

统计学基础

主 编 孔里明 陈 践

中国人民大学出版社

• 北京 •

图书在版编目(CIP)数据

统计学基础/孔里明等主编. —北京: 中国人民大学出版社, 2011. 8
ISBN 978-7-300-14155-8

I. ①统… II. ①孔… 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 (2011) 第 158591 号

统计学基础

主 编 孔里明 陈 践

出版发行	中国人民大学出版社		
社 址	北京中关村大街 31 号	邮政编码	100080
电 话	010-62511242 (总编室)		010-62511398 (质管部)
	010-82501766 (邮购部)		010-62514148 (门市部)
	010-62515195 (发行公司)		010-62515275 (盗版举报)
网 址	http://www.crup.com.cn		
	http://www.ttrnet.com(人大教研网)		
经 销	新华书店		
印 刷	山东高唐印刷有限责任公司		
规 格	185mm×260mm 16 开本	版 次	2011 年 8 月第 1 版
印 张	18.25	印 次	2011 年 8 月第 1 次印刷
字 数	409 000	定 价	30.00 元

前 言

统计学是有关如何收集、整理、分析客观现象数量特征，以便给出正确认识的方法论科学。随着人类社会的进步与发展，统计的应用领域越来越广泛，无论是社会经济还是自然科学，人们对数量分析和统计分析的应用越来越多，要求也越来越高。为了适应市场经济对统计信息和统计知识的需求，把握统计为认识社会规律服务的方向，我们编写了这本教材。

本教材是高等院校经济类、管理类专业的基础课教材，主要讲述了描述统计、推断统计的基本原理及应用方法。在编写过程中遵照我国高等教育发展战略提出的加强基础课程、注重能力和素质培养的要求，写法上完全立足于统计应用，力求通俗易懂，每种方法都从实际问题入手进行讨论，尽可能避免统计公式的推导。书中的统计方法均给出了利用 Excel 软件操作的详细步骤，使读者能轻松完成统计计算。

本教材可作为高等院校经济管理类专业统计学课程的教材使用，对广大实际工作者也极具参考价值。

本教材各章节编写分工如下：第一章：赵欢；第二章：田亚平；第三章、第四章：李文晶；第五章：王丹；第六章：谭文伟；第七章、第八章：马红燕；第九章：陈践。孔里明和陈践负责全书的编排和总纂。

本教材在编写过程中参阅了国内外有关著作和教材，并得到了中国人民大学出版社的大力支持和帮助，在此诚表谢意。由于编者水平所限，书中缺点和错误在所难免，恳请各方同仁批评指正。

编者

2011 年 8 月

目 录

第 1 章 总论	1
1.1 统计概述	1
1.2 统计学的几个基本概念	8
1.3 统计学的分类	14
第 2 章 数据收集与整理	19
2.1 数据的收集	19
2.2 数据的整理	27
第 3 章 统计数据描述	46
3.1 总量指标与相对指标	46
3.2 集中趋势的测定	53
3.3 离中趋势的测定	64
3.4 偏态与峰度的测定	72
第 4 章 时间序列	83
4.1 时间序列的分析指标	83
4.2 长期趋势的测定	95
4.3 季节变动的测定	105
4.4 循环变动的测定	108
第 5 章 指数	116
5.1 统计指数的概念和种类	116
5.2 综合指数	118
5.3 平均指数	124
5.4 指数因素分析	126
5.5 几种常见的经济指数	133
第 6 章 抽样与参数估计	147
6.1 抽样调查的基本问题	147
6.2 抽样误差	152
6.3 抽样分布	157
6.4 参数估计	159
6.5 抽样的组织形式	171
第 7 章 假设检验	187
7.1 假设检验的基本原理	187
7.2 总体均值的假设检验	192

7.3	总体比率的假设检验	201
7.4	正态总体方差的假设检验	205
7.5	假设检验中的其他问题	208
第 8 章	相关与回归分析.....	219
8.1	相关分析	219
8.2	一元线性回归分析	225
8.3	多元线性回归模型	236
第 9 章	方差分析.....	252
9.1	方差分析基础	252
9.2	单因素方差分析	256
9.3	两因素方差分析	262
附 录	280
参考文献	285

第 1 章 总 论

【学习目标】

通过本章的学习，应认识并了解统计的基本概念和特点，明确统计学的研究对象，掌握统计总体与总体单位、标志与指标、变异和变量等几组重要的概念，了解统计学的分类。

1.1 统计概述

1.1.1 统计的含义与发展历史

1. 统计的含义

统计的概念是人类在对事物数量的认识中形成的。英语中的“统计”为 Statistics，源于旧拉丁语 Status（状态），德语 Statistik（国家），新拉丁语 Statisticus（国事），意大利语 Statista（老练的政客），旧意大利语、拉丁语 Status（形势，政体）。汉语中的“统计”有“合计、总计”的意思。早在先秦时期，会、统不分，会计包括统计；在东汉时期称统计为通计；在南北朝时期称统计为总计；至清乾隆十二年（公元 1747 年），开始使用“统计”一词。

现代意义的“统计”一词有三方面含义：统计工作、统计资料和统计学。

(1) 统计工作，即统计实践活动。它是指运用科学的方法，按照预先设计的要求，收集、整理和分析客观事物总体数量方面资料的工作过程，是统计的基础，也是一种社会调查研究活动。一个完整的统计工作包括统计设计、统计调查、统计资料整理和统计分析等几个阶段。参加统计实践的工作人员称为统计工作者。领导、组织并从事统计工作的部门称为统计机构或统计部门。例如，各级统计部门统计所属地区的人口、工业、农业、商业及交通运输业等方面的数据资料，并将这些资料进行汇总、加工整理、分析运用等一系列活动过程，就是统计工作。

(2) 统计资料，也称统计信息。它是统计工作所取得的各项数字资料及有关文字资料的总称。统计资料一般反映在统计表、统计图、统计手册、统计年鉴、统计资料汇编和统计分析报告中。这些统计数字和统计分析资料是分析和研究社会经济问题不可缺少的重要依据。例如，国家统计局每隔一定时期向社会发布的国民经济和社会发展情况的统计公报、统计年鉴等，就是统计资料。

(3) 统计学，也称统计理论。它是研究如何收集、整理和分析统计资料的理论与方

法。统计学目前已经发展成一个涉及范围广泛、内容丰富多彩的学科体系，包括数理统计学、经济统计学、社会统计学和自然科学方面的统计等。

统计工作、统计资料、统计学三者之间存在着密切的联系。统计工作是基础，统计资料和统计学都是在统计实践的基础上产生和发展的。统计资料来源于统计工作，没有统计工作就没有统计资料；同时，统计资料又服务于统计工作，没有一定数量的、积累起来的统计资料，新的统计工作将难以做好。统计学是对统计实践活动的理论抽象和总结，理论来源于实践，但又反过来指导统计实践活动，使统计实践活动更科学、更有效，使取得的统计资料更符合客观实际，更具有使用价值。统计工作的不断发展，不但可以获得更加丰富多彩的统计资料，也会不断丰富统计学，促进统计理论的发展和完善。

统计理论、统计工作和统计资料之间的关系见图 1—1。

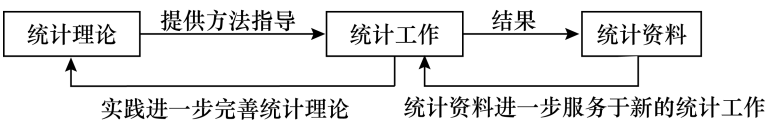


图 1—1 统计理论、统计工作和统计资料关系图

因此，“统计”一词的三种含义是相互联系的，不能将它们分割开来。统计工作与统计学的不断发展和不断丰富过程，也是统计不断实践、认识、再实践、再认识的过程。

本书介绍的统计学基础，仅研究社会经济现象方面，主要阐述的是社会经济统计的基本原理、原则和方法。

2. 统计的产生与发展

(1) 统计实践活动的产生与发展。

统计是为了适应国家管理需要和社会政治经济发展而产生并发展起来的。统计活动的起源很早，在古代奴隶社会，由于国家在赋税、徭役、征兵等方面的需要，就开始了人口、土地等基本国情的登记和计算工作。据历史记载，我国在夏禹时代（公元前 22 世纪）就开始进行人口统计，《书经·禹贡篇》记述了九州的基本状况，人口 1 355 万人。被西方经济学家推崇为“统计学最早的萌芽”。古埃及、古希腊、古罗马的历史中也有类似的记载。古埃及在公元前 3000 年已经有人口、居民财产统计；据说古希腊公元前 600 年就进行过人口普查。古罗马在公元前 400 年建立了人口普查和经常性人口出生、死亡登记制度。这些就是原始形态的统计。

进入封建社会以后，随着人类社会生产的发展，统计的范围逐渐由人口、土地发展到社会经济生活的各个方面。封建君主和政治家日益意识到统计对于治国的重要作用。在我国战国时期，商鞅提出要使国家强盛，必须知道“境内仓、口（府）之数，壮男、壮女之数……”等所谓的十三数，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。秦汉时期，已有地方田亩和户口记录；唐末则有计口授田、田亩鱼鳞册等土地调查和计算；到明清时期，已建立了经常的人口登记和保甲制度等。但由于社会生产力发展仍然很缓慢，所以统计实践仅限于简单的调查登记和汇总计数工作。

长期的封建生产关系阻碍了社会生产力的发展，相应地也阻碍了统计实践的发展。

统计实践的广泛发展始于资本主义社会。17 世纪下半叶，西方进行了工业革命，进入资本主义社会，社会生产力迅速发展，社会分工协作更加精细与复杂，交通、航运、贸易日趋发达，统计实践从国家管理领域扩展到社会经济活动的诸多领域。从 18 世纪起，各资本主义国家都先后设立专业的统计机关，收集各方面统计资料，定期或不定期举行人口、工业、农业、贸易、交通等项调查，出版统计刊物，建立国际统计组织，召开国际统计会议。正是在这样的历史条件下，统计学应运而生。

(2) 统计学的产生与发展。

17 世纪以后，随着统计实践的发展，客观上要求总结丰富的实践经验，使之上升为理论，并进一步指导实践。由于历史条件、研究领域不同，统计学产生了不同的学派。主要有：

①政治算术学派。这一学派产生于 17 世纪中叶的英国，其代表人物是英国学者威廉·配第 (William Petty, 1623—1687)，在他所著的《政治算术》一书中，对当时的英国、法国、荷兰三国的国情国力做了系统的数量对比分析，明确英国的国际地位并不悲观，提出了英国的社会经济发展和方向的道路。尤其重要的是，他对人口、资本、土地、产业的真实情况的表述方法非常独特，即用各国的实际数据资料，用数字、重量、尺度来说话，还提出了用图表形式概括数字资料的理论和方法，这些方法为统计学的产生奠定了良好的基础。威廉·配第做了前人没有做过的从数量方面来研究社会经济现象的工作。正是在这个意义上，马克思称威廉·配第是“政治经济学之父，在某种程度上也可以说是统计学的创始人”。

配第的朋友约翰·格朗特 (John Graunt, 1620—1674) 通过对伦敦 50 多年的人口出生和死亡资料的计算，写出了第一本关于人口统计的著作《对死亡表的自然观察和政治观察》(1662 年)，书中所用的具体数量对比分析方法，对统计学的创立同样起到了极其重要的作用，为统计学作为一种从数量方面认识事物的科学方法开辟了广阔的发展前景。

②国势学派。主要创始人是德国的康令 (H. Conring, 1606—1681) 和阿亨瓦尔 (G. Achenwall, 1719—1772)。国势学主要采用文字记述的形式，把国家重要事项系统地整理并罗列出来，是统计学最早的流派之一，它也叫记述学派。阿亨瓦尔的代表著作是《近代欧洲各国国势学概论》，该书通过研究“国家显著事项”，分析各国的政治经济情况，提出一些治国方略。阿亨瓦尔在大学中开设了一门新课程叫做“国势学”，后人把从事这方面研究的德国学者称为国势学派。由于德文中的“国势”与“统计”一词词源相通，阿亨瓦尔将“统计”一词当作学名来使用。严格地说，这一学派的研究对象和研究方法都不符合统计学的要求，缺乏数量分析的方法与结论，只是登记了一些记述性材料，因此人们又将它称为“有名无实”的统计学派。

政治算术学派和国势学派都以社会经济现象作为研究对象，以社会调查作为研究基础。但政治算术学派注重用数字说话，进行定量分析；而记述学派注重文字表达，进行定性分析。在是否把数量方面的研究作为这门学科的基本特征方面，两个学派互相争论了 200 多年，直到 19 世纪中叶，德国的克尼斯 (K. G. A. Knies) 于 1850 年发表了《独立科学的统计学》论文，提出“国家论”和“统计学”的科学分工，主张把“国家论”命名为“国势学”，把“政治算术”正名为“统计学”，两个学派的争论才告结束。

③数理统计学派。产生于 19 世纪中叶，其创始人为比利时科学家凯特勒 (L. A. J. Quetelet, 1796—1874)，他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，其著作有《论人类》、《统计学的研究》、《关于概率论的书信》和《社会物理学》等。他最先将概率论应用于人口、人体测量和犯罪等问题的研究，完成了统计学和概率论的结合。从此，统计学开始进入更为丰富发展的新阶段，许多学者从各个角度研究统计学，不断增加新内容，相继提出和发展了相关和回归理论、 t 分布以及抽样理论等，使数理统计学很快发展成为一门比较系统、完善的学科。英国学者葛尔登 (F. Galton, 1822—1921) 提出了生物统计学，皮尔逊 (K. Pearson, 1857—1936) 将生物统计一般化而发展为描述统计，爱奇渥斯 (F. Y. Edgeworth, 1845—1926)、鲍莱 (A. L. 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 (R. A. Fisher, 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。国际统计学界称凯特勒为“近代统计学之父”，就在于他发现了大量现象的统计规律和开创性地应用了许多统计方法，促使统计学向新的境界发展。由于这一学派主要在英美等国发展起来，故又称英美数理统计学派。

