占有绝大部分比重。例如,为了解我国钢铁行业的产销、利润等情况,只要选择宝武、首钢、邯钢、包钢等几家特大型钢铁集团进行调查,即可掌握我国钢铁工业的基本情况。

重点调查的单位较少,可设计较多的项目和指标,集中精力掌握详细的数据资料。但需要注意的是,只有当调查任务是掌握调查对象的基本情况,而在总体中的确存在重点单位时,才能进行重点调查。必须指出,根据重点调查获取的数据不能用来推断总体指标。

知识链接 ——重点调查与我国银行业的寡头垄断

2016年上半年,我国18家上市银行营业收入及利润总额分别达到19 534.17亿元及7 265.69亿元,其中工、农、中、建、交五大行营业收入及利润总额分别为13 161.41亿元及5 194.73亿元,占比分别达到67.38%和71.50%。作为全球最赚钱的银行,中国工商银行2016年上半年实现净利润1 502.17亿元,日均净赚8.35亿元。从统计学角度看,我国的银行业是以五大行为主的寡头垄断市场(五大行就是重点调查中的重点单位)。近年来,受去库存、去杠杆、去产能、调结构等宏观经济及政策因素影响,商业银行不良贷款率有所提高,银行业利润增速降低,2016年中国银行还首次出现了利润负增长,其他上市银行利润增速也有不同程度回落,银行业利润增速动辄20%~30%的风光已不复存在。

银行利润增速回落与其说是宏观经济影响所致,从实体经济角度,更是期盼已久的理性回归。银行业获取的超额利润源于行业垄断下对实体经济的盘剥。银行业多赚1分钱,实体经济将多负担1分钱的融资成本,这是一个经济学中的零和博弈。要改变银行业从实体经济“抽血”的扭曲状态,必须打破银行业的垄断,降低银行业的准入门槛,允许民间资本进入银行业,同时进行利率市场化改革,缩小银行利差,将一些经营不善的银行淘汰出信贷市场。

资料来源:南方财富网,2016年9月1日。

考考你

你还能举出重点调查的例子吗?

(五) 典型调查

典型调查是根据调查的目的和任务,在对调查对象全面分析的基础上,有意识地选择若干有典型意义或有代表性的单位进行深入、细致调查的一种非全面调查方式。

典型调查的特点有:

第一,调查范围小,调查单位少,可以多设计一些项目和指标进行深入、细致的调查,用来研究某些复杂的专门问题。

第二,典型调查的目的是通过典型单位来描述或揭示事物的本质和规律,因此调查单位是有意识选择的有代表性和典型性的单位,这些典型单位应具备所研究问题的本质特征或属性。

典型调查的核心在于如何正确地选择典型单位。应当根据调查的目的、任务来选择不同的典型单位。如果为了观察事物的发展趋势,形成科学的预见,可选择新生事物做典型;如果为了总结经验教训,则可选择成功或失败的典型;如果为了了解一般情况或事物发展的一般规律,则可选择中间水平的典型。

知识链接 ——海南澄迈的长寿密码

海南省澄迈县位于海南岛西北部,资源丰富、风光秀丽、生态良好、民风淳朴,素有“青山绿水廿万顷,十里常逢百岁人”之美誉,是中国绿色名县、富硒福地和世界长寿之乡。据第六次全国人口普查数据显示,澄迈县人均预期寿命高达 79.79 岁,比全国平均水平高 6.59 岁。截至 2016 年 11 月,全县 57.5 万总人口中,80 岁以上老人 18 960 人;百岁老人 217 人,占总人口比例为 37.74/10 万;80 岁以上夫妻 911 对,90 岁以上夫妻 77 对,100 岁以上夫妻有 4 对。百岁老人绝对数和占总人口比例均居全国市县第一位。极具岭南特色的长寿文化、优越的生态人居环境及高水准的社会保障体系是澄迈的长寿密码。澄迈的长寿现象引起国内外专家学者的高度关注和浓厚兴趣,纷纷到澄迈调研考察。从统计学角度,这种统计调查即属于典型调查。

资料来源:人民日报,2016 年 12 月 19 日。

重点调查和典型调查虽然都是非全面调查,但它们与抽样调查存在较大区别。抽样调查根据随机原则抽取样本单位,可以根据样本指标推断总体数量特征,而重点调查和典型调查的调查单位不是随机抽取的,调查结果不能用于推断总体指标。

考考你

1. 要推断一批种子的发芽率,采用哪种调查方式好? 说明理由。
2. “麻雀虽小,五脏俱全”说明了哪种调查方式的可行性?

第三节 统计数据的间接来源

鉴于主客观条件的限制,对大多数统计数据使用者来说,获取直接数据往往是不可能的。但可以通过其他渠道获取间接的统计数据,又称第二手数据。间接数据的主要来源有以下两类:

一、公开的统计出版物

第二手数据主要是公开出版或公开报道的数据,一般都是官方数据。有关社会发展和国民经济的基本统计数据,中央和地方政府的统计部门都会定期地通过某种媒体向社会公开发布。目前我国,官方的统计公开出版物主要有中国统计出版社出版的《中国统计年鉴》、《中国统计摘要》、《中国社会统计年鉴》、《中国工业统计年鉴》、《中国农村统计年鉴》、《中国人口统计年鉴》、《中国市场统计年鉴》等,以及各省、市、地区的统计年鉴。国务院各部委以及各行业统计部门也会公开出版一些统计年鉴,如中国财政杂志社出版的《中国财政年鉴》、由国家统计局和环境保护部及其他部委共同编辑完成的《中国环境统计年鉴》等。提供世界各国社会和经济数据的中文出版物也很多,如中国社会科学出版社出版的《世界统计年鉴》、中国财政经济出版社出版的《国外经济统计资料》。

提供世界各国社会和经济数据的外文出版物就更加品种繁多、内容丰富,如世界银行各年度的《世界发展报告》、联合国的《联合国统计年鉴》、联合国粮农组织的《生产年鉴》、国际货币基金组织的《国际金融统计月报》、亚太经合组织的《亚太统计数字》等。

举一反三

你知道还有哪些统计出版物吗?

二、其他渠道

除公开出版的统计刊物上的数据外,还可以通过其他渠道获取一些尚未公开的统计数据,以及广泛分布在各种报纸、杂志、图书、广播、电视传媒中的统计数据。此外,一些民间团体(如民意测验机构、统计事务所、咨询机构等)也会发布一些相关的统计数据。随着计算机网络技术的发展,大量电子版的统计数据公布在各国政府和企事业单位的网站上,利用各种搜索引擎,我们可以方便、快捷地从网络上获取各种数据信息,不仅节省时间,还省去了录入的麻烦。此类数据有些是官方数据,有些是非官方数据。

利用第二手数据对使用者来说既方便快捷、又经济实惠,这也体现了知识经济时代的一大特点:知识产品以接近于零的边际成本向外扩散。但在使用间接数据时,应注意甄别数据的真伪和精确度等,特别要领会指标的含义、计算口径和计算方法,以免误用或滥用,造成分析结果的偏差和错误。在引用第二手数据时,还应注明数据的来源,以体现对别人劳动成果的尊重。

知识链接 武汉——中部中心城市的光荣与梦想

随着国家对东部沿海和西部地区梯次开发的战略实施,辐射华北和渤海湾的北京和天津、辐射长三角的上海、辐射珠三角的广州、辐射成渝的重庆,五大国家中心城市次第崛起,各据一极。中部中心城市则一直处于缺位状态。2016年12月国家发改委出台的《促进中部地区崛起“十三五”规划》中明确提出,支持武汉、郑州建设辐射中部地区的国家中心城市。

如果说湖北省是中部崛起的重要战略支点,九省通衢的武汉作为中部中心城市,即为支点中的支点。除具有深厚的经济文化底蕴及得天独厚的区位优势外,还具有显著的科教优势和人才优势。武汉高校林立、科研院所众多,是科技创新最活跃的城市之一,也是全国五大高教基地。2016年底,武汉市常住人口1 076.62万人,在校大学生人数超过100万,是全球拥有在

校大学生最多的城市。2016年,武汉市地区生产总值11 912.61亿元,同比增长7.8%。按常住人口计算,全年人均地区生产总值111 469元,比上年增长6.2%。在全国省会城市中排名第三。2017年2月,武汉正式提出“百万大学生留汉创业就业工程”;3月,武汉推出专项政策解决留汉大学生住房难问题;5月,再次出台《户口迁移办理工作规定》,大学生落户武汉几乎“零门槛”。社保、创业帮扶、落户等一系列政策措施的相继出台,让大学生有更多获得感与归属感,为武汉建设中心城市、辐射中部地区、促进中部崛起提供了强有力的科技和人才保障。

资料来源:武汉市统计局、武汉市住房保障和建设局、武汉市人力资源和社会保障局、武汉市公安局等官方网站。

考考你

上述数据是官方数据还是非官方数据?对我们阅读者来说是直接数据还是间接数据?

