

第1章 导论

学习目标

1. 理解统计的科学含义；
2. 掌握统计数据类型；
3. 认识统计研究的基本方法；
4. 掌握统计学的基本概念。

案例导入

关于某净水器公司租赁模式研究

在水资源贫乏、水源污染严重、二次供水污染等状况频出的当下，保障饮用水安全刻不容缓。同时，随着人们生活水平的提高，对生活饮用水质量的要求也越来越高，中国水家电正从“饮”时代向“净”时代跨越，净水器市场迎来快速发展的契机。火热的市场表现，让许多企业看中了净水领域的巨大潜力，纷纷进入该市场。由于净水器市场大牌云集，各大厂商卧虎藏龙，整个净水器市场趋近饱和状态。在了解净水器市场存在严重的售后服务不足和营销手段单一这些痛点后，某净水器公司推出一项净水器租赁模式，试图通过对西安市市场研究探寻主力客户，从而掀起一场净水器租赁模式革命，从竞争白热化的净水市场中脱颖而出。

该项目的研究目标包括以下三个方面。

- (1) 通过市场调研了解目前净水器市场的现状以及存在的痛点，了解净水器使用人群的主要特征，划分用户类型。
- (2) 了解每类客户的特征，进行客户画像，确定科蒂洛租赁模式的主力客户。
- (3) 根据租赁主力客户画像，找出影响其租赁使用意愿的因素，提出精确的推广建议，助力科蒂洛租赁模式的落地实施。

资料来源：作者自编。

思考：

- (1) 该项目研究的总体是谁？
- (2) 客户画像分类标准是什么？

通过本章学习希望大家能找出答案。

章节导言

德国的统计学家斯勒兹(Schlozer)曾说过：“统计是动态的历史，历史是静态的统计。”可见，统计学的产生与发展是和社会生产的发展、经济的进步紧密相连的。统计学是关于数据的科学，它提供的是一套有关数据收集、处理、分析、解释，并从数据中得出结论的方法。本章将介绍统计学的一些基本问题，包括统计学的含义、统计数据及其分类、统计中常用的基本概念等。

课程思政

家国情怀塑造

“家国”一词最早出现在《逸周书·皇门解》：“是人斯乃谗贼媢嫉，以不利于厥家国”。在中国语境中“家”字有家长、宗庙等古义，而“国”可以视为由很多小家庭组成的大家庭，由“小家”汇聚而来的国家，内里蕴藏着由小及大、自国而家的逻辑递推关系。“情怀”主要是指个人情感的情绪表达，包括心情、心境、胸怀等。“家国情怀”表达的是个人对家与国的情绪、情感，是对自己国家的一种高度认同感、归属感、责任感和使命感的体现，是一种深层次的文化心理密码。

统计的起源很早，据历史记载，我国在夏禹时代(约公元前 22 世纪)就开始人口统计活动。《尚书·禹贡》记述了九州的基本状况，被西方经济学家推崇为“统计学最早的萌芽”。在中国封建社会，户籍统计和田亩统计等都有很大的发展，其制度、方法和组织都居于当时世界先进水平。

中华民族的统计调查史源远流长，那些被记载的历史数据和统计工具，虽然已退出历史的舞台，但统计调查对数字的严谨和事实的执着，却逐渐烙印在我们的民族基因中，不绝如缕，薪火相传。中国古代统计发展史既是人类统计发展史的重要组成部分，也是推进人类文明不断向前的重要力量。

资料来源：作者自编。

1.1 统计学及其应用领域

1.1.1 统计与统计学

“统计”一词具有多种含义，在不同的场合，可以有不同的解释。作为人类认识客观事物的重要方法，统计因其研究对象和内容具有明显的特殊性，从而形成了一种独立的、有独特方法的工作，更成了一门系统的关于统计方法论的学科。因此，统计有三方面的含义，即统计活动、统计数据 and 统计学。

► 1. 统计活动

统计活动是指利用科学的方法，收集、整理、分析和提供有关社会、经济现象数字资料活动的总称，既包括国家统计局系统的活动，也包括其他机构和个人的统计活动。在古代奴隶社会，由于国家在赋税、徭役、征兵等方面的需要，相应地开始了人口、土地等基

本国情的登记和计算活动，这就是统计活动的雏形。

▶ 2. 统计资料

统计资料也称统计数据，是统计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，表现为各种反映社会经济现象数量特征的原始记录，如统计报表、统计图、统计资料汇编、统计年鉴、统计分析报告等。

▶ 3. 统计学

统计学是阐述统计工作基本理论和基本方法的学科，它研究的是数据资料的收集、整理、汇总、描述和分析，以及在此基础上进行的推断和决策。统计学是一门对统计工作实践的理论概括和经验总结，以现象总体的数量方面为研究对象，阐明统计设计、统计调查、统计整理和统计分析，认识社会的方法论学科。

据历史记载，我国在夏禹时代(约公元前 22 世纪)就开始人口统计活动。《尚书·禹贡》记述了九州的基本状况，被西方经济学家推崇为“统计学最早的萌芽”。在中国封建社会，户籍统计和田亩统计等都有很大的发展，其制度、方法和组织都居于当时世界先进水平。

在地中海沿岸及其他地区，统计活动也有悠久的历史。但从 18 世纪开始，统计才得到快速发展。除了人口、税收、土地等传统内容外，商业、航运、外贸、工业、经济和管理等领域的统计活动也非常普遍。

统计活动历史悠久，统计学也有很长的历史。

统计学最早起源于 17 世纪的英国。其代表人物是约翰·格朗特(John Graunt, 1620—1674)和威廉·配第(William Petty, 1623—1687)，代表作是威廉·配第的《政治算术》(1690)，这里的算术是指统计方法。政治算术学派主张以数字、重量和尺度来说话，用图表形式来概括数字资料，用大量观察和数量分析等方法对社会经济现象进行研究，为统计学的发展开辟了广阔的前景。

19 世纪中叶，产生了数理统计学。比利时科学家凯特勒(Quetelet, 1796—1874)把概率论正式引进统计学，使统计方法得到了质的飞跃，为统计的数量分析奠定了数理基础，由此，数理统计方法研究得到了广泛开展。其后，英国学者葛尔登(Galton, 1822—1921)提出了生物统计学，皮尔逊(Pearson, 1845—1926)将生物统计进行了一般化，发展了描述统计，埃奇沃思(Edgeworth, 1845—1926)、鲍莱(Bowley, 1869—1957)则侧重于描述统计在经济领域中的应用和方法研究，费歇尔(Fischer, 1880—1962)创立了推断统计学。20 世纪 50 年代，又出现了贝叶斯统计学，将统计推断运用于决策问题。

统计工作、统计资料和统计学之间有着密切联系。统计工作与统计资料是过程与成果的关系，统计资料是统计工作的直接成果。就统计工作和统计学的关系来说，统计工作属于实践的范畴，统计学属于理论的范畴。统计学是统计工作实践的理论概括和科学总结，它来源于统计实践，又高于统计实践，反过来又指导统计实践，统计工作的现代化同统计科学研究的支持是分不开的。

统计工作、统计资料和统计学相互依存、相互联系，共同构成了一个完整的整体，这就是我们所说的统计。

1.1.2 统计学研究对象的特点

社会经济统计学的研究对象，是指社会经济现象的总体的数量方面，即社会经济现象

总体的数量特征和数量关系。

社会经济统计是对社会经济现象的一种调查分析活动，它具有以下特点。

▶ 1. 数量性

数量性是指统计的研究对象涉及客观现象的数量方面，包括数量的多少、数量之间的关系。通过分析和研究统计数据资料，可以更好地掌握统计规律，实现统计研究的目的。

▶ 2. 总体性

总体性是指统计的研究对象涵盖客观总体现象的数量方面。例如，人口统计要反映和研究一个国家或一个地区全部人口的综合数量特征，而不是了解和研究某个人的特征，但它是从每个人调查开始的。人口统计是这样，其他统计活动也是这样。