统计发展史表明，统计学是从设置指标研究社会经济现象的数量开始的。随着社会的发展与实践的需要，统计学家对统计方法的不断丰富和完善，统计学也不断发展和演变。从当前世界各国统计研究状况来看，统计学已不仅为研究社会经济现象的数量方面，也为研究自然技术现象的数量方面提供各种统计方法，它既研究确定现象的数量方面，又研究随机现象的数量方面。统计学是一门研究事物数量方面特征与规律的方法论学科。

(3) 我国统计发展简况。

新中国成立前，由于我国是半殖民地半封建社会，统计工作非常落后，统计学基本上照抄照搬西方统计理论，传播的主要是数理统计学派的观点。

新中国成立后，我国全盘接受苏联的统计学即社会经济统计学和统计组织体制，在高度集中的计划经济体制下发挥了重要作用。但同时受苏联影响，数理统计遭到批判，统计学发展缓慢。进入 20 世纪 80 年代，随着中国经济体制向社会主义市场经济转轨，统计也进入了全面改革的现代化新时期。人们突破了以往狭隘的观点，承认社会经济统计学、数理统计学和自然科技方面的统计学都是独立的统计学科，它们可以同时并存，相互借鉴，共同发展。

近年来，社会经济统计学和数理统计学出现了融合的趋势，数理统计方法在社会经济统计中得到了广泛的应用。今天，统计学已划入国家一级学科，随着大统计学学科体系的建立，统计学作为一门独立的科学，其运用已渗透自然科学和社会科学的各个领域。

1.1.2 统计学的研究对象

统计学的性质和研究对象问题，是我国统计理论界长期以来争论较多、分歧较大的问题，我们认为“统计学是一门方法论科学”。

1. 统计研究的对象

统计是从量的方面对社会经济现象进行观察研究的，即统计的认识对象是社会经济

现象的数量方面。虽然统计是研究社会经济现象的数量方面，但它对现象数量方面的研究并不是孤立进行的，而是在质与量的相互联系中研究量的，如果离开了事物的方面，为研究量而研究量，那就不是统计学了。统计研究事物数量方面的目的，在于通过对事物量的方面的观察和量变规律的研究，逐步把握事物的质和对事物的方面的认识。因此，统计对社会经济现象数量方面的认识包括量的规模、现象之间的数量联系、现象数量的变化规律、现象质与量互变的数量界限等，而对事物量的这些方面的研究，都不是仅对个别事物观察所能得到的，必须通过对现象的大规模研究才能有效，因此，统计的研究对象具有如下的特点：

(1) 总体性。

统计对社会经济现象的研究要求具有总体性，是基于满足统计研究的目的来考虑的。但强调总体性的要求，并不排斥统计对社会经济个体现象观察的重视。事实上，统计对总体事物的研究是从对个体的观察开始的。例如，要研究某城市居民的消费水平，目的不在于了解个别居民的消费状况，而是要通过对很多个别居民消费状况的了解，达到对全市居民总体消费水平的认识。

(2) 社会性。

所谓社会性，是指统计研究的对象是社会经济现象的数量表现，具有社会性。统计对象的社会性可以从三个方面进行考察：一是统计的认识对象是社会经济现象的数量方面，因而统计本身也就有了社会性；二是统计认识的主体是社会的人，人的阶级性（社会性）决定了认识立场和认识结论上的社会性；三是一切社会经济活动都和人的利益有关，不同的人群有着不同的利益和利益关系，因此人们相互间的利益分割和利益冲突，必将在统计数据上显示出来。例如，有些集团为了谋求自身利益就可以虚报、瞒报、不报、拒报、篡改、造假统计数据，不能如实反映实际情况。因此，我们说，统计具有社会性，防止对统计数据的人为干扰，是做好统计工作的重要保证。

(3) 数量性。

从统计发展的历史来看，统计的首要特点是其数量性。统计学研究对象的数量性，具体说来，就是通过各种统计指标和指标体系来反映对象总体的规模、水平、速度、比例、效益和趋势等。例如，我国的人口数量构成及其发展趋势、国内生产总值的总量构成、企业内职工工资和劳动生产率增长速度的比例关系，等等。数量性包含三个方面的内容：数量的多少、现象间的数量关系、质与量间的关系。一切客观事物都有质和量两个方面，事物的质与量总是密切联系、共同规定着事物的性质。一定的质规定着一定的量，一定的量也表现为一定的质。但在认识的角度上，质和量是可以区分的，可以在一定的质的情况下，单独地研究数量方面，通过认识事物的量进而认识事物的质。因此，事物的数量是我们认识客观现实的重要方面，通过分析研究统计数据资料，研究和掌握统计规律性，就可以达到我们统计分析研究的目的。

(4) 客观性。

客观性又称为具体性。

统计学所研究的量不是抽象的量，它是现象总体数量特征的客观反映，是现象总体数量特征在一定的时间、地点、条件下的具体反映，即具体数量表现不是主观意志所能

转移的，这是统计与数学的一个重要区别。例如，2010 年我国 GDP 现价总量为 397 983 亿元，全国总人口 134 100 万人，人均 GDP 4 283 美元。数学研究客观世界的空间形式和数量关系时，具有高度的抽象性，它可以撇开所研究客体的具体内容；而统计在研究社会经济现象的数量时，就必须紧密联系被研究现象的具体内容，联系其质的特征。

2. 统计学的研究对象

如前所述，统计学是统计实践活动的理论概括和总结，并反过来指导统计实践活动，因此，统计学的研究对象可以表述为：社会经济总体现象的数量特征及其规律性、统计认识活动过程本身和认识方法。

(1) 统计学的研究对象是社会经济现象的数量方面。

任何事物都有质与量的两个方面，社会经济现象也是如此。统计学正是从数量上来研究社会经济现象，反映社会经济现象的规模、水平、速度、比例关系等各种数量关系，揭示社会经济现象的变化规律。对社会经济现象数量方面的研究涉及社会经济生活各个方面，只有对这些现象的数量进行分析研究，才能获得具体认识，才能掌握其变化规律，才能进行有效地监督和控制。

(2) 统计学是在质与量的联系中研究数量特征和数量关系。

统计学不是抽象地研究纯数量，而是联系一定的质去研究事物的量。一方面，任何量都依存于一定的质，离开质就无所谓量，也无从说核算量。另一方面，任何质都表现为一定的量，只有在研究事物的量的基础上，才能具体深刻地认识事物的本质。统计学在研究社会经济现象数量方面时，既要通过研究现象的数量方面来认识它的本质，又绝不能离开事物的本质抽象地去研究它们的量。统计学所研究的量，是具体事物在一定时间、空间条件下的具体数量表现，它总是和现象的质密切相关。统计学研究事物的量的目的，正是为了深刻地反映现象的性质和内在联系，揭示社会经济现象的变化趋势或规律性。

(3) 统计学研究的是总体现象的数量方面。

统计学研究社会经济现象的目的在于认识社会经济现象总体的规律性。个别现象由于受偶然因素的影响，其数量表现千差万别，只有通过大量观察，充分提取材料，才能真正认识和把握事物的本质特征。统计学正是从个别现象的偶然性去认识总体的必然性，因此，统计学研究的是总体现象的数量方面。

1.1.3 统计学的作用

统计是社会调查研究方法中应用广泛的重要方法之一。列宁曾精辟地指出，统计是认识社会最有力的武器之一。统计之所以能够起到认识社会的作用，是因为统计能够从全部事实的总和中，从事物的内部联系中把握实质，也就是说它能够从数量方面来说明客观事实的现象和发展过程，能够深刻有力地揭示客观事实的内在本质和规律。统计的作用可以归纳为服务和监督两大方面。服务就是提供各项数据资料和分析报告等，统计服务是统计工作永恒的主题；监督是通过对社会经济活动运行过程的检查、分析，达到对社会现象运行态势所进行的实时监控。具体作用体现在以下五个方面。

1. 从宏观上看，统计是国家宏观调控和监督管理的依据

宏观调控是政府的主要职能，政府通过适当的财政政策和货币政策调控国民经济的

运行。政府根据宏观经济的运行状况、社会投资金额的增长速度和货币投放量来调控政策；央行根据房产贷款额的增长量、房产价格和居民可支配收入，通过利率调控房地产业。在制定各种宏观调控政策时，必须有微观经济运行情况的基本统计资料，社会经济统计提供的各种数据可以作为制定宏观调控政策的各种依据。因此，我们说统计是国家宏观调控和监督管理的一项重要基础工作。统计信息是国家编制各类规划的重要依据，只有这样，才能使宏观调控工作建立在科学可靠的基础之上。

2. 从微观上看，统计是企业管理与决策的重要工作

社会发展需要统计，企业经营管理更需要统计。统计是企业管理中的一个重要组成部分，是生产经营决策的依据。统计既是一种认识工具，同时也是一种管理手段。运用统计特有的方法，通过收集、整理，提供企业内部生产、经营活动的基本数据资料，以及企业生产经营相关的社会经济信息资料，为企业的预测工作提供依据；通过统计分析、预测，为企业领导提供各种决策备选方案，有助于企业科学决策；统计信息作为企业信息的主体，还可以通过分析，对企业生产、经营活动过程起监督作用，促进企业经营活动有序、高效运行。

作为企业管理者，只有全面系统地掌握统计知识，学会运用统计方法，才能够及时做出判断、正确分析经济发展变化所表现出来的数量特征以及数据资料所蕴含的经济运行特点和趋势，从而提高依据统计信息进行理性分析、依据统计方法做出科学决策的能力。

3. 日常生活中，统计是为社会提供咨询服务的主要方式

统计信息是各种信息的中心，是最重要的一种信息。统计信息作为社会经济信息的主体，具有信息资源的优势，统计方法作为社会的特有方法，具有特殊重要的作用。统计利用所掌握的丰富的信息资源，运用统计特有的科学方法和技术手段，深入开展综合分析和各项专题研究，为政府、企业、社会公众提供所关心的经济、社会问题数据分析和调研咨询报告。专门的统计调查咨询机构可以接受社会团体、企业集团委托，开展专项社会调查和商品市场信息调研分析，为企业经营决策提供更加完备的统计资料、建议和咨询报告，指导企业更加理性地发展。

4. 科学研究中，统计是进行科学研究的重要方法

科学研究需要掌握大量的背景资料和有关信息，背景资料是进行科学研究的基础，是比较研究的依据。更重要的是，科学研究需要收集第一手的研究资料，为此，研究者必须进行广泛的调查。社会经济统计的调查方法可以为研究者提供问卷设计的方法、统计调查的方法，便于研究者收集原始资料。社会经济统计不仅可以提供信息资料的收集、整理和分析的原理与方法，还可以为科学研究提供必要的数字信息，分析涉及的数量关系及其发展变化，得出有说服力的科学研究成果。

5. 在国际交流中，统计信息是沟通的桥梁

当今世界的经济是开放的经济，互联网技术将世界变成了一个地球村。国际交流是社会经济发展的必要条件，而政治、经济、文化、教育、劳务、投资、科学技术等统计信息的交流是国际交流的主要内容。社会经济统计可以提供多方面的统计信息，对开展对外合作，进行国际交流，有十分重要的作用。

1.2 统计学的几个基本概念

在统计实践中，会经常涉及一些概念，它们是统计学中最常用的基本概念，其核心是围绕统计指标展开的。

1.2.1 统计总体、总体单位

1. 统计总体 (Statistical Population)

(1) 统计总体的含义。

由许多性质相同的个别事物所构成的整体，叫做统计总体，简称总体。例如，欲研究某市商业企业的基本情况，那么由该市所有商业企业组成的全体就是一个统计总体。统计总体具有同质性、大量性和差异性三个特征。

①同质性 (Homogeneity)。即构成总体的各单位必须在某一方面性质相同，这是组成总体的根本条件。如果把不同性质的个别单位结合在一起，它表现的总体特征是模糊不清的。例如，在研究商业银行这一总体中，每一个银行都必须具有“商业”这一共同性质。凡不具备“商业”这一性质的银行都不能进入这一总体中。

同样，总体的同质性也是相对的。在不同的研究条件下，对总体的同质性有不同的规定性。例如，研究某市国有企业基本情况时，总体是由全市国有企业组成的总体，这时同质性只要求把具有“国有企业”特征的单位结合起来组成总体，而不强调是工业企业，还是商业企业，或是其他企业。

②大量性 (Mass)。即总体应该由为数众多的单位构成，仅仅由个别单位或为数极少的单位结合不足以构成总体。大量性是组成总体的基本前提。总体必须具备大量性，是因为构成总体的个别单位的数量表现是多种多样的，只对少数单位进行观察，其结果难以反映总体的本质特征。总体的本质特征只有对众多的个别单位的数量进行综合时才能反映出来。以某市民办院校组成的总体为例，每一民办院校是个别单位，那么全市所有的民办院校组成的全体就具有大量性，把全市每一个民办院校的学生结构进行综合分析，就能说明该市民办院校这一总体的学生结构特征。

