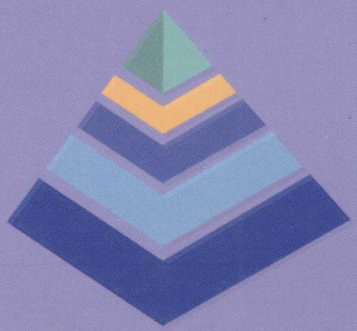
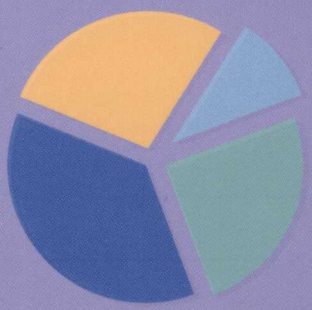
普通高等教育“十三五”规划教材

统计学
STATISTICS

张东光　袁 岩/主编

科学出版社

普通高等教育“十三五”规划教材

统 计 学

主 编 张东光 袁 岩
副主编 李艺唯 王晓红 刘爱芹

科 学 出 版 社

北 京

内 容 简 介

本书是在编者多年教学实践基础上编写而成。本书注意与其他相关课程内容的衔接，强调精炼与实用、理论与实际的结合，主要从应用角度说明统计学的基本理论和方法。全书紧紧围绕“统计与数据”这一主线展开，共分为10章，基本内容包括三个模块。第一至五章主要介绍描述统计学的基本理论与方法，第六至九章介绍推断统计学的基本理论方法及其在相关回归分析中的应用，第十章介绍国民经济主要统计指标。外加两个附录：一是常用统计表，二是 Excel 在统计学中的应用。每章配有思考与练习题，并提供了练习题参考答案。为便于教学，本书配有多媒体教学课件，且配有《统计学学习指导》。

本书可作为经济类和管理类非统计学专业的本科生教材，亦可作为广大实际经济管理工作人员学习和应用统计学知识的参考书。

图书在版编目（CIP）数据

统计学 / 张东光，袁岩主编. —北京：科学出版社，2016
普通高等教育“十三五”规划教材
ISBN 978-7-03-049420-7

Ⅰ. ①统… Ⅱ. ①张… ②袁… Ⅲ. ①统计学-高等学校-教材
Ⅳ. ①C8

中国版本图书馆 CIP 数据核字（2016）第 167658 号

责任编辑：滕亚帆 李 萍 / 责任校对：张凤琴
责任印制：徐晓晨 / 封面设计：华路天然工作室

科 学 出 版 社出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com

北京京华虎彩印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2016 年 8 月第 一 版 开本：787×1092 1/16
2016 年 9 月第二次印刷 印张：18
字数：470 000

定价：39.80 元

（如有印装质量问题，我社负责调换）

前　言

任何事物都是质与量的统一体，由于对事物认识的角度不同，所以形成了不同的学科。从质的角度对事物的认识，形成了各种各样的理论学科，人们称之为实质性科学。从量的角度对事物的具体认识，就形成了统计学。统计学是从数量角度认知事物数量规律的方法论科学。它回答事物的数量是多少，事物间的数量关系是怎样的，事物具有怎样的数量规律等问题。但这种对事物数量规律的认识并不是独立的，而是在实质性科学理论指导下的定量研究，目的在于达到对事物的全面认识。在现代社会中，统计学的应用越来越广泛，特别是在大数据时代，随着计算机工具的应用，那种无聊的“纸和笔的统计学”时代已经过时了。

拥有统计思维，对于认知世界非常重要，你甚至有可能直接得出世界顶级学者过去经过多年研究才能得到的结论。是否拥有这种能力，一定会对人生产生非常重大的影响。早在 1903 年，英国作家、历史学家赫伯特·乔治·威尔斯曾经预言：在未来社会，统计思维将像阅读能力一样成为社会人必不可少的能力。2011 年 2 月，在国务院学位委员会第 28 次会议通过的新的《学位授予和人才培养学科目录(2011)》中，统计学从数学和经济学中独立出来，成为一级学科。这一变革必将对中国统计教育事业的发展和人才培养产生巨大且深远的影响。

统计学发展到今天，随着应用的范围越来越广，统计方法也越来越多，有的是适合各种研究对象的通用方法，有的则是针对特定问题的专门方法。介绍统计学基本理论和方法的教材也层出不穷，且各有所长。

