

“我进行的这项工作所使用的方法，…，即用数字、重量和尺度的词汇来表达我自己想说明的问题。”

“劳动是财富之父，土地是财富之母”

——W. 配第

第一章 总 论

本章是全书的基础，对统计学的学科性质、统计数据类型及其研究方法和统计学中的有关基本概念进行了介绍。具体要求：①理解统计的含义与本质；②对统计学产生与发展的简要历史，特别是对主要学派有所了解；③比较全面地认识统计学的学科性质和作用；④熟知统计数据的各种类型、特征以及计量尺度，掌握统计数据的研究过程和基本方法；⑤对总体、个体、样本、标志、变量、指标和指标体系等统计学的基本概念有比较系统、全面的掌握。

第一节 什么是统计学

一、统计的含义与本质

要了解什么是统计学，首先要了解什么是统计。

在当今信息时代，我们经常可以从各种媒体听到或看到各种各样的数据，例如某国家或地区某年 GDP 多少亿元、增长速度百分之几，某企业某年销售收入多少万元、创利多少万元，某地区某年人口出生率千分之几、人均 GDP 多少元等等。这些数据意味着什么？为什么要得到这些数据？如何得到这些数据？这就是统计的问题。

那么什么是统计？根据理解的角度不同，“统计”一词可以有三种含义：统计活动、统计数据和统计学。统计活动是对各种统计数据进行搜索、整理并做出相应的推断、分析的活动，通常被划分为统计调查、统计整理和统计分析三个阶段；统计数据是通过统计活动获得的、用以表现研究现象特征的各种形式的数；统计学则是指导统计活动的理论和方法，是关于如何搜集、整理和分析统计数据的科学。显然，统计的三种含义以统计数据为核心而紧密联系：统计数据与统计活动是统计成果与过程的关系，统计活动与统计学则是统计实践与理论的关系。

由于任何统计活动都有一定的针对性，所以不难发现，统计就是要围绕研究目的和任务，运用科学的统计方法，去获取真实客观的有关统计数据，做出必要的统计分析，以了解和认识事物的真相。所以，统计的本质就是关于为何统计，统计什么和如何统计的思想。

二、统计学的产生和发展

统计学随统计的产生而产生，而统计则起源很早，随社会生产的发展和国家管理的需要而逐步产生和发展起来。在原始社会，人类最初的一般计数活动蕴藏着统计萌芽。在奴隶社会，统治阶级为了对内统治和对外战争，需要征兵征税，开始了人口、土地和财产的统计。例如，中国公元前 1000 多年的夏朝分为九州，人口 1355 万人，土地 2438 万顷。差不多同一时期的古希腊、罗马等奴隶制国家，也有人口、财产和世袭领地的统计。当然，由于生产力水平所限，奴隶社会的统计只属于初级阶段。到了封建社会，统计有了一定的发展，封建君主和精明的政

治家日益意识到统计对于治国强邦的重要性，统计范围有所扩大。但由于封建经济的封闭割据和保守性，统计活动的范围受到了限制，统计方法也很不完善。到了资本主义社会，随着社会生产力的迅速发展和社会分工的愈益精细，统计得到了很大的发展，除了政府管理的需要外，逐步扩展到工业、农业、贸易、银行、保险、交通、邮电和海关等经济领域，以及社会、科技和环境等领域，并且出现了专业的统计机构和研究组织，统计方法得到了迅速完善和发展，大大提高了统计的认识能力，而电子计算技术的应用为统计活动的现代化进程提供了重要手段。正是在这样的历史背景下，统计学应运而生。从 17 世纪中末期开始，经过 300 余年的发展，形成了今天的统计学。

从统计学的发展过程看，它可以分为三个阶段：古典统计学时期、近代统计学时期和现代统计学时期，贯穿整个过程的主线是统计方法的逐步充实、完善和发展。

（一）古典统计学时期

从 17 世纪末到 18 世纪末，是统计学的萌芽时期，即古典统计学时期。当时有两大流派：国势学派和政治算术学派。国势学派认为统计学是关于国家显著事项的学问，主要通过国家组织、人口、军队、领土、居民职业和资源财产等事项的记述对国情国力进行研究，代表人物是德国的康令（H.Coning, 1606-1681）和阿亨瓦尔（G.Achenwall, 1719-1772）。由于这个学派在进行国势比较分析中偏重事物性质的解释而不注重数量分析，因此尽管它首先提出“统计学”之名而无统计学之实。

政治算术学派主张以数字、重量和尺度来研究社会经济现象及其相互关系，代表人物是英国的威廉·配第（W.Petty, 1623-1687）和约翰·格朗特（J.Graunt, 1620-1674）。威廉·配第的代表作是《政治算术》（1676 年），提出“不用比较级、最高级进行思辨或议论，而是用数字……来表达自己的问题……借以考察在自然中有可见的根据的原因。”该书用数量分析的方法对比了英国、法国和荷兰三国的“财富和力量”，以批驳当时英国国内的悲观论调。他还提出了用图表概括数字资料的理论和方法。马克思称威廉·配第为“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。”约翰·格朗特则是利用大量数据研究社会人口变动规律的创始人，其著作《关于死亡表的自然和政治观察》一书，首次通过大量观察，对新生儿性别比例和不同原因死亡人数比例等人口规律进行了分析，并且第一次编制了初具规模的“生命表”。由于政治算术学派用大量观察法、分类分析法和对比分析法等综合研究社会经济问题，因此虽无“统计学”之名而实为统计学之正统起源。

（二）近代统计学时期

从 18 世纪末到 19 世纪末，是近代统计学时期。这一时期的一个重大成就是大数法则和概率论被引进了统计学之中。之后，最小平方法、误差理论和正态分布理论等相继成了统计学的重要内容。这一时期也曾有两大流派：数理统计学派和社会统计学派。数理统计学派始于 19 世纪中叶，代表人物是比利时的凯特莱（A.Quetelet, 1796-1874），著有《概率论书简》、《社会物理学》等，他主张用研究自然科学的方法来研究社会现象，正式把概率论引入统计学，并最先用大数定律论证了社会生活中随机现象的规律性，还提出了误差理论和“平均人”思想。凯特莱的贡献，使统计学的发展进入了一个新的阶段。

社会统计学派始于 19 世纪末，首创人物为德国的克尼斯（K.G.A.Knies, 1821-1898），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性的

实质性科学，其显著特点是强调对总体进行大量观察和分析，通过研究其内在联系来揭示社会现象的规律。各国专家学者在社会经济统计指标的设定与计算、指数的编制、统计调查的组织和实施、经济社会发展评价与预测等方面取得了一系列重要成果。德国统计学家恩格尔（C.L.E.Engel,1821-1896）提出的“恩格尔系数”，美国经济学家库兹涅茨和英国经济学家斯通等人研究的国民收入和国内生产总值的核算方法等，都是伟大的贡献。

（三）现代统计学时期

从 19 世纪末到现在，是现代统计学时期。这一时期的显著特点是数理统计学由于同自然科学、工程技术科学紧密结合及被广泛应用于各个领域而获得迅速发展，各种新的统计理论与方法、尤其是推断统计理论与方法得以大量涌现，例如英国统计学家卡尔·皮尔逊（K.Pearson,1857-1936）的卡方 χ^2 分布理论，统计学家戈赛特（W.S.Gosset,1876-1937）的小样本 t 分布理论，统计学家费希尔（R.A.Fisher,1890-1962）的 F 分布理论和实验设计方法，波兰统计学家尼曼（J.Neyman,1894-1981）和英国统计学家皮尔逊（E.S.Pearson, 1895-1980）的置信区间估计理论和假设检验理论，以及非参数统计法、序贯抽样法、多元统计分析法、时间序列跟踪预测法等都应运而生，并逐步成为现代统计学的主要内容。现代统计学时期是统计学发展最辉煌的时期。