▶ 3. 具体性

具体性是指统计的研究对象包括自然、社会经济领域中具体现象的数量方面。它不是纯数量的研究，而是具有明确的现实含义。正因为统计的数量是客观存在的、具体实在的，所以它才能独立于客观世界，不以人们的主观意志为转移。

▶ 4. 变异性

统计研究对象的变异性是指构成统计研究对象的总体各单位，除了在某一方面必须是同质的以外，在其他方面又要有差异，并且这些差异并不是由某种特定的原因事先给定的。也就是说，总体各单位除了必须有某一共同标志表现作为它们形成统计总体的客观依据以外，还必须在所要研究的标志上存在变异的表现。否则，就没有必要进行统计分析研究了。

1.1.3 统计学应用领域

统计方法是适用于所有学科领域的通用数据分析方法，只要有数据的地方就会用到统计方法。随着人们对定量研究的日益重视，统计方法应用到自然科学和社会科学的众多领域，统计学已发展成为由若干分支学科组成的学科体系。可以说，几乎所有的研究领域都要用到统计方法，比如政府部门、学术研究领域、日常生活、公司或企业的生产经营管理都要用到统计方法。

当然，统计并不是仅仅为了管理，它是从自然科学、社会科学的多个领域中发展起来的，为多个学科提供了一种通用的数据分析方法。从某种意义上说，统计仅仅是一种数据分析的方法。与数学一样，统计是一种工具，是一种数据分析的工具。利用统计方法可以简化繁杂的数据，比如，用图表展示数据，建立数据模型。有人认为统计的全部目的就是让人看懂数据，其实这仅仅是统计的一个方面，统计更重要的功能是对数据进行分析，它提供了一套分析数据的方法和工具。不同的人对数据分析的理解不大一样，曲解数据分析是常见的现象。在有些人的心目中，数据分析就是寻找支持，他们的心目中可能有了某种“结论”性的东西，或者说他们希望看到一种符合他们需要的某种结论，而后去找些统计数据来支持他们的结论。这恰恰歪曲了数据分析的本质，数据分析的真正目的是从数据中找出规律，从数据中寻找启发，而不是寻找支持。真正的数据分析事先是没有结论的，通过对数据的分析才能得出结论。统计不是万能的，它不能解决你面临的所有问题。统计可以帮助你分析数据，并从分析中得出某种结论，但对统计结论的进一步解释，则需要你的专业知识。比如，吸烟会使患肺癌的概率增大，这是一个统计结论，但要解释吸烟为什么能

引起肺癌，这就不是单靠统计所能做到的了，而是需要有更多的医学知识才行。

1.2 统计数据的数据类型

统计数据可以从不同的角度进行分类。按照计量尺度的不同，统计数据可以分为4种类型：定类数据、定序数据、定距数据、定比数据。

1.2.1 定类数据

定类数据是指使用定类尺度获得的数据。定类尺度也称类别尺度或列名尺度，是最粗略、计量层次最低的计量尺度。这种计量尺度只能按照现象的某种属性对其进行平行的分类或分组。例如，人口按照性别分为男、女两类；企业按照经济性质分为国有、集体、私营、合资、独资企业等。定类尺度只是测量了现象之间的类差别，对各类之间的其他差别却没有反映。因此，使用该尺度对现象进行的分类，各类别之间是平等的并列关系，无法区分优劣或大小，各类别之间的顺序并不重要。虽然定类尺度计量的结果只是表现为某种类别，但为了便于统计处理，特别是为了便于计算机识别，可以对不同类别用不同的数字或编码来表示，例如，在人口统计中按地区分组、民族分组，并用数字作为代号，如北京为“01”，河北为“02”等；“1”表示男性人口，“0”表示女性人口等。这些数字只是给不同类别的一个代码，并不意味着这些数字可以区分大小或进行任何数学运算。对定类尺度的计量结果，通常是通过计算每一类别中各元素或个体出现的频数或频率来进行分析的。