本书是在广泛吸收国内外优秀教材优点的基础上，结合我们的教学经验集体编写而成的，是山东财经大学精品课程教材。本书具有如下特点：内容上，注重与其他课程的衔接；阐述上，强调精炼与实用、理论与实际的结合，摈弃复杂的数学公式推导，主要从应用角度阐明统计学的基本理论和方法；写作上，力求达到通俗易懂，好学好记，学以致用的目的。

全书内容共 10 章：第一～五章主要介绍描述统计学的基本理论与方法；第六～九章介绍推断统计学的基本理论与方法及其在相关回归分析中的应用；第十章介绍国民经济统计的主要指标。外加两个附录：附录 1 是各种常用统计表的信息；附录 2 是 Excel 在统计学中的应用。前者满足统计研究中常用信息的查询，后者则向读者提供一种常见统计软件的操作办法，便于读者在理解统计基本理论和方法的基础上进行实际操作。每章配有思考与练习题，并提供练习题参考答案，供读者练习参考使用。为便于教学使用，本书配有多媒体教学课件，且配有《统计学学习指导》(简称《指导书》)。《指导书》明确了每一章的学习目的和要求，并对本书内容和重要公式进行了系统的梳理，还提供了拓展内容、内容丰富的各种练习题及答案。本书适合 3 学时 1 周的本科教学，可根据不

同学时增减内容。

本书由山东财经大学统计学院长期从事统计学教学的教师共同编写而成，张东光、袁岩两位教授任主编，李艺唯副教授、王晓红教授、刘爱芹副教授任副主编，参加编写的人员还有（按姓氏笔画为序）：田金方、任文东、刘建冰、李杰、吴世国、张伟、栾文英、裴海峰、翟艳敏、薛梅林。本书结构由主编提出，并与副主编共同商定，在各作者提供初稿的基础上，最后由主编对内容进行了修改、编纂。山东财经大学统计学院的三位博士生导师石玉峰教授、赵霞教授、林春艳教授作为审稿人，通读了全书，并提出了宝贵的修改意见和建议。

科学出版社的编辑为本书的出版付出了辛勤的汗水，我们多次沟通、交流，正是由于他们的热情支持感染着作者，才使本书顺利出版。本书写作过程中也得到了山东财经大学统计学院党委书记姜伟等有关领导的大力支持。借此书出版之际，特向为本书出版付出辛勤劳动的诸位领导、同事、编辑们表示衷心的感谢，也向我们参考的书籍作者们表示衷心的感谢。

虽然我们努力工作，对书稿的内容反复斟酌修改，甚至数易其稿，但由于水平所限，书中难免存在不妥之处，恳请使用本书的各位老师、同学和其他读者，将您的建议和意见反馈给我们，我们将进行改正，衷心感谢您的支持和帮助。

作　者

2016 年 4 月于山东财经大学燕山校区

目　录

第一章　统计与数据

统计是认识社会最有力的武器之一，只有真正拥有统计思想，形成统计思维，才能很好地应用统计方法解决实际问题。为了更好地领会统计思想，灵活地应用统计方法来认识客观现象的数量规律，本章首先介绍统计学的基本问题，主要包括统计的含义及特点，统计学的基本要素，统计学的类型，统计数据的计量尺度及其类型、统计数据的表现形式，统计数据的来源等。

第一节　统计与统计学

一、统计的含义

“统计”一词在各种实践活动和科学研究领域中经常出现。然而，不同的人或在不同的场合，对其理解是有差异的。比较公认的看法是，统计有三种含义，即统计工作、统计资料和统计学。

1. 统计工作

统计工作(statistical work)即统计活动，是指搜集、整理和分析统计数据，并探索数据内在数量规律性的活动过程。它的产生与发展已有几千年的历史。统计工作一般说来是由统计设计、统计调查、统计整理和统计分析构成的。统计设计是对所做统计工作的通盘考虑和安排，表现为各种设计方案。统计调查则是根据统计设计的要求，对调查对象中个别事物的特征进行的资料搜集过程，其结果为各种杂乱无章的个体资料，它是统计工作的基础。统计整理是根据统计研究目的，对统计调查搜集的个体资料进行的加工，并使之条理化、系统化的过程，是统计工作的中间环节。它是统计调查的延续，又是统计分析的前提。统计分析是对统计整理的资料进行的数量分析，从而揭示客观现象数量规律的过程，它是统计工作的中心环节。

2. 统计资料

统计资料或称统计数据(statistical data)，是统计工作过程所取得的各种数字资料和其他资料的总称，表现为各种反映客观现象数量特征的原始记录、统计台账、统计表、统计图、统计分析报告、政府统计公报、统计年鉴等各种数字和文字资料。