三、统计学的学科性质

那么统计学到底是一门什么样的学科呢？我们认为，经过 300 余年的演变与发展，统计学成为了一门以各种现象的数量方面作为研究对象、并为这种研究提供方法论的学科。

（一）统计学就其研究对象而言，具有数量性、总体性和差异性的特点

统计学的研究对象是各种现象的数量方面，即各种能体现数量大小、数量关系、数量变动、数量界限和数量规律等特征的统计数据，所以统计学的首要特点就是数量性，就是要通过定量研究、运用各种统计数据来体现所研究现象的数量特征，进而达到认识现象的本质和规律的目的。在社会经济领域，这种数量性通过特有的统计指标及指标体系来加以体现。

那么，统计研究怎样体现现象的数量特征？这就需要强调总体性，就是要从所研究现象的总体出发，通过对现象总体中的构成元素即个体进行大量观察和综合分析，来达到认识现象的总体数量特征的目的，也就是说统计研究的目的不是认识个别事物的数量特征，而是具有综合意义的总体数量特征。例如，我们所说的居民收入水平并不是指某家某户的，而是指一个国家、地区或城市的。当然，对现象总体的认识是以对个体的观察为基础的，只是不能以个体的表现（往往具有特殊性和偶然性）来说明现象的一般性或规律性，因为个体之间存在着差异。

但正是由于个体差异的存在，才引起了人们去了解个体差异背后所可能隐藏的规律的兴趣，所以统计研究的第三个特点就是差异性，就是要从所研究现象总体的各个个体之间的差异中概括出共同的普遍的数量特征，并对差异情况做出必要的反映。例如，一个国家、地区或城市的居民之间的收入差异是客观存在的，高收入者与低收入者之间的差距有时会很大，所以我们只能用某种平均收入指标来反映该国家、地区或城市居民的收入水平，并用基尼系数等指标来反映居民之间的收入差异程度。现象总体的内在差异性 is 统计研究总体数量特征的前提，而这种差异并不是事先已知或由某种固定的因素来确定的，它是由各种非确定性因素，即偶然因素共同作用的结果。

上述三方面的特点相互联系，共同决定了统计研究内容的广泛性。当然，社会经济现象的数量方面是统计学最主要的研究对象，同时也是最复杂、最具有挑战性的部分。

（二）统计学就其学科范畴而言，具有方法性、层次性和通用性的特点

既然统计学的研究对象是现象数量方面，那么很自然就出现了这样一个问题：如何来研究现象的数量？或者说，如何来搜集、整理和分析统计数据？因此，作为一门用以指导统计数据的搜集、整理和分析的科学，首先要为研究现象的数量提供科学的理论、原则和方法，这些理论、原则和方法的系统化，就形成了统计学。所以，方法性是统计学最为显著的学科特点，统计学的每一步发展无一不是统计方法完善、创新和突破的结果。所以，统计学提供给人类的是一种开展定量研究思想和工具。

统计学是一门一级学科，拥有完整、严密的学科体系，具有很强的层次性，其二级学科包括理论统计学、应用统计学、统计学史和统计学其他学科等。其中，理论统计学是关于统计学的一般理论、原则和方法，提供以抽象的数量为研究对象的方法论。应用统计学则是统计方法在社会、经济、自然和工程等各个领域应用的结果，是理论统计学与各应用领域实质性科学理论相结合的产物，提供以具体的数量为研究对象的方法论，例如社会经济统计学、医学卫生统计学和天文气象统计学等，具有学科交叉的性质。理论统计学与应用统计学相互促进，共同发展。

作为方法论科学，统计学的一般理论、原则和方法，在各种需要开展定量研究的领域都具有通用性，除了理论统计学具有普遍的通用性外，应用统计学还在各自的领域内具有特殊的通用性。正因为如此，统计学定量研究的功能才得以充分的发挥，应用领域才得以迅速的拓展。可以说，统计方法的应用如今已经无所不在了。

（三）统计学就其研究方式而言，具有描述性和推断性的特点

统计学研究现象的数量方面，既有描述的方式，也有推断的方式，两者各有侧重并各具特色。所谓描述性，就是指运用各种方法对研究现象进行观察调查，获取数据，然后进行汇总、分类和计算，并用表格、图形和综合指标的方式来加以显示。把各种真实描述所研究现象数量特征的理论和方法加以系统化，就形成了描述统计学。所谓推断性，就是指在随机原理和概率论基础上，根据样本观测结果对总体数量特征做出估计或进行假设检验，以期对不确定事物（随机现象）做出定量的推断。把各种定量推断不确定事物的理论和方法加以系统化，就形成了推断统计学。以事物已发生的事实为依据对其未来发展趋势进行定量预测，也可属于推断统计学的范畴。很显然，描述统计学是基础，推断统计学是其深入和发展，但描述统计与推断统计并不是相互割裂的，而是描述中有推断，推断中有描述，或者说任何描述都是推断，任何推断也都是描述。因此，描述统计与推断统计相辅相成，相互交叉联系，在统计学中占有同等重要的地位。

四、统计学的作用

由统计学的简要发展史可以看出，统计学是一门应用性很强的学科。实践的需要产生和发展了统计学，统计学反过来又被不断地应用于实践。统计学在认识事物、指导生产、经济管理和科学研究等各个方面都发挥着重要作用。

（一）统计学为我们认识自然和社会提供了必需的方法和途径

人们要认识自然、认识社会，离不开各种各样的数据资料并对这些资料做出各种各样的分析研究，这就需用运用一整套的统计理论和方法作为指导。例如，

一年四季如何变化？人口年龄分布有什么特征？经济运行有什么规律？生活习惯（例如吸烟、酗酒）与某种或某些疾病有什么关系？等等，要想知道其中的答案，统计数据以及以此为基础的分析研究是必不可少的。如果说统计是认识自然和认识社会的手段，那么统计学则为运用这种手段提供了理论和方法。

（二）统计学在指导生产活动过程中发挥着重要作用

众所周知，生产的任务是以尽量少的投入生产出数量尽量多且质量尽可能高的产品。然而影响产品产量和质量的因素是很多的，但有主次之分。这就需要对各种因素进行试验和观察，了解各种因素的影响方式和程度，找出各因素的最佳水平和最佳组合，从而确定最佳的生产条件和生产方式并使之始终处于科学的控制之中。这就需要利用统计理论和方法，科学合理地设计和安排试验并做出分析，同时对生产过程不断地进行监测。最佳生产方案设计和最优质量控制，是统计学的一大应用领域，生产控制图、 6σ 理论与方法等的广泛应用充分说明了这一点。

（三）统计学在社会经济管理活动中的作用更为显著

无论从客观上看还是从微观上看，统计活动都具备着搜集信息、提供咨询、实施监督和支持决策的重要职能。宏观上，政府的管理目标是要保持国民经济持续稳定协调发展，实现劳动力的充分就业和物价水平的稳定，做到自然资源的合理开发和生态环境的良好保护，确保社会的安定和人民生活水平的稳定提高。这就需要政府利用统计学所提供的方法去科学合理地搜集数据资料，对国民经济和社会发展状况进行跟踪监测和预警，对各种社会经济问题进行定量模拟和分析，从而为制定和调整政策提供依据。微观上，企业的管理目标是使生产要素达到最佳配置，取得最佳的经济效益，在激烈的市场竞争中保持优势。这就需要企业及时搜集各种市场信息，科学地反映和分析企业的生产经营状况，准确地预测和判断市场变化的趋势，这一切也必须运用各种统计理论和方法。

（四）统计学为科学研究提供了有力手段

在科学研究中，研究的任务是揭示客观事物的规律性，研究的方法一般是先根据若干观察或实验资料提出某种假设或猜想，然后再通过各种途径进行观察或实验加以验证。显然，统计理论与方法在其中起着重要作用。一方面，它有助于集中并提取观察和实验数据中本质性的东西，从而有助于提出较正确的假说或猜想；另一方面，它又能指导研究人员如何去安排进一步的观察或试验，以判定所提出的假说或猜想是否正确。在医学界，人们利用统计方法来研究疾病的原因或影响因素，判断药物或治疗方法的有效性；在考古学界，人们凭借统计方法来推断特定发掘物的历史年代；在心理学界，人们用统计方法分析特定刺激的心理效应；在经济学界，人们用统计方法研究国民经济运行状况和各种决策方案的优劣；在生物学界，人们用统计方法来研究基因定律（如基因分离定律、基因自由组合定律和基因频率稳定性定律等都经过了严格的统计检验）；如此等等，几乎所有的科学研究领域都离不开统计学。历史上许多有关领域的著名专家，往往也是著名的统计学家。