在使用定类尺度对现象进行分类时，必须符合穷尽和互斥的要求。类别穷尽是指在所做的全部分类中，必须保证每一个元素或个体都能归属某一个类别，而不能在其他类别中重复出现。比如，按照自然的二分法，一个人要么是男性，要么是女性，总是有所归属，而且只能属于其中的一个类别。定类尺度是对现象最基本的测度，是其他计量尺度的基础。在形式上，定类尺度具有对称性和传递性两种属性，对称性说明各类之间彼此相对称，传递性则表示在运算上各类量值只具有相等与不相等的性质。这种测定尺度和分组在实际统计活动中使用得很广泛，主要用于计算各组数值占总体数值的比重和众数等，但不能对各类编号进行加减乘除计算。

1.2.2 定序数据

定序数据是指使用定序尺度获得的数据。定序尺度又称为顺序尺度，是对现象之间的等级差别或顺序差别的一种测度。该尺度不仅可以将现象分成不同的类别，还可以确定这些类别的优劣或顺序。或者说，它不仅可以测度类别差，还可以测度次序差。定序尺度的计量结果虽然也表现为类别，但这些类别之间是可以比较顺序的。例如，产品等级就是对产品质量好坏的一种次序测量，它可以将产品分为一类品、二类品、三类品及次品等；考试成绩可以分为优、良、中、及格、不及格等；一个人受教育程度可以分为小学、初中、高中、大学及以上等；一个人对某种现象的态度可以分为非常同意、同意、保持中立、不同意、非常不同意，等等。很显然，定序尺度对现象的计量要比定类尺度精确一些，但它也只是测量了类别之间的顺序，而未测量出类别之间的精确差值。因此，该尺度其计量结

果不仅能对现象分门别类，还能比较大小，但不能进行加、减、乘、除等数学运算。

1.2.3 定距数据

定距数据是指使用定距尺度获得的数据。定距尺度也称为间隔尺度，它不仅可以将现象区分为不同类型并进行排序，并且可以准确地指出类别之间的差距。

定距数据是指具有顺序和距离属性的数据，既可以表示年份这种离散型数据，也可以表示温度这种连续性数据，比如年份的表示、温度用摄氏度或华氏度来度量等。定距数据的计量结果表现为数值，这种数值的每一间隔都是相等的，只要给出一个度量单位，就可以准确地指出两个计量数之间的差值，如考试成绩在80~90分，相差10分。既然能确定类别之间的差距，当然也就可以比较顺序(比如室外温度高于室内温度、不同实验舱内温度有高低之别)和异同。由于定距数据的计量结果表现为数值，并可以计量数值，因而，它不仅具有定类尺度和定序尺度的特性，其结果还可以进行加、减运算。但要特别注意的是，定距数据不能做乘除运算，且定距数据没有绝对零点，即当变量值为0时不是表示没有，如当温度为0时并不是表示没有温度。

1.2.4 定比数据

定比数据是指使用定比尺度获得的数据。它与定距数据属于同一层次，一般可不做区分，其计量的结果也表示为数值。它除了具有上述三种计量尺度的全部特性外，还具有一个特性，那就是可以计算两个测度之间的比值。这就要求定比尺度中必须有一个绝对固定的“零点”，这也是它与定距数据的唯一差别。换言之，定距数据中没有绝对零点，即定距数据的计量值可以为0，这里的“0”是表示一个数值，即“0”水平，而不表示“没有”或“不存在”。例如，一个学生的统计学考试成绩为0分，是表示它的统计学成绩水平为0，并不表示他没有考试成绩或没有任何统计学知识；一个地区的温度为0℃，表示一种温度的水平，并不是没有温度。可见定距数据中的“0”是一个有意义的数值。定比数据则不同，它有一个绝对“零点”。也就是说，在定比数据中，“0”表示“没有”或“不存在”。比如一个人的收入为“0”，表示这个人没有收入；一种产品的产量为“0”，表示没有这种产品等。因此，采用定比数据计量的结果通常不会出现“0”值。在现实生活中，大多数使用的是定比尺度。