3. 统计学

统计学(statistics)是阐述统计工作基本理论和基本方法的科学，是对统计工作实践的理论概括和经验总结。它以客观现象总体的数量方面作为研究对象，阐明统计设计、统计调查、统计整理和统计分析的工作理论与方法，是一门方法论科学。

统计工作、统计资料和统计学之间有着密切联系。统计工作同统计资料之间是工作过程同成果的关系，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，统计工作属于实践范畴，统计学属于理论范畴。统计学是统计工作的理论概括和科学总结，它来源于统计工作，又高于统计工作，反过来又指导统计工作，两者相辅相成，统计工作的现代化同统计科学研究的支持也是分不开的。

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

二、统计的特点

统计研究不同于其他学科的科学研究，它具有自己的特点。

（一）数量性

数量性是统计的基本特点，统计是从数量方面入手认识客观现象的工具。可以说，没有数量就没有统计。常言道："数字是统计的语言"指的就是这个意思。

数量性具体包括三方面的内容：

(1) 数量特征。即研究客观现象的总规模、总水平等。

(2) 数量关系。即研究客观现象的内部结构、比例关系、相关关系等。

(3) 数量界限。即引起客观现象质变的数量。例如，完成计划与未完成计划有质的差别，计划完成程度 100% 就是质与量互变的界限。

（二）总体性

统计学以客观现象总体的数量方面作为自己的研究对象，因此，总体性是统计的又一重要特点。所谓总体性，是指统计从整体上反映和分析客观现象的数量性，而不是着眼于个别事物的数量，因为事物的本质和发展规律只有从整体上观察，才能作出正确的判断。例如，只有对大量的出生人口进行观察才能得出合理的人口性别比例，若只对个别家庭的出生人口进行观察是很难得出正确结论的。

（三）具体性

统计研究的是客观现象在具体时间、地点、条件下的数量，而不是抽象的数量，这是统计学与纯数学的一个重要区别。

任何客观现象都是质与量的统一体。一定的质规定一定的量，一定的量又表现出一定的质。例如，100 万这个数字，在纯数学中只是一个抽象的数字，而统计学中则必须明确它所对应的客观现象是谁，是以什么单位计量的，是什么时间条件下的数量等。

三、统计学的类型

统计研究方法已广泛应用于自然科学和社会科学的众多领域，统计学也已发展成为由若干分支组成的学科体系。由于不同的视角或不同的研究重点，人们常对统计学科体系作出不同的分类。根据统计学是方法论科学这一特点，一般有两种基本的分类：一是从方法的功能看，统计学可分为描述统计学和推断统计学；二是从方法研究的重点看，统计学可分为理论统计学和应用统计学。

（一）描述统计学和推断统计学

描述统计学(descriptive statistics)是研究如何取得反映客观现象特征的数据，并利用统计分组方法对所搜集的数据进行加工整理，以统计图表形式加以显示，从而确定数据的分布形态，进而通过综合计算分析得出客观现象最基本的数量规律性的统计学分支。描述统计学的内容包括统计调查方式和资料搜集方法，统计数据的分组、显示方法，数据分布基本特征的

计算分析方法等。

推断统计学(inferential statistics)是研究如何根据样本数据去推断总体特征的统计学分支，它是在对样本数据进行描述统计的基础上，对总体未知分布或参数作出的概率形式的推断。推断统计学的主要内容包括参数估计和假设检验等。

描述统计学与推断统计学的划分，反映了统计方法发展的前后两个阶段和使用统计方法探索客观现象总体内在数量规律性的不同过程。统计研究过程的起点是个体统计数据，终点是探索出客观现象总体内在的数量规律性。在这一过程中，如果搜集到的是总体数据(如普查数据)，那么运用描述统计就可以达到认识客观现象总体内在数量规律性的目的；如果获得的只是研究总体的一部分数据(如样本数据)，那么要得出总体内在的数量规律性，就需要运用概率论的理论和样本信息，对总体进行科学的推断。显然，描述统计学和推断统计学是统计学的两个重要分支。描述统计学是整个统计学的基础，推断统计学则是现代统计学的核心内容。推断统计学在现代统计学中的作用和地位越来越重要，因为在对现实问题的研究中，所获得的数据主要是样本数据，但这并不等于说描述统计学不重要。如果没有描述统计搜集可靠的统计数据并提供有效的样本信息，再科学的统计推断方法也难以得出切合实际的结论。从描述统计学发展到推断统计学，既是统计学发展的巨大成就，也是统计学发展成熟的重要标志。