本章小结

1. 统计数据是对客观现象进行计量的结果。按照计量的精确程度,可将计量尺度从低级到高级、从粗略到精确分为定类尺度、定序尺度、定距尺度和定比尺度。不同的计量尺度可得到不同的统计数据,不同的统计数据适用不同的统计方法。

2. 从计量尺度计量的结果看,可将统计数据分为两种类型:定性数据和定量数据。采用定类尺度和定序尺度计量的数据称为定性数据,它反映事物的品质特征;采用定距尺度和定比尺度计量的数据称为定量数据,它反映事物的数量特征。

3. 从使用者的角度看,统计数据有两个来源:统计调查和科学试验是数据的直接来源;引用他人调查或试验的数据是数据的间接来源。

4. 统计调查是获取统计数据最主要的直接来源,也是统计工作的基础环节和初始阶段。为保证所收集数据的质量,统计调查开始之前,必须周密设计调查方案,确定调查目的、调查对象、调查单位、调查项目和调查时间,科学设计调查表格。

5. 从统计调查方法看,有直接观察法、报告法、采访法、通信法、网上调查等;从组织方式看,可以是普查、统计报表制度、抽样调查、重点调查或典型调查。在实际工作中,可根据调查的目的任务及不同的调查方式方法的适用情况,灵活选择和结合使用。

6. 对大多数使用者来说,获取直接数据往往是不可能的,而只能获取间接数据。信息时代为统计数据的提供者提供了众多的间接来源:公开的统计出版物,各种报纸、杂志、图书、广播、电视传媒及计算机网络等。

常用术语

- | | |
|--|---|
| 1. 定类尺度 nominal scale | 2. 定序尺度 ordinal scale |
| 3. 定距尺度 interval scale | 4. 定比尺度 ratio scale |
| 5. 数据 data | 6. 变量 variables |
| 7. 品质/属性变量 qualitative/categorical variables | 8. 数值/定量变量 quantitative/numerical variables |
| 9. 连续变量 continuous variables | 10. 离散变量 discrete variables |
| 11. 普查 census | 12. 统计报表 statistical statement |
| 13. 抽样调查 sampling survey | 14. 重点调查 key-point investigation |
| 15. 典型调查 typical survey | |

复习思考题

1. 数据的计量尺度有哪几种? 不同的计量尺度各有什么特点?
2. 列举一些日常生活中常见的数据计量尺度。
3. 统计数据有几种类型? 不同类型的数据各有什么特点?
4. 举例说明数据与变量之间的关系。
5. 从使用者的角度看,统计数据的来源有哪些?
6. 统计调查方案的设计包括哪些内容?
7. 调查时间与调查时限有什么区别? 为什么在普查中要规定统一的调查时间和调查时限?
8. 统计调查的组织形式有哪些? 它们之间有什么区别,各自适用于什么情况?

实践与讨论题

1. 某高校拟对在校生的消费观念、消费支出进行调查,以获取大学生消费支出的数据资料。你认为可采取什么调查组织方式? 说明你的理由并尝试设计调查方案和调查问卷。
2. 某市税务部门欲调研“营改增”对不同行业服务业税负的影响。请问采取哪种调查方式较好? 为什么?

附录

如何设计调查问卷

调查问卷是一种特殊形式的调查表,其特点是用一系列按照严密逻辑结构组成的问题,向被调查者了解具体事实和个体对某一问题的反映和看法。调查问卷已成为统计调查中收集统计数据的一种常用工具。

从形式看,调查问卷可以是表格形式、卡片形式或簿记形式。科学合理地设计问卷是询问调查的关键。一份好的调查问卷必须具备能将问题传达给被调查者和使被调查者乐于、便于回答的功能。

一、问卷的设计原则

(一)主题明确

根据调查目的和任务确定问卷主题,拟订调查提纲,从实际出发设计具体问题,问题的目的明确、重点突出。保证通过应答者的回答能得到充分必要的信息。

(二)结构合理、逻辑性强

问题的排列体现调查提纲要求的逻辑结构,要有一定的逻辑顺序,层次分明,详略得当,符合应答者的思维程序。一般应先易后难,先简后繁,先具体后抽象,先问事实性问题,后问态度和意向方面的问题,最后问理由方面的问题。

(三)通俗易懂、便于回答

问题设计要充分考虑应答者的知识层次和接受能力,语句通俗,语意明确,内容具体,措辞客观,不带任何主观倾向和暗示,语气亲切自然。对敏感性问题的提问要注意技巧和方法,避免直接提问引起应答者的戒备和逆反心理。而且涉及的问题应当是可以测度的,使应答者一目了然,并愿意积极配合,如实回答。

(四)其他注意事项

控制问卷的长度,回答问题的时间控制在 20 分钟左右;在问卷格式编排上,要便于数据的校验、整理和汇总;尊重应答者的私人信息,注意保密等。

二、问卷的组成

调查问卷一般由说明词、主题问句和作业记录三部分组成。

说明词列于问卷的前面,用来说明调查的目的、内容 and 要求,请求被调查者给予合作等;主题问句是用来收集数据资料的一系列问句,是构成问卷的主体;作业记录是问卷执行和完成情况的记录,由调查者最后填写。

三、问题的形式

根据不同的调查内容和要求,问句可以选择以下几种形式:

(一)开放式问句

开放式问句又称自由式问句,即提出问题,但没有备选答案,由应答者自由回答,这种问句的答案可能是漫无边际的,不便于统计汇总和整理分析,因此一般不宜多采用。

例如:您对目前的娱乐类电视节目有什么看法?

(二)单选式问句

这类问句一般有两个以上的备选答案,要求被调查者选择一个。如回答“是”或“否”,容易提问,也容易回答。但只限于收集简单的事实或态度,对于很多复杂问题不适用。

例如:您的性别:()

A. 男 B. 女

(三)多选式问句

这类问句列举了几种可能的备选答案,由调查者选出最符合实际的两个或多个答案。其优点是答案有一定的范围,便于调查者取舍。但由于备选答案不可能穷尽所有的情况,也容易使被调查者做出随意的选择。

例如:您每天看哪些类型的电视节目?()

A. 综合新闻 B. 经济信息 C. 娱乐 D. 体育 E. 法治 F. 益智

(四)顺位式问句

顺位式问句即列出对某一问题不同层次的答案,由被调查者排出次序,表明自己的态度或倾向。

例如:您对以下电视节目的喜好程度从高到低分别是:()

A. 综合新闻 B. 经济信息 C. 娱乐 D. 体育 E. 法治 F. 益智

(五)标度式问句

这是一种直接测定被调查者主观感受和意见强弱程度的问句。常用的形式有两种:

一种是简单分级标示,在一个横轴上标示出一个方向的强弱程度,由被调查者根据自己的感觉和意见在上面圈注。这种形式只适用于对一事物只有一种看法的情况。

例如:你对体育类电视节目的态度:()

A. 很感兴趣 B. 比较感兴趣 C. 一般 D. 不感兴趣

另一种称为两极间分级标示,即用一个横轴标示出具有两个对立方向的不同强弱程度,由被调查者圈注。这种形式适用于对一事物有两种极端态度的情况。

例如:你认为这次中韩足球对抗赛中,国家足球队队员的表现:()

A. 很好 B. 好 C. 较好 D. 一般 E. 较差 F. 差 G. 很差

应用案例

《财富生活》杂志读者问卷调查

说明词:

在这里,您可以看到:财富巨子、商界新锐讲述创富故事,分享成功之道;最新国际、国内财经资讯集萃,趋势分析;城市精英参考的理财信息;下一季的流行宣言和历经岁月积淀的经典奢华;领略最具新奇特色的玩乐体验和品味独到的生活方式。生活原是一场阅读,打开《财富生活》,您会看到不同的人生。为了更好地为您提供服务,请您填写以下内容。您的意见和建议会得到我们的重视,并被迅速反馈到杂志中。

参与填写《财富生活》调查表的前 100 位读者将收到一份精美的礼品。

问卷主体:

1. 年龄: A. 1~25 岁 B. 26~30 岁 C. 31~35 岁 D. 36~45 岁 E. 46 岁以上
2. 学历: A. 专科 B. 本科 C. 本科以上
3. 婚姻: A. 已婚 B. 未婚

4. 平均月收入: A. 5 000 元以下 B. 5 000~10 000 元 C. 10 000~15 000 元
D. 15 000~20 000 元 E. 20 000 元以上
5. 平均月消费: A. 3 000 元以下 B. 3 000~5 000 元 C. 5 000~8 000 元
D. 8 000~10 000 元 E. 10 000 元以上
6. 您是从何种途径购买、阅读到《财富生活》杂志的?
A. 报刊亭 B. 地铁 C. 书店 D. 写字楼 E. 机场 F. 酒店 G. 超市
H. 其他
7. 您对本期杂志的整体感觉:
A. 非常好 B. 还不错 C. 一般 D. 较差 E. 非常差
8. 下期您还会购买或阅读《财富生活》杂志吗?
A. 肯定会 B. 先翻阅一下再说 C. 很难说 D. 不会
9. 您手中这本杂志的传阅范围是:
A. 1~3 人 B. 4~6 人 C. 7~9 人 D. 10 人以上
10. 这本杂志吸引您的是:
A. 封面人物 B. 排版设计 C. 专题创意 D. 文章内容 E. 图片质量
11. 您感兴趣的版块是:
A. 封面故事 B. 财富管理 C. 风尚 D. 休闲生活 E. 名家专栏
F. 文化娱乐
12. 您感兴趣的栏目是:
A. 财富大家 B. 财富宝贝 C. 观点 D. 财经资讯 E. 名编时评
F. 特别策划 G. 品牌故事 H. 新视界 I. 名车 J. 私家地理
K. 健康运动 L. 写食主义 M. 零点看板 N. 张望世界 O. 私藏
P. 胡说有道 Q. 碟语轻歌 R. 人物 S. 声色 T. 悦读
U. 话题
13. 您最喜欢的本期文章是:

14. 您最不喜欢的本期文章是:

15. 您最喜欢的本期版式(图片)是:

16. 您最不喜欢的本期版式(图片)是:

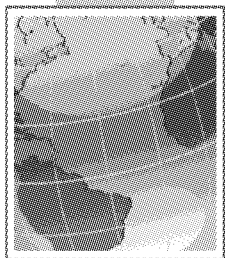
17. 您认为还应当增加或删除什么内容或栏目:

18. 您对《财富生活》杂志还有什么意见和建议:

背景介绍:《财富生活》是由中国科学技术协会主管、中国科技新闻学会主办、由《硅谷》杂志社(深圳)编辑出版的一本时尚杂志,读者群主要集中于高知、高薪、高品位的白领阶层。此次读者调查的目的在于了解读者的基本情况及对杂志的看法,以更好地贴近读者,办出自己的特色。

资料来源:《财富生活》,2006 年第 3 期。

I 表格与图形法



第三章 统计数据整理



学习目的

- 了解数据整理在统计工作中的作用、数据整理的内容、数据审核与汇总的技术。
- 领会统计分组的概念、关键,统计分组体系及次数分布的类型和特征。
- 明确统计数据的表现形式——统计表和统计图的基本内容。
- 掌握统计分组、分配数列和统计表(统计图)的编制(绘制)方法。



重点难点提示

- 重点:统计分组、变量数列的编制、统计数据的显示方法——统计表和统计图。
- 难点:统计分组、变量数列的编制、次数分布图的绘制。

经过统计调查及其他途径收集到的统计数据,往往是原始零乱和不系统的。必须按照统计研究的目的任务对数据进行加工整理,使其条理化和系统化,统计分组是统计整理最基本的方法。统计分组后形成的分配数列是以统计表格的形式来表现统计数据的,借助计算机的统计软件,还可以生成诸多直观形象的统计图。表格和图形法是描述统计学中最基本的内容和方法。本章介绍数据整理的一般问题、统计分组及统计数据整理结果的表现载体——统计表和统计图。

第一节 数据整理的一般问题

一、数据整理(data handling)的概念和作用

统计数据整理有广义和狭义两方面的含义。

狭义的数据整理是对统计调查所收集到的第一手数据进行分类和汇总,使其系统化、条理化、科学化,以得出反映事物总体综合特征的资料的工作过程,称为汇总性整理;广义的数据整理还包括对第二手数据进行的再整理。本节所指的数据整理是第一层含义,即狭义的统计整理。

统计调查取得的原始数据是分散杂乱、不系统的,只能表明各个被调查单位的个体情况,反映事物的表面现象或某一侧面,不能说明事物的全貌和总体特征。因此,只有对这些数据进行加工整理才能认识现象的整体特征及其内在联系。统计数据整理,是统计调查的继续,也是统计分析的前提,是感性认识到理性认识的连接点,承前启后,在整个统计工作中具有重要的

作用。

二、数据整理的程序

(一)数据整理方案的设计

科学合理地制订整理方案,是保证统计整理有计划、有组织地进行的首要步骤,是统计设计在整理阶段的具体化。统计整理方案的主要内容包括:选择整理组织形式,确定分组和分类目录,选定统计资料审核办法,制定整理表式等。

(二)数据的审核与检查

数据审核就是检查数据中是否有错误。从不同渠道取得的数据及不同类型的数据在审核的内容和方法上都有差异。

对于通过调查取得的第一手数据,应主要从完整性和准确性两个方面去审核。完整性审核主要是检查调查单位或个体是否有遗漏,所有的调查项目是否填写齐全等。准确性审核主要是检查数据是否有错误,是否存在异常值等。对于异常值要仔细进行鉴别:如果异常值属于记录时的错误,在分析前应予以纠正;如果异常值是正确的数值,则应予以保留。

对于通过其他渠道取得的第二手数据,应着重审核数据的适用性和时效性。数据使用者首先应弄清数据的来源、口径以及有关的背景材料,以便确定这些数据是否符合分析研究的需要,不能盲目地生搬硬套。此外,还要对数据的时效性进行审核,对于有些时效性较强的问题,如果所取得的数据过于滞后,就可能失去研究价值。

(三)数据的排序

数据排序是按一定标准将数据排列,以便研究者通过浏览数据发现一些明显的特征或趋势,找到解决问题的线索。除此之外,排序还有助于对数据检查纠错,以及为重新归类或分组等提供方便。

对于定性数据,如果是字母型数据,排序有升序、降序之分,但习惯上升序用得更多,因为升序与字母的自然排列相同;如果是汉字型数据,排列方式很多,如可按汉字的首位拼音字母排序,这与字母型数据的排序完全一样,也可按姓氏笔画排序。对于数值型数据的排序只有两种,即递增或递减。

排序后的数据也称为顺序统计量。无论是定性数据还是定量数据,排序均可借助于计算机很容易地完成。

(四)数据分组

数据分组是根据所设计的统计整理方案的要求,按事先确定的汇总组织形式和汇总的具体方法,将数据进行分组汇总,计算各组单位数和总体单位总数,计算各组的指标数值和总体的指标总量。在统计整理阶段,大量的、具体的工作就是对原始资料进行分组汇总和计算。数据分组是数据整理的基本方法。

(五)数据的显示——统计图表

通过统计整理以后的统计资料包括两方面的内容,即数据资料和相关的文字资料。其中,数据资料的条理化、科学化往往通过编制统计表或绘制统计图显示出来,因此,统计表和统计图成为显示统计数据的重要工具。

(六)数据的积累与保管

从长期动态的角度,统计数据整理还包括对数据的积累与保管,以利于统计数据的纵向对比,反映现象发展的规律及趋势。

第二节 数据分组

一、数据分组的概念与作用

(一)概念

数据分组(data grouping)就是根据统计研究目的和客观现象的内在特点,按照选定的某个或几个标志,将被研究的总体数据分成若干部分的科学分类。

数据分组应遵循“相同者合并,不同者分开”的原则,将相同性质的总体单位划分在同一组内,而不同性质的总体单位划分在不同的组,使组与组之间具有差别性,而同一组内又保持相对的同质性。例如,在整理人口资料时,可按性别标志分成男、女两组,还可按年龄、文化程度等标志分成若干组。

(二)统计分组的作用

1. 区分现象的类型

复杂的客观现象之间往往存在着各种差异,通过统计分组将它们划分为性质不同的类型,以体现出现象之间质与量上的差异。它是统计工作中应用最广泛、最主要的分组。

2. 反映总体的内部结构

将客观现象总体按照某一标志分成若干组成部分后,计算各组成部分占总体的比重,揭示总体内部的构成,反映部分与总体之间、部分与部分之间的数量关系。

3. 揭示现象之间的依存关系

通过统计分组,可以揭示现象之间的数量关系,从而更加全面、深刻地认识事物。例如,一国财政支出结构表明,用于科教经费方面的支出所占比重越大,国内生产总值和财政收入增长越快。这说明经济增长与科教发展之间有密切的相互促进作用,也使我们更加深刻地领会科学技术是第一生产力的内涵和我国的“科教兴国”战略。

4. 反映事物的数量特征和发展规律

通过统计分组将零星分散的数据资料进行整理,可发现其规律和特点。通过统计调查取得的资料,往往是零星分散的,甚至是杂乱无章的,统计分组的任务就是将它们整理得既有条理,又能反映事物的特点和规律。例如,通过对某班英语期中考试成绩进行分组,发现大部分学生的成绩都处于70~80分这一组中,越往两端,人数越少,符合统计学中的正态分布规律。

二、数据分组的关键

(一)数据分组的关键——选择分组标志和划分各组界限

统计分组是基本的统计方法之一,分组的好坏直接关系到统计整理能否得到正确、客观的统计资料,也关系到能否得出正确的结论。数据分组的关键就是正确地选择分组标志。选定了分组标志,还要进一步在分组标志变异的范围内,划定各个相邻组之间的性质界限和数值界限。如果划不清各组的界限,就将失去分组的意义。

(二)如何选择分组标志

1. 根据研究目的选择分组标志

同一现象由于研究目的不同,选择的分组标志也就不同,应选择与研究目的有密切关系的

标志作为分组标志,才能使统计分组提供符合要求的分组资料。例如,要研究工业企业经济类型的构成,就要选择所有制这个标志。要研究人口的老龄化程度,就应选择年龄作为分组标志。

2. 选择能够反映现象本质的标志

客观现象的标志多种多样,有的标志能够揭示现象的本质特征,是有决定意义的重要标志;有的则是非本质、无关紧要的标志。只有选择那些能够揭示现象本质特征的重要标志作为分组标志,才能得到反映现象本质特征的分组资料。例如,在对比研究各国经济发展程度和富裕程度时,既可选择社会总产值、国内生产总值(GDP)、国民生产总值(GNP)等总量指标,也可选择人均 GDP、人均 GNP 等分析指标,但总量指标只能反映经济发展规模和总量,不能反映经济发展强度,而人均 GDP、人均 GNP 等指标能反映一个国家或地区一定时期内人均创造的社会财富,因此,应选择人均 GDP 或人均 GNP 作为分组标志。

3. 考虑现象所处的历史条件或经济条件

社会经济现象随着时间、地点、条件的不同而经常发生变化。同一分组标志在某一时期适用,在另一时期就不一定适用;在某一条件下适用,在另一条件下就不一定适用。某一标志在一定历史条件是重要的分组标志,但时过境迁,可能失去其重要意义。因此,在选择分组标志时,应考虑到现象所处的具体历史条件或经济条件,做到与时俱进。

知识链接 ——人均 GDP 与国家贫富排行榜

根据国际货币基金组织(IMF)官方网站 2017 年 4 月发布的数据显示,2016 年世界各国(地区)人均 GDP 最高的是卢森堡,达到103 199美元,排名第二至十位的分别是瑞士(79 242 美元)、挪威(70 392美元)、中国澳门特区(67 079美元)、爱尔兰(62 562美元)、卡塔尔(60 787 美元)、冰岛(59 629美元)、美国(57 436美元)、丹麦(53 744美元)及新加坡(52 961美元)。中国大陆 2016 年 GDP 总量达112 182.81亿美元,世界排名第二,仅次于美国(185 691亿美元);人均 GDP 8 113美元,世界排名第 74 位,与发达国家相比还有较大差距。

资料来源:IMF 中文主页(<http://www.imf.org/external/chinese/>)。

考考你

在对某市居民的生活质量进行调研时,按家庭人均收入分组和按家庭人均支出分组哪个更合适?

三、数据分组方法

根据分组标志的不同特征,统计总体可以按品质变量(品质标志)分组,也可以按数值变量(数量标志)分组。

(一)按品质变量分组

按品质变量分组,是指选择反映客观现象属性差异的品质标志作为分组标志。例如,人口按性别可分成男、女两组。

按品质变量分组有简单和复杂两种情况。简单的分组,标志一经确定,组数和组限都非常明显。而复杂的分组,组与组的界限往往不易划清,组数、组限难以确定,如国民经济部门

分类、产品分类、职业分类等。在统计工作中,对于比较复杂的分组,则由上级统计机关或各业务主管部门统一编制标准的分类目录,如商品分类目录、工业产品分类目录、工业部门分类目录等。

知识链接 ——我国国民经济核算中的三大产业分类

根据国家统计局最新发布的《国民经济行业分类》(GB/T 4754—2017)的规定,我国国民经济核算中的三大产业具体划分如下:第一产业是指农、林、牧、渔业(不含农、林、牧、渔中的服务业)。第二产业是指采矿业(不含开采辅助活动),制造业(不含金属制品、机械和设备修理业),电力、热力、燃气及水生产和供应业,建筑业。第三产业即服务业,是指除第一产业、第二产业以外的其他行业。主要包括:批发和零售业,交通运输、仓储和邮政业,住宿和餐饮业,信息传输、软件和信息技术服务业,金融业,房地产业,租赁和商务服务业,科学研究和技术服务业,水利、环境和公共设施管理业,居民服务、修理和其他服务业,教育、卫生和社会工作,文化、体育和娱乐业,公共管理、社会保障和社会组织,国际组织,以及农、林、牧、渔业中的农、林、牧、渔服务业,采矿业中的开采辅助活动,制造业中的金属制品、机械和设备修理业。

资料来源:国家统计局官方网站。

(二)按数值变量分组

按数值变量分组,是指选择反映客观现象数量差异的数量标志作为分组标志。例如,职工按工资收入分组、工业企业按产值分组或按计划完成程度分组等。

按数值变量分组,能够准确地反映客观现象在数量上的差异,但有时却难以反映其性质上的差异。哲学告诉我们,量变到一定程度会引起质变,根据质量互变规律,在按数量标志分组时,必须找到量变到质变的数量界限,然后按这些数量界限进行分组,才能反映事物的本质特征。如工业总产值按计划完成程度分组,100%就是一个量变到质变的数量界限,低于100%为未完成计划,100%为完成计划,超过100%为超额完成计划。

根据数值变量的变化特征,可供选择的分组方式有单项式分组和组距式分组两种。

1. 单项式分组

单项式分组即按每个具体变量值对总体所进行的分组,单项式分组中每组只有一个变量值。如工人按看管设备台数分组(见表3-1)。

表3-1 某车间工人看管设备台数统计表

按工人看管设备台数分组(台)	工人数(人)
1	20
2	40
3	15
4	15
合 计	90

单项式分组只适用于离散型变量,且在变量值不多、变动范围有限的条件下采用。当离散型变量变动范围比较大、统计单位数又很多的情况下,若采用单项式分组,把每一变量作为一

组,则必然会使分组的组数过多。各组次数过于分散,不能反映总体内部各部分的性质和差异,从而失去了统计分组的真正意义。对于连续型变量,由于其变量值无法一一列举,不能采用单项式分组。在上述情况下,就需要采用组距式分组方法。

2. 组距式分组

组距式分组是按变量值的一定范围对现象总体所进行的分组。在现象总体的变动范围内,将其划分为若干个区间,各区间内的所有变量值作为一组,其性质相同,组与组之间的性质相异。与单项式分组相比较,各组的变量值不是某一个具体的点值,而是一段区间。如对我国土地海拔状况的分组(见表 3-2)。

表 3-2 我国土地海拔状况分组表

按海拔分组 (米)	面 积 (万平方公里)	占总面积比重 (%)
500 以下	241.70	25.18
500~1 000	162.50	16.93
1 000~2 000	239.90	24.99
2 000~3 000	67.60	7.04
3 000 以上	248.30	25.86
合 计	960.00	100.00

资料来源:国家统计局编《中国统计年鉴 1998》,中国统计出版社,第 6 页。

组距式分组一般在变量值变动幅度较大的条件下采用。在组距式分组中,涉及组限、组数、组距、组中值等分组要素。

组限(class boundaries),是用来表示各组之间界限的变量值,是决定事物质量的数量界限。

组限中,每一组中的最小变量值为下限,最大变量值为上限。如表 3-2 中,左栏数据都是组限,在第二组中,“500 米”是下限,“1 000 米”是上限。

组限的表达形式与变量的特点密切相关。连续型变量分组,相邻组的组限必须重合。如表 3-2 中各组的组限 500 米、1 000 米、2 000 米、3 000 米,既作为前一组的上限,又作为后一组的下限,数值重叠。那么这些变量值应归属于哪一组?统计上一般按“上限不在内”的原则进行处理,也就是说,作为上限的变量值应归属于下一组,如“1 000 米”应归属于第三组,其余类推。

离散型变量分组,相邻组的组限既可间断,也可重合,一般采用间断式表达(见表 3-3)。

表 3-3 某企业职工年龄构成

按年龄分组(岁)	人口数(人)
18~35	226
36~49	134
50 以上	40
合 计	400

组距(width of class interval),是指一组变量值的区间长度,也就是每一组上限与下限之间的差量。公式为:组距=上限-下限。

在表 3-2 中,第二组的组距=1 000-500=500(米)。组距式分组中,既有下限又有上限的组称为“闭口组”;有上限无下限或有下限无上限的组,称为“开口组”,一般用“×××以下”或“×××以上”表示。如表 3-2 中的“500 米以下”和“3 000 米以上”两组。

在组距式分组中,根据各组的组距是否相等,可以分为等距分组和异距分组。每组组距都相等的分组称为等距分组,各组组距不尽相同的称为异距分组。采用等距分组还是异距分组,要根据研究的目的和现象的特点来决定。等距分组时,由于组距相等,各组次数分布不受组距大小的影响,便于比较,有利于现象间依存关系的研究。异距分组时,各组次数的多少受组距大小影响,为了比较则需要借助于次数密度指标(这一指标将在后面介绍)。

组数,是指分组的个数。

在研究总体一定的情况下,组数的多少和组距的大小是紧密联系的。一般来说,组数和组距成反比关系,即组数少,则组距大;组数多,则组距小。如果组数太多,组距过小,会使分组资料烦琐、庞杂,难以显示总体内部的特征和分布规律;如果组数太少,组距过大,可能会失去分组意义,达不到正确反映客观事实的目的。在确定组距和组数时,应注意保证各组都能有足够的单位数,组数既不能太多,也不宜太少,应以能充分、准确体现现象的分布特征为宜。在等距分组条件下,组数等于全距除以组距。

组中值(class midpoint),是指各组变量值的代表性水平。

组限重合式组的组中值=(下限+上限)/2

组限非重合式组的组中值=(本组下限+后一组下限)/2

缺下限组的组中值=上限-(邻组组距/2)

缺上限组的组中值=下限+(邻组组距/2)

四、数据分组体系

数据分组按其分组标志的多少及交叉状况,可以区分为简单分组、复合分组两大体系。

(一)简单分组

简单分组,就是对被研究现象总体只按一个标志所进行的分组。如表 3-3 中用“年龄”一个标志对职工进行分组就属于简单分组。

(二)复合分组

复合分组就是对被研究的现象总体按两个或两个以上的标志结合起来所进行的分组。具体而言,它是先按一个标志分组,然后再按另一个标志将已分好的各组又划分为若干小组。复合分组的排列形式既可以是层叠式,也可以是交叉式。参见表 3-4、表 3-5。

表 3-4 某国高等学校按学科分组的在校学生数(层叠式分组)

指 标	在校学生(人)
总 计	3 174 362
本 科	1 986 125

续表

指 标	在校学生(人)
专 科	1 188 237
理 学	332 178
本 科	192 847
专 科	139 331
经济学	483 446
本 科	286 001
专 科	197 445
医 学	271 137
本 科	192 220
专 科	78 917
...	...
...	...