统计学的最主要奠基人费歇（R. A. Fisher）曾经说过：“给 20 世纪带来了人类进步的独特方面的是统计学。”我们也相信，统计学将为人类社会的进步做出更大的贡献。

第二节 统计数据类型与研究方法

一、统计数据类型

既然统计学是关于统计数据的科学，那么统计数据有那些类型呢？大致上可以从以下四个角度进行分类。

（一）统计数据按照所采用的计量尺度不同，可以分为定性数据与定量数据两类

定性数据是指只能用文字或数字代码来表现事物的品质特征或属性特征的数据，具体又分为定类数据与定序数据两种。定类数据是对事物进行分类的结果，表现为类别，由定类尺度计量而成。例如，人口按照性别分为男与女两种类别，人的消费按照支出去向分为衣、食、住、行、烧、用、医、文、娱、健等类别，都属于定类数据。为了便于统计处理（计算机录入等计数处理），常用数字代码来表示各个类别，例如分别用 1、0 表示男性与女性，分别用 1、2、3、4、5、6、7、8、9、10 等表示衣、食、住、行、烧、用、医、文、娱、健等。要注意的是，这时的数字没有任何程度上的差别或大小多少之分，只是符号而已。定序数据是对事物按照一定的排序进行分类的结果，表现为有顺序的类别，由定序尺度计量而成。例如，学生的考试成绩表示为优、良、中、及格、不及格，课题成果的鉴定等级表示为 A、B、C，消费者对某产品的满意程度表示为很满意、满意、一般、不满意、很不满意，等等，都属于定序数据。同样，定序数据也可以用数字代码来表示，例如学生的考试成绩可以分别用 5、4、3、2、1 来表示优、良、中、及格、不及格。这时，数字代码能体现一种顺序或程度的不同，但还不能体现事物之间或不同结果之间（例如及格与不及格之间，很满意与满意之间）的具体数量差别。定序数据所包含的信息量大于定类数据。

定量数据是指用数值来表现事物数量特征的数据，具体又分为定距数据与定比数据两种。定距数据是一种不仅能反映事物所属的类别和顺序，还能反映事物类别或顺序之间数量差距的数据，由定距尺度计量而成。例如，两位学生的考试成绩分别为 85 分和 55 分，不仅说明前者良好、后者不及格，前者高于后者，而且还说明前者比后者高 30 分。再如，某日甲、乙、丙三地的最高气温分别为 30℃、20℃和 10℃，说明该日甲与乙之间最高温的温差等于乙与丙之间的温差，都是 10℃。但要注意的是，定距数据一般只适合于进行加减计算而不适合于乘除运算，例如气温 30℃与 10℃相比，并不能说前者的暖和程度是后者的 3 倍，因为气温可以是 0℃或 0℃以下，而 0℃或 0℃以下并不代表没有温度。这种情况称为不存在绝对零点的现象，类似的还有企业利润等。定比数据是一种不仅能体现事物之间数量差距，还能通过对比运算、即计算两个测度值之间的比值来体现相对程度的数据，由定比尺度计量而成。只要是反映存在绝对零点的现象（即 0 就代表没有）的数据，都是可以进行对比运算的定比数据，例如企业销售收入 3 亿元，人的身高 176 厘米、体重 65 公斤，物体的长度 30 厘米、面积 600 平方厘米、容积 9000 立方厘米，水稻的平均亩产 400 公斤/亩，某地区的人均国内生产总值 25000 元/人、第三产业比重 48%等，都是定比数据。定比数据是包含信息量最多的数据，绝大多数统计数据都属于这一类。

定性数据在一定条件下可以转化为定量数据。例如对于定类数据，通过计数的方法可以计算出各类别的频数及在总体中的比重；对于定序数据，在一定的假设下，可对定序的数字代码进行统计计算（例如计算平均数、标准差等），计算结果在假设范围内有意义。在统计处理与统计分析中，如何使定性数据尽量客观反映实际并提供尽可能多的信息，是一个非常重要的统计问题。

（二）统计数据按照其表现形式不同，可以分为绝对数、相对数和平均数三类

绝对数是用以反映现象或事物绝对数量特征的数据，它以最直观、最基本的形式体现现象或事物的外在数量特征，有明确的计量单位。例如人的身高 176 厘米、体重 65 公斤，地区的人口数 500 万人、属地面积 11000 平方公里、国内生产总值 1250 亿元，企业销售收入 15 亿元、利润 2.1 亿元等，都是有明确计量单位的绝对数。绝对数是表现直接数量标志或总量指标的形式。

相对数是用以反映现象或事物相对数量特征的数据，它通过另外两个相关统计数据的对比来体现现象（事物）内部或现象（事物）之间的联系关系，其结果主要表现为没有明确计量单位的无名数，少部分表现为有明确计量单位的有名数（限于强度相对数）。具体地，相对数又包含结构相对数（例如某地区三大产业比重分别为 12%、48% 和 40%），比例相对数（例如新生儿男女性别比 107:100），比较相对数（例如 A 地区的人均 GDP 是 B 地区的 1.2 倍），动态相对数（例如某地区 GDP 的发展速度为 109%），强度相对数（例如某地区的人口密度为 300 人/平方公里，人口出生率为 11%）和计划完成程度相对数（例如企业产量计划完成程度为 120%）等六种。相对数是表现相对指标的形式。

平均数是用以反映现象或事物平均数量特征的数据，体现现象某一方面的一般数量水平。例如某班级同学的平均年龄 19 岁，某年某企业职工的平均月收入 1500 元，某年某地区居民的平均月消费支出 800 元等，都是平均数。具体地，平均数可以按计算方式不同分为算术平均数、调和平均数、几何平均数等数值平均数与众数、中位数等位置平均数，按时间状态不同分为静态平均数与动态平均数。平均数是表现平均指标的形式。

通过各种尺度计量而成的统计数据，最终都可以归结为绝对数，相对数和平均数这三大表现形式。

（三）统计数据按照其来源不同，可以分为观测数据与实验数据两类

观测数据是通过统计调查或观测的方式而获取的反映研究现象客观存在的数量特征的数据，这类数据是在没有人为控制的条件下获得的。有关社会经济现象的统计数据几乎都是观测数据，例如前面提到的各种统计数据。

实验数据是在人为控制的条件下，通过实验的方式而获得的关于实验对象的数据。自然科学研究中的数据大都属于实验数据，例如生物实验数据、产品性能实验数据、药物疗效实验数据等，都属于这类数据。随着实验方法在经济等领域的应用，逐步形成了实验经济等学科，在经济等领域出现了许多实验数据。

（四）统计数据按照其加工程度不同，可以分为原始数据与次级数据两类

原始数据是指直接向调查对象搜集的、尚待加工整理、只反映个体特征的数据，或通过实验采集的原始记录数据。原始数据是统计数据搜集的主体。

次级数据也称为加工数据或二手数据，是指已经经过加工整理、能反映总体数量特征的各种非原始数据。次级数据又包括直接根据原始数据整理而来的汇总数据和根据各种已有数据进行推算而来的推算数据。如果次级数据已能满足有关分析和研究需要，我们就不应再去搜集原始数据，以免造成浪费。次级数据的来源包括各种统计年鉴、有关期刊和有关网站等。

（五）统计数据按照其时间或空间状态不同，可以分为时序数据与截面数据两类

时序数据是时间序列数据的简称，是对同一现象在不同时间上搜集到的数据（即空间状态相同，时间状态不同），描述的是现象某一方面（或某几方面）的

数量特征随时间而变化的情况,例如把我国 1979 年以来的 GDP 数据按时间先后顺序加以排列,就形成了我国 GDP 的时序数据。

截面数据是对一些同类现象在相同或近似相同的时间上搜集到的数据(即空间状态不同,时间状态相同),描述的是在相同时间状态下同类现象的数量特征在不同空间状态的差异情况,例如我国某年各省、市、区的 GDP 数据,就是截面数据。