(二) 理论统计学和应用统计学

理论统计学(theoretical statistics)是以抽象的数量为研究对象，研究一般的数据搜集、整理和分析数据方法的统计学。理论统计学以数学中的概率论为基础，对统计方法加以推导证明，其中心内容是以归纳方法研究随机变量的一般规律，如统计分布理论、参数估计和假设检验理论、方差分析理论、相关与回归分析理论等。理论统计学的特点是计量不计质，具有通用方法论的理学性质。理论统计学是统计方法的理论基础，没有理论统计学的发展，统计学也不可能发展成为像今天这样一个完善的科学知识体系。

将理论统计学的方法原理应用于各个具体研究领域，就形成了应用统计学(applied statistics)。因此，应用统计学是以各个研究领域的具体数量为研究对象的统计学。应用统计学重在研究一般统计方法的应用，也包括各研究领域实质性科学理论的应用，它既需要进行定量分析，又需要结合研究的客观现象理论进行定性分析。所以，应用统计学是在定性分析基础上的定量研究。例如，统计方法在经济领域的应用形成了经济统计学及其若干分支，在管理领域的应用形成了管理统计学，在社会学研究和社会管理中的应用形成了社会统计学，在人口学中的应用形成了人口统计学，等等。应用统计学除了包括各领域通用的方法，如参数估计、假设检验、方差分析等，还包括某领域所特有的方法，如经济统计学中的指数分析法、现代管理决策法等。应用统计学着重阐明这些方法的统计思想和具体应用，而不是统计方法数学原理的推导和证明。

第二节　统计学的基本要素

一、总体、个体和样本

(一) 总体和个体

统计工作就是通过对所研究的对象进行观测取得其数据资料，并对这些数据资料加以整

理和分析研究的过程。构成统计工作研究对象的全部事物所组成的整体，称为统计总体(population)，简称为总体或母体；而总体中的每个个别事物则称为个体。总体中全部个体的数量称为总体容量，通常用N表示。

在实际研究中所遇到的总体，一般有下列两种：一种总体是由自然物体所组成的总体。例如，要研究全国人口状况，则全国人口是总体，每一个人是个体。又如，要了解某地区的工业生产情况，则该地区的全部工业企业构成总体，每个工业企业是个体。另一种总体是由变量值所组成的总体。例如，要研究某企业职工的平均工资，则该企业每个职工的工资水平的集合构成总体，每个职工的工资水平是个体。又如，要了解某个射击运动员的运动水平，则该射击运动员的每次射击结果的集合构成总体，每次射击结果是个体，等等。这两种不同类型的总体，分别属于不同的研究对象和目的。一般来说，由自然物体所组成的总体能够满足多方面的研究需要，而由变量值所组成的总体主要是满足对该变量的研究需要。在推断统计学中主要使用变量值所构成的总体，并且主要关心变量值的分布，称为分布总体。

如果总体中只包括有限个个体，即总体容量是一个有限数，则称为有限总体；如果总体中包括有无数个个体，即总体容量为无穷大，则称为无限总体。例如，全国人口、某地区工业企业、某企业职工人数都是有限总体；而宇宙中的星球、海洋中的鱼等则可看作无限总体。

确定统计总体就是确定统计活动的研究对象及范围，这需要根据统计研究的目的来进行。研究目的不同，统计总体往往也不同。例如，研究目的是了解工业行业的生产经营状况，则总体就是该行业的全部工业企业所组成的集合；而假若研究目的只是了解工业行业的职工生活情况，则总体就是工业行业的全体职工所组成的集合。

需要特别指出的是，在实际应用中，有时总体中的个体是很不明显的，要区分个体往往十分困难。例如，要考察某地所生产的小麦的出粉率，则总体是该地区所生产的全部小麦，而个体却很不明确；又如，要考察某一段河流的水质污染情况，则总体就是该段河流中的全部水，而个体也很不明确。在上述情况发生的条件下，一般是将每个观察单位看作一个个体，而观察单位的大小以及计量方法则根据观察手段而定。比如，或许将每一千克小麦看作一个个体，将每立方米水或者每升水看作一个个体。

(二) 样本

样本(sample)是指从总体中随机抽取并作为总体的代表的那一部分个体所组成的子集。构成样本的个体数目称为样本容量，简称为样本量，通常用n表示。虽然样本中的个体数量相对于总体而言只是较少的一部分，但样本是从总体中随机抽取并用来代表总体的，基于这种关系，总体又可称为母体，而样本则称为子样。样本是由总体中的一部分个体构成的，假如我们将总体看作由研究对象的所有个体组成的集合，则样本就是总体的一个子集。