对定距数据和定比数据的区别，有些教材中形象地将定距数据比喻成从桌面上开始测量的高度，定比数据则是从地面上开始测量的高度。定比尺度中由于“0”表示不存在，其数值不仅可以比较大小、计算数值，还可以计算数值之间的比值。例如，甲的月平均工资收入为2000元，乙为1000元，可以得出甲的收入为乙的2倍。而定距数据由于不存在绝对零点，我们只能比较数值差，而不能计算比值。比如我们可以说21℃与7℃之差为17℃，而不能说21℃比7℃热3倍。可见定距尺度只能进行加、减运算，而定比尺度则可以进行加、减、乘、除运算。

上述4种计量尺度对现象的测量层次是由低级到高级、由粗略到精确逐步递进的。高层次计量尺度的统计数据具有低层次计量尺度的统计数据的全部特性，但不能反过来。显然，我们可以很容易地将高层次计量尺度的测量结果转化为低层次测量尺度的测量结果，比如将考试成绩的百分制计分转化为五等级分制。如表1-1所示，给出了上述四种计量尺度统计数据的数学特性。

表 1-1 四种计量尺度统计数据的数学特性

数学特征	计量尺度			
	定类数据	定序数据	定距数据	定比数据
分类(=、≠)	√	√	√	√
排序(<、>)		√	√	√
间距(+、-)			√	√
比例(×、÷)				√

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更多的数学特性，能运用的统计分析方法越多，分析时也就越方便，因此，应尽可能使用高层次的计量尺度的统计数据。

按照统计分析时采用的方法不同，又可以将上述 4 种统计数据综合为定类数据(名义数据)、定序数据和计量数据，即将定距数据和定比数据合称为计量数据。这种分类一般表现在统计分析软件中。

1.3 统计研究的基本方法

1.3.1 大量观察法

大量观察法是根据研究的目的和要求，从总体中抽取足够多的单位进行统计研究，从而达到认识总体数量规律性的一种方法。统计学研究的对象个体间存在较大的差异性，而统计所要研究的是撇开它们的差异性而寻求它们的共性。因此，我们不能任意抽取个别或少数单位进行观察，而应运用大量观察法对同类社会经济现象进行调查和综合分析，使次要的、偶然的因素作用相互抵消，从而排除其影响，以研究主要的、共同起作用的因素所呈现的规律性。例如，某电视台的一个节目组想知道观众对其节目的反应和评价，那么就不能以某个个人的观点做出结论。因此，他们只有进行大量的问卷调查，掌握观众的普遍意见，才能得出有一定可靠性的结论。

1.3.2 统计分组法

统计分组法是在统计整理阶段常用的方法。它是根据事物内在的性质和统计研究任务的要求，将总体各单位按照某种标志划分为若干组或部分的一种研究方法。统计分组法是为了更加有效地研究事物，找出事物总体内部差异，通过总体中不同组的特点来综合反映总体的特征。例如，人口按年龄分组、企业按经济类型分类、国民经济按部门分类等。

1.3.3 综合指标法

综合指标法是指反映所研究现象的数量特征和数量关系的指标数值，并通过分解和对比分析，以研究总体的差异和数量关系。综合指标法通过对大量原始数据进行整理汇总，计算各种综合指标，以显示研究对象在具体时间、地点以及各种因素共同作用下所表现的

规模、水平、集中趋势和差异程度等，概括地描述总体的综合特征和变动趋势。常用的综合指标有总量指标、相对指标、平均指标、变异指标、动态指标和统计指数等。

1.3.4 统计模型法

统计模型法则是综合指标法的扩展。它是根据一定的理论和假定条件，用数学方程去模拟现实客观现象相互关系的一种研究方法。利用这种方法可以对在客观现象和过程中存在的数量关系进行比较完整和近似的描述，凸显所研究的综合指标之间的关系，从而简化客观存在的复杂的其他关系，以便利用模型对所关心的现象变化进行数量上的评估和预测。