③差异性 (Diversity)。或称变异性，即构成总体的各单位除了某一方面或几方面性质相同以外，必须在其他方面存在着差别，这是统计的主要内容之一。例如，某领域的职工总体中各单位间有男女的性别属性差异，有 20 岁、21 岁、22 岁、23 岁、24 岁、25 岁、26 岁等年龄标志数值的差异。这种差异是普遍存在的，所以才对大量的个别单位进行调查和统计。如果总体中每一个别单位在所有方面表现共同性，就没有必要对个别单位进行统计调查，而只研究其中某一个别单位就可以说明总体特征了，实际上这种情况几乎不可能出现。因此可以说差异性构成了总体研究的主要内容。

组成总体的三个特征是密切相关的。同质性是组成总体的根本条件，大量性是组成总体的基本前提，差异性总体研究的主要内容。

(2) 统计总体的分类。

按总体单位是否有限，可划分为有限总体和无限总体。一个统计总体所包括的单位

数如果是有限的，称为有限总体。社会经济统计所研究的大多是有限总体。像某市商业企业、全国人口等一些总体，不论总体包括的个别单位如何多，但其数量总是有限的。一个统计总体所包括的单位数如果是无限的，称为无限总体。例如，大量连续生产的零件组成的总体，只要这种零件连续生产下去，这一总体就会无限扩大下去，可以成为无限总体。对有限总体的调查方式可以是全面调查，也可以调查其中一部分单位；而无限总体只能进行非全面调查，据以推断总体。

按单位标志的属性，可分为变量总体和属性总体。凡是反映品质标志单位组成的总体称为属性总体；凡是反映数量标志组成的总体称为变量总体。

2. 总体单位 (Population Unit)

构成总体的每一个别单位，被称为总体单位。原始资料最初就是从总体单位取得的，总体单位是各项调查项目的直接承担者。了解某市商业银行的基本情况，首先要做的是对每个商业银行的属性和数量加以登记；其次才是经过汇总和综合来说明该市商业银行这一总体的基本情况。统计上所讲的单位同人们日常生活中所说的单位并不完全相同，它可以是一个人，一家企业，一所学校，一件产品，等等。我们说“全国高等学校”是一个总体，那么，全国的每一所高校都是总体的一个总体单位，这时的“单位”同人们日常生活中所说的单位，其意义是相同的。但如果把“全国总人口”作为一个总体，那么，每一个具有中华人民共和国国籍的公民都是其中的总体单位，这时的“单位”同人们平常所说的单位就不一样。

3. 总体和总体单位之间的关系

总体和总体单位之间体现着全体和个体、整体和局部的关系。总体和总体单位的确定，要视研究问题的需要和研究范围的大小而定。统计总体和总体单位是多种多样的。而且统计总体与总体单位不是固定不变的，总体与总体单位具有相对性，随着研究任务的改变而改变。这与研究目的和要求有关。例如，要了解某一地区国有工业企业的生产经营情况（研究目的），总体是该地区的全部国有工业企业，该地区每一个国有工业企业是总体单位。而要了解全国各个地区国有工业企业的生产经营情况，总体则是全国各个地区全部国有工业企业，而各个地区的全部国有工业企业为总体单位。

确定总体与总体单位，必须注意两个方面：第一，构成总体的单位必须是同质的，不能把不同质的单位混在总体之中。例如，研究工人的工资水平，就只能将靠工资收入的工人列入统计总体的范围。同时，也只能对工人的工资收入进行考察，对工人由其他方面取得的收入就要加以排除，这样才能正确反映工人的工资水平。第二，总体与总体单位具有相对性，随着研究任务的改变而改变。同一单位可以是总体，也可以是总体单位。例如，要了解全国工业企业职工的工资收入情况，那么全国工业企业职工是总体，每位职工是总体单位。如果旨在了解某个企业职工的工资收入情况，则该企业的所有职工就成了总体，每位职工就是总体单位了。

1.2.2 标志和指标

1. 标志及标志表现

标志是说明总体单位特征的名称或概念，有数量标志和品质标志之分。说明总体单位数量特征的名称或概念，叫做数量标志；说明总体单位属性特征的名称或概念，叫做

品质标志。品质标志表现的是具体属性，只用文字而不用数字描述。数量标志表现为具体的数值。例如人口总体中的每一个人（单位）都有年龄、性别、身高、体重、民族、受教育程度等方面的特征，这些特征使得某个人同其他人区别开来。这里，年龄、身高、体重等是数量标志，而性别、民族和受教育程度等是品质标志。各种标志都有自己的标志表现，如在人口研究中某人的年龄 30 岁、体重 56 公斤，等等，这里的 30 和 56 都是数量标志的表现，也叫标志值；品质标志表现为对特征加以描述的文字，比如民族“汉”，性别“男”，等等。

需要指出的是，标志和标志表现是两个不同含义的概念。标志是从不同的具体属性和具体标志值中抽象出来的名称，用来说明总体单位之间的变异情况。标志表现则是标志的具体内容。品质标志的表现是具体属性，数量标志的表现是具体标志值。实际上，统计调查中拟定的调查项目就是统计标志。例如，我们经常填写的信息表中，姓名、籍贯、民族、文化程度等项目，就是品质标志；所填写的具体情况，如张三、陕西、汉族、大学等就是品质标志表现，即属性。而其他项目如年龄、工龄、工资额等项目，就是数量标志，所填写的 30 岁、10 年、200 元等就是数量标志表现，即标志值或称变量值。

2. 统计指标

（1）统计指标的概念。

统计指标是用来说明总体特征的数量概念或名称。如人口数、土地面积、总产值、销售额等概念。

（2）统计指标的构成要素。

一个完整的统计指标应包括指标名称、指标数值、指标计量单位、指标所属的时间和空间范围及计算方法等要素。但人们在理论上使用的统计指标通常只是一个指标名称，如国内生产总值、耕地面积、居民储蓄额、人口密度等。

统计指标就其完成形态而言，由以下要素构成：

①定性范围，包括指标名称和指标含义。指标含义要明确总现象的质的规定性，包括时间标准和空间标准。例如，GDP，其指标含义是国内生产总值，是指在一定时期内（一个季度或一年），一个国家或地区的经济中所生产出的全部最终产品和劳务的价值，常被公认为衡量国家经济状况的最佳指标。指标含义比较复杂，而指标名称是它的表现形式。

②定量方法，包括计量单位和计量方法，是指标含义的量化规范。例如，总人口的计量单位是“人”，计算方法是全国各地区人口加上现役军人的人口总数。

③指标数值，即按定性范围、定量方法经过调查取得的具体数值，如 2010 年末（即 12 月 31 日 24 时）我国 GDP 达到 397 983 亿元。

（3）统计指标的类别。

统计指标按其表现形式的不同，可分为绝对数指标、相对数指标和平均数指标。绝对数指标反映现象的总规模和总水平，又被称为总量指标，如人口总数、总产出、工资总额、增减量等。相对数指标反映现象之间的对比关系，如发展速度、经济比例等。平均数指标反映现象总体内部各单位的一般水平，也可以反映发展的平均水平和平均速

度,如粮食平均亩产量、平均年龄、平均发展速度等。

按反映现象性质的不同,指标可以分为数量指标和质量指标。数量指标反映现象的绝对量的多少,如工业总产值、利税总额等。质量指标反映现象间的数量关系,如各种相对数和平均数等。这个划分是我们计算综合指数的基础。

按反映现象时间的不同,指标可以分为时期指标和时点指标。时期指标反映现象在一定时空条件下所达到的总规模或总水平,具有可加性、连续登记和指标数值与时间长短有关等特点,如企业的产品产量、销售收入等。时点指标指某种现象在某一时刻所达到的状态。如企业的库存余额、职工人数及银行存款余额等。

按反映现象内容的不同,指标可以分为总体单位总量指标和总体标志总量指标。例如,要计算某企业职工的平均工资,则职工人数为总体单位总量指标,工资总额为总体标志总量指标。

按反映现象范围的不同,指标分为总体指标和样本指标。总体指标诸如总体总量、总体平均数、总体成数、总体方差。样本指标反映样本的数量特征,如样本平均数、样本成数、样本标准差等。

按反映现象的状况不同,分为静态指标和动态指标。静态指标反映既定时间上现象的规模、水平、数量关系等,如一般平均数、绝对数和许多相对数等。动态指标反映现象在不同时间内发展变化的情况,如序时平均数、发展速度、增长速度等。每一类别的统计指标都有更详细的分类研究,这将在本书的不同章节分别予以阐述。

(4) 统计指标的特点。

统计指标具有如下特点:

①同质事物的可量性。没有质的规定性不能成为统计指标,有了质的规定性不能用数量表示,也不能成为统计指标。有些抽象度较高的社会经济概念难以量化,是不能直接用来作为指标名称的,必须将其分解为可以量化的概念才能成为统计指标,因此统计指标是可量的。

②量的综合性。统计指标反映的是总体的量,它是许多个体现象的数量综合的结果。如果某一数量化概念用来反映个体或总体单位特征,那就不能成为统计指标。因此,总体特征的数值是个体数值综合结果。总体性和综合性的含义是一致的,总体特征需要对个体数值综合,也只有对个体数值综合才能说明总体特征。

③具体性。统计指标的数值是具体时间、具体地点限定的数值,它不同于抽象的数学数值,它是社会经济现象的具体数量反映。

④客观性。统计指标的数值既是具体的,也是客观世界的写照。它不能杜撰编造,也不能歪曲夸大。统计指标的数值必须真实准确,及时反映客观世界。它不同于计划指标,计划指标用来说明未来要达到的预期目标,同最终发生的状况不一致是允许的,也是难以避免的。

(5) 统计指标的作用。

统计指标与统计工作有着直接紧密的连带关系。统计指标,就是“统计”一词的第二种含义——统计资料。统计工作的成果就是统计指标。统计指标是统计工作的目的,统计工作是获取统计指标的必要实践活动,可见统计指标在统计活动中的重要作用。各

种独特的作用总括起来可以归纳为以下两点：

①从认识的角度讲，它是记录社会经济现象变化发展情况的工具，同时，又是反映社会经济现象数量规律的手段。这一作用如同机器的仪表一样，把机器的运转状态表现在仪表中，供操作人员观察。

②从社会管理和科学研究的角度讲，它提供以数量表现的事实，是进行社会管理和科学研究的基本依据。无论是在制定政策和计划时，或者是企业经营活动进行调节和管理时，或是经济学家进行经济理论研究和探索时，都要从事实出发，以客观数量为依据，因此，统计指标是管理活动和科学研究的必要工具。

3. 统计指标和标志的关系

指标和标志是一对相关的概念，它们有区别也有联系。

(1) 统计指标和标志的区别。

①指标和标志的概念不同。标志是说明总体单位属性的，一般不具有综合的特征。指标是说明总体的综合数量特征的，具有综合的性质。

②划分种类不同。统计指标分为数量指标和质量指标，它们都是可以用数量来表示的。标志分为数量标志和品质标志，它们不是都可以用数量来表示的，品质标志只能用文字表示。

(2) 统计指标和标志的联系。

①统计指标和数量标志都是数量化概念，是同一的，只有在说明不同对象时才有区别。如学生人数，既可以是统计指标，也可以是数量标志，具体的划分要根据学生人数是要表现总体特征还是要表现总体单位特征而定。

②统计指标数值是总体的数量特征，数量标志的标志值是总体单位的数量特征。统计指标数值是由各单位的标志值汇总或计算得来的。数量标志可以综合为数量指标和质量指标，品质标志只有对它的标志表现所对应的单位加以总计才能形成统计指标。总体单位的某一标志往往是总体某一统计指标的名称。

③随研究目的不同，指标与标志之间可以互相转化，二者体现这样的关系：指标在标志的基础上形成，同时指标又是确定标志的依据。如果原来总体转变成总体单位了，那么原来说明总体的统计指标也就变成总体单位的标志了。

单个统计指标只能反映总体某一方面的数量特征，说明现象某一侧面的情况。客观现象是错综复杂的，要反映其全貌，描述现象发展的全过程，只靠单个指标是不行的，需要设立统计指标体系。

统计指标体系 (Statistical Indicator System) 就是各种相互联系的统计指标所构成的一个有机整体，用来说明所研究现象各个方面相互依存和相互制约的关系。例如，工资总额=平均工资×职工人数，商品销售额=商品销售量×商品销售价格，等等。

统计指标体系因各种现象本身联系的多样性和统计研究的目的不同而分为不同的类别。根据所反映现象的范围内容不同，统计指标体系可以分为综合性统计指标体系和专题性统计指标体系。

综合性统计指标体系是较全面地反映总系统及其各个子系统的综合情况的统计指标体系，如国民经济和社会发展统计指标体系，专题性统计指标体系则是反映某一个方面

或问题的统计指标体系，如经济效益指标体系就是专题性统计指标体系。

1.2.3 变量

在统计研究中，变量（Variables）是一个非常重要的概念。一般来说，变量是现象发展变化的数量化概念，或者说是现象本身所固有的、随条件变化而变化的量。我们前面讲过的所有的统计指标和所有的数量标志，都是变量。变量的具体数值表现就是变量值，比如 2009 年末（即 12 月 31 日 24 时）全国总人口 1 369 202 232 人。“人口数”就是变量，“人口数”的变量值是 1 369 202 232 人。变量值的具体差异叫做变异。社会经济统计学一方面归属经济学科门类，当研究的内容属于经济理论范畴时，就给出数量标志以便与品质标志相对应；另一方面又和一定的数学方法相联系，当研究的内容和数学公式有关的时候，就用变量这一数学范畴命名。