样本具有如下特点：

(1) 样本中的每个个体都必须取自于总体的内部。

(2) 样本具有不唯一性。总体是唯一确定的，而样本则是不唯一确定的，一般情况下，从一个总体中可以抽取多个样本容量相同的不同样本。

(3) 样本是总体的代表。抽取样本的目的在于用它的信息来推断总体特征。样本对总体的代表性高低直接影响到用样本信息推断总体特征的误差大小。一般情况下，样本的代表性与样本容量的大小、抽样方法以及抽样技术等方面有关。如何提高样本的代表性来减少抽样

误差，这是统计学需要研究解决的重大课题之一。

(4) 样本抽取具有随机性。从总体中抽取样本，不受调查者主观因素的影响，总体中的每个个体被抽中与不被抽中完全是由随机因素决定的。因此，组织抽样必须保证总体中的每一个个体都有一定的概率被抽中或不被抽中。

二、变量与变量值

变量(variable)是统计学中的一个常用概念，无论是对客观现象特征的描述还是推断，都离不开变量。变量的概念有广义与狭义之分。广义变量是对客观现象特征描述的概念。凡是客观现象的特征取值或类别在一个以上者，均可以定义为变量。它可以用数字表示，如年龄、收入和消费支出等；也可以用类别表示，如反映人口特征的性别、反映产品合格与不合格的产品质量、人们的宗教信仰和文化程度等，它们的取值都是用文字表示的。狭义变量是指仅用具体数字表示的变量。

变量的具体表现称为变量值，广义变量的变量值以数字和文字表示，而狭义变量的变量值只能用数字表示。

变量具有以下特征：

(1) 变量是用于描述总体或个体特征的名称。

(2) 一个变量通常有多个变量值，变量与变量值不是一一对应的关系。

(3) 变量的取值有两个方面，一是在不同时间上取值，如历年职工工资水平；二是在不同空间上取值，如某一时期内不同行业或地区的职工工资水平。

实际工作中，为了满足不同的研究需要，变量有多种不同的分类方法，常见的主要有以下六种：

(1) 变量按其反映现象特征的不同，一般分为属性变量和数字变量两种。属性变量是反映现象品质特征的名称，包括分类变量和顺序变量；数字变量也可称为数值变量，它是反映现象数量特征的名称。例如，性别、民族是属性变量，而产值、职工人数则是数字变量。

(2) 变量按其取值是否连续，可分为离散型变量和连续型变量。凡变量的取值只能是整数而不会出现小数时，这样的变量称为离散型变量，例如，职工人数、设备台数、家庭人口数等。凡变量取值在整数之间还可以取无限的数值，即变量的取值是连续不断的，这样的变量称为连续型变量。例如，身高、体重、收入和支出等。

(3) 变量按其取值变动是否具有确定性，可分为确定性变量和随机变量。凡变量取值的变动具有确定性、方向性的，称为确定性变量。例如，每个工业企业的职工人数、设备台数等都是确定的，并随企业规模增大而增大。凡变量的取值变动没有确定的方向性，而具有一定偶然性的，称为随机变量。例如，一只股票的价格，由于受宏观政策、基本面情况、技术面情况、行业情况以及各种客观环境等因素的影响，表现出很大的不确定性，因此，股票价格就是一个随机变量。

(4) 变量按其在因果关系中所处的地位不同，可分为因变量和自变量。因变量是受其他因素影响的结果性变量，通常作为研究的对象来对待，又称为被解释变量；自变量是影响因变量变化的各种原因性变量，又称为解释变量。例如，用居民收入解释其消费支出时，收入为自变量，消费支出为因变量。

(5) 变量按其是否由研究对象体系范围内决定，可分为内生变量和外生变量。内生变量

是由研究对象体系范围决定的，外生变量是由研究对象体系范围之外决定的。外生变量数值的变化影响内生变量的数值变化，但它并不受内生变量数值变化的影响。例如，研究农产品的供求关系时农产品的供应量、需求量和价格等都是在农产品市场范围内决定的，都是内生变量，而土地资源、降雨量、农业投资和科技投入等都是在农产品市场范围以外决定的，都是外生变量。内生变量与外生变量是建立经济计量模型的重要概念。

(6) 变量按其取值是否具有客观性，可分为实在变量和虚拟变量。凡取值是客观实际存在的变量，称为实在变量或实体变量。虚拟变量则是为了满足统计研究的需要，对客观现象的各类属性表现人为规定的数字，又称为工具变量。例如，男性定为 1，女性定为 0；合格定为 1，不合格定为 0，等等。虚拟变量在建立经济计量模型中往往都会用到。