表 3—5

某国高等学校按学科分组的在校学生数(交叉式分组)

单位:人

	总 计	理 学	经济学	医 学	...
合 计	3 174 362	332 178	483 446	271 137	...
本 科	1 986 125	192 847	286 001	192 220	...
专 科	1 188 237	139 331	197 445	78 917	...

第三节 分配数列和次数分布

一、分配数列

(一)分配数列的概念

将以分组标志确定的组别依次排列,同时列出各组的次数或频率所形成的数列称为次数分布数列或频数分布数列,又称分配数列(frequency series)。

编制分配数列是统计整理的一种重要形式,也是统计描述和统计分析的一种重要方法。

分配数列可以反映总体中所有单位在各组间的分布状态、分布特征以及结构状况,并可据此研究总体某一标志的平均水平及其变动规律。

(二)分配数列的种类与编制

1. 分配数列的种类

依据分组标志的不同,分配数列可以分为品质数列和变量数列。

按品质变量将总体分组而形成的分配数列称为品质分配数列,简称品质数列。品质数列由两个要素构成:一是分组标志及用文字表示的各组名称,简称组别;二是各组分配的次数或频率(见表3—6)。

表 3—6

某企业员工性别分布表

按性别分组	员工人数(人)	比重(%)
男	187	46.75
女	213	53.25
合 计	400	100.00

按数值变量将总体分组而形成的分配数列称为变量分配数列,简称变量数列。变量数列也由两个要素构成:一是分组变量及由变量值表示的各组名称;二是各组分配的次数或频率。

次数(frequency),指分布在各组中的单位数,又称频数。

频率(relative frequency),是各组次数与总次数的比值,又称比率。

变量数列按数列中每组变量值的多少及取值范围,可分为单项数列和组距数列。

单项式变量数列,就是每组只有一个变量值的变量数列,在变量值的变动幅度较小时使用(见表 3—1 和表 3—8)。

组距式变量数列,就是由变量值的一定区间来表现的数列(见表 3—7)。

变量数列按数列中变量值的类型不同,可分为连续变量数列和离散变量数列。

连续变量数列,是指数列中的变量是连续不断变化的,相邻的两个变量值之间可作无限分割。例如,以人的身高、体重、收入或商品销售额等变量编制的变量数列。由于连续变量是连续不断变化的,不能一一列举它的数值,适宜于编制组距数列。

离散变量数列,是指数列中各变量值之间是以整数位断开的。或者说,数列中的变量值的单位不能无限细分。例如,以学校数、商业网点数、机器台数、职工人数等变量编制的数列。由于离散变量数列中的变量值可以用整数一一列举,当变量值变动范围较小时,可以编制单项变量数列(见表 3—8);当变量值的变动范围较大、变量个数较多时,则需要编制组距变量数列。

表 3—7

某企业职工年龄分布表

按年龄分组(岁)	职工人数(人)	比重(%)
20 以下	52	13.00
20~30	68	17.00
30~40	127	31.75
40~50	89	22.25
50 以上	64	16.00
合 计	400	100.00

表 3-8

某企业工人日产量分组表

按日产量分组(件)	工人人数(人)	比重(%)
5	14	7.00
6	27	13.50
7	56	28.00
8	63	31.50
9	29	14.50
10	11	5.50
合 计	200	100.00

2. 分配数列的编制

由于品质数列相对简单,以下以变量数列为例进行介绍。

变量数列编制时的注意事项:

第一,组距最好为 5 或 10 的倍数,以保持数列的规范整齐。

第二,最小组的下限应略低于最小变量值,最大组的上限略高于最大变量值;当存在极端值时,可以采用开口组,以保证所有数据不遗漏。

第三,离散型变量分组,相邻组的组限可以间断,也可以重叠;连续型变量分组,相邻组的组限必须重叠。

第四,相邻组组限重叠时,临界点的数据按“上限不在内”的原则归组。

变量数列编制的基本步骤为:

第一步:将原始资料按数值大小依次排列,计算全距(range),全距等于数列中的最大变量值与最小变量值之差,反映数据的最大变异程度。

第二步:根据变量的类型确定分组方法(单项式分组或组距式分组)。

第三步:确定组数和组距。两者成反比关系,当组数确定后,在等距数列中可以用以下公式近似计算组距:组距=全距÷组数。

确定组数和组距要考虑尽量反映总体分布的规律,考虑组距内的同质性、组间的差异性,组距一般不能太大。

第四步:确定组限(第一组的下限要小于或等于最小变量值,最后一组的上限要大于最大变量值;连续型变量分组,相邻组的组限必须重叠;离散型变量分组,相邻组的组限可以间断,也可以重叠)。

第五步:汇总出各组的单位数,计算频率,并编制统计表。

[例 3-1] 对某市居民家庭人均食物类消费支出情况进行抽样调查,得到 50 户家庭人均月食物类消费性支出(单位:元)。资料如下:

518	608	674	831	1 314	409	662	780	991	476	605
638	920	296	1 270	542	666	626	375	593	700	1 020
620	564	617	665	240	490	662	695	650	499	599
789	450	198	605	677	890	638	530	647	889	300
858	671	1 160	580	861	756					

以上资料杂乱无章,很难看出分布特征,需要对其进行统计整理。先将这些数据从小到大

顺序排列:

198 240 296 300 375 409 450 476 490 499 518
530 542 564 580 593 599 605 605 608 617 620
626 638 638 647 650 662 662 665 666 671 674
677 689 695 700 756 780 789 831 858 861 890
920 991 1 020 1 160 1 270 1 314

经过初步整理,可以看出,最大值为 1 314 元,最小值为 198 元,全距为 1 116 元(1 314—198)。由于变量值变动幅度很大,且为连续变量,因此不宜编制单项式数列,只能编制组距式数列。编制组距式数列的关键是确定组距与组数。根据数据分布的特点,本例将数据分为 5 组,采用异距数列形式。

对极端值采用开口组处理,最终得到如下分配数列(见表 3—9)。

表 3—9 某市居民家庭人均月食物类消费支出分布表

人均月食物类消费支出(元)	家庭数(户)	比重(%)
300 以下	3	6.00
300~500	7	14.00
500~700	26	52.00
700~1 000	10	20.00
1 000 以上	4	8.00
合 计	50	100.00

从表 3—9 可以看出,不同家庭人均月食物类消费支出水平的分布特征被清晰地表现出来了,每月支出在 500~700 元的家庭最多,占 52.00%;支出在 300 元以下的家庭最少,仅占 6.00%。

二、次数分布

将统计数据按其分组标志进行分组的过程,实际上就是将变量值在各组中的次数进行分布的过程。次数分布就是观察值按其分组标志分布在各组内的次数。

(一)次数分布的种类

性质不同的现象呈现不同的次数分布,大致可以归纳为以下三种类型:

1. 钟形分布

钟形分布又称丘形分布,它的特点是“两头小,中间大”,即越靠近中间的变量值,分布次数越多;越远离变量值中点,分布次数越少,形态如钟或山丘。根据两侧的次数分布是否对称,钟形分布可分为对称分布和非对称分布。

对称分布又称正态分布,它是以变量值的中点为对称轴,两侧变量值分布的次数随着离变量值中点距离增大而逐渐减少,减少的次数基本相等(见图 3—1)。大多数社会经济现象符合正态分布的特征。例如,大量观察法下人群中身高、体重的分布一般属于正态分布。

非对称分布又称偏态分布。图形有不同方向和程度的偏斜。如向右偏斜的右偏分布(见图 3—2)和向左偏斜的左偏分布(见图 3—3)。

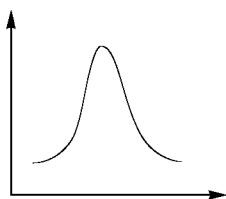


图 3-1 正态分布

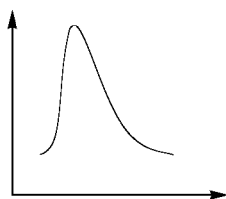


图 3-2 右偏分布

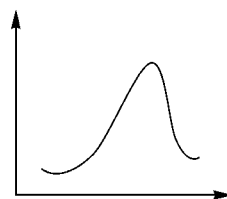


图 3-3 左偏分布

举一反三

举例说明还有哪些社会经济现象符合正态分布的特征?

2. U 形分布

U 形分布,是指较大和较小的变量值出现的次数都偏多,而中间变量值出现的次数偏少,表现出“两头大,中间小”的次数分布特征。其次数分布曲线形如 U 字,如图 3-4 所示。例如,人口按年龄分布的死亡次数分布呈 U 形分布。

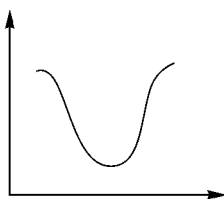


图 3-4 U 形分布

3. J 形分布

J 形分布的特征是“一边大,一边小”。即次数随着变量值的变化大多数集中在某一端的分布。其曲线形如英文字母的“J”字,具体有两种类型。

(1)正 J 形分布。表现为次数随着变量值的增大而增多,大部分变量值集中分布在右边,如图 3-5 所示。

(2)反 J 形分布。表现为次数随着变量值的增大而减少,如图 3-6 所示。

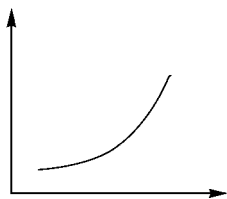


图 3-5 正 J 形分布

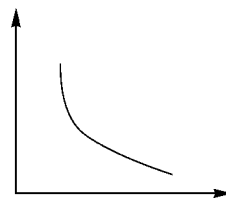


图 3-6 反 J 形分布

举一反三

还有哪些社会经济现象符合 J 形分布的特征?