有时,时序数据与截面数据可以结合起来,成为平行数据(即时间、空间状态都不同),例如列出历年各省、市、区的 GDP 数据,就成为了平行数据。

二、统计数据研究过程

统计数据研究过程,也即统计研究过程,包括以下四个基本环节:统计设计,数据搜集,数据整理和数据分析与解释。

(一) 统计设计

统计设计就是制定统计数据研究方案的环节,是关于以后各环节的总体安排。统计设计要在有关学科理论指导下,根据研究问题的性质、目的和任务,科学地确定统计研究的总体对象,明确所要搜集数据的种类,确定相应的统计指标及其体系并给出统一的定义和标准,确定统计数据搜集、整理、推断和分析的基本方法,规定研究工作的进度安排和质量要求,拟定研究工作的资源配置和组织实施方式等。统计设计对于统计数据研究质量至关重要,要求设计者不仅掌握系统的统计学理论和方法,而且具有所研究领域的有关知识和理论素养。

(二) 数据搜集

数据搜集就是按照统计设计的要求,有针对地获取所需的统计数据的环节,也称为统计调查环节。也就是说,要通过统计观测或实验的方式、方法去搜集各种各类计算统计指标所需的原始数据,以及其他已经存在的各种相关数据。数据搜集是否准确、及时、完整,直接影响到统计分析的质量。

(三) 数据整理

数据整理就是对通过统计观测或实验所获得的原始数据,进行必要的系统化处理,使之条理化、综合化,成为能反映总体特征的统计数据的环节,也称为统计整理环节。数据整理也包括对已有数据的再加工和深加工。数据整理的手段有统计分组、汇总和计算等,整理结果表现为统计图、统计表或统计指标。

(四) 数据分析与解释

数据分析是在数据整理的基础上,围绕统计设计所确定的研究任务,运用各种统计方法对数据进行各种统计分析,得出某些有用的定量结论的环节,也称为统计分析环节。数据分析实质上就是对数据的深加工整理,是整个统计研究的核心,也是统计研究的最终目的。在这个环节,既要用到描述统计方法,又要用到推断统计方法。数据解释则是对整理和分析的数据或有关数量结果进行说明,即说明为什么会得出这些数据,这些数据的含义分别是什么,从中能得出哪些具有规律性的结论,需要进一步探讨哪些问题等等。数据解释是对数据分析的深化。

三、统计数据研究方法

统计数据研究的基本方法有大量观察法、统计分组法、综合指标法、统计推断法和统计模型法。

(一) 大量观察法

这是统计数据搜集环节(即统计调查环节)的基本方法,即要对所研究现象总体中的足够多数的个体进行观察和研究,以期得到具有规律性的总体数量特征。大量观察法的数理依据是大数定律,即虽然每个个体受偶然因素的影响作用

不同而在数量上存有差异，但对总体而言可以相互抵消而显现出稳定的规律性，因此只有对足够多数的个体进行观察，观察值的平均结果才会趋于稳定，建立在大量观察法基础上的统计数据才会给出具有普遍意义的结论。统计学中的各种观测调查方法都属于大量观察法。

（二）统计分组法

由于所研究现象本身的复杂性、差异性及多层次性，需要我们对所研究现象进行分组或分类研究，以期在同质的基础上探求不同组或不同类之间的差异性。统计分组法在整个统计研究过程中占有重要的地位，在统计调查环节可以通过统计分组法来搜集不同类的原始数据，并可使抽样调查的样本代表性得到提高（即分层抽样方式）；在统计整理环节可以通过统计分组法使各种数据得到分门别类的加工处理和储存，并为编制分布数列提供基础；在统计分析环节则可以通过统计分组法来划分现象类型、研究总体内在结构、比较不同类或组之内的差异（显著性检验）和分析不同变量之间的相关关系。统计分组法有传统分组法、判别分析法和聚类分析法等。

（三）综合指标法

统计研究现象的数量方面主要是通过统计综合指标来反映的，就是运用统计指标来综合反映现象总体的数量特征，常见的综合指标有总量指标、相对指标和平均指标。综合指标法在统计学、尤其是社会经济统计学中占有十分重要的地位，是描述统计学的核心内容。如何最真实客观地记录、描述和反映所研究现象的数量特征和数量关系，是统计指标理论研究的一大课题。

（四）统计推断法

在统计研究过程中，我们所观察的往往只是所研究现象总体中的一部分个体，掌握的只是具有随机性的样本观测数据，而认识总体数量特征才是统计研究的目所在，这就需要我们根据概率论和样本分布理论，由样本观测数据来推断总体数量特征——参数估计或假设检验。这种由样本来推断总体的方法就叫统计推断法。统计推断法已在统计研究的许多领域得到应用，除了最常见的总体指标的推断外，统计模型参数的估计和检验、根据时间序列所做的外推预测等，也都属于统计推断的范畴，都存在着误差和置信度的问题。

（五）统计模型法

在以统计指标来反映现象总体数量特征的同时，我们还经常需要对相关现象之间的数量变动关系进行定量研究，以了解某一（些）现象数量变动与另一（些）现象数量变动之间的关系及变动的影响程度。在研究这种数量变动关系时，需要根据具体的研究对象和一定的假定条件，用合适的数学方程来进行模拟，这种方法就叫做统计模型法。统计模型的三个基本要素是：变量、数学方程和模型参数。运用统计模型法，可以使统计分析更具广度和深度，提高统计的认识能力。统计学提供了各种线性的和非线性的、简单的和复杂的统计模型构建方法。

上述各种方法之间是相互联系、互相配合的，共同组成了统计学方法体系。

第三节 统计学的基本概念

一、总体与样本

（一）总体

1.总体的含义与特征

我们对所研究现象数量特征的认识是着眼于总体的。所谓总体，就是统计研究的客观对象的全体，是由所有具有某种共同性质的事物所组成的集合体，有时也称为母体。组成总体的每个个别事物就称为个体，也称为总体单位。总体中个体数量的多少，称为总体容量或总体单位总数。

作为总体，应该具有大量性、同质性和差异性三个特征。大量性是指总体中的个体数必须是充分多的，同质性是指总体中的每个个体都必须具有某种共同属性或特征，差异性则指个体的属性或特征在某些方面又必须存在一定的差异。统计研究总体的数量特征，大量性是条件，同质性是基础，差异性的前提。

2.总体的分类

(1) 总体按照其个体数量是否有限可以分为有限总体和无限总体两类。有限总体是指由有限多个个体所组成的总体，即总体容量是有限的，例如某市的企业总体、人口总体等。无限总体是指由无穷多个个体所组成的总体，即总体容量是无限的，例如，宇宙中的星球总体、大气中的悬浮颗粒总体等。无论是有限总体还是无限总体，都需要运用统计方法去研究其数量特征。

(2) 总体按照其存在形态不同可以分为具体总体和抽象总体两类。具体总体是指由现实存在的各个具体事物所组成的总体，例如客观存在的全国人口总体、某批产品总体、某市企业总体等。抽象总体也称为设想总体，是由想象中存在的各个假定事物所组成的总体，例如某条件下生产的产品总体、某型号步枪连续射击的射程总体、某特殊类型的消费者总体等。当然，抽象总体并不是无根据的假设，而是对具体总体作某种抽象的结果，是在某种假设条件下存在的各个事物的全体。对于抽象总体，原则上可以假设它包含无穷多个个体，可以把它看成为无限总体。一般地，抽象总体是具体总体的延伸和抽象化，而具体总体则可看作为抽象总体的组成部分。在现实中，对研究总体抽象化既有利于各种数据的处理与使用，又能在许多场合更深入地揭示出事物的本质。例如，若考察某一批产品的质量，则研究的是具体总体，结果是表明这一批产品的质量高低；而若考察某种工艺条件下的该种产品的质量，则研究的是抽象总体，结果不仅能表明产品本身的质量，而且更重要的是还可以说明这种工艺条件的性能及先进性。

(3) 总体按照其个体能否计数可以分为可计数总体和不可计数总体两类。可计数总体是指能对其所包括的个体进行计数且计数结果能加总的总体，例如人口总体，每个个人是可计数的，而计数的结果即人数是可相加的；工业企业总体、某批产品总体等也都如此。可计数总体的特征是，它所包含的个体具有相同的计量单位，可以计算总体单位总数。不可计数总体是指对其所包括的个体不可计数或计数结果不能加总的总体，例如零售商品总体，虽然每件商品都具有商品的共性，但由于各自的使用价值形态和计量单位不同，所以在商品的件数上是不能直接相加的。然而，零售商品总体的物价水平和销售数量的总变动情况却是统计研究的内容之一。不可计数总体的特征是，它所包含的个体通常不具有相同的计量单位，不能计算总体单位总数。

(4) 总体按照其个体是否人为划定可以分为自然总体和人为总体两类。自然总体是由自然确定的个体所组成的，即个体是明确的、易定的，例如人口总体中的个人、企业总体中的企业、家庭总体中的家庭等都是自然个体。人为总体是由人为确定的个体所组成的，其个体往往不明显或难以确定，例如在考察某种小麦的出粉率时，总体是全部该种小麦，但个体显然不能是每一粒小麦，那么该以一公斤小麦还是以一百公斤小麦或一吨小麦作为一个个体，并没有明确的规定。再如，要研究林区的木材储藏量，也不能以每一棵树作为个体，但应该以多大面

积的区域作为一个个体也没有明确的规定。对于个体不明显的现象,要根据研究对象的具体情况和研究目的不同恰当地加以确定。

在实践中,还经常需要对被研究的总体进行分组或分类研究,尤其是要对总体中的某特定组或特定类进行分析研究。这时,总体中的一个组或类,就被称为一个研究域或一个子总体。例如在研究消费者的购买力时,对某特定类型的消费者群(例如老年消费者群、儿童消费者群、学生消费者群、妇女消费者群等)进行特别的研究,就是对消费者总体中的一个子总体进行研究,这在市场营销学中称为市场细分研究。子总体具有和原总体同样的性质。

3.总体与个体的关系

总体与个体的关系不是一成不变的,其可变性体现为三方面:一是总体容量随着个体数的增减可变大或变小;二是随着研究目的不同,总体中的个体可发生变化,例如要研究某市的居民身体素质,则总体是该市所有人口,若要研究该市的居民家庭生活水平和消费结构,则总体是该市所有居民家庭;三是随着研究范围的变化,总体与个体的角色可以变换。例如,在研究某地区某校学生的学习状况时,总体由该校所有学生构成,即学校是总体。而若要研究该地区所有学生的构成状况,则总体由该地区的所有学校所构成,学校则成为了个体。这说明,个体与总体要根据研究目的和对象范围而定。

(二) 样本

1.样本的含义

样本是与总体相对应的概念,在统计学中占有重要地位,几乎所有的统计理论和办法都建立在样本之上,而统计实践研究现象的数量方面也总是由样本开始的。所谓样本,就是从总体中抽取的一部分个体所组成的集合,也称为子样。样本中所包含的个体数,称为样本容量或样本单位数。

从理论上讲,样本可以大到与总体容量相同(这时,总体就是一个特殊的样本),也可以小到只包含一个个体。但在实践上,样本的大小总是处于总体容量与1之间。因此,样本是一个来自于总体的有限小总体。一般地,当样本容量大于30(或50)时,称为大样本,样本容量小于30(或50)时,称为小样本。

从一个总体中随机抽取一定容量的样本,可以有很多种样本构成的可能,也即可以从同一总体中抽取很多个容量相同但个体不同(或个体抽取的顺序不同)的样本。因此,样本不具有唯一性,除非样本就是总体本身。我们把从一个总体中最多可以抽取的不同样本数,称为样本个数,具体要依总体容量、样本容量和抽样方法而定。

2.样本与总体的关系

样本和总体的关系是多方面的。首先,总体是所要研究的对象,而样本则是所要观测的对象,样本是总体的代表和缩影。前面已经指出,统计研究现象的数量是着眼于总体的,但由于很多情况下不能进行全面观测研究,所以只能从样本着手,期望通过对样本数量特征的认识来达到对总体特征认识的目的,也就是通过样本的观测来研究总体。因此,样本是总体的代表和缩影。但由于样本的内在结构与总体的内在结构不完全一致,因此样本不可能完全代表总体,这是由个体特征之间的差异性所决定的。为了保证样本的代表性,样本的抽取必须是随机的,即总体中的每个个体都有一定的、事先已知的非零概率被抽取到样本中来。

其次,样本是用来推断总体的。对样本进行观测的目的是要对总体数量特征做出估计或判断,即通常所说的以样本推断总体。这种推断,从逻辑上看属于不完全的归纳推理,因为由个体到一般的归纳仅限于样本,而对于总体则需要推断。

因此，由样本推断总体，其结果不是必然的，而是或然的，即具有一定的不确定性。在推断中，大样本和小样本所使用的方法有所不同，其理论依据也有所差别。一般地，大样本方法适用于所有总体，小样本方法适用于某些特殊总体。

第三，总体和样本的角色是可以改变的。随着考察角度的改变，一定的研究总体也可成为另一意义上总体的一个样本，这说明总体与样本的角色是可以改变的。例如，一个国家的所有人口是一个总体，如果对所有人口进行调查就属于全面观测，即人口普查。但如果要从历史上动态考察该国人口变化规律，则一定时点上的人口总体即具体总体就成了一个样本，而动态上的人口总体才是与之相对应的总体。因此从广义上看，对总体的观测也是一种样本观测，统计观测、搜集的数据都是样本数据。所以，统计学就是关于样本的科学。

二、标志和变量

（一）标志

1.标志的含义

无论是研究总体还是观测样本，都必须从认识个体特征出发。用以描述或体现个体特征的名称，在统计学上就称为标志，而标志在每个个体上的具体结果则称为标志表现。例如，人口总体中的个人，性别、年龄、职业、身高、文化程度、民族、收入等都是标志；而某个人是男性、35岁、教师、170cm、大学毕业、汉族、月收入1000元等，分别是上述各个标志的标志表现。再如，工业企业总体中的企业，其标志有行业、资产总额、职工人数、年增加产值、年利润额等，某企业属于食品行业，资产总额6000万元，职工300名，年增加值5000万元，年利润额1000万元等分别是上述标志的标志表现。对于任何一个标志，总体或样本中的个体数有多少，其标志表现就有多少。

2.标志的分类

（1）标志按其结果的表示方式不同可以分为品质标志和数量标志两种。品质标志表明个体的属性特征，其结果一般只能用文字表述而不能用数值表示，即只能表现为定性数据。例如前述的个人性别、职业、文化程度和民族等，都属于品质标志。数量标志表明个体的数量特征，其结果以数值表示，即表现为定量数据。例如前述的个人年龄、身高、收入等，都属于数量标志。不难发现，品质标志表现需要采用定类尺度或定序尺度来计量，数量标志表现需要采用定距尺度或定比尺度来计量。

（2）标志按其在每个个体上的表现结果是否相同可以分为不变标志和可变标志两种。不变标志是指在每个个体上的具体表现完全相同的标志，例如高校学生总体中，身份是不变标志，因为大家都是高校学生；再如某地区服装加工企业总体中，行业是不变标志，因为各企业都从事服装加工活动。很显然，不变标志是相关个体集成总体的基础，总体的同质性也正体现在要求至少具有一个不变标志。可变标志则是指在每个个体上的具体表现不相同的标志，例如高校学生总体中，出生地、年龄、身高、专业等都是可变标志，因为它们在不同学生身上的体现是不相同的；某地区服装加工企业总体中，职工人数、年增加值等也都是可变标志。这种个体标志表现不相同的现象，在统计学上称为变异。总体的差异性也正体现在至少要有个可变标志。

（3）标志按其表现个体特征的直接程度不同可分为直接标志和间接标志两种。直接标志也称为第一性标志，它直接表明个体的属性特征或数量特征。例如企业的所属行业、职工人数、年产量，个人的性别、年龄、民族等标志，都是直接标志。一般地，品质标志都是直接标志。间接标志也称为第二性标志，它是

通过两个或两个以上数量标志计算后（通常是对比）间接表明个体数量特征的标志。例如企业的职工平均工资、人均产量等标志，都属于间接标志，因为它们分别是企业工资总额标志与职工人数标志之比、企业产量标志与职工人数标志之比。很显然，间接标志是以直接标志为基础的。一般地，间接标志都是数量标志。

（二）变量

1.变量的含义

变量是统计学中经常涉及到的一个概念。从狭义上看，变量是指可变的数量标志，例如人的年龄、身高，企业的职工人数、产量等都是变量，因为这些标志在不同个体上的值是不同的，是可变的。因此，变量是可变数量标志的抽象化。变量的具体数值，就称为变量值，也称标志值。

从广义上看，变量不仅指可变的数量标志，也包括可变的品质标志，因为可变的品质标志在各个个体上的表现结果也是不同的，只不过在作为变量处理时所用的方法有所不同（如前述的定类尺度和定序尺度）。因此，可变标志就是变量。

2.变量的分类

（1）变量按其反映数据的计量尺度不同，可以分为定性变量和定量变量两种。反映定性数据的变量就是定性变量，又包括反映定类数据的定类变量和反映定序数据的定序变量。反映定量数据的变量就是定量变量，又包括反映定距数据的定距变量和反映定比数据的定比变量。

（2）变量按其所受影响因素不同，可以分为确定性变量和随机性变量两种。确定性变量是指受确定性因素影响的变量，也即影响变量值变化的因素是明确的、可解释的或可人为控制的，因而变量的变化方向和变动程度是可确定的。例如，企业职工工资总额不外乎受职工人数和平均工资两个因素影响，这两个因素都是可人为控制的，对工资总额影响的大小和方向是确定的。随机性变量是指受随机因素影响的变量，也即影响变量值变化的因素是不确定的、偶然的，变量受随机因素影响的大小和方向是不确定的。例如农作物产量的高低，受土壤、水分、气温、光照、施肥、管理等多种因素影响，而水分、光照、气温等的变化是非确定的或非人所能控制的，因而农作物产量是随机性变量，不是确定性变量。但是，随机性变量也蕴藏着一定的规律性，通过大量观测可以揭示这种规律性。例如通过大量观测发现，随着施肥量的适当增加和管理水平的提高，农产量会呈上升趋势。正因为如此，通过大量观测或试验来发现随机变量的变动规律，成了统计学方法研究的主要任务之一。通常，自然现象的变量大多属于随机性变量，而社会经济现象的变量既有确定性变量，也有随机性变量。其中许多社会经济现象变量既受确定性因素影响，也受随机因素影响，因而对其加以观测研究的难度也更大。

（3）变量按其数值的变化是否连续，可以分为离散型变量和连续型变量两种。离散型变量是指只能取整数值的变量，即变量的变化是不连续的、间断的。例如人数、企业数、机器台数、货币面值等，都是离散型变量，它们都只能取整数。离散型变量只能采用计数的方法来取得其数值。连续型变量是指可以在一定区间内取任意实数值的变量，即变量的变化是连续的、不间断的。例如人的身高、体重，企业的总产值、利润率等，都是连续型变量，它们都能取任意实数。连续型变量要采用测量或计量的方法来取得其数值。很显然，定类尺度和定序尺度只能用来计量离散型变量，而定距尺度和定比尺度既可用于计量离散型变量又可用于计量连续型变量。

由于个体与标志紧密相依，个体是标志的承担者，因而一旦个体和所要研究的标志确定，那么也可以把该标志（即变量）的所有可能取值所组成的集合称为

总体，对所观察到的部分标志值（变量值）所组成的集合称为样本。可见，由个体所组成的总体（样本）可以转换为以变量值所组成的总体（样本），但前者是根本性的和决定性的。

三、统计指标与指标体系

（一）统计指标

1.统计指标的含义与构成要素

统计指标简称指标，是反映现象总体数量特征的概念及其数值。例如，浙江省 2005 年的国内生产总值为 13365 亿元、人口总数为 4434 万人等，都属于统计指标。

统计指标由指标名称和指标数值两个基本部分组成。指标名称反映所研究现象的实际内容，是对现象本质特征的一种概括，是对总体数量特征的质的规定性。确定统计指标必须以一定的理论为依据，例如经济统计指标的理论依据是经济学。但并不是所有的理论概念或范畴都是统计指标，这要看它能否数量化。例如，商品销售额、国内生产总值等是统计指标，而商品、国民经济等则不是统计指标。因此，作为统计指标的概念与理论概念是有区别的。指标数值是所研究现象实际内容的数量表现，是对总体本质特征的量的规定性，是对个体特征综合和计算的结果。

由于所研究现象的范围是可变的，其发展过程是动态的，因此每个统计指标数值都必须有明确的空间界限和时间界限。同时，为了使同一指标在不同空间和时间上的数值具有可比性，必须确定统一的指标计算方法。并且，为了使指标数值意义明确，还必须明确的计量单位。这样，统计指标就涉及到指标名称、计算方法、空间限制、时间限制、具体数值和计量单位六个要素。例如浙江省 2005 年地区生产总值 13437.85 亿元，指标名称是地区生产总值，计算方法是根据不同产业部门、不同支出构成的特点和资料来源情况而采用不同的方法，空间限制是浙江省，时间限制是 2005 年，具体数值是 13365，计量单位是亿元。由此可以看出，统计指标具有数量性、综合性和具体性三个特点。

2.统计指标与标志的关系

统计指标与标志既有区别，又有联系。区别主要有两个方面：首先是指标和标志说明的对象不同，指标说明总体的特征，标志则说明个体的特征。其次是指标和标志的表现形式不同，指标是用数值来表现的，而标志则既有只能用文字来表现的标志，又有用数值来表现的数量标志。当然，这个区分是相对的。

联系也有两个方面：首先，标志是计算统计指标的依据，即统计指标数值是根据个体的标志表现综合而来的。例如，对品质标志可根据定类尺度或定序尺度计算各类个体数及其所占的比重（如一批产品中合格品数及合格率），对数量标志则可根据定距尺度或定比尺度计算各种总量指标、平均指标和相对指标等。其次，由于总体与个体的确定是相对的、可以换位的，因而指标与标志的确定也是相对的、可以换位的。这样，指标与标志在许多场合并不需要严格区分，例如企业人数、企业总产量、企业总产值等，既是指标也是标志。因此，指标与标志同属于变量的范畴。

3.统计指标的分类

（1）统计指标按其计算的范围不同，可以分为总体指标和样本指标。总体指标根据（有限）总体中所有个体的标志表现综合计算而得，反映总体数量特征；样本指标则仅根据总体中部分个体的标志表现综合计算而得，反映样本数量特征。总体指标也称总体参数，对于某一确定的总体，任何一个总体指标的数值是

惟一的，但在非全面观测的情况下是未知的。样本指标也称样本统计量，对于所抽的一个样本来说，任何一个样本指标都有一个可知的数值，但由于样本是随机抽取的、非惟一的，因此样本指标的数值随样本不同而不同，样本指标是随机变量。统计研究的一大任务，就是要用可知但非惟一的样本指标数值去推断惟一却未知的总体指标数值。