三、参数和统计量

(一) 参数

用来描述总体特征的概括性数字度量，称为参数。

参数(parameter)是研究者想要了解的总体的某种特征值。对单总体而言，参数通常有总体均值、总体标准差、总体比例等。在统计中，总体参数通常用希腊字母表示。通常总体均值以 μ 表示，总体标准差以 σ 表示、总体比例以 π 表示。

参数的真实数据往往是未知的，但它是一个常数，其取值具有唯一性。比如，某一地区所有人口的平均年龄，一个城市所有家庭的收入差异，一批产品的合格率等，它们的数值都是唯一确定的，但这些数据通常又是未知的。正因未知，才需要进行抽样调查，然后根据样本信息推断出总体参数，所以参数可以推断。

(二) 统计量

用来描述样本特征的概括性数字度量，称为统计量(statistic)。

统计量是样本的函数，并且不包含未知的参数，它是由样本决定的，因此统计量是随机变量，它的取值具有随机性，但只要抽取特定的样本，它的数值就变为可知了。对单总体而言，统计量有样本均值、样本标准差、样本比例等。统计量常用小写英语字母来表示。通常用 $\bar{x}$ 表示样本均值、s 表示样本标准差、p 表示样本比例。它们分别对应待估计的总体均值 μ、总体标准差 σ、总体比例 π。

除了样本均值、样本标准差、样本比例这些统计量，还有一些是为统计分析的需要而构造出来的统计量。比如，用于假设检验的 Z 统计量、t 统计量、F 统计量、χ^2 统计量等。

关于统计量的抽样分布将在第六章介绍，参数估计及其参数假设检验将分别在第七章和第八章说明。

第三节　数据的类型

一、数据的计量尺度

统计数据是对客观现象进行计量的结果，对于不同的事物，我们能够计量或者测度的程度是不同的。有些事物只能对其属性进行分类，如性别、文化程度、商品的品牌和质量等级等；有些则可以用比较精确的数字来加以计量，如年龄、收入、受教育年限、商品的价格和

重量等。显然，采用数字计量比起分类计量对于事物的计量更为准确一些。

根据计量学的一般分类方法，按照对事物计量的精确程度，可将数据的计量尺度由低级到高级、由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。采用不同的数据计量尺度可以得到不同类型的统计数据，而不同类型的统计数据又适用于不同的统计分析方法。

（一）定类尺度

定类尺度(nominal scale)也称类别尺度或者列名尺度，它是按照事物的某种属性对其进行平行的分类或分组的一种测度。

生活中，用这样的计量尺度来测度的事物很多，如性别、民族、国籍、商品的品牌等；该计量尺度是最粗略、计量层次最低的计量尺度，是其他计量尺度的基础，主要特征体现为：

(1) 只能区分事物的类别，也就是说定类尺度只具有“等于”和“不等于”的数学特征。例如，用性别来划分人群，就可以分为男性和女性；用品牌来划分手机，就可以分为摩托罗拉、诺基亚、飞利浦、松下和其他品牌，可以看出，各类之间都是平等的并列关系，人们无法说明男性大于女性，摩托罗拉大于飞利浦，而只能说这个人是男性而不是女性，某手机是摩托罗拉而不是飞利浦。

(2) 对事物的区分必须符合穷尽和互斥的要求。类别穷尽是指在全部分类中，必须保证每一个个体都能归属于某一类别，不能有所遗漏；类别互斥则是指每一个个体只能归属于一个类别，而不能在其他类别中重复出现。例如，一个人要么是男性，要么是女性，总有所归属，而且只能属于其中的一个类别；同样，一部手机只能是众多手机品牌的一个，而且只能是其中一个。

(3) 对定类尺度的数据进行分析的统计量主要是频数或者频率(frequency)。虽然定类尺度计量的结果只能表现为某种类别，但是为了在统计处理中的方便，特别是便于计算机识别，通常都用数字对不同的类别进行编码。例如，用 1 表示男性，0 表示女性；用 1 表示摩托罗拉，2 表示诺基亚，3 表示飞利浦，4 表示松下，5 表示其他品牌。在一次分析中，可通过计算这些不同编码出现的次数，即频数或者频率，来对我们感兴趣的问题进行描述。例如，在一次民意测验中，参与投票的男性有多少名，女性有多少名；或者在一次评选我最喜爱的手机品牌活动中，摩托罗拉、诺基亚、飞利浦、松下得到的选票各是多少。

（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或顺序差别的一种测度。例如，受教育水平、职称、商品的质量等级、人们对某一事物的态度等都是用这种计量尺度来测度的。其计量精度要优于定类尺度，主要特征体现为：