1. 自变量和因变量（Independent Variables and Dependent Variables）

自身变化会引起其他变量变化的量，叫自变量；受其他变量影响而变化的量，叫因变量。比如分析影响 GDP 的因素，我们可以考虑利率、通货膨胀率等，这里的利率、通货膨胀率就是自变量，GDP 就是因变量。这类概念多在相关和回归分析中使用。

2. 确定性变量和随机性变量（Certainty Variables and Random Variables）

确定性变量是指影响变量值变化的因素是确定、可控的变量。例如，圆的面积随半径的长短而变化，变化关系是确定的，因此圆的面积是确定性变量。随机性变量是指变量值的变化受某种或某几种不确定性因素的影响，其变化不是沿着一定的方向发展，而是带有很大的偶然性。例如，按随机原则从总体中选取容量一定的样本，每一次都会得到不同的结果，因此，样本指标是个随机变量。又如，测量同一个螺母的内径（尺寸），不同的人可能得到不同的结果，这个“内径尺寸”就是随机变量，等等。随机变量在抽样理论、数理统计中经常使用。

3. 连续变量和离散变量（Continuous Variables and Discrete Variables）

变量的连续性或离散性，是以变量值是否可以无限分割为标准的。凡是一个变量相邻的两个变量值之间可以继续分割，取得新的变量值，那么，这样的变量称为连续变量。比如道路的长度、耕地面积、人的平均身高、劳动生产率、粮食总产量，等等，都属于连续变量，它们通常需要用计算或测量的方法取得变量值。凡一个变量相邻的两个变量值之间不可能再分割出新的变量值，即变量取值只能取整数的自然数，这样的变量称为离散变量，如人数、企业数、产品件数，等等。离散变量通常以点数的方法取得变量值。这两种概念多在变量数列分析中使用。

4. 其他变量类型

分类变量（Categorical Variables）的变量值是定性的，表现为互不相容的类别或属性。分类变量可分为无序分类变量和有序分类变量两类。无序分类变量（Unordered Categorical Variable）是指所分类别或属性之间无程度和顺序的差别。对于有序分类变量（Ordinal Variable），应先按等级顺序分组，清点各组的观察单位个数，编制有序变量（各等级）的频数表，所得资料称为等级资料。例如，产品的质量可以分为特等品、一等品、二等品、三等品、等外品、次品、废品，就属于有序变量。变量类型不是一成不变的，根据研究的目的，各类变量之间可以进行转化。

1.3 统计学的分类

1.3.1 描述统计学与推断统计学

统计学研究如何取得反映客观现象的数据，并通过图表形式对所收集的数据进行加工处理和显示，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一门学科。另外，观察者以数据的形态建立一个用以解释其随机性和不确定性的数学模型，以之来推断研究中的步骤及母体，这种用法被称作推断统计学。这两种用法都可以被称作为应用统计学。另外也有一个叫做数理统计学的学科，专门用来讨论这门学科背后的理论基础。

统计学可分为描述统计学和推断统计学。

1. 描述统计学

描述统计学（Descriptive Statistics）产生于 20 世纪 20 年代。生物学家达尔文在研究生物的遗传变异等时，采集了生物、地质、地理等方面的大量标本和化石，在收集、整理和权衡事实的过程中，借助了简单的统计方法，引起了很多人的兴趣。此后，一些生物学家兼科学家将生物进化和统计研究相结合，使生命科学的研究从定性分析迈向定量分析，开创了生物统计学派。

从生物统计学派生发展而来的描述统计学以大样本近似分布为基础，重点研究数据的收集整理和对数据进行直观的描述。其内容包括统计数据的收集方法、数据的加工处理方法、数据的显示方法、数据分布特征的概括与分析方法等。例如，通过对于数据资料的图像化处理，将资料摘要变为图表，以直观了解整体资料分布的情况。通常会使用的工具是频数分布表与图示法，如多边图、直方图、圆形图、散点图等。通过分析数据资料，以了解各变量内的观察值集中与分散的情况。运用的工具有：集中量数，如平均数、中位数、众数、几何平均数、调和平均数等；变异量数，如全距、平均差、标准差、相对差、四分差等。

2. 推断统计学

推断统计学（Inferential Statistic）诞生于 1920 年前后，主要代表人物是戈瑟特（W. S. Gosset, 1876—1937）和费希尔（R. A. Fisher, 1890—1962）。推断统计学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的基础上，对统计总体的未知数量特征做出以概率形式表述的推断。

在推断统计中，测量样本的集中趋势与离散趋势都是变量的无偏估计值，但是以平均数、变异数、标准差的有效性最高。数据的次数分配情况，往往会呈现常态分配。为了表示测量数据与常态分配偏离的情况，会使用偏态、峰度这两种统计数据。为了解个别观察值在整体中所占的位置，需要将观察值转换为相对量数，如百分等级或标准分数。

1.3.2 理论统计学与应用统计学

根据统计学研究方法和统计方法的应用范围不同，统计学分为理论统计学（Theo-

ry Statistics) 和应用统计学 (Applied Statistics)。理论统计学是把研究对象一般化、抽象化, 以概率论为基础, 从纯理论的角度, 对统计方法加以推导论证, 中心的内容是统计推断问题, 实质是以归纳方法研究随机变量的一般规律。例如, 统计分布理论、统计估计和假设检验理论、相关与回归分析、方差分析、时间序列分析、随机过程理论等。理论统计学的发展, 推动和完善了统计学学科。理论统计学是统计方法论的理论基础, 没有理论统计学, 统计学科的发展将止步不前。

应用统计学是指统计学的一般理论和方法在社会、自然、经济、工程等各个领域的应用以及应用中的具体方法, 它是统计学和其他学科之间形成的交叉学科, 也是理论统计学发展的源泉。例如, 将统计学的方法应用于生物学的研究, 形成了生物统计学; 应用于医学研究, 形成了医疗卫生统计学; 应用于商业, 形成了商务统计学; 应用于旅游业, 形成了旅游统计学; 应用于现代金融服务业, 形成了金融统计学等。

此外, 统计学作为一门方法论科学, 能够和其他学科交叉融合, 形成新的统计学分支。如社会经济统计学、农业统计学、管理统计学、商业统计学、贸易统计学、人口统计学等。



本章小结

本章主要介绍了统计的含义及统计学的基本概念, 统计的研究对象、特点及作用, 统计学的分类等内容, 使读者对统计有一个初步的了解, 为更好地理解和应用统计方法奠定了基础。

统计的含义主要有三种, 即统计工作、统计资料和统计学。统计的研究对象是社会经济现象总体的数量方面, 通过对社会经济现象总体数量方面的研究, 以便认识社会经济现象的现状、本质、现象间的数量关系和发展变化的趋势与规律。它具有数量性、总体性、具体性和社会性四个特点。

统计学常用的基本概念包括总体与总体单位、标志与指标、指标体系、变异与变量等。其关系可用图 1—2 描述。

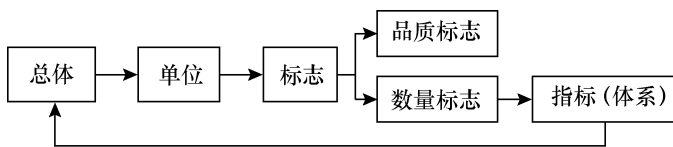


图 1—2 统计学常用概念之间的关系图

统计学可分为描述统计学和推断统计学, 根据统计学研究方法和统计方法的应用范围不同, 统计学分为理论统计学和应用统计学。



思考与练习题

思考题

1. 怎样理解社会经济统计学的研究对象?
2. 品质标志和数量标志有什么区别?

3. 试举例说明标志与指标的区别和联系。
4. 举例说明什么是变量和变量值。
5. 为了分析本班同学的学习情况，请你设计一些指标。

练习题

一、单项选择题

1. 统计是从（ ）方面入手认识现象的。
A. 数量 B. 质量 C. 数量和质量 D. 以上都不对
2. 构成统计总体的个别事物称为（ ）。
A. 调查总体 B. 调查单位 C. 总体单位 D. 标志值
3. 统计总体是由客观存在的（ ）的集合。
A. 不同性质基础上结合起来的许多个别事物
B. 同一性质基础上结合起来的许多个别事物
C. 随意基础上结合起来的许多个别事物
D. 有选择的基础上结合起来的许多个别事物
4. 要了解全系 6 个班 240 位学生的学习情况，则总体单位是（ ）。
A. 全系 6 个班 B. 每一个班 C. 240 位学生 D. 每一个学生
5. 标志是说明（ ）特征的名称。
A. 总体 B. 总体单位 C. 总体单位量 D. 总体单位质的
6. 工业企业的设备台数、产品产值（ ）。
A. 是连续变量
B. 是离散变量
C. 前者是连续变量、后者是离散变量
D. 前者是离散变量、后者是连续变量
7. 数量指标的表现形式是（ ）。
A. 绝对数 B. 相对数 C. 平均数 D. 小数
8. 构成统计总体的个别事物称为（ ）。
A. 报告单位 B. 标志值 C. 品质标志 D. 总体单位
9. 几位学生的某门课成绩分别是 67 分、78 分、88 分、89 分、96 分，“学生成绩”是（ ）。
A. 品质标志 B. 数量标志 C. 标志值 D. 数量指标
10. 对某市百货商店工作人员进行调查，总体单位是（ ）。
A. 该市某一百货商店
B. 该市各百货商店
C. 该市某一百货商店的所有工作人员
D. 该市百货商店每位工作人员
11. 下列分组中，（ ）是按照数量标志分组的。
A. 企业按年生产能力分组
B. 企业工人按性别分组

- C. 人口按民族分组
 - D. 家庭按城镇分组
12. 标志是说明总体单位特征的名称, ()。
- A. 它分为品质标志和数量标志两类
 - B. 品质标志具有标志值
 - C. 数量指标具有标志值
 - D. 品质标志和数量标志都具有标志值

二、多项选择题

1. 统计学研究对象的特点, 除具有总体性和数量性之外, 还表现有 ()。
- A. 具体性
 - B. 同质性
 - C. 差异性
 - D. 数据的随机性
 - E. 范围的广泛性
2. 在对工业企业生产设备调查中 ()。
- A. 全部的工业企业是调查对象
 - B. 工业企业的全部生产设备是调查对象
 - C. 每台设备是调查单位
 - D. 每台设备是调查单位也是填报单位
 - E. 每个工业企业是填报单位
3. 下列分组中, () 是按品质标志分组的。
- A. 职工按工龄分组
 - B. 科技人员按职称分组
 - C. 人口按民族分组
 - D. 企业按经济类型分组
 - E. 人口按地区分组
4. 下列标志中, 属于数量标志的有 ()。
- A. 性别
 - B. 职称
 - C. 工资
 - D. 年龄
 - E. 工龄
5. 下列变量中属于连续变量的有 ()。
- A. 公司利润
 - B. 网络用户数
 - C. 存款余额
 - D. 机器设备价值额
 - E. 学生人数

三、判断题

1. 统计工作的研究对象是客观现象总体的数量方面。()
2. 社会经济统计的研究对象是社会经济现象总体的各个方面。()
3. “2010 年甲企业的职工工资总额 260 万元” 是数量指标。()
4. “1999 年某省拥有普通高等学校 32 所” 是质量指标。()
5. “男或女” 是品质标志 “性别” 的变异。()
6. 当 “总体单位” 变为 “总体” 时, “数量指标” 就变为 “指标”。()
7. 总体和总体单位是固定不变的。()

8. 变量是指有变异的指标和数量标志。()
9. 推断统计学是整个统计学的基础，而描述统计学是在此基础上发展而来的。()
10. “某城市每个家庭拥有的汽车数” 是一个离散型变量。()

第 2 章 数据收集与整理

【学习目标】

通过本章的学习，了解统计数据的来源、统计数据收集的原则和方法、统计整理的概念和意义以及统计整理的步骤；理解各种统计调查组织方式的特点和应用范围、分组标志的选择和分组的方法；掌握问卷设计的基本技术，能根据某一实际问题初步设计统计调查问卷，并掌握统计分组的方法，能根据实际资料进行统计分组，编制分布数列和统计表。

2.1 数据的收集

2.1.1 统计数据来源

所有的统计数据如果追本溯源都来自于调查或实验。但是从使用者的角度看，统计数据的来源主要有直接来源和间接来源两个渠道。

通过调查或实践活动直接获取的第一手统计数据即原始统计数据，称为统计数据的直接来源。例如，通过调查来收集某市工业企业职工工资情况所获得的第一手资料，就属于数据的直接来源。

如果与研究内容有关的资料是别人调查获得的，或者有一些是历史遗留的统计数据，现在只是对这些原有的信息重新加工整理，使之成为进行统计分析可以使用的数据，我们称其为统计数据的间接来源。例如，统计部门公布的统计信息、定期发布的统计报表、定期出版的各类统计年鉴等，这些都属于统计数据的间接来源。间接来源的统计数据又称为第二手统计数据或次级统计数据，虽然具有收集方便、成本较低的特点，但是第二手资料并不是专门为所研究的问题设计的，相关性不够，因而在使用时要特别谨慎。

2.1.2 统计数据收集的原则和方法

1. 统计数据收集的原则

(1) 准确性原则。准确性原则就是如实反映客观事物，这是保证统计资料质量的首要环节。如果数据不真实，必将给统计各个阶段的工作带来消极的影响。

(2) 及时性原则。及时性原则强调所收集数据的时效性，即要求在数据收集的规定时间内，尽快提供资料。如果统计数据收集得不及时，就会贻误统计整理、分析的时

间，使统计工作失去应有的作用。

(3) 系统性原则。系统性原则是指收集的数据应该有条理、合乎逻辑，便于整理汇总。

(4) 完整性原则。完整性原则是指被调查单位不重复、没有遗漏，所列调查项目数据收集齐全。如果数据收集不齐全，就有可能无法反映出所要研究现象的全貌，甚至会做出错误的判断。