(2) 统计指标按其反映现象的内容不同，可以分为数量指标和质量指标两种。数量指标也称为总量指标，它是反映现象总体某一方面绝对数量特征的指标，表明现象所达到的总规模、总水平或工作总量。例如人口数、企业数、总产量、总产值、土地面积、投资额等，都属于数量指标。数量指标的计量单位有实物单位、价值单位和时间单位三种，其中实物单位又有自然单位、度量衡单位、双重单位和复合单位等。数量指标按照其反映现象内容的不同，可以分为总体标志总量和总体容量（即第二章要介绍的总体总频数）两种：总体标志总量是总体中所有个体的某个标志的标志值之和，也即某变量的所有变量值之和；而总体容量则是总体所包含的个体数，也即某变量的变量值个数。显然，总体标志总量中的标志值（变量值）与总体容量中的个体是一一对应的，所以总体容量越大（小），总体标志总量一般也就越大（小）。在通常情况下，数量指标主要是指总体标志总量。数量指标按照其反映现象时间状况的不同，又可以分为时期指标与时点指标两种：时期指标是反映现象在一定时期内累计达到的总量，其数值大小与时间长短有直接关系，不同时间上的数值可以累加，数值需要通过连续登记取得，例如企业产量、地区 GDP 等指标；时点指标是反映现象在某一时刻（时刻、瞬间）所达到的总量，其数值大小与时间长短无直接关系，不同时间上的数值不可以累加，数值通常不需要通过连续登记取得，例如企业人数、地区居民存款余额等指标。数量指标的具体结果就是前述的绝对数。

质量指标是反映现象总体内在对比关系或总体间对比关系的指标，表明现象所达到的相对水平、平均水平、工作质量或相互依存关系。例如人口性别比例、职工平均工资、产品合格率、人均土地面积、产值增长速度、资金利润率等，都属于质量指标。质量指标又可分为相对指标和平均指标两种。相对指标是反映事物内部或相关事物之间相对数量关系的指标，是两个有联系的统计指标对比的结果，包括结构相对指标（总体中部分总量与总体总量之比）、比例相对指标（总体中某部分总量与其他部分总量之比）、比较相对指标（两个同类指标之比）、动态相对指标（同一指标在不同时间之比）、强度相对指标（两个性质不同但有联系的总量指标之比）和计划完成程度相对指标（实际指标与计划指标之比）等；平均指标是反映变量分布集中趋势或中心位置的指标，表明变量的一般数量水平，包括算术平均指标、几何平均指标、调和平均指标、众数指标和中位数指标等（将在第三章介绍）。平均指标和相对指标中的一部分强度相对指标有计量单位，其他相对指标则没有具体的计量单位。相对指标和平均指标的具体结果分别是前述的相对数和平均数。

由于数量指标（主要是总体标志总量）的数值大小一般与总体容量大小有关，所以又称为外延指标；而质量指标的数值大小一般与总体容量大小无直接关系，所以又称为内涵指标。在三大类统计指标中，总量指标是基础，相对指标和平均指标由总量指标派生而来。

(3) 统计指标按其反映现象的时间状态不同，可以分为静态指标和动态指标两种。静态指标是反映现象总体在某一时刻或相对静止时间上数量特征的指标，包括一般的总量指标、静态相对指标和一般平均指标。动态指标是反映现象

总体在不同时期或时点上发展变化情况的指标，包括增长量指标、动态相对指标和序时平均指标等。

4.统计指标的设计

统计指标的设计是一个重要的问题，应着重考虑以下几个方面：

（1）科学确定指标的名称和涵义。统计指标作为科学的概念决不单凭想象就可产生，必须从事物的普遍联系中，以对客观现象本质规律进行充分认识为基础，是人们对客观现实认识加以抽象概括的结果。在确定了指标名称后，还需要对指标涵义做出明确解释，指出它的内涵，即质的规定性是什么，有什么作用和功能，有什么优缺点等，以便不同指标之间能从根本上相互区别。通常，统计指标的定义方法有提要法、示算法、穷举法和限定法等，可根据不同情况加以选择。

（2）科学确定指标的计算范围和计算方法。统计指标都需要量化，因此必须依据对指标本质属性的分析，明确划分指标的计算范围和界限，对于可能产生的疑问应做出必要的规定和解释。计算范围通常包括时间范围、空间范围和口径范围，其中口径范围的确定最为复杂。例如，职工的工资包括哪些项目（是否包含奖金、津贴、劳保福利）？就属于口径范围。确定指标的口径范围就是明确指标的外延。与此同时，对指标的计算方法必须明确规定，尤其是对一些涉及要素较多、较复杂的指标。例如，国内生产总值指标就可以分别从生产、分配和使用角度分别采用生产法、收入法和支出法等方法进行核算，对每一种方法都必须做出具体的说明，以便于取得所需的资料并进行比较分析。

（3）确定指标的数据来源和量化尺度。要根据指标的内容、性质不同，确定不同的数据来源，选择不同的数据搜集方式和方法。在数据搜集过程中，要根据标志或变量的性质不同，采用不同的量化尺度。尤其是对于品质标志，对量化的标准要有明确的规定。作为一种延伸，还应对如何解释指标的结果做出说明。例如，基尼系数 0.35 说明什么，相关系数 -0.9 表明什么（取值范围如何），哪些指标的数值越大越好，哪些指标的数值越小越好等，都应有必要的规定和解释。

（4）确定合适的计量单位。统计指标的计量单位通常有实物单位、货币单位、劳动单位和相对比较单位等，要根据指标类型和数量特征的不同加以选择确定。有时还使用其他计量单位，例如考试成绩和体操比赛中的“分”等。同一指标常常有多个计量单位可以使用，如果不明确规定或说明，就不知道指标数值的实际意义。例如国内生产总值的计量单位，可以是亿元，也可以是万元、千元等，应该选定其中合适的一个。

（二）统计指标体系

1.统计指标体系的含义

任何一个总体都有多方面的数量特征，而一个统计指标只能说明某一方面的数量特征。因此，要想比较全面了解所研究现象总体的数量特征，就应该设计和使用一系列相互联系的统计指标。这种反映同一总体多个方面数量特征的、一系列相互联系的统计指标所形成的体系，就称为统计指标体系。例如，对于工业企业总体，要想比较全面了解工业企业的生产经营情况，就必须从劳动、设备、原材料、资金等基本要素和生产、销售等环节出发，通过职工人数、固定资产总额、流动资金总额、总产值、增加值、产品销售率、销售收入、利润额、利润率、劳动生产率、资金周转速度等一系列指标来反映，这些指标就构成了反映工业企业生产经营状况的统计指标体系。

统计指标体系由一系列统计指标构成，但并不是单个指标的简单组合，而是各个指标之间相互联系、相互制约的。通过指标体系，不仅可以使我们认识现象

的角度更多、视野更广、看问题更全面，而且还可以从指标的相互联系中再现客观现象之间的内在联系获取更多的信息。

2. 统计指标体系的表现形式

统计指标体系有多种表现形式，根据统计指标之间的联系关系，大致有以下几种：一是数学等式关系，即若干统计指标之间可以构成一个等式关系，例如“国民生产总值=国民生产净值+固定资产折旧”，“非金融资产+金融资产=金融负债+自有资金”等。二是相互补充关系，即各个指标相互配合、相互补充，从不同方面共同来说明现象的数量特征，例如前述的反映工业企业生产经营状况的统计指标体系。三是相关关系，即各个指标之间存在着一定的相关关系，例如人均国民生产总值、人均收入和人均消费支出三个指标之间就存在较高的相关关系，共同反映一个国家或地区的经济生活水平。四是原因、条件和结果关系，即若干指标中有的是原因，有的是条件，有的则为结果，例如工业企业总产值、工人劳动生产率和资金装备率三者之间，工业总产值的增加是工人劳动生产率提高的结果，而提高工人劳动生产率的条件是提高资金装备率。

3. 统计指标体系的设计

统计指标体系的设计，应该根据研究的目的，确定需要观测和分析的几个方面，并分清主次，然后每个方面各设置所需的统计指标，用系统论的观点把它们组织起来，共同构成统计指标体系的框架。一般地，统计指标体系的设计应遵循以下几个原则：

（1）目的性原则。即应该紧紧围绕研究目的，满足科学研究、宏观经济管理、社会组织管理等的不同需要。对于同一总体，若研究目的不同，则所设计的指标体系就不同。

（2）科学性原则。即应该符合科学的原理，能比较全面、准确、客观地描述和反映所研究现象的数量特征，从而了解和认识其本质性的东西。这就必须与所研究现象所属领域的专业理论紧密联系，科学地设置和选择各有关指标。