(1) 不仅能区分事物的类别，而且能够比较类别间的优劣和顺序，不仅具有“等于”和“不等于”的数学特征，而且具有比较的数学特征。例如，把人们按照受教育水平分为小学及以下、初中、高中、大学及以上，或者把手机按照质量分为一等品、二等品、三等品。非常明显，划分到大学及以上这一类的人受教育水平是高于高中这一类人的，高中又高于初中；同样，一等品的手机质量优于二等品，二等品优于三等品等。

(2) 对事物的区分同样要求符合穷尽和互斥。

(3) 对定序尺度的数据进行分析的统计量主要是频数、频率、累计频数(cumulated frequency)和累计频率。同样，人们需要对定序尺度计量的结果进行编码以方便计算机的读取和统计运算。但与定类尺度不同的是，在分析中，人们不仅可以计算得到不同受教育水平的人数，而且还可以计算在某一水平之下或者之上的人数。例如，受教育水平在高中及以下的人数就可以通过将受教育水平在小学及以下、初中和高中的人数累加得到，即累计频数。

(三) 定距尺度

定距尺度(interval scale)也称为间隔尺度，是对事物类别或者次序之间间距进行的一种测度。

常见的用定距尺度来测度的有考试成绩、各种心理测试的得分、某个地区的温度等。它是一种较定类尺度和定序尺度更为高级，更为精确的一种计量尺度，其主要特征体现为：

(1) 不仅能区分事物的类别、进行排序、比较大小，而且还可以精确地计量类别间的差异。定距尺度的计量结果为数值，并且可以计算差值，即又具有了加和减的数学特征，因而它的计算也就超越了只能比较相等或者大小的运算范畴，进而可以进行加、减运算了。例如，对一组被访者的统计学考试成绩排序，就可以知道这组被访者中谁的成绩最高，谁的最低，不同的被访者之间的成绩差是多少。

(2) 没有绝对零点。这里的“0”表示一个数值，即“0”水平，而不表示“没有”和“不存在”。例如，一个被访者的统计学考试成绩为 0 分，并不表示他没有考试或者没有任何统计学知识，而是这次考试的成绩为 0 分。

(四) 定比尺度

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

在日常生活中，大多数情况下使用的都是定比尺度。例如，年龄、收入、某地区每年的失业人数、犯罪人数等。它是与定距尺度属于同一层次的一种计量尺度，但其功能要比定距尺度强一些，其主要特征体现为：

(1) 除了能区分类别、排序、比较大小、求出大小差异外，还可以计算两个测度值之间的比值。同定距数据一致，定比尺度得到的结果也表示为数值，所不同的是，它不仅可以进行定距尺度所能够进行的所有运算，而且在此基础上还增加了乘、除的数学运算功能。例如，在一次调查中，调查者记录了一组被访者的月收入数据，如果其中 A 被访者的收入是 2 000 元，B 被访者的收入是 8 000 元，我们不仅可以得出 A 和 B 在收入方面是有差异的，B 的收入高于 A，并且二者的差值为 6 000 元，同时还可以得出 B 的收入是 A 的 4 倍。

(2) 具有绝对零点，即“0”表示“没有”或“不存在”，是一个没有意义的数值。例如，被访者的收入如果为 0，那就表示该被访者没有收入；同样，某地区 2014 年 4 月份的失业人数为 0，也就表示了该地区在这一时期没有任何人失业。

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，但反之则不是。显然，研究者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计量尺度的计量结果。例如，将考试成绩由百分制计分转化为优、良、中、及格、不及格的五等级计分。表 1.1 给出了上述四种计量尺度数学特征的比较。

表 1.1　四种计量尺度数学特征的比较

数学特性 \ 计量尺度	定类尺度	定序尺度	定距尺度	定比尺度
分类(=，≠)	√	√	√	√
排序(<，>)		√	√	√
间距(+，−)			√	√
比值(×，÷)				√

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，所运用的统计分析方法也就越多，分析的内容越丰富，也越方便。因此，在实际工作中，应该尽可能地使用高层次的计量尺度来对事物进行测度。

