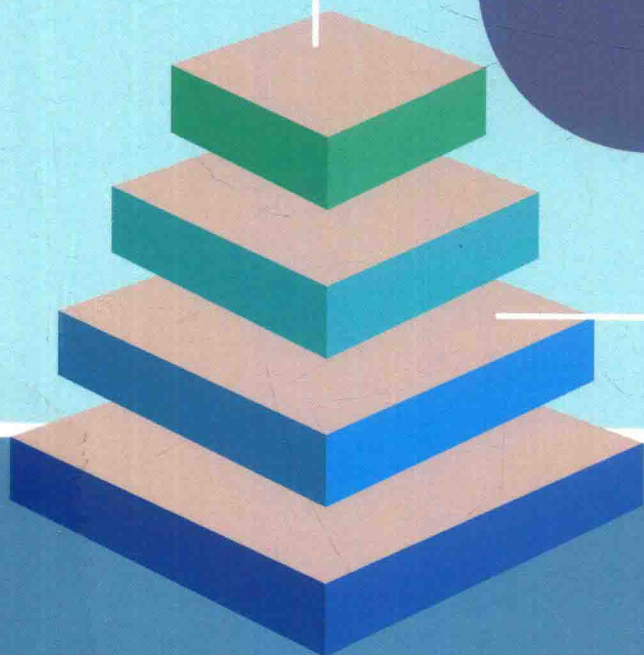
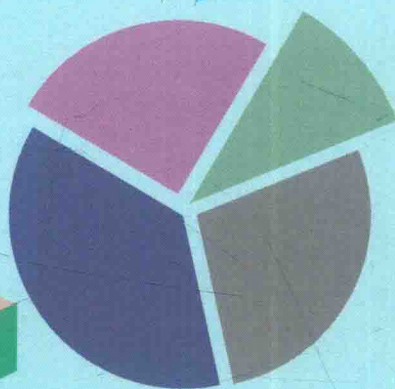


统计学原理

丁东洋 周丽莉 / 编



非外借



科学出版社

统计学原理

丁东洋 周丽莉 编

科学出版社

北京

内 容 简 介

本书系统介绍了统计学的基础理论及应用, 主要内容包括数据的收集、数据的整理与显示、数据分布特征的描述、抽样分布与推断、 χ^2 检验与方差分析、相关与回归分析、时间序列分析与统计指数等。另外, 对 SPSS 的基本应用在部分章节有所介绍, 同时选择了典型的例题和习题揭示统计学的应用。

本书通俗易懂, 可作为高等院校相关专业统计学课程的教材或教学参考书。

图书在版编目 (CIP) 数据

统计学原理 / 丁东洋, 周丽莉编. —北京: 科学出版社, 2017.3

ISBN 978-7-03-051012-9

I. ①统… II. ①丁… ②周… III. ①统计学

IV. ①C8

中国版本图书馆 CIP 数据核字 (2016) 第 302079 号

责任编辑: 滕亚帆 田轶静 / 责任校对: 郑金红
责任印制: 张 伟 / 封面设计: 华路天然设计工作室

科学出版社 出版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.science.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

*

2017 年 3 月第 一 版 开本: 787×1092 1/16

2017 年 3 月第一次印刷 印张: 14 3/4

字数: 340 000

定价: 41.00 元

(如有印装质量问题, 我社负责调换)

前 言

英国著名作家威尔斯 (H. G. Wells, 1866~1946) 曾说过, “统计思维总有一天会像读与写一样成为一个有效率公民的必备能力”。这一论断在今天早已成为现实。尤其是以各个不同领域的数据和变量为具体研究对象的应用统计学, 充分体现了统计方法应用的广泛性, 统计学已经成为必备的工具和技能之一。因而本书的定位也是面向非统计专业的经济管理类专业用书。

统计学是经济管理类专业的核心课程。而从课程设置方面看, 非统计专业的统计学教学课时一般为 48 课时, 在培养的过程中, 相应的教学模式主要是培养学生掌握通用的统计方法和理论, 并且能够熟练掌握常用的统计软件完成相关的应用方法, 进而达到适应自身所在领域数据分析工作的培养目标。多年的教学经验告诉我们, 一方面, 统计学教材中内容过多, 在有限的学时内无法完成教学; 另一方面, 需要一定的课时讲解统计软件以强调和实现统计学的应用性。尤其对于非统计专业的学生, 其中一些专业通常不会再开设一门讲解统计软件的课程, 而且统计软件的学习需要跟随学习的内容做大量的实际演练, 所以在统计教材相应的章节中简要介绍统计软件的操作在一定程度上能够提高学习的效率, 在带领学生学习软件操作的同时, 达到统计方法实用性的培养目标。