（3）可行性原则。即应该确保每个指标都能准确计算出其数值，也即要使所选择的每个指标都有可靠的数据来源。同时，还要考虑指标的计量手段和计算方法是否简便可行，应使之具有可操作性。

（4）灵活性原则。即应该从客观实际需要出发，结合条件和需要，灵活地加以确定，并要注意着眼于未来。反映同一研究目的统计指标体系，应根据环境的变化和事物发展的新特点、新趋势，不断地加以调整和完善。当然，也要注意保持指标体系的历史可比性。

（5）层次性原则。即应该体现所研究现象本质内涵的层次性，区分大系统、小系统和子系统等的范围并弄清不同层次系统之间的关系，从而使不同层次系统都有相应的指标体系来反映其数量特征，以助于人们既认识所研究现象总括的数量特征，又认识更详细、更具体的分系统的数量特征。

（6）联系性原则。即应该使所包含的各个指标之间相互联系，避免信息的重复遗漏。一方面，应在若干都能反映现象某一方面数量特征的指标中，选择最有代表性的指标；另一方面，所研究现象的各个方面都应有一定的指标来描述和反映，共同构成一个完整的体系。既不能把互不相干的指标捆在一起组成指标体系，又不能全由高度相关的指标来组成指标体系。

（7）协调性原则。即应该使各指标之间相互协调一致，在计算方法、计算口径、计算内容等方面相互衔接，不可出现相互矛盾和抵触的现象。同时，要注意与其他环节或其他方面的统计指标紧密配合。

本章小结

1.统计一词包含三个含义：统计数据、统计活动和统计学。统计的本质是关于“为何统计，统计什么和如何统计”的思想，就是围绕研究目的和任务，运用科学的统计方法，去获取真实客观的统计数据，做出必要的统计分析，以了解和认识事物的真相。而统计学则是关于如何搜集、整理和分析统计数据的科学。

2.统计学的产生与发展大致经历了三个时期：古典统计学时期、近代统计学时期和现代统计学时期。在古典统计学时期有德国的国势学派与英国的政治算术学派之分，在近代统计学时期有德国的社会统计学派与比利时的数理统计学派之争，在现代统计学时期则以推断统计发展为主要特征。

3.统计学的研究对象是现象的数量方面，即统计数据。统计学的学科性质从研究对象上看具有数量性、总体性和差异性的特点，从学科范畴上看具有方法性、层次性和通用性的特点，从构成内容上看具有描述性和推断性的特点。

4.统计数据可以从多个不同角度进行分类。根据计量尺度不同可以分为定性数据与定量数据，根据表现形式不同可以分为绝对数、相对数与平均数，根据来源不同可以分为观测数据与实验数据，根据加工程度不同可以分为原始数据和加工数据，根据时间或空间状态不同可以分为时序数据与截面数据。

5.统计研究过程即统计数据研究过程，大致包括统计设计、数据搜集、数据整理与分析、数据解释四个环节，研究方法主要有大量观察法、统计分组法、综合指标法、统计推断法和统计模型法。

6.总体是由客观存在的、所有具有某种共同性质的事物所组成集合体，具有大量性、同质性和差异性的特征。构成总体的每个个别事物称为个体。总体有有限总体与无限总体、具体总体与抽象总体、可计数总体与不可计数总体、自然总体与人为总体之分。总体与个体的关系不是一成不变的。

7.样本是由来自于总体的一部分个体所组成的有限小总体。样本中所包含的个体数，称为样本容量。从总体中最多可以获得的容量为 n 的不同样本数，称为样本个数。样本是总体的代表，样本是用来推断总体的，样本与总体的角色是可以改变的。

8.标志是说明个体特征的名称。标志有品质标志与数量标志、可变标志与不变标志、直接标志与间接标志之分。

9.从广义上说，变量就是可变的标志。变量有定性变量与定量变量、确定性变量与随机性变量、离散型变量与连续型变量之分。

10.统计指标是反映现象总体数量特征的概念及其数值，有指标名称、计算方法、空间限制、时间限制、具体数值和计量单位六个要素。指标与标志既有区别，又有联系。统计指标有总体指标与样本指标、数量指标与质量指标、静态指标与动态指标之分。

11.统计指标体系是由反映同一总体多个方面数量特征的、一系列相互联系的统计指标所形成的体系。其表现形式有数学等式关系、相互补充关系、相关关系和条件、原因、结果关系等。

12.统计学在认识事物、指导生产、经济管理和科学研究等方面都具有重要的作用。

练习与思考

一、判断题

1.统计学是一门关于统计数据的搜集、整理和分析的方法论科学。

2.统计学起源于德国的国势学派。

- 3.描述统计与推断统计的区别在于前者简单,后者复杂。
- 4.数量指标根据数量标志计算而来,质量指标根据品质标志计算而来。
- 5.任何统计数据都可以归类于绝对数、相对数或平均数中的一种。
- 6.统计学可以被理解为关于样本的科学。
- 7.从广义上说,可变标志、指标都是变量。
- 8.无论数量指标或质量指标,其数值大小都与总体容量(或样本容量)有关。
- 9.任何总体,其所包含的个体必须至少具备一个可变标志和一个不变标志。
- 10.电话号码是数量标志。

二、单项选择题

- 1.统计学的研究对象是()。
A. 各种现象的内在规律 B. 各种现象的数量方面
C.统计活动过程 D.总体与样本的关系
- 2.某班3名男生的身高分别为172cm、176cm和178cm,这3个数是()。
A.标志 B.变量 C.变量值 D.指标
- 3.以一、二、三等来表示产品质量的优劣,那么产品等级是()。
A.质量指标 B.品质标志 C.数量标志 D.数量指标
- 4.下列哪个变量不能采用定比尺度计量?()
A.企业职工人数 B.企业产品产量 C.企业销售额 D.企业利润额
- 5.下列哪个指标不属于质量指标?()
A.企业职工平均工资 B.企业利润率
C.企业产品合格率 D.企业增加值
- 6.要了解某市30所中学的学生眼睛视力状况,则个体是()。
A.每所中学 B.全部中学 C.每名学生 D.每名学生的眼睛视力

三、简答题

- 1.统计的含义与本质是什么?
- 2.什么是统计学?有哪些性质?
- 3.统计数据有那些分类?不同类型的数据有什么不同特点?试举例说明。
- 4.定性数据在统计处理上要注意什么问题?
- 5.如何正确理解描述统计与推断统计的关系?
- 6.统计研究的基本过程如何?常用的统计方法有哪些?
- 7.总体、样本、个体三者关系如何?试举例说明。
- 8.如何理解总体的大量性、同质性和差异性?
- 9.如何理解标志、指标、变量三者的含义?试举例说明。
- 10.品质标志、数量标志、质量指标、数量指标四者关系如何?试举例说明。
- 11.什么是统计指标体系?有哪些表现形式?试举例说明。
- 12.统计学与数学有什么不同?

人物介绍

威廉·配第(William Petty, 1623-1687):英国古典政治经济学创始人,统计学家。威廉·配第出生于英国的一个手工业者家庭,从事过许多职业,从商船上的服务员、水手到医生、音乐教授。他13岁进入卡昂大学学习,1648年进入牛津大学研修医学,1649年获得医学博士学位,后来担任牛津大学解剖学教授。1650年到英国驻爱尔兰军队中去当医生。1651年他成了爱尔兰总督小克伦威尔的侍从医生。后逐渐离开医学,从政从商,发展为贵族。他头脑聪明,学习勤奋,

敢于冒险，善于投机，晚年成为拥有大片土地的大地主，还先后创办了渔场、冶铁和铝矿企业。威廉·配第的主要贡献是最先提出了劳动决定价值的基本原理，并在劳动价值论的基础上考察了工资、地租、利息等范畴，他把地租看作是剩余价值的基本形态。马克思称他为“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”。威廉·配第的代表著有《赋税论》，《政治算术》等。