2. 统计数据收集的方法

在实际调查中，收集数据常用的方法有直接观察法、报告法、询问法和问卷调查法等几种。

(1) 观察法。观察法即调查人员亲临现场，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，道路车辆通过状态数据采集、超市商品的购买者行为数据采集等。此法取得的资料具有较高的准确性，但耗用的人力、物力和时间比较多，同时有些社会经济现象还不能用此法进行测量。如对农民或职工家庭收支情况资料的收集，一般不宜直接计量和观察。因此，其应用受到一定的限制。

(2) 报告法。报告法是基层单位根据上级的要求，收集相应的数据资料，并逐级上报给有关部门。当前，我国各企业、机关向上级填报统计报表，采用的就是报告法。报告法是各地方、各部门、各单位按照有关法规的规定，必须对国家履行的一种义务。报告法具有统一项目、统一表式、统一要求和统一上报程序的特点，其取得的资料一般比较可靠和准确。现行的统计报表形式就是采用报告法收集数据资料的。

(3) 询问法。询问法是调查者根据调查提纲向被调查者询问，由被调查者答复以取得统计数据资料的一种方法。它又可以分为个别访问法和开调查会两种。个别访问法是由调查人员向被调查者逐一询问来收集资料的方法，其优点表现在：调查人员对调查项目有统一的理解，能按统一口径逐项询问而取得资料，可保证调查资料的准确性。但该方法需要花费大量的人力和时间，不适合进行全面调查。开调查会是指邀请了解情况的人参加座谈会，以此来收集资料的方法。采用此法可以共同商讨、相互启发、相互核实，能深入了解实际情况，取得的资料比较准确可靠。这种方法的不足在于：参加会议的人容易受权威人士或第一发言人的影响，以致出现信息偏差。尽管如此，只要我们注意扬长避短，询问法仍是有用武之地的，因而在典型调查中被广泛采用。

2.1.3 统计调查方案

统计调查是一项细致复杂的工作，在组织调查之前必须设计周密可行的调查方案，以保证统计调查有组织、有计划地进行。一个完整的调查方案应该包括以下几个方面的内容。

1. 确定调查目的和任务

这是设计统计调查方案的首要问题。确定调查目的，就是要明确调查要解决什么问题，这样才能进一步确定调查的内容和范围以及采用的方式和方法。目的不明，任务不清，就无法确定向谁调查，调查什么，怎样调查，整个调查工作就会陷入盲目混乱，造成人力、物力、财力的浪费。

2. 确定调查对象和调查单位

确定调查对象和调查单位，是为了回答向谁调查、由谁来具体提供调查资料的问

题。确定调查对象和调查单位,必须以调查目的为依据。

调查对象就是在某项调查中需要进行调查研究的现象的总体,它是由性质相同的许多调查单位组成的。调查单位就是所要调查现象的总体中的个体,它是调查项目的承担者,是统计信息源。例如,调查目的是要了解某工商银行所属各办事处的职工人数、存款贷款余额等情况,这时该工商银行所管辖的所有办事处就构成了调查对象,每一个办事处则为一个调查单位。

明确调查单位还需要把它与报告单位相区别。报告单位也称填报单位,它是负责向上级报告和提交统计资料的单位。报告单位一般在行政上、经济上具有一定独立性,而调查单位可以是人,也可以是企事业单位,还可以是物。根据调查目的,调查单位与报告单位有时一致,有时不一致。如工业企业普查,每个工业企业既是调查单位又是报告单位;而工业企业生产设备状况的普查,调查单位是工业企业的每台生产设备,而报告单位则是每个工业企业。一般来说,在统计调查中,要明确规定调查单位和报告单位,这是为了说明向谁调查统计资料,由谁提交统计资料,以避免在调查单位和报告单位不一致时产生矛盾。

3. 拟定调查提纲和调查表

调查提纲由调查项目构成。调查项目就是调查中所要登记的调查单位的各种标志,包括品质(属性)标志和数量标志。确定调查提纲所要解决的问题是:向调查单位调查什么?用什么标志反映调查单位的特征?在调查中涉及哪些调查项目?这些都应根据调查目的和调查单位的特点而定,并贯彻“少而精”的原则进行处理。

调查表是指将调查项目按照一定的顺序编制而成的统计表格。它主要用于统计调查阶段,是收集原始资料的基本工具,且便于填写和汇总整理。调查问卷是调查表的一种方式。在统计调查中,为了解被调查者对某些事物的认识、看法和观点,了解民情、民意,常常采用调查问卷形式收集原始资料。

4. 确定调查时间和方法

调查时间包括两个方面的含义:首先,调查时间是指调查资料所属的时间,如果调查的是时期现象,就要明确规定调查资料的起止日期,如果要调查的是时点现象,就要明确规定统一的标准时点。例如,第二次全国经济普查规定的普查时期为2008年1月1日至12月31日,普查标准时点为2008年12月31日24时。其次,调查时间是指调查期限,即整个调查工作的起止时间,调查期限包括收集资料及报送等全部工作所需要的时间。为了保证资料的时效性,对调查期限的规定不宜过长。调查方法,包括调查的组织形式和收集资料的具体方法,主要根据调查的目的、要求和调查对象的特点而定。

2.1.4 统计调查问卷

采用问卷进行调查始于20世纪30年代的美国,最初应用于政治选举、商业推销和经济预测等方面,后逐步成为调查研究中收集资料的一种主要方式。我国也广泛采用调查问卷的方式来研究社会经济领域里的现象和问题,现在已将调查问卷纳入了统计制度的范围,成为统计调查的一个重要组成部分。

1. 调查问卷的基本类型

问卷作为收集资料的工具,在使用过程中并非只有一种模式。调查者的研究目的、

调查内容、调查方式的不同,决定了调查问卷的形式不尽相同。按问卷填写方式划分,可将调查问卷分为自填式问卷和访问式问卷,这是调查问卷的两种基本类型。这两类问卷在设计上有所区别,使用过程中各有优缺点。

(1) 自填式问卷。自填式问卷是指通过邮寄或分发等方法,由被调查者自己填写的问卷。在这种情况下,被调查者可以不受其他影响,如实表达自己的意见,尤其是对敏感性问题的调查,自填式问卷往往可以得到较为可靠的资料。同时,这种问卷使用了标准化词语,每个被调查者所面临的都是完全相同的问题,因而不存在调查人员对问卷的主观随意解释和诱导,从而可以避免调查人员的偏见。但这类问卷还存在一些不足:问卷填写的答案可能会含糊不清、被调查者对某些问题拒绝回答、无法得知被调查者是否独立完成答案及其回答问题的环境等,这些都会影响问卷质量。

(2) 访问式问卷。访问式问卷是指由调查人员通过现场询问,根据被调查者口头回答的结果代为填写的问卷。这类问卷的应答率高、可控性强,调查人员可以设法确保被调查者独立回答问题,并能控制对方按问题的设计顺序回答,从而保证应答的完整性。同时,调查人员还可以观察被调查者的态度及其回答问题的环境,有利于进一步分析、判断相关问题。但这类问卷也存在一些不足,如费用高、容易受调查人员的影响、匿名性差、当被调查者对调查人员的某些举止有偏见或不理解时会导致差错或有意说谎、调查人员对被调查者的意思没有正确理解或记录失当也有可能出差错等。另外,运用这类问卷调查,由于调查人员知道被调查者的一些基本情况,有时会给被调查者带来心理压力,甚至出现拒答的情况。

2. 调查问卷的基本结构

调查问卷的主要内容是关于调查事项的若干问题和答案,但仅有这些内容是不够的。一份完整的调查问卷,通常由题目、说明信、被调查者基本情况、调查事项的问题和答案、填写说明与解释等五个主要部分构成。

(1) 题目。

题目是调查问卷的主题。调查问卷的题目应该准确、醒目、突出。要能准确而概括地表达问卷的性质和内容;观点新颖,句式构成上要富于吸引力和感染力;言简意赅,明确具体。

(2) 说明信(又称封面信)。

说明信一般在问卷的开头,是写给被调查者的一封短信。这是调查者与被调查者的沟通媒介,目的是让被调查者了解调查的意义,引起被调查者足够的重视和兴趣,争取他们的支持与合作。说明信要说明调查者的身份、调查的中心内容及要达到的目的和意义、选样原则和方法、调查结果的使用和依法保密的措施与承诺等,有时还需要将奖励的方式、方法及奖金、奖品等有关问题叙述清楚。说明信必须态度诚恳,口吻亲切,以打消被调查者的疑虑,取得真实资料。写好说明信,取得被调查者的合作与支持,是问卷调查取得成功的必要保证。

(3) 被调查者的基本情况。

被调查者的基本情况是对调查资料进行分类研究的基本依据。一般来说,被调查者包括两大类,一类是个人,另一类是单位。如果被调查者是个人,其基本情况包括姓

名、性别、民族、年龄、文化程度、职业、职务或技术职称、个人或家庭收入等项目；如果被调查者是企事业单位，则包括单位名称、经济类型、行业类别、职工人数、规模、资产等项目。若采用不记名调查，被调查者的姓名可在基本情况中省略。

(4) 调查事项的问题和答案。

调查事项的问题和答案是调查问卷最主要、最基本的组成部分。调查资料的收集主要通过这一部分完成，它也是问卷的目的所在。通常在这一部分既提出问题，又给出回答方式。从形式上看，有开放式问题与封闭式问题之分；从内容上看，有背景问题、行为问题、态度问题与解释性问题之别。问题的内容决定于调查目的和调查项目，这里仅就问题的形式予以阐述。

开放式问题是指问题对每一个应答者（被调查者）是同一的，但不确定任何答案，应答者可以自由地围绕提出的问题，写下描述性的情况和意见。开放式问题的优点是：应答者不受任何定式的约束，可以自由发表意见，获得的资料往往比较丰富而生动。其不足之处是：答案五花八门，复杂多样，难以定量处理，而且受应答者表述能力的影响较大，由此会造成一些调查性误差。

封闭式问题是指问题不仅对每一位应答者相同，而且每一个问题都已预先分列了若干答案，由应答者在其中选择符合自己实际情况的答案。封闭式问题的优点是：问题清楚具体，应答者容易回答，答案标准，整齐划一，填写方便，容易整理，适于定量分析。其不足之处是：由于事先规定了预选答案，应答者的创造性受到约束，不利于发现新问题；应答者在对预选答案不理解、不满意或随便选择的情况下，会影响调查结果的正确性。

由于两种问题各有优缺点，为了弥补它们的不足，在实际操作中经常是将两种问题形式结合使用，从而形成一种优势互补的调查问卷。

(5) 填写说明和解释。

填写说明和解释包括填写问卷的要求、调查项目的意义、被调查者应注意的事项等，其目的在于明确填写问卷的要求和方法。

除了上述五个基本部分以外，问卷的最后也可以写上几句短语，表示对被调查者的感谢，或征求被调查者对问卷设计和问卷调查的意见和感受。如果是访问式问卷，还可以加上作业证明记载，其主要内容包括调查人员姓名、调查时间等，便于明确调查人员的责任，有利于检查并修正调查资料。

3. 问卷设计应注意的问题

问卷设计是一项十分复杂又需要耐心的工作，即使是很有经验的研究人员，在进行这项工作时也要反复推敲，否则就达不到调查目的。因此，问卷设计必须注意下列问题：

问卷上所列问题应该都是必要的，可要可不要的问题不要列入。

所问问题应是被调查者熟悉且易于回答的，避免出现被调查者不了解或难以回答的问题。回答全部问题所用的时间不宜超过半小时。

注意询问语句的措辞和语气，一般应注意以下几点：问题要提得清楚、明确、具体、简短；明确问题的界限和范围，问句的词义要清楚；避免引导性问题或带有暗示性问题的出现。

属于年龄、收入等涉及隐私的问题，最好采用间接提问的方法，不要直接询问“您今年多大年纪”或“您每月的收入是多少”，可以给出一定范围，如“21~30岁、31~40岁”等或“500~1 000元、1 000~1 500元”等。

问卷上所拟答案要有穷尽性，避免重复或相互交叉。问卷上拟定的答案要编号。

问卷纸张质地要良好，不易破损，字迹印刷清晰，留作填写说明的空白处要大，页数较多时要装订成册。

2.1.5 统计数据收集的组织形式

统计数据收集的组织形式主要有统计报表和专门调查两大类。

1. 统计报表

统计报表是按照全国统一规定的表格形式、统一规定的指标内容、统一规定的报送程序和报送时间，由填报单位自下而上逐级提供统计资料的一种统计调查方式。统计报表所包括的范围比较全面、分组比较齐全、指标内容和调查周期相对稳定，因此，统计报表是统计数据收集组织形式中的一种重要方式。

统计报表担负着为计划的制订及其执行情况的检查提供资料的任务，这就决定了统计报表必须以全面调查为主，非全面调查为辅。

(1) 统计报表的类型。

统计报表可按不同的标志划分为以下几种类型。

①月报、年报和其他时间报表。统计报表按报送周期不同，分为日报、旬报、月报、季报、半年报和年报统计报表，其中以月报和年报统计报表为主。周期短的统计报表，投入的人力、物力、财力就要多，因此，指标项目可以少一些、粗一些；周期长的统计报表，指标项目可以多一些、细一些。月报、年报的周期较长，它的内容比较详尽；日报、旬报周期较短，其内容只限于填报少量最主要的指标。统计报表的报送原则是，凡一年、半年报告一次能满足需要的，就不用季报、月报；月报能满足要求的，就不用日报、旬报。