1.3.5 归纳推理法

归纳推理法是由个别到一般、由事实到概括的推理方法。综合分析法通过综合指标概括反映总体一般的数量特征，有异于总体各单位的标志值，但又必须从各单位的标志值中归纳而来。归纳推理法可以使我们从具体的事实得出一般的知识，扩大知识领域，增长新的知识，所以是统计研究中常用的方法。

1.4 统计中的几个基本概念

1.4.1 总体和总体单位

► 1. 总体和总体单位的概念

统计总体是指由客观存在的，具有某种共同性质的许多个别现象所组成的整体，简称总体。构成总体的个别现象称作总体单位，也称为个体。例如，要调查了解某镇农民家庭收入情况，那么该镇所有农户就构成了统计总体，每个农户则是总体单位。再如，要调查了解某工业企业职工的情况，那么该企业所有职工构成了总体，每个职工则是总体单位。

总体单位可大可小，如果研究全世界的所有国家地区，那么所有国家地区就是总体，而每一个国家就是总体单位，如果研究全国大学本科学生，全国所有的大学本科学生就是总体，而每一个大学本科学生就是总体单位。

总体单位数也是常用到的概念，是指一个总体中总体单位的数目，常用 N 表示。

► 2. 总体的特征

总体有三个特征，即同质性、大量性和变异性。

所谓同质性，是指构成总体的各个单位必须在某一(些)方面是相同的。例如，调查了解某镇工业生产情况，那么构成总体的每个单位必须有以下共同性，即每个总体单位必须是工业企业并且是该镇所属。同质性是构成总体的首要条件。

大量性是指统计总体必须由足够多的单位组成，仅是个别单位或少数单位不能形成总体。因为统计研究的目的是从数量方面揭示社会经济现象的本质特征和规律性，只有从大量现象的普遍联系中才能表现出来，个别单位和少数单位的特征难以反映现象的本质和规律性。

变异性是指构成总体的各个单位在某些方面是相同的，但在其他方面则各不相同，具

有质的差别和量的差别，这种差别称为变异。例如，企业职工的社会成分是相同的，而每个职工的年龄、工龄、性别、文化程度、工资等都是各不相同的。统计研究的目的就在于通过分析每个单位的差别来认识总体的一般性和规律性。

▶ 3. 总体的分类

统计总体按其包括的单位数的限度，可分为有限总体和无限总体。一个统计总体中包括的单位数如果是有限的或者可数的，称为有限总体。例如，某厂职工人数、某市工业企业个数、某市人口总数等都是有限总体。一个统计总体中包括的单位数如果是无限的或者在一定的条件下是无法明确的，则称为无限总体。例如，某种野生动物的数量、空气中的灰尘等都是无限总体，由于无限总体的数量不能确定，因此无法进行全面调查研究。

1.4.2 样本和样本容量

▶ 1. 样本

样本是指根据研究的目的和要求，从总体中抽取部分总体单位(即个体)组成的集合。在一般情况下，由于总体数量非常大，直接研究总体特征比较复杂。此时，从总体中抽取部分代表性的单位组成样本，利用样本信息对总体进行推断更方便。

▶ 2. 样本容量

样本容量是指样本中个体的数目。由于样本是从总体中抽取出来的，因此样本容量的单位数要小于总体的单位数。样本容量又称“样本数”，是指一个样本的必要抽样单位数目。在组织抽样调查时，抽样误差的大小直接影响样本指标代表性的大小，而必要的样本单位数目是保证抽样误差不超过某一给定范围的重要因素之一。因此，在抽样设计时，必须决定样本单位数目，因为适当的样本单位数目是保证样本指标具有充分代表性的基本前提。

1.4.3 标志和标志值

标志是说明总体单位特征的名称。标志的具体表现称为标志表现。例如，职工的年龄有18岁、19岁、20岁等，其中“年龄”是标志，“18、19、20”则是标志的具体表现，称为标志值。再如，职工的性别有男和女，其中“性别”是标志，“男、女”则是标志表现。标志具体分类如下。