本书的编写是依据统计学教学大纲, 在总结多年的教学经验基础上, 参阅大量的国内外相关书籍完成的。从章节的结构上看, 本书在参照已有教材的同时, 做了一定的调整。一是按照统计学的数据收集、整理、分析和解释的框架设置章节, 二是按照变量间关系的组合选择统计方法的讲授, 例如, 分析定性变量和定性变量间关系的 χ^2 检验, 分析定性变量和定量变量间关系的方差分析, 分析定量变量和定量变量间关系的回归分析等内容在本书中都有涉及。从内容的难度上看, 本书属于初级统计学的水平, 数学方法及概率论知识缺乏的同学也能够阅读和掌握书中方法。

在此感谢南昌大学教材出版资助和南昌大学廉政研究中心的资助。特别感谢南昌大学公共管理学系白茹、冯春丽、梁议心、王超、杨瑞等同学在数据收集和习题整理等方面所做的工作。

由于编者水平有限, 书中难免有不完善之处, 诚望读者批评指正。

编 者

2016 年 9 月

目 录

第一章 导论	1
第一节 什么是统计学	1
第二节 统计学的基本概念	4
思考与练习	7
第二章 数据的收集	8
第一节 变量和数据	8
第二节 数据的来源	11
第三节 调查问卷	20
第四节 SPSS 的基本说明	24
思考与练习	28
第三章 数据的整理与显示	29
第一节 数据的整理	29
第二节 数据的显示	36
第三节 SPSS 的数据文件管理	44
思考与练习	48
第四章 数据分布特征的描述	50
第一节 分布的集中趋势	50
第二节 分布的离散趋势	60
第三节 分布的形状	65
第四节 SPSS 的描述统计	66
思考与练习	75
第五章 抽样分布与推断	77
第一节 抽样分布	77
第二节 总体参数估计	89
第三节 假设检验	100
第四节 SPSS 的假设检验	108
思考与练习	113
第六章 χ^2 检验与方差分析	116
第一节 χ^2 检验	116
第二节 方差分析	119
第三节 SPSS 的 χ^2 检验与方差分析	132
思考与练习	140

第七章 相关与回归分析	143
第一节 相关分析	143
第二节 一元线性回归分析	149
第三节 多元线性回归及非线性回归	158
第四节 SPSS 的相关与回归分析	163
思考与练习	168
第八章 时间序列分析	170
第一节 时间序列分析概述	170
第二节 时间序列的水平指标分析	174
第三节 时间序列的速度指标分析	179
第四节 时间序列的分解分析	182
思考与练习	193
第九章 统计指数	195
第一节 统计指数概述	195
第二节 综合指数的编制与应用	198
第三节 平均指数的编制与应用	203
第四节 指数体系与因素分析	209
思考与练习	216
参考文献	218
附录	219

第一章 导 论

第一节 什么是统计学

一、统计学的含义和发展历程

统计学 (statistics) 是研究如何收集数据、整理、分析、解释数据,并由数据得出结论的学科。数据的收集是取得数据的过程。数据的整理是对数据进行加工处理的过程,主要是指用统计图表展示数据。数据的分析和解释是核心内容,指的是通过统计描述和统计推断的方法探索数据的内在规律性,进而对结果进行解释说明,并从数据分析中得出客观结论。

统计的本质就是关于统计目的、统计对象和统计方法的思想,从 17 世纪中末期开始,经过三个多世纪的发展,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

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,主要通过通过对国家组织、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)。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)。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。

(二) 近代统计学时期

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定律和概率论被引进统计学。之后,最小二乘法、误差理论和正态分布理论等相继成为统计学的重要内容。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)。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。

(三) 现代统计学时期

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,特别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 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,波兰统计学家奈曼(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。

在中国,由于其独特的历史环境,一般将统计分为文科意义上的“统计学”和理科意义上的“数理统计学”。这是由于新中国成立初期的统计学被认为主要用于国民经济核算和计划经济服务,随着统计分析工具应用的推广和深入,文科意义上的统计学已经逐渐淡化。尤其是2011年中华人民共和国国务院学位委员会在新的研究生专业目录中将统计学上升为一级学科,标志着中国的统计学与国际通用的词汇“Statistics”基本一致。

二、统计学的应用领域

一般来说,统计学的研究对象是数据,而数据可以来源于自然科学领域,也可以来源于社会科学领域,因而统计学的应用领域非常广泛。美国著名统计学家萨维奇(L. J. Savage, 1917~1971)曾经说过,“统计学基本上是寄生的。靠研究其他领域内的工作而生存。这不是对统计学的轻视,这是因为对很多寄主来说,如果没有寄生虫就会死”,这句话形象地说明了统计学学科的应用广泛性。在很多领域,统计学与实质性科学相结合,形成了众多的交叉学科或边缘学科。如生物统计学、商务统计学、统计物理学、统计天文学等。有些学科在名称中没有“统计”字样,但其方法主体都是统计,如经济计量学、数据挖掘等。总体上看,各统计分支学科大体都基于一个共同的内核,这就是统计学的基本思想和基本方法。而在各个领域,统计学的研究对象具有相同的特点。

(1) 数量性。数量性是统计学研究对象的基本特点。统计数据是客观事物量的反映,通过数据来测度事物的类型、量的顺序、量的大小和量的关系。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。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,可以达到我们统计分析研究的目的。

(2) 总体性。统计研究虽然是从个别入手,对个别单位的具体事实进行观察研究,但其目的是达到认识总体数量特征。自然、社会经济现象的数据资料和数量对比关系等一般是

在一系列复杂因素的影响下形成的。在这些因素中，有起着决定和普遍作用的主要因素，也有起着偶然和局部作用的次要因素。由于种种原因，在不同的个体中，它们相结合的方式和实际发生的作用都不可能完全相同。所以，对于每个个体，就具有一定的随机性质，而对有足够多的个体的总体来说又具有相对稳定的共同趋势，显示出一定的规律性。统计研究对象的总体性，是从个体的实际表现的研究过渡到对总体的数量表现的研究的。例如，进行居民入户调查，虽然是对具体的每个调查户进行观察，但其目的是要反映一个地区居民的收入和消费等状况。

(3) 变异性。统计研究的是同质总体的数量特征，其前提是总体各单位特征表现存在着差异，而这些差异不是由某些特定的原因事先给定的。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，还必须要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。例如，在校学生作为统计数据资料对象，每个学生在性别、年龄、成绩等方面是会有不同表现的。这样，统计分析研究才能对其表现出来的差异探索统计规律性。

三、统计学的分科

统计方法已被应用到自然科学和社会科学的众多领域，统计学也发展成为由若干分支学科组成的学科体系。从统计方法的构成来看，统计学可以分为描述统计学 (descriptive statistics) 和推断统计学 (inferential statistics)；从统计研究的重点来看，统计学可以分为理论统计学 (theoretical statistics) 和应用统计学 (applied statistics)。

(一) 描述统计学和推断统计学

描述统计学是指收集由试验或调查所获得的资料，进行整理、归类，计算出各种用于说明总体数量特征的数据，并用图形或表格的形式将它们显示出来。统计描述的主要内容包括数据的收集、数据的加工处理、数据的显示方法、数据分布特征的概括与分析。

推断统计学是指利用概率论的理论，根据试验或调查获得的样本信息科学地推断总体的数量特征。统计推断可以用于对总体数量特征的估计，也可以用于对总体某些假设的检验，所以，统计推断方法包括参数估计法和假设检验法。

描述统计学和推断统计学的划分一方面反映了统计方法发展的前后两个阶段，另一方面也反映了应用统计方法探索客观事物数量规律性的不同过程。统计研究过程的起点是统计数据，终点是探索出客观现象内在的数量规律性。在这一过程中，如果收集到的是总体数据 (如普查数据)，则经过描述统计之后就可以达到认识总体数量规律性的目的；如果所获得的只是研究总体的一部分数据 (样本数据)，那么要找到总体的数量规律性，就必须应用概率论的理论并根据样本信息对总体进行科学的推断。

显然，描述统计和推断统计是统计方法的两个组成部分。描述统计是整个统计学的基础，推断统计则是现代统计学的主要内容。由于在对现实问题的研究中，所获得的数据主要是样本数据，所以推断统计在现代统计学中的地位和作用越来越重要，已成为统计学的

核心内容。当然,这并不等于说描述统计不重要,如果没有描述统计收集可靠的统计数据并提供有效的样本信息,即使再科学的统计推断方法也难以得出切合实际的结论。描述统计学到推断统计学的发展过程,既反映了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

理论统计学主要研究统计学的一般理论和统计方法的数学理论。由于现代统计学用到了几乎所有方面的数学知识,从事统计理论和方法研究的人员需要有坚实的数学基础。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今天这样一个完善的科学知识体系。

应用统计学是研究如何应用统计方法去解决实际问题的。统计学是一门收集和分析数据的科学。由于在自然科学及社会科学研究领域中,都需要通过数据分析来解决实际问题,因而,统计方法的应用几乎扩展到所有的科学研究领域,如农业、动物学、人类学、考古学、审计学、晶体学、人口学、牙医学、生态学、经济学、教育学等。

第二节 统计学的基本概念

一、统计总体与样本

统计总体就是根据一定的目的确定的所要研究对象的全体。统计总体是由客观存在的、具有某种共同性质的许多个别单位所构成的整体。构成总体的这些个别单位称为总体单位。例如,所有的工业企业就是一个总体,这是因为在性质上每个工业企业的经济职能是相同的,即都是从事工业生产活动的基本单位,统计上称之为同质性。这些工业企业的集合就构成了统计总体。对于该总体,每一个工业企业就是一个总体单位。统计总体的另一个特点是大量性,统计总体是由具有某种共同性质的许多个别单位所构成的整体,那么,所研究的单位就不是个别的或少量的。

总体可以分为有限总体和无限总体。总体所包含的单位数是有限的,称为有限总体,如在某一时间段内进行的人口普查、库存物资普查。总体所包含的单位数是无限的,称为无限总体,如连续生产的某种产品的生产数量、大海里的鱼资源数等。对有限总体可以进行全面调查,也可以进行非全面调查。但对无限总体只能抽取一部分单位进行非全面调查,据以推断总体。

样本是从总体抽取出的,由部分单位组成的集合体,是总体的代表。在抽样推断中,总体又被称为母体,相应地,样本也被称为子样。抽取样本应注意样本的单位必须取自总体,这是因为抽取样本的目的是推断总体。而且一个总体可以抽取许多样本,样本个数的多少与抽样方法直接相关。

二、统计标志

(一) 标志

总体各单位所具有的属性或特征称为标志。从不同角度考察,每个总体单位可以有許多特征,如每个职工可以有性别、年龄、民族、工种等特征,这些都是职工的标志。总体是由单位构成的,单位是标志的承担者。进行统计研究就是从观察、登记标志的表现入手的,并对标志的表现进行综合以揭示总体的数量特征,由此可见,标志是统计研究的基础。

(二) 标志的分类

(1) 按照标志在总体中各单位的具体表现是否相同,可分为不变标志和可变标志。当一个标志在各个单位的具体表现都相同时,这个标志称为不变标志;当一个标志在各个单位的具体表现有可能不同时,这个标志称为可变标志或变异标志。例如,在教师总体中,职业这一标志在各单位的表现都是相同的,都是教师,所以职业就是不变标志。一个总体中,至少要有一个不变标志,才能把各单位结合成为一个总体。而教师的性别、年龄、民族等则是变异标志。变异标志是统计研究的主要内容,因为如果标志在各总体单位之间的表现都相同,那就没有进行统计研究的必要了。因而总体的同质性是问题研究的基础,而总体的变异性则是问题研究的本体。

(2) 按照标志所反映单位的不同特征,可以分为品质标志和数量标志。品质标志表示事物质的特性,是不能用数值表示的,如职工的性别、民族、工种等。数量标志表示事物量的特性,是可以数值表示的,如职工年龄、工资、工龄等。品质标志主要用于分组,将性质不相同的总体单位划分开来,便于计算各组的总体单位数,计算结构和比例指标。数量标志既可用于分组,也可用于计算标志总量及其他各种质量指标。

三、统计指标

(一) 统计指标的含义

统计指标是反映统计总体数量特征的概念与相应数值的。例如,2015年我国国内生产总值为67.67万亿元,这就是统计指标。每项指标都由两项要素构成,即指标的名称和指标的取值。指标的名称是对所研究现象本质的抽象概括,即对研究现象总体数量特征的质的规定性。指标的数值是反映所研究现象在具体时间、地点、条件下的规模、水平或比率关系等,所以指标数值必须包括时间状态、空间范围、计量单位、计算方法的限定,不得随意变动。

指标是反映总体的数量特征,标志是反映单位的属性特征,两者的关系如同总体与单位的关系一样,总体是由单位构成,而指标是通过各标志的具体表现综合汇总而得的。由个体过渡到总体,由标志过渡到指标,是人们对现象认识的深化和发展。只有对大量个体

标志的综合而形成统计指标，并通过统计指标获得个体单位难以显现的信息，才能揭示现象的本质属性和数量特征。

(二) 统计指标的分类

统计指标按其说明总体内容的不同分为数量指标和质量指标。

(1) 数量指标。是指说明总体外延规模的统计指标，如人口数、企业数、工资总额、商品销售额等。数量指标所反映的是总体的绝对数量，具有实物的或货币的计量单位，其数值的大小，随着总体范围的变化而变化，它是认识总体现象的基础指标。

(2) 质量指标。是指说明总体内部数量关系和总体单位水平的统计指标。例如，人口的年龄构成、性别比例、农业——轻工业——重工业比例、平均单产、平均工资等。它通常是用相对数和平均数的形式表现的，其数值的大小与范围的变化没有直接关系。

统计指标按其作用和表现形式的不同，可分为总量指标、相对指标和平均指标。总量指标又分为实物指标、劳动指标和价值指标三种。这些统计指标的涵义、内容、计算方法和作用各不相同，将在第九章中详细叙述。

(三) 统计指标体系

统计指标是反映总体数量特征的概念，但单个指标只能说明现象的某一侧面的情况，若要全面反映客观现象的全貌，描述现象发展的全过程，就必须设置统计指标体系，利用统计指标体系，以全面刻画现象的本质特征及其现象各方面相互依存、相互制约的关系。

统计指标体系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。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。

根据所研究问题的范围大小，可以建立宏观统计指标体系和微观统计指标体系。宏观统计指标体系就是反映整个现象大范围的统计指标体系。例如，反映整个国民经济和社会发展的统计指标体系。微观统计指标体系就是反映现象较小范围的统计指标体系，如反映企业或事业单位的统计指标体系。介于这两者之间的可以称为中观统计指标体系，如反映各地区或各部门的统计指标体系。

根据所反映现象的不同范围和内容，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，例如，经济效益指标体系就是专题性统计指标体系。

统计指标体系也可以指若干个统计指标之间的联系表现为一个方程关系。例如， $\text{工资总额} = \text{平均工资} \times \text{职工人数}$ ； $\text{商品销售额} = \text{商品销售量} \times \text{商品销售价格}$ ；等等。统计指标体系对于统计分析和研究具有重要的意义。通过一个设计科学的统计指标体系，可以描述现象的全貌和发展的全过程，分析和研究现象总体存在的矛盾以及各种因素对现象总体

变动结果的方向和程度,也可以对未来的指标进行计算和预测,对未来现象发展变化的趋势进行预测。

思考与练习

一、思考题

- (1) 怎样理解统计的不同含义?
- (2) 统计学的含义是什么? 怎样理解统计学与统计数据的关系?
- (3) 统计学的研究对象是什么? 统计学研究对象有哪些特点?
- (4) 怎样认识描述统计学与推断统计学之间的联系与差别?
- (5) 怎样区别统计总体与总体单位?
- (6) 什么是统计标志? 标志有哪些分类? 品质标志和数量标志有什么区别?
- (7) 统计指标的含义及其特点是什么?
- (8) 标志与指标的区别与联系是什么?
- (9) 什么是数量指标、质量指标?
- (10) 什么是统计指标体系? 有哪些分类?

二、练习题

(1) 某班级共有 50 名学生,其中男生 30 人、女生 20 人。男生平均身高是 170cm,女生平均身高是 160cm。学校想了解该班全体学生的身体状况,分别对每名学生的肺活量、身高、体重、视力、跳远进行了测量。

回答以下问题:

- 1) 描述其统计总体和总体单位。
 - 2) 指出哪些是品质标志? 哪些是数量标志?
- (2) 要调查某地区销售的电视机品牌情况,试指出总体、总体单位是什么?

第二章 数据的收集

要从数量上认识客观对象，就必须通过调查或实验来收集数据。要收集什么数据，采用哪种方法去收集数据，首先需要界定变量，明确数据类型。

第一节 变量和数据

一、变量

统计中的标志和指标都是可变的，例如，人的性别有男女之分，各时期、各地区、各部门的工业总产值各有不同等，这种差别叫做变异。变异就是有差别的意思，包括质的差别和量的差别。变异是统计的前提条件。

变量就是可以取不同值的量，这是数学上的一个名词。在社会经济统计中，变量包括各种数量标志和全部统计指标，它都是以数值表示的，不包括品质标志。变量就是数量标志的名称或指标的名称，变量的具体数值表现则称为变量值。例如，职工人数是一个变量，因为各个工厂的职工人数不同。某工厂有 852 人，另一工厂有 1686 人，第三个工厂有 964 人，等等，都是职工人数这个变量的具体数值，也就是变量值。要注意区分变量和变量值。如上例，852 人、1686 人、964 人三个变量值的平均数，不能说是三个“变量”的平均数，因为这里只有“职工人数”这一个变量，并没有三个变量。以整数数值变化的变量，称为离散型变量；也可以有连续数值变化的变量，即可以用小数值表示的变量，称为连续型变量。离散型变量的各变量值之间是以整数位断开的，如人数、机器台数、工厂数等，都只能按整数计算；连续型变量的数值是接连不断的，相邻的两数值之间可作无限分割，如身高、体重、年龄等。

变量值按是否连续可分为连续变量与离散变量两种。在一定区间内可任意取值的变量叫连续变量，其数值是连续不断的，相邻两个数值可作无限分割，即可取无限个数值。例如，生产零件的规格尺寸、人体测量的身高、体重、胸围等为连续变量，其数值只能用测量或计量的方法取得。可按一定顺序一一列举其数值的变量叫做离散变量，其数值表现为断开的。例如，企业个数、职工人数、设备台数、学校数、医院数等，都只能按计量单位计数，这种变量的数值一般用计数方法取得。

二、数据

(一) 数据的计量层次

统计数据是对客观现象进行计量的结果。对客观现象进行计量，就必须弄清楚数的计

量尺度问题。根据对研究对象计量的不同精确程度,将计量层次由低到高,由粗略到精确分为四个层次:定类尺度、定序尺度、定距尺度和定比尺度。

(1) 定类尺度。将数字作为现象总体中不同类别或不同组别的代码,这是最低层次的尺度。在这种情况下,不同的数字仅表示不同类(组)别的品质差别,而不表示它们之间量的顺序或量的大小。这种尺度的主要数学特征是“=”或“≠”。例如,将国民经济按其经济类型,可以分为国有经济、集体经济、私营经济、个体经济等,并用(01)代码表示国有经济,(02)表示集体经济,(03)表示私营经济,(04)表示个体经济。并且用(011)代表国有经济中的国有企业,(012)代表国有联营企业;用(021)表示集体经济中集体企业,(022)表示集体联营企业;用(031)表示私营经济中的私营独资企业,(032)表示私人合伙企业,(033)表示私营有限责任公司;用(041)表示个体经济中的个体工商户,(042)表示个人合伙等。其中两位代码表示经济大类,而三位代码则表示各类中的构成。不同代码反映同一水平的各类(组)别,并不反映其大小顺序。各类中虽然可以计算它的单位数,但不能反映第一类的一个单位可以相当于第二类的几个单位等。

(2) 定序尺度。定序尺度不但可以用数表示量的不同类(组)别,而且也反映量的大小顺序关系,从而可以列出各单位、各类(组)的次序。这种尺度的主要数学特征是“>”或“<”。例如,对合格产品按其性能和好坏,分成优等品、一等品、合格品等。这种尺度虽然也不能表明一个单位一等品等于几个单位二等品,但却明确表示一等品性能高于二等品,而二等品性能又高于三等品等。定序尺度除了用于分类(组)外,在变量数列分析中还可以确定中位数、四分位数、众数等指标的位置。

(3) 定距尺度。定距尺度也称间隔尺度,是对事物类别或次序之间间距的计量,它通常使用自然或度量衡单位作为计量尺度。定距尺度是比定序尺度高一层次的计量尺度。定距尺度的计量结果表现为数值,可以进行加或减的运算,但却不能进行乘或除的运算,其原因是在等级序列中没有固定的,有确定意义的“零”值。例如,我们不能说 40°C 比 20°C 暖和2倍,温度中没有真实零,而且 0°C 可以转换为 32°F 。

(4) 定比尺度。定比尺度是在定距尺度的基础上,确定可以作为比较的基数,将两种相关的数加以对比,而形成新的相对数,用以反映现象的构成、比重、速度、密度等数量关系。由于它是在比较基数上形成的尺度,所以能够显示更加深刻的意义。定比尺度的主要数学特征是“÷”或“×”。例如,将某地区人口数和土地面积进行对比来计算人口密度指标,说明人口相对的密集程度。甲地区人口可能比乙地区多,但甲地区的土地更广阔,用人口密度指标就可以相对说明甲地区人口不是多了,而是少了。又如将一个国家(地区)的国内生产总值与该国(地区)居民对比。计算人均国内生产总值,可以反映国家(地区)的综合经济能力。

上述四种计量尺度对事物的计量层次是由低级到高级,由粗略到精确逐步递进的。高层次的计量尺度具有低层次计量尺度的全部特性,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计量尺度的测量结果,比如将考试成绩的百分制转化为五等级分制。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。

(二) 数据的类型

(1) 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。统计数据是采用某种计量尺度对事物进行计量的结果,采用不同的计量尺度会得到不同类型的统计数据。从上述四种计量尺度计量的结果来看,可以将统计数据分为四种类型:分类数据、顺序数据、定距数据和定比数据。前两类数据说明的是事物的品质特征,不能用数据表示,其结果均表现为类别,也称为定性数据或品质数据(qualitative data);后两类数据说明的是现象的数量特征,能够用数值来表现,因此也称为定量数据或数值型数据(quantitative data)。由于定距尺度和定比尺度属于同一测度层次,所以可以把后两种数据看成同一类数据,统称为定量数据或数值型数据。

区分测量的层次和数据的类型是十分重要的,因为对不同类型的数据需要采用不同的统计方法来处理和分析。比如,对定类数据,通常计算出各组的频数或频率,计算其众数和异众比率,进行列联表分析和 χ^2 检验等;对定序数据,可以计算其中位数和四分位差,计算等级相关系数等非参数分析;对定距或定比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、进行参数估计和检验等。我们所处理的大多为数量数据。

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,因为后者具有前者的数学特性。比如,在描述数据的集中趋势时,对定类数据通常是计算众数,对定序数据通常是计算中位数,但对定距和定比数据同样也可以计算众数和中位数。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,因为低层次数据不具有高层次测量数据的数学特性。比如,对于定距和定比数据可以计算平均数,但对于定类数据和定序数据则不能计算平均数。理解这一点,对于选择统计分析方法是十分有用的。

(2) 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数量特征,有明确的计量单位。例如,人的身高 176cm、体重 65kg、地区的人口数 500 万人、属地面积 11000km²、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(如某地区三大产业比重分别为 12%, 48%和 40%),比例相对数(如新生儿男女性别比 107:100),比较相对数(如 A 地区的人均 GDP 是 B 地区的 1.2 倍),动态相对数(如某地区 GDP 的发展速度为 109%),强度相对数(如某地区的人口密度为 300 人/km²,人口出生率为 11%)和计划完成程度相对数(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。

(3) 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, 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多属于实验数据, 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。

(4) 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。

(5) 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, 就成了平行数据。

第二节 数据的来源

统计数据收集是根据统计研究预定的目的要求和任务, 运用科学的调查方法与手段, 有计划、有组织地向客观实际收集数字资料的过程。从统计工作的性质看, 统计数据的收集是社会调查的组成部分; 从统计工作过程的阶段性看, 统计数据的收集处于统计工作过程的基础阶段。统计数据主要来源于两种渠道: 一是来源于直接的调查和科学试验, 对使用者来说, 这是统计数据的直接来源, 称为第一手或直接的统计数据; 二是来源于别