②邮寄报表和电信报表。统计报表按报送方式不同，分为邮寄报表和电信报表。传统的电信报表主要采取电报、电话的方式报送。随着信息技术、网络技术的快速发展，信息传递的手段不断现代化，统计报表已发展为以计算机网络报送为主、其他形式为辅的报送方式。

③基层报表和综合报表。统计报表按填报单位不同，分为基层报表和综合报表。基层报表主要由基层企、事业单位填报，所提供的原始资料是统计的基础资料。综合报表是由主管部门根据基层报表逐级汇总填报的统计报表，汇总后得到各级基本统计指标。填报基层报表的单位称为基层填报单位，填报综合报表的单位称为综合填报单位。

④国家报表、部门报表和地方报表。统计报表按实施范围不同，分为国家报表、部门报表和地方报表。国家报表是根据国家的有关统计调查项目和统计调查计划制定的统计报表，也叫国民经济基本统计报表。这种统计报表是从整个国民经济的角度出发制定，并按照国民经济的部门来划分，如农业、工业、建筑业、固定资产投资、国内贸易、劳动工资、物价等。这些报表在全国范围内的各行各业实施，主要用来收集整个国民经济和社会发展情况的基本统计资料。部门报表是根据有关部门的统计调查项目和统

计调查计划制定的统计报表,实施范围限于各业务主管部门系统内,一般用来收集各级主管部门所需要的专门统计资料。地方报表是根据有关地方统计调查项目和统计调查计划相应制定的统计报表,其实施范围是各省、市、自治区,主要用来满足地方的专门需要。部门和地方报表都是国家统计报表的补充。

(2) 统计报表的内容。

统计报表一般应包括以下内容:

①报送目录。报表目录是指报送的报表名称、报表的填报单位、调查对象、报送时间和报送程序等事项的一览表。目录的作用是使填报单位了解在什么时间、用什么方式、向什么单位报送什么报表。

②报表表式。报表表式是指统计报表的具体格式。报表格式要求简单明了,每张表式要明确规定出表名、表号、填报单位、报送日期、主栏项目、纵栏项目、表下补充资料、填报单位负责人和填报人签章以及制表部门等。

③填表说明。填表说明是指填写报表时应遵守的各种规定和应注意的问题。填表说明主要包括指标解释、计算方法、包括范围和有关事项的具体说明等。

2. 专门调查

(1) 普查。

普查是为特定目的而专门组织的一次性或周期性的全面调查。如全国人口普查、工业普查、科技人员普查等,其主要用来调查属于一定时点或一定时期内的社会现象总量。普查的目的在于收集那些不能够或不适宜用其他统计调查方式收集的统计数据,以掌握重要的国情、国力数据。

普查是一种全面调查。普查对总体中的所有个体都进行调查登记,直接获得总体参数的具体数值,因此采用这种调查方式收集的资料具有全面系统、准确可靠的优点,但由于组织普查工作较为繁重、涉及面广、时间较长,且需要耗费大量人力、物力,不可能经常性连续进行,只能按照需要采取一次性或周期性的方式展开。周期性普查是我国政府统计调查工作的基础。我国统计制度规定,每逢年份的末尾数字为“0”的年份进行全国人口普查,每逢末尾数字为“5”的年份进行全国工业普查,每逢末尾数字为“6”的年份进行全国农业普查,每逢末尾数字为“3”和“8”的年份进行全国经济普查。

普查的组织原则是:第一,统一规定调查资料所属的标准时点,避免因为自然变动或机械变动而产生收集资料的重复或遗漏。例如,我国第一次至第四次人口普查统一规定以7月1日零时作为标准时点,第五次、第六次人口普查统一规定以2000年及2010年11月1日零时为标准时点。第二,在普查范围内,各调查单位和各调查点尽可能同时进行调查,并尽可能在最短期限内完成,以便在方法上、步调上保持一致,保证普查资料的真实性和时效性。第三,调查项目一经统一规定,不能任意改变或增减,以免影响综合汇总,降低普查资料的质量。同一种普查,每次调查项目的规定也应力求一致。此外,普查工作要尽可能按一定周期进行,以利于对历次普查资料进行对比分析。

(2) 抽样调查。

抽样调查是一种非全面调查,它是根据随机原则的要求,从总体中抽取部分个体构成样本,并依据样本信息推断总体数量特征的一种调查方法。抽样调查虽然是非全面调

查,但它的目的却是取得反映总体情况的信息资料。抽样调查具有以下三个显著特点:第一,调查采用了随机性原则;第二,用样本指标推断总体指标;第三,抽样误差事先可以控制。

抽样调查是一种具有广泛适用性的调查。抽样调查可以用于破坏性检验场合的调查和推断。例如,对一些产品的使用寿命、抗压、防震等能力的检验等,只能采用非全面调查的方法。抽样调查还可以用于较大规模总体或无限总体的调查。在总体单位数无限或虽然有限却无法一一观察的情况下,要研究总体的数量特征,就难以进行全面调查,而只能以抽样调查方法推断总体。抽样调查还可以用于检查和补充全面调查数据。全面调查涉及范围广、参加人员多,往往存在着较大的登记性误差。为了提高数据的准确性,常在全面调查后再抽取一部分单位进行一次抽样调查,利用抽样调查的数据估计全面调查的误差,并据此对全面调查数据进行调整和补充。

(3) 重点调查。

重点调查是一种非全面调查,它是在调查对象中选择一部分重点单位进行的调查。所谓重点单位,是指在总体中具有举足轻重地位的那些单位,虽然它们在调查对象的全部单位中只占一小部分,但其标志总量在被研究总体的全部标志总量中却占绝大比重。因此,对这些重点单位的标志进行调查,就可以在数量方面说明总体在该标志总量方面的基本情况。例如,中国移动、中国电信和中国联通等几个大通信企业,虽然在全国的通信企业中只是少数,但它们的营业收入却占全国电信企业营业收入的绝大部分,所以,对这几个重点企业进行调查,就可以了解我国电信企业的基本情况。正因为如此,重点调查可以节省大量的人力、物力和财力,并能使调查工作做得更加细致、及时。此外,重点调查的组织也较灵活,既可以组织专门调查,也可以运用统计报表形式进行调查。

重点单位的选择要根据调查任务来确定。一般来说,选出的单位应尽可能少些,而其标志值在总体标志总量中所占比重应该尽可能大些。另外,被选中单位的经营管理制度应比较健全,统计力量应比较充实,统计基础应比较牢固,这样才能准确、及时地取得资料。重点单位的选择要客观。由于重点单位的选择是着眼于这些单位的标志值在总体标志总量中的比重,而不是这些单位的技术、管理或其他方面是否有特定意义,所以,重点单位的选择不应带有主观因素。值得注意的是,虽然重点单位的标志值在总体标志总量中占有很大比重,掌握了他们的情况,就基本掌握了总体特征,但这些情况毕竟不能完整反映总体总量,而且重点调查的资料也不具备推断总体总量的条件。因此,重点调查只是为了获得反映总体基本情况的统计资料,而不宜推断总体。

(4) 典型调查。

典型调查也是一种非全面调查。它是根据调查的目的与要求,在对被调查对象进行全面分析的基础上,有意识地选择若干具有典型意义或有代表性的单位进行深入、细致的调查研究,借以认识事物发展规律的一种调查方式。典型调查大体上可以分为两种:一种是对个别典型单位进行“解剖麻雀”式调查,其目的在于通过个别单位特征说明现象的一般情况或规律性;另一种是“划类选典”式的典型调查,即在对调查总体进行分类的基础上选择一部分典型单位,由这一部分典型单位构成一个代表性的样本总体,然

后对这个样本总体进行调查，根据其调查结果，从数量上对总体进行统计推断和统计分析。

进行典型调查的主要目的不在于取得社会经济现象的总体数值，而在于了解与有关数字相关的具体情况，以补充全面调查的不足。典型调查的优点是调查范围小，调查单位少，灵活机动，节省人力、物力和财力等。其不足是在实际操作中选择真正有代表性的典型单位比较困难，而且还容易受人为因素的干扰，从而可能导致调查结论有一定的倾向性。

上述各种统计调查方式各有其优越性,也各有其局限性,在实际工作中应根据不同情况灵活运用,也可以把几种调查方式结合使用。我国将建立以必要的周期性普查为基础,以经常性的抽样调查为主体,同时辅以统计报表、重点调查、科学推算等多种方法综合运用的统计调查方法体系。

2.2 数据的整理

2.2.1 统计整理的概念和意义

1. 统计整理的概念

统计整理是根据统计研究的任务与要求，对统计调查所收集到的原始数据进行科学综合与加工，使之系统化、条理化的工作过程。

如为了了解某市工业企业总产值的基本情况，对该市工业企业进行调查。通过调查取得该市 100 个工业企业 2005 年 12 月总产值资料如下（单位：万元）：40，65，30，90，125，230，…160。这些资料只能反映每个国有企业的情况，必须对这些资料进行整理，才能得到全市工业企业总产值分布情况。先将这些数据资料按总产值的多少进行整理，可得到表 2—1 的结果。

表 2—1 **2005 年 12 月某市工业企业总产值分布表**

总产值 (万元)	0~50	50~100	100~150	150~200	200~250	合计
企业数 (个)	10	24	35	21	10	100

2. 统计整理的意义

统计整理是统计工作的第二个阶段，它是统计调查的继续，又是统计分析的前提，在整个统计工作中起着承前启后的作用。

统计调查所获取的资料主要是反映总体单位特征的原始资料, 这些资料比较分散、杂乱, 只能表明各个被调查单位的具体情况, 不能说明事物的全貌和总体情况; 这些资料只能反映事物的表面现象, 不能深刻地说明事物的本质, 难以揭示事物的发展规律。经过统计整理, 可以实现从个别单位的数量特征到总体数量特征的第一次综合, 得到描述现象总体数量特征的最初始的综合指标。

2.2.2 统计整理的步骤

统计整理大体可分为以下几个步骤:

1. 制定统计整理方案

统计整理方案是指导整个工作的基本文件，在方案中明确规定各种统计分组及统计指标设置。统计整理方案具体表现为一系列综合表式、填表说明等。在统计整理方案之外，还要制订一个整体的工作计划，说明统计整理工作完成的期限、组织领导、培训、资料的审核和汇总方法等。

2. 审核、纠正统计数据资料

由于通过不同的调查环节，统计资料所取得数据的准确程度也不相同，差错在所难免。汇总前如不进行严格审核，势必会影响汇总结果的准确性。对原始资料的审核主要包括资料的完整性、及时性和准确性三个方面的内容。发现统计数据差错后，要分别就不同情况及时纠正与处理。

3. 进行科学的统计数据分组与汇总

这是统计整理的中心环节。对原始资料进行统计分组和计算汇总，就是按照一定的组织形式和方法对原始资料进行统计分组和统计汇总，计算出各组的单位数和总体的单位总数，计算出各组的指标数值和总体的指标数值。

4. 编制统计表和绘制统计图

根据统计整理的结果编制成统计表或绘制成统计图，简明扼要地反映现象在数量方面的具体表现和相互联系。

2.2.3 统计分组

1. 统计分组的概念

统计分组就是根据统计研究的目的和现象总体的内在特征，按照一定的标志将统计总体划分为若干个性不同而又有联系的几个部分的一种统计方法。总体的这些组成部分称为“组”。例如，研究某一地区人口状况时，可按年龄这一标志将人口划分为不同年龄组。

统计分组对总体而言，是“分”，即将总体划分为性质不同的几个组成部分；对个体而言，是“合”，即将总体中性质相同的单位组合在一起，形成一个组。通过统计分组，使同一组内各单位的性质相同，不同组的各单位性质相异，即“组内同质性，组间差异性”。

2. 统计分组的原则

在进行统计分组时，应遵循以下三个原则：第一，穷尽原则，即在分组时，要把所有观测数据都包括在内；第二，互斥原则，即组与组之间的界限要分明，不能有交叉，在遇到与分组标志值相同的数据时，要按照“上限不在内”原则处理；第三，在定性分析的基础上科学分组，如在对考试成绩进行分组时，可将 60 分作为分组界限，如 60 以下和 60 以上，而不能做如 55~60 这样的分组。

3. 统计分组的作用

统计分组在统计整理中具有重要的作用，主要表现在以下几个方面：

(1) 统计分组可以划分社会经济现象的类型。

统计分组的根本作用在于区分现象质的差别。统计分组的过程就是区别事物性质的过程。大量的社会经济现象存在复杂多样的类型，各种类型各有其自身的特点和发展规

律。借助于统计分组,把社会经济现象总体划分为性质不同的类型,就可以研究各种类型的数量特征,评价彼此间的关系和发展变化规律。例如,依据生产资料所有制性质和国家有关法规,经济类型可分为国有经济、集体经济、私营经济、个体经济、联营经济、股份制经济、外商投资经济、港澳台投资经济、其他经济九大类。

(2) 统计分组可以研究现象的内部结构。

统计分组后,统计总体被划分为若干个性质的组成部分,通过计算各组数值在总体总量中所占的比重,可以分析研究总体内部各个组成部分的性质、结构和比例关系,表明现象总体的本质特征。