▶ 1. 标志按其具体表现的性质不同可以分为品质标志和数量标志

品质标志是表明现象质的特征，通常不能用数量表示，如职工的性别、民族、文化程度等。数量标志是表明现象量的特征，一般用数量表示，如职工的年龄、工龄、工资等。

▶ 2. 标志按其变异情况可以分为不变标志和可变标志

不变标志即标志的具体表现都是相同的或一致的。不变标志是统计总体同质性的体现，一个统计总体至少存在一个不变标志。可变标志即标志的具体表现不尽相同。可变标志是统计总体变异性的体现，如各个企业的职工人数、增加值、资产总值、利税额等都可能存在差异，这些标志便是可变标志。

1.4.4 统计指标及其类型

统计指标是指表明统计总体数量特征的概念及其数量表现。例如，我国2022年国内

生产总值 1 210 207.2 亿元，它是根据一定的统计方法对总体各单位的标志表现进行登记、核算、汇总而成的统计指标，说明我国国民经济这个数量特征。这个数量指标的名称是“国内生产总值”，指标的数值是“1 210 207.2 亿元”。

► 1. 统计指标体系设计

统计指标体系是指由众多有联系的指标组成的有机的整体。其中各个指标之间不是孤立的，而是有联系的，并且组合在一起，能共同说明总体的某一特征的整体。

统计指标体系设计的内容包括指标体系框架和指标名称、含义、内容、计算范围、空间范围等。统计指标设立原则包括以下两点。

(1) 明确统计研究目的。指标设立必须以满足统计研究目的为前提，这是一个方向性的问题。在设立统计指标之前必须清楚调查研究的目的是什么，要向着什么方向前进。

(2) 体现统计研究对象的本质。设立指标时要熟知统计研究对象的特点，使设立的统计指标体现统计研究对象的本质，为以后的统计分析奠定基础。

► 2. 指标与标志的区别和联系

指标与标志的区别：一是指标说明总体某一综合数量特征，而标志说明总体单位特征；二是指标都可以用数量表示，而标志既可以用数值来表示，也可以用文字来表示。

指标与标志的联系：一是许多统计指标的数值是由总体单位的数量标志汇总得到的；二是标志和指标之间是可以相互转化的，因研究范围和总体不同而定。

► 3. 统计指标的分类

在实际工作或统计研究中，会经常用到统计指标。对统计指标通常有两种理解和使用方法。一是概念性指标，它是用来反映现象总体数量状况的基本概念。例如，年末全国人口总数、全年国内生产总值、国内生产总值年度增长率等。二是具体的统计指标，它是反映现象总体数量状况的概念和数值。

概念性的统计指标包含 3 个要素，即指标名称、计量单位和计算方法，例如，国内生产总值、社会消费品零售总额、居民消费价格指数(CPI)、工业品出厂价格指数(PPI)、国民可支配收入等。具体的统计指标是反映总体数量特征的概念及其指标数值，一般包含 6 个要素，除了概念性统计指标的 3 个要素之外，还包括时间限制、空间限制、指标数值等 3 个要素。

(1) 统计指标按它所说明的总体现象内容的特征，可以分为数量指标和质量指标。

数量指标是反映总体某一特征的绝对数量。这类指标主要说明总体的规模、工作总量和水平，一般用绝对数表示。例如，2021 年中国粮食播种面积为 1.18 亿公顷，较 2020 年增加了 863 千公顷等。质量指标是反映总体的强度、密度、效果、结构、工作质量等，例如，人口密度、劳动生产率、资金利润率等。这类指标一般用平均数、相对数表示。质量指标的数值并不随总体范围的大小而增减。例如一个 100 万人口的城市第三产业在国民生产总值所占的比重也可能小于某个 30 万人口的城市第三产业在国民生产总值中所占的比重。