(二) 累计次数和频率

累计次数和频率也是说明总体中各单位分布特征的指标,用以说明总体中在某一变量值水平上下总共包含的总体单位次数和频率。按累计方向不同,累计次数和频率的计算方法有两种:

(1)向上累计。它是将各组次数和频率由变量值低的组向变量值高的组逐组累计。在组距数列中,说明各组上限以下总共包含的总体单位次数和比率有多少。

(2)向下累计。它是将各组次数和频率由变量值高的组向变量值低的组逐组累计。在组距数列中,说明各组下限以上总共包含的总体单位次数和比率有多少。

累计次数和累计频率可以更简便地概括总体各单位的分布特征。现以表 3—10 为例说明。

表 3—10 某班学生考试成绩分布表

按成绩分组 (分)	学生人数 (人)	频率 (%)	向上累计		向下累计	
			次数(人)	频率(%)	次数(人)	频率(%)
50~60	3	7.5	3	7.5	40	100.0
60~70	8	20.00	11	27.5	37	92.5
70~80	16	40.0	27	67.5	29	72.5
80~90	8	20.0	35	87.5	13	32.5
90~100	5	12.5	40	100.0	5	12.5
合 计	40	100.0	—	—	—	—

累计次数的特点是:向上累计或向下累计次数之和均等于总体单位总数,而累计频率之和等于 100%。

考考你

表 3—10 中,向上累计中的 27 人、67.5%和向下累计中的 29 人、72.5%的含义是什么?

(三) 简单次数分布图的绘制

1. 单项式数列次数分布图的绘制

以变量值为横轴,以次数为纵轴,在坐标系中描出各组变量值和相应的分配次数所对应的坐标点,用折线连起,即得到分布曲线图。具体图形见本章第四节。

2. 组距式数列次数分布图的绘制

(1)等距数列次数分布图。以横轴代表变量值,以纵轴代表次数,在坐标轴上标出各组组限与各组次数,以各组组距为宽,各组次数为高,绘出对应矩形,各组矩形连在一起构成次数分布柱形图。将各组矩形上端的中点(即各组组中值与各组次数的交点)连成一条折线,形成次数分布曲线图。

(2)异距数列次数分布图。由于异距数列的次数分布受变量值和组距两个因素的影响,必须消除组距大小不等的影响,将不等组距换算成统一标准组距次数(标准组距通常是组距数列中的最小组距),换算公式如下:

$$\begin{aligned}\text{某组标准组距次数} &= \text{该组单位组距次数} \times \text{标准组距} \\ &= (\text{该组次数} / \text{该组组距}) \times \text{标准组距} \\ &= \text{该组次数密度} \times \text{标准组距}\end{aligned}$$

次数密度(frequency distribution density),也称频数密度,是指单位组距内分布的次数。

以横轴代表各组变量值,以纵轴代表各组次数密度,绘制其次数分布图。具体图形见本章第四节。

第四节 统计图表

一、统计表

(一)统计表的概念与作用

把经过汇总整理得出的系统化的统计数据资料按一定顺序填列在一定的表格内,形成的就是统计表。统计表的作用表现在:

- (1)能使大量的统计数据条理化、系统化,更清晰地描述数据之间的相互关系。
- (2)采用表格形式使得统计数据的显示直观形象、简明易懂。
- (3)利用统计表便于计算和比较表内各项统计指标。
- (4)利用统计表易于检查数据的完整性和正确性。

(二)统计表的构成

从内容上看,统计表由主词和宾词两部分组成。主词是统计表所要说明的总体及其分组;宾词是用来说明总体的统计指标。在通常情况下,表的主词排列在表的左边,列于横栏;表的宾词排列在表的右边,列于纵栏。但有时为了更好地编排表的内容,也可将主宾词更换位置或合并排列。

从构成要素看,统计表包括总标题、横标目、纵标目和指标数值四个基本要素。

- (1)总标题。它是统计表的名称,用以概括说明整个表的内容,一般位于表的上方居中。
- (2)横标目。它是横行内容的名称,代表统计表所要说明的总体及其分组,通常也称主词。
- (3)纵标目。它是纵栏内容的名称,用来说明总体情况的统计指标名称,通常也称宾词。
- (4)指标数值。它是各项指标的具体数值,内容由横标目和纵标目限定,其数字可以是绝对数、相对数或平均数。

另外,为了补充统计表中未说明的问题,统计表往往还附有一些说明,包括资料来源、指标计算方法、填报单位、填表人、填表日期等。现以表3-11为例说明统计表的结构。

表 3-11

2016 年中国国内生产总值及构成

项 目	增加值	
	产值(亿元)	比重(%)
第一产业	63 671	8.56
第二产业	296 236	39.81
第三产业	384 221	51.63
合 计	744 127	100.00

资料来源：国家统计局《2016 年国民经济和社会发展统计公报》。

(三)统计表的种类

统计表按照总体分组情况不同,可分为简单表、简单分组表和复合分组表三类。

1. 简单表

主词未经任何分组的统计表称为简单表。简单表的主词一般按时间顺序排列,或按总体各单位名称排列。通常是对调查来的原始资料初步整理所采用的形式。如表3-12为按时间顺序排列形成的简单表,即为第五章中的时间序列;表3-13为按总体各单位名称排列的简单表。

表 3-12

2013~2016 年各季度我国商业银行整体不良贷款情况统计表

季 度	不良贷款余额(亿元)	不良贷款率(%)
2013 年第一季度	5 265	0.96
2013 年第二季度	5 395	0.96
2013 年第三季度	5 636	0.97
2013 年第四季度	5 921	1.00
2014 年第一季度	6 461	1.04
2014 年第二季度	6 944	1.08
2014 年第三季度	7 669	1.16
2014 年第四季度	8 426	1.25
2015 年第一季度	9 828	1.39
2015 年第二季度	10 919	1.50
2015 年第三季度	11 863	1.59
2015 年第四季度	12 744	1.67
2016 年第一季度	13 921	1.75
2016 年第二季度	14 373	1.75
2016 年第三季度	14 939	1.76
2016 年第四季度	15 123	1.74

数据来源：根据银监会官网资料整理。

表 3—13

2016 年某集团公司下属企业销售量统计表

厂 别	销售量(件)
甲	5 000
乙	7 000
丙	6 000
合 计	18 000

2. 简单分组表

主词按某一个标志进行分组的统计表称为简单分组表(见表 3—10)。

3. 复合分组表

主词按两个或两个以上标志进行复合分组的统计表称为复合分组表(见表 3—5)。

(四)编制统计表的注意事项

统计表编制的总要求是:简练、明确、实用、美观、便于比较。统计表的编制应注意如下事项:

第一,标题。统计表的总标题、项目、指标要简明扼要,能准确反映内容,使人一目了然,便于分析。

第二,线条的绘制。表的上下端用粗线,左右两边不封口;纵栏之间用细线分开,横行之间可以不加线。如果横行过多,也可以每 5 行加一细线。

第三,合计栏设置。统计表纵列若需合计时,一般应将合计列在最后一行;各横行若需合计时,可将合计列在最前一栏或最后一栏。

第四,指标数值。表中数字应填写整齐,对准位数。数字为零时要写出“0”;不应填写数字时用符号“—”表示;缺失某项数字时用符号“…”表示;统计表数字部分一般不留空白。如果某项数字与邻项数字相同,仍应填写数字,不得用“同上”、“同左”等字样代替。

第五,计量单位。如果整张表格的计量单位只有一个,通常列在表的右上角;如果计量单位较多,应列在相应指标栏内。

第六,注解或资料来源。必要时,在统计表下应加注解或说明,以便查考。

第七,统计表纵栏较多时,为便于阅读,可编栏号。习惯上对非填写数量数据的各栏分别以(甲)、(乙)、(丙)、(丁)……的次序编栏,对填写数量数据的各栏分别以(1)、(2)、(3)、(4)……的次序编栏。对各栏数字之间有一定计算关系的,也可用数学等号表示,如(3)=(2)÷(1)。

知识链接 —— 学术论文中的三线式表格

在学术专著、学术期刊论文或学位论文中,一般要求表格应采用标准格式——三线式表格排版,即表格中的线条只有三条横线,上下两条横线加粗,纵栏之间无需用竖线分开,具体格式如表 3—14 所示。

表 3-14 2011~2016 年湖北省地区生产总值及增长率

年份	2011	2012	2013	2014	2015	2016
地区生产总值 (亿元)	19 632	22 250	24 792	27 379	29 550	32 298
增长率(%)	13.8	11.3	10.1	9.7	8.9	8.1

二、统计图

(一)统计图的概念

所谓统计图,就是以点、线、面积、形状等方法描述、显示统计数据的形式。利用统计图显示统计资料形象具体、简明生动、通俗易懂、一目了然,可以给人以深刻的印象。它可以表示现象之间的对比关系,揭示总体结构及其变化,描述现象发展变化的趋势,分析现象之间的依存关系,说明现象在地域上的分布状况等。

(二)统计图的常见类型

1. 条形图

条形图是用条形的长短或高矮来显示各变量值分布次数多少的统计图。至于具体的形状,可以是线条,也可以是立体的圆柱、方柱或锥体。它最适用于显示离散型变量的次数分布。

2. 直方图和折线图

直方图用来显示连续变量的次数分布。其图形与条形图相似,但两者有根本区别。直方图每个条形的左边界为该组的下限,右边界是上限,中点为组中值。由于连续型变量组距数列组限是重合的,因此,直方图的条形是相互连接的。若某一组次数为零,则该组的条形不存在,但位置必须留出来,不能将其他组的条形移过来。

绘制直方图时,将所研究的变量放在横轴上,频数、相对频数或百分比频数放在纵轴上。每组的频数、相对频数或频率在图上就是一个长方形,长方形的底在横轴上,宽度是组距,长方形的高就是对应的频数或频率(见图 3-7)。

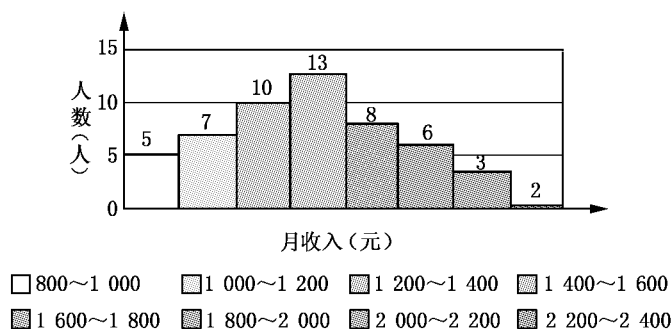


图 3-7 直方图

根据直方图,我们还可以做折线图。折线图是线图的一种,它是将直方图的各条形中点用线连接起来得到的,通常连线要延伸到高于和低于数列两端各一级的中点与横轴相交,这样,折线就与横轴围成一个多边形,折线图因此也称次数多边形图(见图 3-8)。

当对数据所分的组数很多时,组距会越来越小,这时所绘制的折线图就会越来越光滑,逐

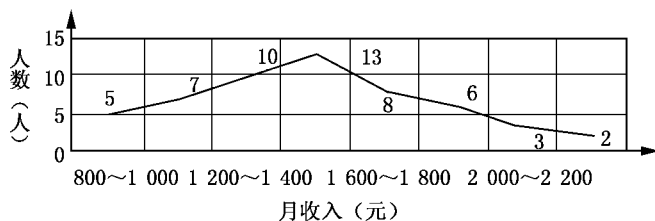


图 3-8 折线图

渐形成一条平滑的曲线,也即第七章中的概率分布曲线。

3. 圆形图

圆形图也称饼图,是以圆形的分割来表示总体的分组及结构情况。它最适宜用来显示定类变量的次数分布,因为它没有起点和终点,图中的各部分看不出顺序,这与定类变量各类别间是平等关系的特征一致。圆形图可以是平面的,也可以是立体形式的(见图3-9)。

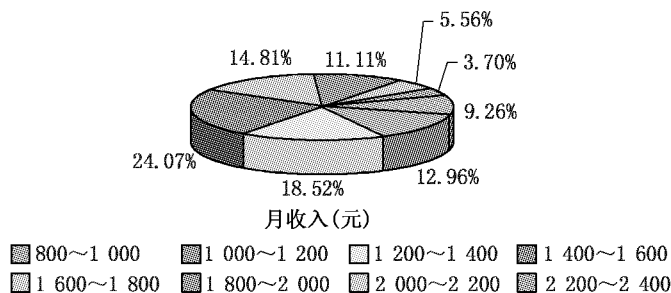


图 3-9 饼图

以上介绍的只是主要的统计图形式。除此之外,在一些特定的领域要使用特殊的统计图,如股市分析的蜡烛图、箱形图等,在此不一一介绍。统计图种类很多,我们可针对不同的数据类型和不同的研究目的,选择合适的图形。

本章小结

1. 统计数据整理是统计数据调查的继续和数据分析的前提,是感性认识上升到理性认识的连接点。数据整理的前提是审核,分组汇总是核心。按品质变量分组的关键是界定各类型组的性质差异;按数值变量分组的关键是正确确定各组的数量界限,即组数与组限。
2. 统计分组必须遵循穷尽性和互斥性两个原则。数值变量分组有单项式和组距式两种。组距式分组又有离散型与连续型、等距与异距分组之分,对于组距式分组要计算组距、组数、组中值。
3. 分配数列由两个要素构成:一是组别,二是各组次数或频率。
4. 次数分布主要有钟形分布、U形分布和J形分布。
5. 统计表和统计图是显示统计数据的两种重要形式。统计表的结构从形式上看,可分为总标题、横标目、纵标目和指标数值;从内容上看,可分为主词和宾词两部分。统计图主要有条形图、直方图、圆形图等。

常用术语

- | | |
|---|---------------------------|
| 1. 数据整理 data handling | 2. 数据分组 data grouping |
| 3. 次数 frequency | 4. 频率 relative frequency |
| 5. 频数分布 frequency distribution | 6. 全距 range |
| 7. 组距 width of class interval | 8. 组限 class boundaries |
| 9. 组中值 class midpoint | 10. 分配数列 frequency series |
| 11. 次数密度 frequency distribution density | 12. 统计表 statistical table |
| 13. 统计图 statistical graphics | 14. 直方图 histogram |
| 15. 饼图 pie chart | |

复习思考题

1. 统计整理的内容和程序有哪些?
2. 什么是统计分组? 统计分组的作用体现在哪里?
3. 统计分组的关键是什么?
4. 如何编制变量数列?
5. 常见的次数分布有几种类型?
6. 什么是累计次数和累计频数?
7. 什么是统计表? 其构成如何?
8. 什么是统计图? 常见的统计图有哪些?
9. 某班级 40 名学生年龄如下, 试编制变量数列:

22	17	18	16	17	18	23	17	17	20
17	17	18	17	16	21	20	19	18	17
17	18	20	24	19	16	18	20	16	17
16	18	17	17	16	17	17	16	21	17

附录

Excel 在数据整理中的应用

一、用 FREQUENCY 函数进行统计分组

用 FREQUENCY 函数进行统计分组, 首先要将数据输入 Excel 表的一列中, 并选定放置分组结果的区域。现以 50 名工人的日产量资料为例进行说明。

某车间 50 名工人日产零件数资料如下(单位: 件)

148	140	127	120	110	104	148	135	129	123
116	109	132	135	129	123	110	108	148	135
128	123	114	108	132	124	120	125	116	118
125	137	107	113	132	140	137	119	119	127
129	119	124	130	118	107	113	122	128	114

先将 50 名工人的日产量数据输入 A3 到 A52 单元格中, 如附表 3-1 所示, 并选定 C3 到 C7 单元格, 然后按照以下步骤进行操作:

第一步, 从“插入”菜单中选择“函数”项, 或单击“常用”工具栏的“fx”按钮, 弹出“粘贴函数”对话框, 在对话框左侧“函数分类”列表中选择“统计”, 再在右侧的“函数名”列表中选择 FREQUENCY(当前窗口没有, 可拖动滚动条寻找), 确定(回车)进入 FREQUENCY 对话框。

第二步,在 FREQUENCY 对话框的上端并列有两个框:

在 Data_array 框中输入待分组的频数分布(对话框中称“频数分布”)的原数据区域,本例可输入 A3:A52。既可以用鼠标拖选,也可以键盘输入。

在 Bins_array 框中输入分组的组限。FREQUENCY 要求按组距的上限分组(帮助信息称“分段点”),不接受非数据字符的分组标志(如“……以上”、“不足……”之类)。本例可输入 109、119、129、139、150。输入时注意,由于分组结果要给出一组频数,故必须以数据公式的形式输入,即在输入数据的两端加大括号 {}, 各数据(即上限)之间用分号隔开。

输入完毕,即在框下给出频数分布 6,13,17,9,5(后面的 0 表示没有其他)。

第三步,按 Shift+Ctrl+Enter 组合键,即将频数分布 6,13,17,9,5 记入指定的 C3 至 C7 单元格内(注意:回车和按“确定”按钮都无效)。

分组操作结果如附表 3-1 所示。

附表 3-1

	A	B	C	D	E	F	G	H
1					由低到高累计		由高到低累计	
2		按产量分组	频数(人)	频率(%)	频数	频率	频数	频率
3	148	100~110	6	12	6	12	50	100
4	140	110~120	13	26	19	38	44	88
5	127	120~130	17	34	36	72	31	62
6	120	130~1140	9	18	45	90	14	28
7	110	140~150	5	10	50	100	5	10
8	104	合 计	50	100				

取得频数分布后,可使用公式输入与函数相结合的方法继续计算频率、累计频数和累计频率。其方法是:

(1)首先合计频数。单击 C8 单元格,输入“=SUM(C3:C7)”,回车确认,得出结果为 50 人(SUM 为求和函数)。或选定 C3 至 C7 单元格,点击“常用”工具栏的“ Σ ”按钮,即得到这一栏的合计数。

(2)计算频率。单击 D3 单元格,输入“=C3*100/50”(* 为乘法符号),回车得出本组频率 12%;然后,使用填充柄功能(点击 D3,将鼠标移至单元格右下角的小黑方块上,鼠标变成黑十字形)按住鼠标左键向下拖至 D7,松开鼠标即得各组的频率;最后,使用 SUM 函数或按“ Σ ”按钮,得到 D8 的频率总和 100。

(3)计算由低到高累计频数和频率。单击 E3 单元格,输入“C3”,回车得出 6,再单击 E4 单元格,输入“=E3+E4”,回车得出 19;然后,利用填充柄功能按住鼠标左下拖至 E7,松开鼠标即得各组的由低到高累计频数。由低到高累计频率的操作方法与此相同。