二、数据的类型

(一) 分类数据、顺序数据、数值型数据

根据测度数据计量尺度的不同，可以将统计数据分为分类数据、顺序数据、数值型数据。

分类数据(categorical data)只能归于某一类别的非数字型数据，是由定类尺度计量形成的。

顺序数据(rank data)只能归于某一有序类别的非数字型数据，是由定序尺度计量形成的。

数值型数据(metric data)是按定距和定比计量尺度测量的具体值。

区分数据的类型是十分重要的，因为对于不同类型的数据，人们将采用不同的统计方法来处理和分析。例如，对分类数据通常计算出各组的频数或者频率，计算其众数(mode)和异众比率(variation ratio)，进行列联表分析(crosstab analysis)和 χ^2 检验等；对顺序数据，可以计算其中位数和四分位差，计算等级相关系数等；对数值型数据还可以用更多的统计方法进行处理，如计算各种统计量、进行参数估计和假设检验等。

(二) 定性数据和定量数据

根据数据所说明现象特征的不同，可以将统计数据分为定性数据和定量数据。

定性数据(nominal data)是说明现象品质特征表现的具体类别，通常用文字表示，其结果表现为类别，这类数据是由定类尺度和定序尺度计量形成的。定量数据(quantitative data)是说明现象数量特征的，通常用数值来表示，这类数据是由定距尺度和定比尺度计量形成的。

对于不同类型的数据，可以采用不同的统计方法来处理和分析。对于定性数据可以用数字代码来表示各类别或顺序，如用 1 表示男性，用 0 表示女性等，同时，采用分组法可以计算分析各组的频数或频率。对于定量数据则可用更多的统计方法去处理，现实中我们处理的大多数是定量数据。

这里需要特别指出的是，适用于低层次测量数据的统计方法，也适用于较高层次的测量数据。例如，对于分类数据和顺序数据可以计算众数和中位数，对于数值型数据也可以。反之，适用于高层次测量数据的统计方法，则不能用于较低层次的测量数据。例如，对于数值型数据可以计算均值，对于分类数据和顺序数据则不行。理解这一点，对于选择合适的统计分析方法是十分有用的。

（三）截面数据与时间序列数据

根据数据所表现的时空特征不同，可以将其分为截面数据与时间序列数据。

截面数据(cross-section data)是指在相同或近似相同的时点上采集的数据，用于描述现象在某一时点的变化情况。通常情况下，这类数据是在不同的空间上获得的。比如，2014年我国各地区城市居民人均可支配收入、农村居民人均纯收入都是截面数据。

时间序列数据(time series data)是指在不同时间上采集到的数据的集合，同描述现象随时间变化的情况。通常情况下，这类数据是按时间顺序搜集到的，其中常用的为年度、季度和月度数据等。比如，2010～2015年我国国内生产总值数据就是时间序列数据。

关于截面数据和时间序列数据的分析，将在后续的章节内容中进行专门介绍。

三、数值型数据的表现形式

数值型数据从表现形式上一般分为绝对数、相对数和平均数三大类。

（一）绝对数

绝对数(absolute number)是反映统计研究对象某一方面绝对数量的数据。其主要功能是用来描述研究对象的规模大小或水平高低，如人口数、财政收入、利润总额等。其数值的表现形式为以不同计量单位计量的绝对数。

绝对数常见的计量单位有：实物单位、货币单位和劳动单位。

1. 实物单位

实物单位是根据事物的自然属性特点确定的计量单位。以实物单位计量的绝对数称为实物量，它能够具体、形象地表现现象总体数量，但不易于进行综合、汇总。

2. 货币单位

货币单位即价值尺度。以货币单位计量的绝对数称为价值量。价值量可以弥补实物量的局限性，对不同总体可以进行综合与汇总，概括性强，但是它的抽象性明显，不宜从该数据中直接判断总体的实物形态。

3. 劳动单位

劳动单位是使用劳动时间表示的计量单位。以劳动单位计量的绝对数称为劳动量，如使用工日、工时计量的工作总量。劳动量在人力资源统计及生产周期较长的工业企业工作量统计中具有重要作用。

绝对数按其所反映的时间状况不同，可以分为时期数据和时点数据两类。

时期数据是反映研究对象在某一段时间内累计发生的数值总量，如全年社会商品零售总额、季度工业增加值、年新增人口数等都属于时期数据。时点数据是反映研究对象在某个时点上所表现的数值总量，如年初(末)人口数、月初(末)商品库存数、年初(末)固定资产占有额等都属于时点数据。

时期数据与时点数据具有不同的特点。具体表现为：①时期数据值的大小与其所属时期长度有直接关系，而时点数据值的大小与其所统计的时间间隔长短没有直接关系；②时期数据的前后各时期上的数值直接相加有实际意义，而时点数据前后各时点上的数值直接相加没有实际意义。正确区分时期数据和时点数据，对于进行时间序列数据分析和研究具有重要的意义。一般来说，时期数据应注意其所反映的时间长度，而时点数据则应注意其反映的时