(2) 统计指标按其具体内容和作用可分为总量指标、相对指标和平均指标。

总量指标是反映总体现象规模的统计指标，它表明总体现象发展的结果。例如：2022 年国内生产总值(GDP)为 1 210 207.2 亿元，2022 年年末国家外汇储备 31 277 亿美元，2022 年全国粮食总产量达到 6.87 亿吨等。总量指标按其所反应的时间状况不同又分为时期指标

和时点指标。时期指标又称时期数，它所反映的是现象在一段时期内的总量，如产品总产量、商品零售额等。时期数通常可以累积，从而得到更长时期内的总量。时点指标又称时点数，它所反映的是现象在某一时刻或瞬间上的总量，如年末人口数、股票价格和股票价格指数等。时点数通常不能累积，各时点数累计后没有实际意义。

相对指标即“相对数”，是用两个有联系的指标进行对比的比值来反映社会经济现象数量特征和数量关系的综合指标。相对指标其数值有两种表现形式：无名数和复名数。无名数是一种抽象化的数值，多以系数、倍数、成数、百分数或千分数表示。复名数主要用来表示强度的相对指标，以表明事物的密度、强度和普遍程度等。例如，人均粮食产量用“千克/人”表示，人口密度用“人/平方公里”表示等。

平均指标又称平均数或均值，它反映的是现象在某一空间或时间上的平均数量状况，如平均工资、平均成本等。

1.4.5 变量及其类型

在统计中，把说明现象某种特征的概念称为变量，变量的具体表现称为变量值。标志和指标都属于变量的范畴。统计数据就是变量的具体表现。变量就是可变的数量标志。例如，商业企业的职工人数、商品流转额、流动资金占用额等数量标志，这些变动的数量标志就称作变量。

变量值就是变量的具体表现，也就是变动的数量标志的具体表现。例如，企业的职工人数是一个变量，甲企业职工人数 100 人，乙企业职工人数 150 人，丙企业职工人数 200 人等，100 人、150 人、200 人都是职工人数这个变量的变量值(标志值)。

变量按其取值的不同，可以分为离散变量和连续变量。离散变量可以取有限个值，并且其取值都以整位数断开，可以一一列举，如人数、产品数量等；连续变量可以取无穷多个值，其取值连续不断，不能一一列举，如年龄、身高、粮食产量等都是连续变量。在社会经济问题研究中，当离散变量的取值很多时，也可以将离散变量当作连续变量来处理。

变量按其性质不同，可以分为确定性变量和随机变量。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。比如对于一个生产企业，利润就是销售收入除去成本和税金，在销售收入、成本、税金发生后，企业所得利润就是一个确定的数值。随机变量是指变量值的大小变化没有一个确定的方向，带有偶然性。比如，检查一批灯泡的使用寿命，每只灯泡的使用寿命就是随机变量。随机变量比确定性变量普遍得多，随机变量带有相应的概率模型。

导入案例分析

应用本章所学知识，根据某净水器研究目标，该项目研究总体为西安市居民，样本是西安市购买过净水器的居民。客户画像是以样本为研究对象，其用户类型，可以按品质标志分类，例如性别、居住地点、职业；可以按数量标志进行分类，例如年龄、收入。

| 本章小结 |

本章主要讲解统计的三个含义：统计数据、统计活动和统计学。而统计学学科从研究对象上看具有数量性、总体性和差异性的特点，统计数据可以从多个不同角度进行分类。按照计量尺度的不同，统计数据可以分为4种类型：定类数据、定序数据、定距数据、定比数据。统计数据研究过程，大致包括统计设计、数据收集、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。

统计学包括几组基础概念：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的集合体，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之分。样本是由来自总体的一部分个体所组成的有限小总体。样本所包含的个体数，称为样本容量。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之分。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。统计指标是反映现象总体数量特征的概念及其数值。指标与标志既有区别，又有联系。

| 复习与思考 |

1. 简述品质标志与数量标志的区别。
2. 品质标志与质量指标有何不同？品质标志可否汇总为质量指标？
3. 举例说明总体、样本、统计量、变量这几个概念。

| 在线课堂 |



拓展知识：
农夫山泉的大数据挖掘