(4)计算由高到低累计频数和频率。单击 G3 单元格,输入“=C8”,回车得出 50,再单击 G4 单元格,输入“=G3-C3”,回车得出 44;然后,利用填充柄功能按住鼠标左键向下拖至 G7,松开鼠标即得各组的由高到低累计频数。由高到低累计频率的操作方法与此相同。

二、使用直方图工具进行分组、计算频率和绘制统计图

使用“直方图”工具,可以将调查所得的数据分组,计算调查各组中出现的频数和频率,并绘制次数分布直方图和累计频率折线图。现仍以 50 名工人的产量资料为例,说明“直方图”工具的使用方法。

首先,将调查数据输入 Excel 表的 A 列中,并将之排序(点击“常用”工具栏中的排序按钮,也可以使用“数据”菜单中的“排序”选项进行排序),排序结果如附表 3-2 所示。

附表 3—2

工人日产量频数分组表

	A	B	C	D	E
1	日产量	分组上限	分组	频率(人)	累积(%)
2	104	109			
3	107	119			
4	107	129			
5	108	139			
6	108	150			
7	109				
8	110				
9	110				

其次,为了将调查单位按组归类,需要输入分组的边界值。在“直方图”工具中,要按组距的上限输入,但不接受非字符(“……以上”或“不足……”之类)。本例将“109、119、129、139、150”五组输入 B 列 2~6 行。

输入完原始数据和分组界限后,使用“直方图”工具进行如下操作:

第一步,在“工具”菜单中单击“数据分析”项,从弹出的“数据分析”对话框中选择“直方图”并确定打开。

第二步,在“直方图”对话框中的“输入区域”框中输入原始数据的区域“A2:A51”;在“接受区域”框中输入接受调查单位按组归类的区域“B2:B6”,若这里不输入分组界限,则系统将在最小值和最大值之间建立一个平滑分布的分组。

由于第一行是标志行,因此,需要选定“标志”复选框。

第三步,在“输出区域”框中键入输出表左上角的单元格行列号,本例为“C2”。如果需要同时给出次数分布直方图,可单击“图表输出”复选框;如果需要输出累计频率,则单击选定“累积百分率”复选框,系统将在直方图上添加累计频率折线。

选定以上各项后,回车确认,即在 B 列右侧给出一个 3 列的分组表和一个直方图,表和图中的“频率”实际上是频数,“累积(%)”实际上是累计频率,图形实际上是条形图,因为在 Excel 中不能画出各直方图相连的直方图。

三、用“图表向导”工具绘制统计图

利用附表 3—3 的统计资料进行分析:

附表 3—3

各类型工业企业历年引进外资总额比较表

单位:亿元

年 份	总 计	国有企业	集体企业	其他企业
1981	4 237	3 289	948	0
1991	9 716	6 302	3 117	297
2001	23 924	13 064	8 552	2 338
2011	98 520	30 400	42 215	25 905

首先,选定要绘制统计图的资料,本题用鼠标选定 C、D、E 三列数据,作为绘图的依据。然后,单击“常用”工具栏中的“图表向导”按钮,或在“插入”菜单中单击“图表”选项,屏幕弹出“图表向导——4步骤之 1——图表类型”对话框,需要分四步进行操作:

第一步,在“4 步骤之 1”对话框中,有“标准类型”和“自定义类型”两个选项卡,系统常设为“标准类型”选项卡。在对话框的左侧有 14 种图表类型的名称(未显示部分可以用滚动条查看),单击其中某一类型,在对话

框的右侧就显示其若干“子图表类型”(不同的格式)。本例选“柱状图”的“簇状柱状图”(右侧类型中的第一个,选定后反白显示),选定后即在右侧的下方出现这种图形的特点(本例显示“簇状柱状图,比较相交于类别轴上的数值大小”),并提示用户按住下面条形按钮可以查看图形示例。

如果对“标准类型”不满意,可以打开“自定义类型”选项卡,在卡的左侧列有 20 种图表类型,右侧是各种类型的图形示例,并说明其特点,以使用户选择。

选定图表类型后,单击屏幕底行的“下一步”按钮,即进入“图表向导——4 步骤之 2——图表源数据”对话框。

第二步,在“4 步骤之 2”对话框中,有“数据区域”和“系列”两个选项卡:

(1)“数据区域”选项卡中,上部是根据选定数据绘制的图形,底部有“数据区域”和“系列产生在”两项设置。首先,在“数据区域”框中列出选定的数据区,经检查发现有错误可以重新输入正确的数据区(本例为“=Sheet1!\$C\$1:\$E\$7”,经检查无误)。然后,在“系列产生在”项有“行”和“列”两个选项(本例中“系列”指各个柱形)。其中,选择“行”(使其前面的○变成⊙),则所绘制图形的横轴(横坐标)上按列标题分组(本题为不同年份的数值);选择“列”,则横轴上按行标题分组(本例为不同年份),各柱形表示统计表中各列的数值(本题为不同类型企业的引进外资数值)。在两种组合中,为突出不同类型企业引外资数值变动的幅度,本例选择了“列”,选定后即在屏幕上部的图形中显示出来。

(2)关于“系列”选项卡中的操作,由于本例在选定数据区域时连同列标题一起选定,所以有“系列”选项卡左侧的“系列”项下列出了国有企业、集体企业、其他企业三个列标题,右侧两个框中分别显示出了各列的标题位置和数值区域。如果未选定列标题,将在“系列”项下列出“系列 1”、“系列 2”、“系列 3”三项,可首先单击“系列 1”,在右侧的“名称”框中输入“国有企业”,在“值”框中输入“=Sheet1!\$C\$2:\$C\$5”,“系列 1”即改为“国有企业”,并在上部图例中显示出来。其他两组也照此操作。此外,如需添加系列,可按下面的“添加”按钮,输入新的系列名称和数值区域。如需删除某一系列,可先单击该系列,再按下“删除”按钮。最下面还有“分类(X)轴标志”,是对横轴的分类标示,本例在上框中输入“1981,1991,2001,2011”四个年份,以替换原来横轴上的 1、2、3、4。

完成以上操作后,单击屏幕底部的“下一步”按钮,即进入“图表向导——4 步骤之 3——图表选项”对话框。

第三步,在“4 步骤之 3”对话框中,并列有“标题”、“坐标轴”、“网格线”、“图例”、“数据标志”、“数据表”6 个选项,各选项操作如下:

(1)打开“标题”选项卡,分别输入图表标题、分类(X)轴标题、数值(Y)轴标题,输入后即在右侧的图形中显示出来。

(2)打开“坐标轴”选项卡,再次确定是否显示“分类(X)轴”和“数值(Y)轴”,并确定“分类(X)轴”采用“自动”、“分类”、“时间刻度”中的哪个方式。

(3)打开“网格线”选项卡,确定采用哪种网格线的显示方式。本例为使图面清晰简洁,只选择了 Y 轴的“主要网格线”,选定后即在右侧图形上显示出来。

(4)打开“图例”选项卡,确定是否显示图例和图例的放置位置。

(5)打开“数据标志”选项卡,确定是否在各柱形上标注 Y 轴数据或分类,是否在标注旁加图例标识。

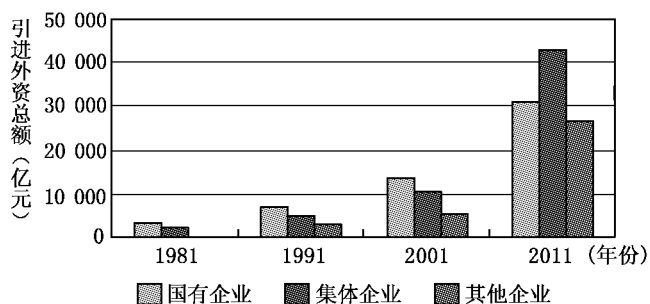
(6)打开“数据表”选项卡,确定是否“显示数据表”和“显示图例项标示”。本例在前面已经列有数据表,故在此选择不显示。

以上操作完成后,按“下一步”按钮,进入“图表向导——4 步骤之 4——图表位置”对话框。

第四步,在“4 步骤之 4”对话框中,确定图表以什么方式显示。“作为新工作表插入”是在工作簿中新建单独的图表,“作为其中的对象插入”是在已有的工作表中插入图表。后面方框中输入要放置图表的工作表名。本例选择“作为其中的对象插入”。最后按“完成”按钮,将绘成的图表嵌在工作表中,结束操作。本例的图如附图 3-1 所示。

如果对图的外观不满意,可以进行调整。

若是对图表的尺寸和位置不满意,可以单击图表中的某一项(总标题、横轴标题、纵轴标题、图例、图形),



附图 3-1 本例统计图

或图表中的空白处(选中整个图表),则选定部分出现一个虚线框,点击框内并按住鼠标左键拖动,可以将之放在满意的位置;移动鼠标至框线上 8 个黑方块(缩位框)中的任一个,当鼠标变成双向箭头时,按住鼠标左键拖动,可以改变其大小和尺寸。

若是对图表的格式不满意,可以双击图表中的某一项(总标题、横轴标题、纵轴标题、图例、坐标轴),或图表中的空白处(选中整个图表),就会弹出一个格式对话框,对话框中包括“图案”、“字体”、“属性”三项,可以在该对话框中对图表进行满意的改动。