(3) 统计分组可以分析现象之间的相互依存关系。

任何现象都不是孤立的,通过对某一被定为原因的现象的分组来观察另一被定为结果的现象在各组的分布及其变化情况,可以知道这两个现象之间是否确有因果关系和依存关系。

4. 分组标志的选择

分组标志就是分组的依据或标准。统计分组的关键是选择分组标志和划分各组界限,其中选择分组标志是统计分组的核心问题。在选择分组标志时,必须遵循以下原则:

(1) 根据统计研究的目的选择分组标志。

例如,要研究全国大学生的构成情况,可以选择性别、年龄、民族等作为分组标志;要研究全国大学生的健康状况,可以选择身高、体重、肺活量等作为分组标志。

(2) 根据事物内部矛盾选择反映事物本质的分组标志。

事物的标志是多种多样的,有些标志是主要的标志,能够反映事物的本质,而有的是次要标志。例如,要研究某市教师的生活水平状况,在教师的“工资”、“其他收入”、“家庭成员的平均收入”等标志中,最能反映教师生活水平状况的分组标志应当是“家庭成员的平均收入”,因为教师的生活水平不仅受工资和其他收入的影响,更重要的是受所负担家庭成员多少的影响。因此,在分组时,要从统计研究的目的出发,从若干标志中选择最能反映事物本质特征的标志。

(3) 根据被研究事物所处的具体条件选择分组标志。

例如,研究工业企业规模与劳动生产率之间的关系时,需要按企业规模进行分组,而反映企业规模的标志有职工人数、生产能力、固定资产价值等,究竟选择哪一种作为分组标志,需要根据具体条件而定。对于劳动密集型产业,应采用职工人数作为分组标志来反映企业生产规模的大小;对于技术密集型产业,反映企业生产规模大小就要选用固定资产价值或生产能力作为分组标志。

5. 统计分组的方法

(1) 按品质标志分组或按数量标志分组。

按品质标志分组是指选择反映事物属性差异的品质标志作为分组标志,并在品质标志的差异范围内划分各组界限,将总体划分为性质不同的组成部分。如职工按性别、文化程度、工种分组,企业按经济类型、地区分组,学生按生源、性别分组等都是按品质标志分组。

按数量标志分组是指选择反映事物数量差异的数量标志作为分组标志，并在数量标志的差异范围内划分各组界限，将总体划分为若干性质不同的组成部分。如人口按年龄、身高、体重分组，企业按总产值、职工人数、资金利润率分组等。

(2) 简单分组和复合分组。

简单分组就是按一个标志对总体进行分组，如学生按学习成绩分组，企业按生产规模分组。

复合分组是指对总体按两个或两个以上的标志层叠起来进行的分组，即将总体先按一个分组标志进行分组，在此基础上按另一个分组标志把已分成的各个组成部分再分成若干小组。例如，学生先按性别分组，在此基础上，再按年龄分组就是复合分组。进行复合分组时，应根据统计分析的要求确定分组标志的主次顺序。

2.2.4 次数分布

1. 次数分布的概念

在统计分组的基础上，将总体的所有单位按组归类整理，所形成的总体单位数在各组之间的分布，称为次数分布。分布在各组的总体单位数叫次数或频数。各组次数与总体总次数之比称为频率或比重。次数分布实质上就是反映总体中所有单位在各组间的分布状态或分布特征，因此也称为次数分布数列，简称分布数列或分配数列。

次数分布有两个构成要素：一是统计分组，二是各组的单位数。次数分布具有两个特征：第一，各组的频率介于 0 到 1 之间；第二，各组频率之和等于 1。

2. 次数分布的种类

根据分组标志特征的不同，次数分布可以分为品质分布数列和变量分布数列。

(1) 品质分布数列。

按品质标志分组所形成的分布数列，称为品质分布数列，简称品质数列。品质分布数列由各组名称和次数构成，如某企业工人性别构成情况（见表 2—2）。

表 2—2 某企业工人性别构成情况

按性别分组	职工人数（人）	比重（%）
男	650	65
女	350	35
合计	1 000	100

(2) 变量分布数列。

按数量标志分组所形成的次数分布数列称变量分布数列，简称变量数列。变量分布数列由各组变量值和各组次数构成，如某班学生统计学成绩分布（见表 2—3）。

表 2—3 某班学生统计学成绩分布

按成绩分组	人数（人）	比重（%）
60 以下	4	10
60～70	6	15
70～80	16	40

续前表

按成绩分组	人数（人）	比重（%）
80～90	10	25
90 以上	4	10
合计	40	100

变量数列按变量的表示方法不同，又可分为单项变量数列和组距变量数列两种。

①单项变量数列。

单项变量数列是按数量标志分组后，用一个变量值代表一个组所形成的数列，简称单项数列，如某车间第一季度工人平均日产量（见表 2—4）。

表 2—4 某车间第一季度工人平均日产量

按平均日产量分组（件）	工人数	
	绝对数（人）	比重（%）
7	8	16
8	18	36
9	19	38
10	5	10
合计	50	100

单项数列适用于变量值个数较少，变量值变动范围较小的离散型变量。

②组距变量数列。

组距变量数列是按照数量标志分组后，用变量值变动的一定范围代表一个组所形成的数列，简称组距数列，上述表 2—3 即为组距变量数列。组距变量数列适用于变量值个数较多，变动范围较大的离散型变量以及所有的连续型变量。

在组距数列中，表示各组界限的变量值叫组限，其中较小的变量值称为下限，较大的变量值称上限，如表 2—3 中 60、70、80、90 都是组限，第二组中 60 是下限，70 是上限。各组上限与下限之间的距离叫组距，上限与下限之间的中点值称组中值。即：

组距＝上限－下限

组中值＝(上限＋下限)/2

如表 2—3 中，第二组组距为 10 分（70－60），组中值为 65 分 [(60＋70)/2]。

在编制组距数列时，常常使用“某某以上”或“某某以下”，这样不确定组限的组，称为开口组（如表 2—3 所示），否则称闭口组。

开口组的组中值按下列公式计算：

缺下限的最小组的组中值＝上限－相邻的组距/2

缺上限的最大组的组中值＝下限＋相邻的组距/2

如表 2—3 中，第一组的组中值为 55 分（60－10/2），第五组的组中值为 95 分（90＋10/2）。需要指出的是，组中值不为 0。

组距数列根据各组的组距是否相等可以分为等距数列和异距数列。如果各组的组距

相等，称为等距数列；如果各组的组距不相等，称为异距数列。编制组距数列时，采用等距数列还是异距数列，要根据统计研究的目的和现象的特点来决定。一般在社会经济现象性质差异变动比较均衡的条件下使用等距数列。如学生的学习成绩分组，工人的工资、工龄分组等。等距数列能清楚反映总体的分布特征，而异距数列则能比较准确反映总体内部各部分之间的性质差异。如社会经济现象的分布存在明显的偏斜状况时或其标志变异范围较大时常采用异距数列。

3. 累计次数分布

所谓累计次数分布，是将变量数列各组的次数和频率逐组累计相加而形成的次数分布，它有向上累计和向下累计两种计算方法。

向上累计，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，各累计数的意义是各组上限以下总共所包含的总体次数和比率有多少。

向下累计，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，各累计数的意义是各组下限以上总共所包含的总体次数和频率有多少。如表 2—5 所示。

表 2—5 某班学生统计学原理考试成绩分布表

学生成绩（分）	人数（人）	比例（%）	向上累计		向下累计	
			人数（人）	比例（%）	人数（人）	比例（%）
60 以下	4	10	4	10	40	100
60～70	6	15	10	25	36	90
70～80	12	30	22	55	30	75
80～90	15	37.5	37	92.5	18	45
90～100	3	7.5	40	100	3	7.5
合计	40	100				

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。当所关心的是标志值比较小的次数分布情况时，通常用向上累计，以表明在这些数值以下的总体单位数及其所占比例。当所关心的是标志值比较大的次数分布情况时，通常用向下累计，以表明在这些数值以上的总体单位数及其所占比例。

4. 次数分布类型

社会经济现象的次数分布曲线多种多样，人们通过长期观察和总结，将其归纳为三种类型，分别是：钟形分布、J 形分布和 U 形分布。

(1) 钟形分布。

钟形分布的特征是“两头小，中间大”，即靠近中间的变量值分布的次数多，靠近两边的变量值分布的次数少，其曲线宛如一口古钟，如图 2—1 所示。

图 2—1 中，其分布特征是以标志变量中心为对称轴，左右两侧对称，两侧变量值分布的次数随着与中间变量值距离的增大而渐次减少，在统计学中，称这种分布为对称分布。而图 2—2 为非对称分布，它们各有不同方向的偏态，即左偏态分布和右偏态分布。在实际工作中，许多社会经济现象总体的分布都趋于对称分布中的正态分布。正态分布是描述统计中的一种主要分布，它在社会经济统计分析中具有重要意义。

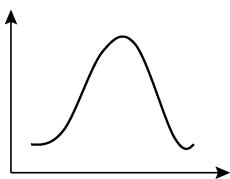


图 2—1 钟形分布

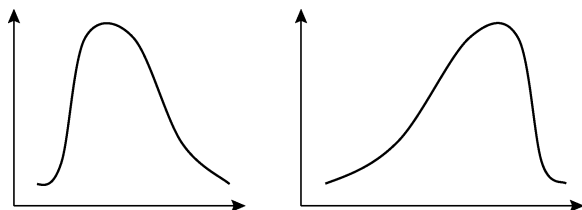


图 2—2 偏态分布

(2) J 形分布。

J 形分布有两种类型：一种是次数随着变量的增大而增多，如投资按利润率大小分布；另一种呈反 J 形分布，即次数随着变量的增大而减少，如随着产品产量的增加，产品单位成本下降。J 形分布如图 2—3 所示。

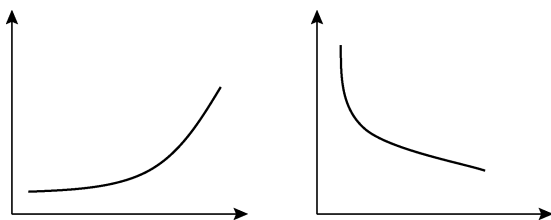


图 2—3 J 形分布

(3) U 形分布。

U 形分布的形状与钟形分布相反，靠近中间的变量值分布次数少，靠近两端的变量值分布次数多，形成“两头大，中间小”的 U 形分布。如人口死亡率分布就属于这种类型。U 形分布如图 2—4 所示。

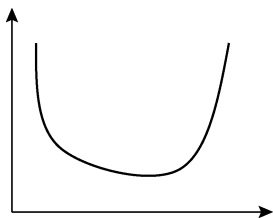


图 2—4 U 形分布

上述各种次数分布类型中，钟形分布在社会经济现象中最为常见，也最符合人们认识问题的习惯。例如，一个班级学生的考试成绩，差的和好的总是少数，居于中游者人数最多。再如农作物单位面积产量的分布，机械零件公差分布，一般社会居民收入的

分布等，基本上都表现为钟形分布或接近钟形分布的分布特征。了解这些分布类型，将有助于我们进一步认识事物的本质及其发展变化的规律性。

2.2.5 变量数列的编制

根据一定的统计资料编制变量数列，要注意以下问题：

1. 确定形式

根据变量的类型和变量值的变动范围来确定是编制单项数列还是编制组距数列。对于离散型变量，如果变量值个数少，变动范围小，可编制单项数列。如果变量值个数多，变动范围大，则编制组距数列。对于所有的连续变量则宜编制组距数列。

2. 确定组距和组数

首先确定全距。所谓全距，是指变量值中最大值和最小值之间的距离，然后再依据大多数变量值集中的区间和变量值的分散程度来考虑组距和组数的问题。组数和组距是相互制约的，组数越多，组距就越小；反之，组数越少，组距就越大。全距、组数、组距的关系可用公式表达如下：

组距 = 全距 / 组数

在等距数列中，组数可用下列公式计算：

$K=1+3.322\times\lg n$

其中 K 表示组数， n 表示数据个数。该公式称为斯德吉斯（H. A. Sturges）组数公式。

3. 确定组限

确定组限要考虑以下几点：应有利于体现总体单位分布的规律；最小组的下限要略低于最小变量值，最大组的上限要略高于最大变量值。

组限的表示方法有两种：重叠和不重叠。按连续变量分组划分组限时，相邻两组的组限必须重叠，即相邻两组的上限与下限用同一数值表示，如身高 160~165，165~170 两组。为避免分组时对总体单位重复计数，一般遵循“上限不在内”的原则。如有人身高为 165cm 时，应计入“165~170”这一组。按离散变量分组划分组限时，相邻两组的上限和下限通常以两个确定的不同数值来表示，即相邻两组的组限不重叠。如企业按职工人数分组可表示为 100 人以下，101~500 人，501~1 000 人，1 000 人以上。但在实际工作中，为了保证不重复、不遗漏总体单位，对于离散变量也可采用重叠式组限。如上例，也可表示为 100 人以下，100~500 人，500~1 000 人，1 000 人以上。

4. 变量数列编制的步骤

举例说明变量数列的编制步骤。例如，某生产小组 40 名工人日产量分为高、中、低 3 组，每人的某日产量如下（单位：件）：

660	700	450	460	460	480	640	490	540	500	650	550	560	560
560	590	580	560	570	570	580	590	590	550	590	560	590	600
590	600	640	480	650	490	750	660	450	560	750	710		

第一步：将原始资料按大小顺序排列。

450 450 460 460 480 480 490 490 500 540 550 550 560 560
560 560 560 560 570 570 580 580 590 590 590 590 590 600
600 640 640 650 650 660 660 700 710 750 750

第二步：确定组距和组数。

上述资料排序后，最小值为 450 件，最大值为 750 件。

全距 = 750 - 450 = 300

组数 = 3

组距 = 全距 / 组数 = 300 / 3 = 100

第三步：确定组限。

组限为：500、600

第四步：编制变量数列。

表 2—6 某生产小组工人日产量分布情况

按日产量分组（件）	人数（人）	比重（%）
500 以下	8	20
500~600	20	50
600 以上	12	30
合计	40	100

通过编制的组距数列，从表 2—6 可以很清楚地看出：日产量在 500 件以下的工人有 8 人，占总数的 20%；日产量在 600 件以上的工人有 12 人，占总数的 30%；日产量在 500~600 件的工人最多，有 20 人，占总数的 50%。该生产小组工人日产量呈明显的正态分布特征。

2.2.6 统计资料的汇总

统计资料汇总就是在编制次数分布数列后，汇总计算各组及总体的单位数和各组及总体的标志总量。统计资料汇总是统计整理的中心内容。统计资料汇总的组织形式一般有逐级汇总、集中汇总和综合汇总。

逐级汇总就是按照一定的统计组织管理系统，自下而上地逐级汇总本系统、本地区、本单位的调查资料。我国的定期统计报表一般采用这种组织形式，即县级—市级—省级—国家级行政主管部门逐级汇总资料。逐级汇总能及时审核和纠正统计调查资料，但逐级汇总层次较多，容易发生差错，且汇总需要的时间较长，会导致资料的时效性较差。

集中汇总就是将全部调查资料集中到组织调查的最高机关或其指定的机构进行汇总。比如快速物质普查通常由基层单位将统计数据直报国家统计局，由国家统计局直接汇总全国数据。集中汇总不经过中间环节，尽可能排除某些地方对统计数据的人为干扰，并可大大缩短汇总时间。但集中汇总不能及时满足地方或基层领导的需要，出现差错也不易查找和纠正。

综合汇总即将逐级汇总和集中汇总两种汇总形式结合起来使用。它是根据各级管理工作的需要，对一部分地区或主管部门所需要的基本资料，实行逐级汇总，以满足各级

主管部门的需要，而将需要在全国范围内进行加工的资料，实行集中汇总，以保证资料的及时使用。综合汇总兼有两种汇总形式的优点，弥补了相互的不足，但耗费人力、物力、财力较大。

统计资料的汇总技术有手工汇总和计算机汇总两种。

2.2.7 统计表和统计图

数据整理的结果最终表现为统计表和统计图。

1. 统计表

统计表是以纵横垂直交叉的直线绘制的用来表现统计数据资料的表格。统计表能清楚地表述统计资料的内容，使人一目了然。利用统计表便于计算，也便于分析各项目之间的关系。

统计表从外形结构看，一般包括表头、行标题、列标题和数据。表头就是统计表的名称，一般位于表的上端中央。行标题就是横行的名称，用来表示总体各组的名称，一般写在表的左方。列标题就是纵栏的名称，用来表示统计指标的名称，一般写在表的上方。数据就是各行标题与列标题的交叉处所填列的数值，是统计表的主要内容。统计表的注释一般放在表的下端，说明统计资料的来源和调查方法。

统计表的构成如表 2—7 所示。

表 2—7		2005—2007 年我国国内生产总值			单位：亿元	
行标题	年份	第一产业	第二产业	第三产业	合计	列标题
	2005	22 420	87 365	73 433	183 217	数据
	2006	24 040	103 162	84 721	211 923	
	2007	28 095	121 381	100 054	249 530	
本表按当年价格计算。 (资料来源：中国统计年鉴2008，北京：中国统计出版社，2008)						注释

统计表在设计时应注意下列事项：

- (1) 在使用电子计算机的条件下，横向宽度要符合打印机的型号；在手工整理的条件下，统计表应设计成纵横交叉线条组成的长方形表格，长与宽之间保持适当的比例。
- (2) 线条的绘制。表的上下两端应以粗线绘制，表内纵横线以细线绘制，表格的左右两端一般不画线，采用“开口式”。
- (3) 合计栏的设置。统计表各纵栏需要合计时，一般应将合计列在最后一行；各横行需要合计时，可将合计列在最前一栏或最后一栏。
- (4) 标题设计。无论是总标题，还是行标题和列标题都应简明扼要，简要表述统计资料的内容及所属的时间和空间范围。
- (5) 指标数值。表中数字应填写整齐，对准位数。当数字为 0 或因数字太小而忽略不计时，可填写为“0”；当缺某项数字资料时，可用符号“...”表示；不应有数字时用符号“—”表示；如有相同数字应全部重写，切忌写“同上”、“同左”等字样。
- (6) 计量单位。统计表必须注明数字资料的计量单位。当全表只有一种计量单位时，可以把它写在表头的右上方，如果表中各栏的指标数值计量单位不同，可在行标题

后添一列计量单位。

(7) 注解与资料来源。为保证统计资料的科学性与严肃性,在统计表下,应注明资料来源,以便考证。

2. 统计图

统计图是以直观、形象的图形将数据的分布特征呈现出来的重要辅助工具。

常用的统计图有柱形图、饼图、直方图和折线图。

柱形图是以一簇宽度相等、相互分离的条状图形的长度(或高度)来表示次数分布的统计图。柱形图中条状图形的长度(或高度)所表示的数据可以是次数,也可以是次数的相对比例。柱形图图形为落在一个直角坐标系中的条状或柱状图形,这个直角坐标系可以是二维的,也可以是三维的。相应的条状或柱状图形为平面或立体的。直角坐标系的原点的具体数值可以为0,也可以为其他特定数值,这需要根据实际情况而定。图2—5是根据表2—7绘制的柱形图。

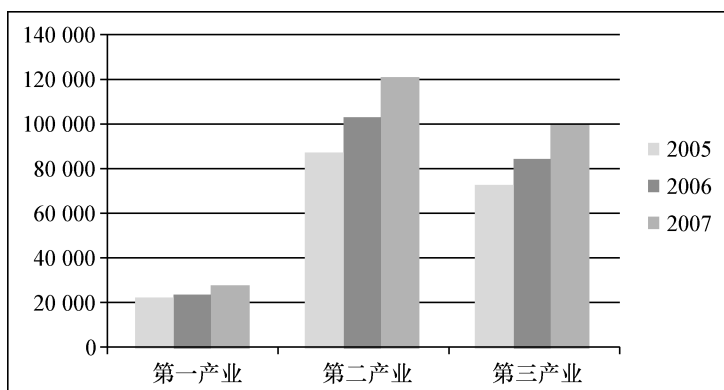


图 2—5 2005—2007 年我国国内生产总值柱形图

饼图是以同一圆形内一簇扇形的面积的大小来表示数值分布的统计图。图2—6是根据表2—7中我国2007年国内生产总值数据绘制的饼图。

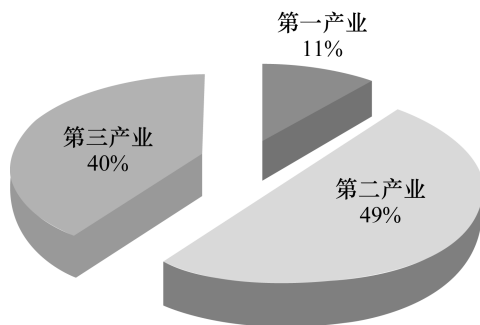


图 2—6 2007 年我国国内生产总值饼图

直方图是以各组的组距为宽,以各组的次数为高,在直角坐标系的第一象限依次绘制一系列矩形来表示次数分布状态的统计图。图2—7是根据表2—6绘制的直方图。

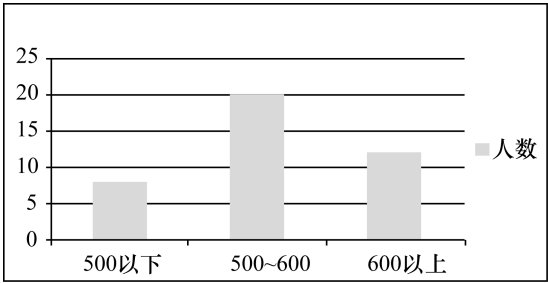


图 2—7 某生产小组工人日产量分布直方图

折线图是将各组的组中值和次数在直角坐标系的点，用一条折线联系起来，以反映次数分布状态的统计图。

附录：Excel 应用

某学期某班 35 名学生的统计学考试成绩的原始数据如下：

92 87 75 74 65 91 98 78 84 83 78 96 62 84 76 52 69 84
81 88 70 91 89 87 75 79 82 86 79 82 64 90 56 84 89

要求：对该班学生的统计学考试成绩进行等距分组、计算组中值和次数分布，编制统计表和统计图。

解：运用本章介绍的数据整理方法，按以下步骤进行计算。

第一步，计算取值范围。

可以利用 Excel “分析工具库” 中的 “描述统计” 工具，直接计算出原始数据的取值范围及其他测度。在 “工具” 菜单中，单击 “数据分析” 命令，调出 “分析工具” 列表框。然后如图 2—8 所示，在 “分析工具” 列表框中，选中需要使用的 “描述统计” 工具。



图 2—8 调用 “描述统计” 工具

在 “描述统计” 对话框（见图 2—9）中的 “输入区域” 中输入原始数据所在的单元格，在本例中 35 名学生的统计学考试成绩的原始数据位于 A1 到 A35 单元格；“分组方式” 选择 “逐列”；在 “输出选项” 中，若希望运算结果输出在当前工作表上，选定 “输出区域”，并在窗口中输入输出计算结果的起始位置。在本例中选择了运算结果输出在当前工作表上，输出运算结果的起始位置为 D12。若需要将运算结果输出到新的工作表，可选定相应选项。最后选定含有取值范围数值的 “汇总统计”。



图 2—9 “描述统计”对话框

用鼠标左键单击“确定”，完成取值范围的计算，计算机输出计算结果如图 2—10 所示。其中“区域”即取值范围为 46 分，也可以由图中的“最大值”98 分和“最小值”52 分，计算出取值范围 $R=98-52=46$ 分。

文件(F) 编辑(E) 视图(V) 插入(I) 格式(O) 工具(T)				
宋体 12 B I U				
A1				92
A	B	C	D	
1	92			
2	87	列1		
3	75			
4	74	平均		80
5	65	标准误差		1.834772
6	91	中位数		82
7	98	众数		84
8	78	标准差		10.85465
9	84	方差		117.8235
10	83	峰度		0.381472
11	78	偏度		-0.81631
12	96	区域		46
13	62	最小值		52
14	84	最大值		98
15	76	求和		2800
16	52	观测数		35

图 2—10 输出的计算结果

第二步，确定组数和组距。

由确定组数的经验公式，可计算出参考的组数 $H=1+\lg 35/\lg 2=6$ 。由于人们习惯

于将考试成绩分为 5 类，所以分成 5 组比较合适。

组数确定之后，可以根据取值范围、组数和组距之间的数量关系，计算出组距的取值 $d=46/5=9.11$ 分，取整之后组距为 10 分，这也符合人们的习惯要求。

第三步，计算次数分布。

计算次数分布，就是按照分组将原始数据一一分配到各组中，然后计算落在各组中数据的个数。这里，采用了 Excel 的 COUNTIF 函数。

COUNTIF 是计算给定区域内满足特定条件的单元格数目的函数。COUNTIF 函数的语法为 COUNTIF (range, criteria)。其中，range 为需要计算其中满足条件的单元格数目的单元格区域；criteria 为确定哪些单元格将被计算在内的条件，其形式可以为数字、表达式或文本。在本例中，需要计算其中是否满足条件的单元格即为原始数据所在单元格 A1 到 A35；评判标准为各分组的上限和下限。为了简洁起见，本例中的评判标准为小于各组上限的单边条件，满足了数据分组“上限不在内”的约定，计算出来的结果是向上累计的人数。

首先，给定 COUNTIF 函数的评判标准，在本例中为小于各组上限。

其次，给出 COUNTIF 函数的相关参数数值，在对应“60 分以下”组的“累计人数”一栏中，即单元格 G4 中，编写的 COUNTIF 函数“=COUNTIF (A1 : A35, “<60”)”，单元格 A1 到 A35 为待分组的原始数据；“各组取值”各栏给出了对应各组的取值上限，其中，“60 分以下”组的取值上限为 60 分。

再次，用鼠标拖动填充柄经过需要填充相同 COUNTIF 函数公式的单元格 D20 到 D24，释放鼠标按键，完成各组向上累计人数的运算。向上累计人数是累计次数数值，通过对累计次数计算的逆运算，可以计算出各组次数数值。

最后，在各组次数的基础上，将各组人数除以全班总人数，可以计算出各组人数占全班总人数的比重，即次数的相对形式——频率。在对应于“60 分以下”组的“频率”一栏中，即单元格 F20 中，编写计算公式“=E20/D24”，单元格 E20 为该组次数数值 2 人。运算结果是该组人数占全班总人数的比重为 0.06（见图 2—11）。

文件(F) 编辑(E) 视图(V) 插入(I) 格式(O) 工具(T) 数据(D) 窗口(W) 帮助						
宋体 12 B I U 背景色 格式刷 求和 平均值 最大值 最小值 条件格式 数据有效性 数据透视表 数据查询 数据工具 窗口 帮助						
E27	fx					
	A	B	C	D	E	F
17	69					
18	84					
19	81		各组上限	累计频数	频数	频率
20	88		60	2	2	0.06
21	70		70	6	4	0.11
22	91		80	15	9	0.26
23	89		90	29	14	0.40
24	87		100	35	6	0.17
25	75					

图 2—11 COUNTIF 函数的应用