

统计学基础

主编◎孔里明 陈 践



TONGJIXUE
JICHU

统计学基础

主 编 孔里明 陈 践

中国人民大学出版社
· 北京 ·

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。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。

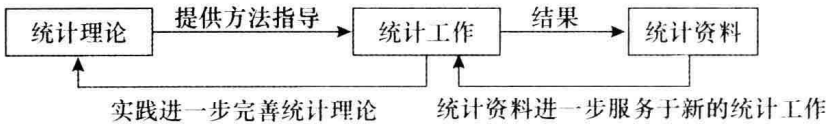


图 1—1 统计理论、统计工作和统计资料关系图

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、研究领域的不同,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

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。

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。指标含义要明确总体现象的质的规定性，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。

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。()

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

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

统计报表是按照国家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

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: