

21世纪普通高等院校系列规划教材

TONGJIXUE

统计学

主 编 张春国
副主编 甘伦知



西南财经大学出版社
Southwestern University of Finance & Economics Press

TONGJIXUE

统 计 学

主 编 张春国
副主编 甘伦知



西南财经大学出版社
Southwestern University of Finance & Economics Press

图书在版编目(CIP)数据

统计学/张春国主编. —成都:西南财经大学出版社,2013. 1
ISBN 978-7-5504-0975-0

I. ①统… II. ①张… 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(2013)第 014955 号

统计学

主 编:张春国

副主编:甘伦知

责任编辑:孙 婧

助理编辑:高 龄 罗艳君 李继琳

封面设计:杨红鹰

责任印制:封俊川

出版发行	西南财经大学出版社(四川省成都市光华村街 55 号)
网 址	http://www.bookcj.com
电子邮件	bookcj@foxmail.com
邮政编码	610074
电 话	028-87353785 87352368
照 排	四川胜翔数码印务设计有限公司
印 刷	四川森林印务有限责任公司
成品尺寸	185mm×260mm
印 张	15.25
字 数	345 千字
版 次	2013 年 1 月第 1 版
印 次	2013 年 1 月第 1 次印刷
印 数	1—3000 册
书 号	ISBN 978-7-5504-0975-0
定 价	30.00 元

1. 版权所有,翻印必究。
2. 如有印刷、装订等差错,可向本社营销部调换。
3. 本书封底无本社数码防伪标志,不得销售。

前言

统计作为研究客观事物数量的手段,被广泛地运用于自然、社会、经济、管理以及科学研究等诸多领域。“统计学”作为经济类以及管理类各专业的基础课程,其研究的内容主要涉及数据的搜集、数据的整理及图表展示、统计数据的分析预测等方面。遵照面向 21 世纪教学内容及课程体系改革的精神,为进一步拓宽专业口径、加强基础课程教学、全面提高教育教学质量、培养理论知识与动手能力并重的素质型人才,我们编写了这本《统计学》教材。由于学时的限制,经济与管理类各专业不可能开设太多的统计学课程,为此,我们认真总结了多年的教学经验,对统计学课程改革进行探索,并将其融入本教材的编写中。

全书包含了描述统计和推断统计的基本内容,共分为 10 章。它包括概论,统计数据的搜集、整理和显示,描述分析的基本指标,概率和抽样分布,参数估计,假设检验,方差分析,相关与回归分析,时间数列分析以及统计指数。其中第 1 章、第 2 章、第 3 章、第 9 章、第 10 章由张春国老师编写,第 4 章、第 5 章、第 6 章、第 7 章、第 8 章由甘伦知老师编写。

在教材编写过程中,我们注重吸收统计改革实践和统计理论研究的最新成果,充分考虑社会实践活动对统计方法的要求,将统计学理论与统计方法应用有机结合,尽可能将数据处理与大众化的 Excel 运用相结合,使读者能够学习到较为系统、实用的数据处理方法与技术,增强教材的实用性。在教材的内容体系上,尽可能做到语言通俗易懂、知识由浅入深、内容前后呼应、内在逻辑性强等特色;同时力求少而精,理论联系实际,努力做到概念准确、条理清晰、层次分明。目的是使读者具备基本的统计思想,掌握基本的数据处理方法,培育和提高读者应用统计方法解决各种实际问题的能力。

感谢甘伦知老师在本书编写过程中提出的宝贵意见!感谢四川理工学院经济与管理学院给予的大力帮助!感谢西南财大出版社有限责任公司对本书的大力支持,特别要感谢编辑孙婧同志对书稿的细心编校!由于水平有限,书中难免存在错误和疏漏之处。敬请读者提出宝贵意见,以使我们在共同努力下,对《统计学》教材和统计教学进行完善和改进。

张春国于四川理工学院

2012 年 11 月

目 录

第一章 概论	(1)
第一节 统计的含义	(1)
第二节 统计数据类型	(9)
第三节 统计学中的几个基本概念	(10)
第二章 统计数据的搜集、整理和显示	(15)
第一节 统计调查	(15)
第二节 统计数据的整理	(23)
第三节 频数分布	(31)
第四节 统计数据的图示	(37)
第五节 统计图与统计表的设计	(47)
第三章 描述分析的基本指标	(52)
第一节 总量指标	(52)
第二节 相对指标	(55)
第三节 平均指标	(61)
第四节 离中趋势的测定	(76)
第五节 变量分布的偏度和峰度	(84)
第四章 概率和抽样分布	(88)
第一节 事件及其概率	(88)
第二节 随机变量的概率分布	(90)
第三节 大数定律与中心极限定理	(98)
第四节 抽样分布	(99)

第五章 参数估计	(104)
第一节 点估计	(104)
第二节 总体均值的区间估计	(106)
第三节 总体比例的区间估计	(113)
第四节 总体方差的区间估计	(115)
第五节 样本容量的确定	(117)
第六章 假设检验	(120)
第一节 假设检验的基本问题	(120)
第二节 总体均值的检验	(123)
第三节 总体比例的检验	(130)
第四节 总体方差的检验	(132)
第七章 方差分析	(135)
第一节 单因素方差分析	(135)
第二节 双因素方差分析	(138)
第八章 相关与回归分析	(145)
第一节 相关分析	(145)
第二节 一元线性回归分析	(148)
第三节 多元线性回归分析	(154)
第四节 可线性化的非线性回归模型	(157)
第九章 时间数列分析	(160)
第一节 时间数列的意义和种类	(160)
第二节 时间数列水平分析指标	(163)
第三节 时间数列速度分析指标	(172)
第四节 现象发展的趋势分析	(176)

第十章 统计指数	(196)
第一节 统计指数的概念与种类	(196)
第二节 综合指数	(198)
第三节 平均指数	(202)
第四节 指数体系与因素分析	(205)
第五节 平均指标指数及其因素分析	(210)
第六节 几种常用的经济指数	(215)
附表	(222)
附表1 标准正态分布表	(222)
附表2 卡方分布表	(223)
附表3 t 分布表	(226)
附表4 F 分布表	(228)
附表5 随机数字表 (部分)	(229)
附表6 F 分布表续表 ($\alpha=0.05$)	(230)
附表7 F 分布表续表 ($\alpha=0.1$)	(232)
附表8 随机数字表 (部分)	(233)
主要参考文献	(235)

第一章 概论

【教学导读】

类似以下问题在社会生活中经常遇到,有些问题需要由统计来回答,而有些问题则是由统计提出但统计本身不能解答,但都离不开统计。

1. 中国的综合国力在世界上排第几位?

2. CPI(消费物价指数)、PPI(生产者物价指数)、PMI(采购经理指数)、GDP(国内生产总值)是什么?与我有关吗?

3. 在校大学生中未婚同居的比例有多大?

4. 为什么说“吸烟有害健康”?

5. 有人说“左撇子”聪明,你信吗?理由呢?

……

可见,现代人都需要具备一定的统计知识。

本章将告诉我们什么是统计,统计研究的对象是什么,统计又是怎样研究其对象的。学习本章需要明确统计研究的对象及其基本方法,了解统计产生及发展的历史;熟悉统计数据类型,掌握统计学中的几个基本概念,特别要理清各概念之间的相互关系。

第一节 统计的含义

一、统计的含义

人们都知道,一个企业需要对其各方面经济活动产生的数据特别是收入、成本、费用进行记录,以反映其经营成果;一个国家也需要对人口、资源、教育以及社会经济活动等各方面的数据进行记录,以反映经济社会发展的状况。这就是人们通常理解的“统计”,统计就是与数字打交道。

“统计”一词来源已久,其含义也有差异。汉语中“统计”原为合计或汇总计算的意思;英语中的“统计”最早出自拉丁语“Stutus”,指的是各种现象的状态或状况,后来“统计”一词在英语中演变为“Statistics”,即统计学和统计资料。事实上,“统计”除了上述两层含义外,还包括统计工作。

1. 统计工作

统计工作是指统计人员对社会、经济以及自然现象的数量进行搜集、整理、分析以及开发利用等一系列工作的总称。这里面包含三层意思:一是从事统计工作的主体是统计

人员,包括企业统计人员、事业单位统计人员、政府机构统计人员等;二是统计工作研究的对象是社会、经济以及自然现象的数量方面,社会经济生活中,只要是管理、决策需要的数据都可能成为统计工作研究的对象;三是统计工作的主要过程包括统计数据的调查、统计数据的加工汇总、统计数据的分析利用等。

2. 统计资料

统计工作在各个阶段取得的成果称为统计资料,主要表现为各种数据。它包括统计调查阶段取得的原始资料、统计加工汇总获得的综合资料以及统计分析阶段形成的分析数据。数量性是统计资料的基本特征。

【例 1.1】国家统计局(www.stats.gov.cn/)2012年2月22日发布:初步核算,2011年国内生产总值471 564亿元,比上年增长9.2%。其中,第一产业增加值47 712亿元,增长4.5%;第二产业增加值220 592亿元,增长10.6%;第三产业增加值203 260亿元,增长8.9%。第一产业增加值占国内生产总值的比重为10.1%,第二产业增加值比重为46.8%,第三产业增加值的比重为43.1%。我国2006年到2011年国内生产总值及其增值速度见图1-1。

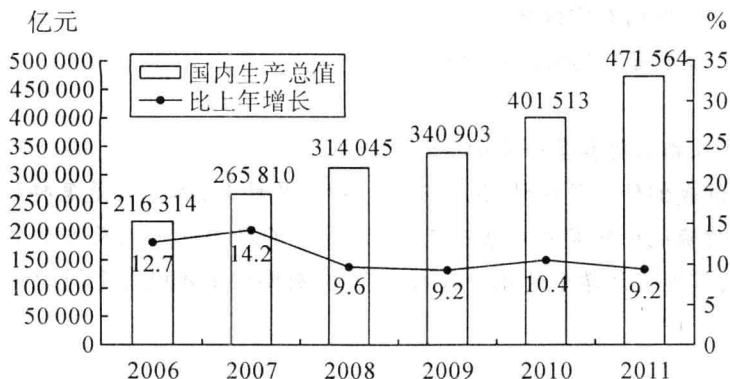


图 1-1 2006—2011 年国内生产总值及其增值速度

3. 统计学

统计学是指统计工作的经验积累到一定程度形成的知识体系,它是系统阐述数据搜集、整理、分析以及解释等一系列理论、方法的科学,为我们研究数据提供科学的指导。《不列颠百科全书》对统计学的定义为:“统计学是关于搜集和分析数据的科学和艺术。”这里所指的“艺术”是指统计方法应用的技巧,它在展现统计数据时应该体现出美感和实用。

当前,统计学已经应用到经济、金融、资源、人口、文化教育、体育、科研、医药卫生、心理学等社会生活的诸多领域。

统计工作、统计资料、统计学三者的关系:统计工作是基础,统计工作与统计资料之间是实践活动与成果的关系;统计工作与统计学之间是实践与理论的关系。

二、统计的研究对象及特点

统计工作的研究对象是社会、经济以及自然现象的数量方面。就是说统计研究的落

脚点是客观事物的各种数量,具体体现在统计调查是搜集具体事物的数量,统计加工整理是为了获得调查对象的综合数量,统计分析需要计算各种分析指标并运用指标去发现问题、分析问题和解决问题。

统计工作的研究对象具有以下四个特点:

1. 数量性

统计主要从以下三个方面研究客观现象的数量。

(1) 数量多少。用来反映统计的结果,所有的统计活动都必然产生相应的结果。如一个国家或地区的人口数量、粮食产量、国内生产总值(GDP)等。

(2) 现象之间的数量关系。它是指相互关联的现象之间的数量对比关系或数量依存关系。数量对比关系是在相互联系的现象之间进行数量对比的结果,如全部人口中的男女性别比例,GDP与人口总数量之间的比率,净利润与投资额之间的比率等。数量依存关系则是某种现象的数量变动引起另一种现象的数量发生相应的变动,比如人均GDP与全国商品零售总额之间的关系,父母身高与子女身高之间的关系等。

(3) 现象质与量之间互变的界限。统计研究现象的数量不是最终目的,而是通过研究现象的数量来反映事物的本质。那么事物的数量积累到什么程度才会发生质变,这是统计关注的问题。比如某地区居民贫困线标准,高考成绩的一本大学分数线、二本大学分数线等。

2. 总体性

统计研究由大量相同性质的个别事物集合而成的整体,这样的整体事物叫总体。通过对大量事物的观察、研究,综合地反映客观事物的发展水平、速度、构成和比例关系,研究总体的综合数量特征,而不是研究个别事物(现象)的数量。但是对个体事物进行调查是统计研究的起点。

3. 具体性

具体性也叫客观性。它是指统计活动所研究的各种客观现象在特定条件下的具体数量,是客观存在的,不是抽象和虚拟的,更不能编造。比如中国的总人口、国土面积以及GDP都是一个个客观存在的数量,即使对这些数据的统计存在误差,但不能否认这些数量的客观存在。因此,统计研究的数量有别于纯数学研究的抽象数量。

4. 变异性

变异性是指统计在研究总体事物的过程中,组成总体的个别单位事物之间是存在差异的,统计正是通过研究这种有差异的个别事物来反映总体的综合特征。总体内部事物之间的差异是必然的,也是客观的。正是这种变异性构成了统计研究的前提条件。

三、统计学研究的基本方法

基本的统计研究方法主要有大量观察法、实验设计法、描述统计方法和推断统计方法四种。

1. 大量观察法

大量观察法是指对所研究事物的全部或足够多的单位进行观察的方法,它是统计研究最基本的方法。运用大量观察法,是由研究对象的大量性和复杂性所决定的。大量的、

复杂的客观事物是在诸多因素的影响和作用下形成的,各单位的特征及其数量表现有很大的差别,不能任意选取个别或少数单位进行观察。只有在对所研究对象进行定性分析的基础上,确定调查对象的明确范围,观察全部或足够多数的调查单位,使得影响个体的偶然因素相互抵消,才能显示出现象的一般特征,最终认识客观事物的本质及其规律。

普查、抽样调查、统计报表调查、重点调查、典型调查等调查形式都采用大量观察法。

2. 实验设计法

实验设计法是指设计合理的实验程序,通过控制实验条件,以搜集某种专门的数据,比如汽车碰撞实验、新药临床对比实验等。通过实验搜集到客观、真实的数据,以此说明某种客观事物或研究方法是否符合预定的研究目的和要求。大量科学研究的数据来自实验设计。

3. 描述统计方法

描述统计方法是指对研究对象的基本状况、特征、数量关系、演变趋势及其规律性进行研究的一种方法。它主要包括统计分组法、综合指标法和统计模型法。

(1) 统计分组法:按照某一标准将总体性事物划分为性质不同的几个部分(类型)。通过统计分组,便于对统计资料进行汇总,获取基本的综合数据。同时,也可以在统计分组的基础上,研究现象的内部结构以及数量之间的相互关系,为统计分析奠定基础。

(2) 综合指标法:对大量观察所搜集到的数据进行汇总,获得反映总体数量特征的基本指标,并在此基础上计算其他各种分析指标,以显示现象在特定时间、地点、条件综合作用下所表现出的结果。利用这些综合指标,可以研究事物的状况、特征以及事物之间的数量关系,以帮助人们在实际工作中发现问题,找寻解决问题的办法。

(3) 统计模型法:根据客观事物的内在联系,利用观察到的统计数据,判断其分布特征、趋势或者是数量之间的相互关系形态,建立相应的数学模型,以研究事物的发展趋势或者事物之间的数量关系。“建模”已成为统计研究复杂客观事物数量关系的一种重要方法。

4. 推断统计方法

推断统计方法是指在一定置信度(把握程度)下,根据样本资料的特征对总体信息做出估计和预判的方法。推断统计是现代统计学的重要方法之一。

(1) 参数估计法:当总体参数未知时,利用样本数据计算样本统计量,并以此为依据对总体参数可能的取值或取值区间做出估计的一种方法。比如电视节目的收视率、居民家庭收支数据等都是参数估计的结果。

(2) 假设检验法:在对总体情况不了解的情况下,可以事先对总体做出某种假设,然后根据样本实际观察的资料对所做假设进行检验,来判断这种假设的真伪,以决定行动的取舍。比如采购商可以利用此法对供应商提供商品的质量进行假设检验,以决定是否下单采购。

四、统计学的产生和发展

统计学的产生与统计实践活动的发展是密不可分的,统计作为一种社会实践活动,是随着记数活动而产生和发展起来的。统计发展的历史可以追溯到远古的原始社会,在

那时,人们按部落居住在一起,打猎、捕鱼后就要算算有多少人、多少食物,以便进行分配,这是统计的雏形。我国夏禹时代就有了人口数据和土地数据的记载,这说明在夏朝就已经有了统计的萌芽。为了赋税、徭役和兵役的需要,世界各国历代都有田亩和户口的记录,在奴隶社会和封建社会主要是对人口、土地和财产进行统计。进入17世纪后,资本主义在欧洲大陆开始有了较大的发展,统计实践活动也进入各个领域,人们不断总结统计工作的经验,将统计实践活动上升到理论高度,逐步形成系统的理论和方法,成为一门系统的统计科学。从统计学的产生和发展过程来看,统计学大致可以划分为古典统计学、近代统计学和现代统计学三个阶段。

1. 古典统计学时期

17世纪中叶至18世纪中叶是古典统计学时期,在这一时期,统计学理论初步形成,并出现了两大主要学术派别——国势学派和政治算术学派。

(1) 国势学派

13世纪至16世纪中叶,欧洲各国资本主义出现不平衡发展,各国的国情也不一致,欧洲各主要国家都深感有调查国内外情况的必要,从意大利开始,各国相继进行本国和国外的历史沿革、地理条件、国家典章制度、财政收入、军事实力、居民风俗习惯、国家工商业、交通运输等国情调查。当各国调查资料积累到一定数量时,开始有学者将之汇编成册,并开始初步研究。德国学者芒斯特(Sebastian Münster,1489—1552年)所编的《世界志》,是第一部反映这些国家财富调查的科学统计著作。随后,意大利、英国等国也都有学者编著了有关世界各国或本国国情方面的著作。这些著作一方面为当政者或者准备从政者提供必要的国内外知识,另一方面也为国家制定方针政策以及资本家在国内外经商、争夺海外市场和开拓殖民地提供一定的国内外知识。这些著作主要用简单的文字记述方法来论述和反映国情方面的知识,缺少必要的比较、因果等分析方法。后人称之为“国势论”。

进入17世纪中叶以后,德国学者赫尔曼·康令(Hermann Conring,1606—1681年)把国势学从法学、史学和地理学等学科中独立出来,并在大学开设了“国势学”课程,内容是各国社会秩序、立宪、行政、人口、土地、财政、国家组织与结构等所谓“国家显著事项”,目的在于“授人以政治经营所需的知识”。在其所著的《国势学》一书中,康令除了采用前人的记述方法外,还采用了分组法、比较法。尽管康令在该著作中没有具体采用数据计算和比较,只采用笼统的比较说法,但其研究方式得到当时很多学者的欢迎和后来学者的继承,并形成了一个独有的派别——国势学派,也称记述学派。施梅兹尔(Martin Schmeitzel,1679—1747年)在耶拿、哈莱两所大学讲授的《政治统计学讲义》继承了康令的学说,且开始使用“Statisticum”(统计的)一词,但他对国势学的发展无实质上的贡献。

国势学派的主要继承人和代表人物哥特弗莱德·阿亨瓦尔(Gottfried Achenwall,1719—1772年)在其著作《近代欧洲各国国势学纲要》一书讲述了“一国或多数国家的显著事项”,主要采用对比分析的方法研究了国家组织、人口、军队、领土、居民职业、资源财富和国情国力,比较了各国实力的强弱,为德国的君主政体服务。1749年,阿亨瓦尔在其著作《欧洲最主要各国新国势学概要》一书中,为自己所研究的这门学科,首创了一个新的德文词汇——Statistik,即统计学,把他老师施梅兹尔提出的形容词“Statisticum”(统计

的)名词化为 Statistika(统计学)。“统计”一词最初来源于中世纪拉丁语的 Status,表示各种现象的状态和状况,由这一词根所组成的意大利语 Stato,表示国家或者各国的国家结构及国情等意思。国势学后来被正式命名为“统计学”。由于国势学派只是以文字对国情进行记述,只是用比较级和最高级的词汇对事物的状态进行描述,偏重事物性质的解释,不研究事物的计量分析方法,不注重数量计算和数量对比,没有借用数据资料去探索客观现象的规律性,从而获得了“有统计学之名,无统计学之实”的评价。但随着资本主义市场经济的不断发展,对事物量的计算和分析显得越来越重要,该学派后来发生了分裂,分化为图表学派和比较学派。

(2) 政治算术学派

16世纪末到17世纪中,英国连续几次遭遇瘟疫,居民生活卫生状况出现严重危机,加之荷兰、法国又趁机频频进攻,英国人开始对国家的现状和前途产生了担忧,悲观情绪不断加重。为此,为了救国救民,一些爱国人士开始大量搜集数据资料,并用图表、分组、对比等方法撰写文章和著作来证明英国能够战胜荷兰和法国并走向富强。约翰·格朗特(John Graunt,1620—1674年)和威廉·配第(William Petty,1623—1687年)是其中的杰出代表,也是政治算术学派的主要代表人物。

约翰·格朗特以1604年的“死亡公报”为研究资料,1662年发表了《对死亡表的自然观察和政治观察》一书。该书以人口问题为中心,通过对男女性别比例、各种原因的死亡人数占死亡总人数的比例关系等客观现象的数量方面进行分析,揭示出一系列的统计规律。他提出一个要在多年内形成的规律,需要进行多次观察,这体现了统计学中的大数法则的观点。正是因为格朗特在该书中认识到大量观察法的重要性,很多统计学家将该书誉为“真正统计科学的肇端”。

威廉·配第1676年完成了《政治算术》一书。在该书中,他对统计学研究的目的、任务和对象作出了初步的阐述,并认为社会现象和自然现象一样存在着客观规律,而“政治算术”则可以作为探索规律和法则的手段。他在该书中利用实际资料,运用数字、重量和尺度等统计方法对当时的英国、荷兰、法国之间的“国富和力量”进行数量上的计算和比较,做出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,从而为统计学的形成和发展奠定了方法论基础。因此,马克思称配第是“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”。

由于配第论述问题所涉及的理论和方法都被称为“政治算术”,而不是“统计学”。而以格朗特和配第为代表的政治算术学派都是以数量分析为特征,并且研究客观现象数量关系,就其内容和方法来看应该是统计学的正统起源。因此,政治算术学派的观点是“有统计之实,无统计学之名”。

2. 近代统计学时期

近代统计学大约是从18世纪末到19世纪末,这一时期的统计学主要有数理统计学派和社会统计学派两大学派。

(1) 数理统计学派

随着殖民地的开拓、航海运输业的繁荣,保险业也得到了较快发展。保险业由于带有赌博性质,具有极大的风险性,保险业的发展也促进了赌博数学和概率论的发展。18世

纪,瑞士的贝努里(Jokob Bernoulli,1645—1705年)较全面论述了概率原理,并将概率论建立在数学基础上创立了贝努里定理。由于概率论的产生和发展,越来越多的人开始运用概率论来研究政治、经济、法律和道德等社会问题。其中最具有代表性的是法国数学家兼统计学家拉普拉斯(Pierre Simon Laplace,1749—1827年)和比利时统计学家凯特勒(Adolphe Quetelet,1796—1874年)。

拉普拉斯在《概率论分析》一书中,把古典概率论引入统计学领域,使统计学逐渐形成一门成熟的科学。除此之外,拉普拉斯也发展了概率论,1786年在《关于巴黎人口出生、婚姻和死亡》的文章中,采用了抽样法,即以部分资料去推断全部资料的计算方法,从而促进了统计学方法论的发展。

凯特勒完成了统计学和概率论的结合,他所著的《社会物理学》论述了社会生活中随机偶然现象也贯彻着必然的规律性,要认识这些规律性必须通过实证的大量观察法,并且可以用概率计算表明规律的存在形式。他在著作中把自然科学的研究方法引入社会领域,并用概率论、大数法则、误差法则、正态分布等数理方法研究社会现象,为统计学开辟了广阔的道路。凯特勒也把德国的国势学派,英国的政治算术派和意大利、法国的古典概率派加以统一、改造并融合成具有近代意义的统计学。由于凯特勒正式将概率论引入统计学,使统计学发生了质的飞跃,从古典统计学进入近代统计学。按其贡献可以认为他是古典统计学的完成者,近代统计学的先驱,也是数理统计学派的奠基人,因此,有人推崇他为“近代统计学之父”。在凯特勒的影响和推动下,欧洲各国统计出现了以下积极变化:各国陆续建立国家统计机构,各国纷纷开展全方位的国情调查,民间统计学术团体开始形成,国际统计合作得以开展。

(2) 社会统计学派

自凯特勒后,统计学的发展开始变得丰富而复杂起来。由于运用领域不同,统计学的发展呈现出不同的方向和特色。19世纪后半叶,社会统计学派在德国异军突起,该学派由德国大学教授克尼斯(K. G. A. Knies,1821—1889年)首创,主要代表人物为德国统计学家恩格尔(C. I. E. Engel,1821—1896年)和梅尔(G. V. Mayr,1841—1925年)。克尼斯在其1850年出版的著作《作为独立科学的统计学》中提出了“国家论”与“统计学”科学分工的主张。他认为国家论是用文字记述的国势学的科学命名,统计学则是用数值研究社会经济规律的政治算术的科学命名。梅尔在1877年出版的《社会生活中的规律性》一书中指出,统计学的研究对象是总体而不是个别现象,这是因为社会现象具有复杂性和整体性,必须对总体进行大量观察和分析,研究其内在联系,才能揭示现象的内在规律。他强调统计学是一门具有特殊研究方法的实质性社会科学,主张将统计方法与统计科学结合起来。恩格尔将统计过程分为统计调查、整理和分析三个阶段,并通过对比英、法、德和比利时等国工人家庭的调查,撰写了《比利时工人家庭的生活费》一书,提出著名的恩格尔定律。

19世纪后半叶到20世纪初,社会统计学派在欧洲大陆占有优势,并在世界各国传播,对美、苏、日、中等国家产生了不同程度的影响。1882年起,社会统计学派与数理统计学派展开了关于实质性科学和方法论科学的争论。由于社会经济发展要求统计学提供更多的统计方法,社会科学本身的细化和量化也要求统计学能提供更有力的调查、整理和分析

资料的方法,这使社会统计学派日益重视方法论的研究,再加上受数理统计学派的影响,社会统计学派的性质逐渐发生了变化,由原来的实质性科学向方法论科学转变。但是,社会统计学派仍然强调在统计研究中必须以事物的质为前提和认识事物的重要性,这同数理统计学派的计量不计质的方法论性质是有本质区别的。

3. 现代统计学时期

现代统计学是从20世纪初到现在。19世纪末到20世纪初,进化论和能量守恒定律的出现促进了描述统计的完善,也使描述统计学派发展到巅峰。直到20世纪30年代,费舍尔(R. A. Fisher, 1890—1962年)的推断统计学才促使数理统计进入现代范畴。

(1) 欧美数理统计学

自然科学的迅速发展,促进了数理统计学的发展。20世纪20年代以后,细胞学的发展使统计学迈进了推断统计的新阶段,20世纪50年代,推断统计的基本框架已经建成,并逐渐成为20世纪的主流统计学。1907年,英国的戈塞特(William Sealy Gosset, 1876—1937年)提出了小样本 t 统计量理论;英国的费舍尔对小样本理论进一步发展,给出了 F 统计量、极大似然估计、实验设计和方差分析等方法;英国的高尔顿(Francis Galton, 1822—1911年)提出了相关与回归思想,明确给出相关系数的计算公式;英国的皮尔逊(Karl Pearson, 1857—1936年)发展了拟合优度检验,还给出了卡方统计量及其极限分布;波兰的尼曼(Jerzy Splawa Neyman, 1894—1981年)创立了区间估计理论,并和E·皮尔逊发展了假设检验理论;美国的瓦尔德(A. Wasld, 1902—1950年)等提出了统计决策理论和多元分布理论;美国的威尔科克森(Frank Wilcoxon)发展了一系列非参数统计方法,开辟了统计学的新领域。数理统计学的快速发展使其从统计学中分离出来自成一派。该学派由于是在英美等国发展起来,故又被称为欧美数理统计学派。

20世纪50年代以后,计算机的出现是统计学的全面发展阶段。由于受计算机和新兴科学的影响,统计学越来越依赖计算技术,成为数量分析的方法论科学。这一时期统计学的研究和应用范围越来越广,使得在现代统计学史中很难找到权威性的代表人物。当今的统计学家只能限制在有限的专业领域内从事某方面的研究,这是现代统计学的主要特点,如科克伦(W. G. Cochran, 1909—1980年)的实验设计理论、安得森(Th. W. Anderson)的复变数分析等。

(2) 社会经济统计学

十月革命胜利后,由于受社会统计学派的影响,苏联的大多数统计学家都认为统计学是一门实质性的社会科学。他们将统计学定义为在质与量的密切联系中研究大量社会现象的数量方面,研究社会发展规律在具体地点及时间条件下的数量表现的具有鲜明阶级性的社会科学。这一定义对我国及东欧的社会主义国家都产生了很大影响,并最终形成了以马克思主义政治经济学为理论基础的社会经济统计学派。该学派是在反虚无论、反消亡论、反万能科学论和反数学形式主义中形成的,认为数理统计学是数学,不是统计学,强调统计学的阶级性。

20世纪50年代以后,是现代统计学的全面发展阶段。由于计算机科学和新兴科学的发展,统计学越来越依赖于计算技术,成为数量分析的方法论科学。统计学的研究和应用范围也越来越广,这使得现代统计学的发展有几个明显趋势:第一,随着现代数学的发

展,越来越多的数学方法在统计学中得到广泛应用;第二,统计学不断向其他学科领域渗透,使得以统计学为基础的边缘学科不断形成;第三,伴随着计算机技术在统计学中的应用,大量数理方法得以普及,并成为实证分析的主要工具;第四,统计学的作用和功能已从描述向推断、预测及决策方向发展。它已从一门实质性的社会性学科,发展成为方法论的综合性学科。

第二节 统计数据类型

统计数据有广义和狭义两种理解。狭义的统计数据仅指统计调查阶段获得的各种资料,广义的统计数据是指统计调查、整理、分析各个工作阶段所获得的所有资料。多数场合下,统计数据都指狭义内涵。此处的分类也是针对狭义的统计数据。按照不同的标准,统计数据主要有三种分类方法。

一、统计数据按计量功能分类

统计数据按计量功能不同分为分类数据、顺序数据、间距数据和比率数据四种。

1. 分类数据

分类数据也叫类型数据,用来描述事物的非数字型数据,用文字表现。例如人口的性别有男、女两类,企业所属的行业大类有工业、农业、建筑、商品流通、交通运输、金融等类型。分类数据只能用来对事物进行分类,表现事物的类别,不能对事物进行排序,也不能用数字表现。

2. 顺序数据

顺序数据也叫排序数据,用来描述事物的非数字型数据,用文字表现。例如有些科目的考试成绩有优、良、中、及格、不及格五等,大学教师的职称有教授、副教授、讲师、助教四级。这类数据不仅可以用来对事物进行分类,还可以表现事物的高低、等级,但不能用数字表现。

3. 间距数据

间距数据是用来描述事物的数字型数据,用数值表现。该类数据不仅可以描述事物的类型、顺序,还可以计算事物之间的差距,但不能计算比率。温度是典型的间距数据,学号、手机号等表示号牌的数据也是间距数据。

间距数据如果取值“0”,表示“有”的意思。比如温度在摄氏0度表示有温度,0号车也有号牌。

4. 比率数据

比率数据用来描述事物的数字型数据,用数值表现。该类数据不仅可以描述事物的类型、顺序、计算数据之间的差距,还可以计算数据之间的比率。绝大多数的数值型数据都是比率数据,比如月工资、销售收入、职工人数等。

这类数据如果取值“0”,表示“没有”的意思。

二、统计数据按搜集方法分类

统计数据按搜集方法分为观测数据和实验数据两类。

1. 观测数据

观测数据是通过统计报表或专门组织的调查而搜集到的数据。它是在没有对事物进行人为控制的条件下而得到的、对客观事物进行直观反映的统计数据。有关经济与社会统计的数据几乎都属于观测数据。

2. 实验数据

实验数据是通过设计专门的实验方案,并且在实验过程中人为控制实验对象而搜集到的数据。比如,对一种新药疗效进行的临床实验,对一种新的农作物品种进行的田间实验等。自然科学领域中的绝大多数数据都来自实验结果。

三、统计数据按时间状况分类

统计数据按时间状况不同分为截面数据和序时数据两类。

1. 截面数据

截面数据是搜集事物在相同或近似相同时间上的数据,描述事物之间在某一时间下的差异。比如,2011年我国各省、市、自治区的国内生产总值数据。

2. 序时数据

序时数据是搜集同一事物在不同时间上的数据,描述现象随时间变化的情况。比如,2006—2011年我国国内生产总值数据。

第三节 统计学中的几个基本概念

统计学中涉及众多概念,有几个概念出现的频率较高,在这里加以集中介绍,主要包括统计总体、样本、标志、变量、统计指标等。

一、统计总体和样本

1. 统计总体

统计总体简称总体,是指统计研究对象的全体,它是由许多具有同一性质的个别事物组成的整体。比如研究中国的人口情况,由所有中国人组成的整体就是总体;研究国有企业的经济效益,由所有的国有企业组成的整体就是总体;研究某批次产品的质量情况,则该批次的所有产品就是总体。

构成总体的个别事物叫总体单位。前例中人口总体的总体单位就是每一个人,国有企业总体的总体单位是每一户国有企业,产品总体的总体单位就是每一件(台)产品。根据研究目的不同,总体单位可以是人、物、机构等实物单位,也可以是行为、事件、时间等非实物单位。

一般情况下,总体是依据研究目的来确定的。在某一研究目的下,研究对象范围内涉