

普通高等教育“十三五”经济与管理类专业核心课程规划教材

统计学——原理、方法及应用

主 编 汪 朋
副主编 苟凌滨 苏 婕

赠 送
电子课件



西安交通大学出版社
XI'AN JIAOTONG UNIVERSITY PRESS

普通高等教育“十三五”经济与管理类专业核心课程规划教材

统计学——原理、方法及应用

主 编 汪 朋

副主编 苟凌滨 苏 婕



西安交通大学出版社
XI'AN JIAOTONG UNIVERSITY PRESS

图书在版编目(CIP)数据

统计学:原理、方法及应用/汪朋主编. —西安:
西安交通大学出版社,2016.3
ISBN 978-7-5605-8285-6

I. ①统… II. ①汪… III. ①统计学 IV. ①C8

中国版本图书馆 CIP 数据核字(2016)第 030653 号

书 名 统计学——原理、方法及应用
主 编 汪 朋
责任编辑 李逢国

出版发行 西安交通大学出版社
(西安市兴庆南路 10 号 邮政编码 710049)
网 址 <http://www.xjtupress.com>
电 话 (029)82668357 82667874(发行中心)
(029)82668315(总编办)
传 真 (029)82668280
印 刷 陕西元盛印务有限公司

开 本 787mm×1092mm 1/16 印张 24.125 字数 587 千字
版次印次 2016 年 2 月第 1 版 2016 年 2 月第 1 次印刷
书 号 ISBN 978-7-5605-8285-6/C·132
定 价 44.80 元

读者购书、书店添货、如发现印装质量问题,请与本社发行中心联系、调换。
订购热线:(029)82665248 (029)82665249
投稿热线:(029)82668133
读者信箱:xj_rwjg@126.com

版权所有 侵权必究

内 容 提 要

本书定位于统计学基础，介绍了统计学的基本原理、方法及应用。同其他教材相比，本书具有以下几个特点：

(1) 注重统计学原理的讲述，淡化统计方法本身的数学背景。

(2) 通过案例来驱动统计学原理与方法的学习和应用。

(3) 结合统计软件，实现统计方法的应用。

(4) 在各章后面安排了类型和数量较多的思考与练习题。

本书可以作为理、工、经济、管理、人文社会科学专业以及其他领域的统计学教材，也可以作为其他人员学习统计方法的参考书。

前言

Foreword

统计学是关于数据的学科,而数据每时每刻都在产生,遍布于自然界与人类社会的每一个角落。因此,统计学是每个领域、每个人都会用到的科学,不论你是否接受或是否意识到,统计学事实上已经渗透到我们的工作、学习和生活中。在大学,统计学是许多专业的研究生、本科生必修的课程。实际上,随着信息化和大数据时代的到来,不仅是大学生,对于任何人而言,学习一些统计学知识,掌握基本的统计学原理、方法和统计分析工具,都是十分重要和必要的。因此,统计学被称为 21 世纪最有发展前途的学科之一。长期以来,统计学一直被教育部列为诸多专业的核心课程。

在多年的统计学学习、教学和应用实践中,我们体会到只有真正理解了统计的一些基本原理,才能很好地应用统计方法解决实际问题。因此本书定位于统计学基础,介绍统计学的基本原理、方法及应用,其基本框架是:第一章为总论,介绍统计学的形成、内容体系与分支、研究对象、研究方法以及一些基本的概念,这一章是全书的总纲,同时也是全书的基础;第二章和第三章介绍描述统计,研究如何搜集数据、整理数据、显示数据并描述现象的基本数量特征;第四章和第五章介绍动态数据描述和分析的方法,这在社会经济统计中有较大的应用;第六章和第七章介绍推断统计,研究抽样推断中参数估计和假设检验两大基本问题;第八章和第九章介绍方差分析和回归分析,这是统计学中较为基础且应用极为广泛的两类统计模型。

同其他教材相比,本书具有以下几个特点:

(1)注重统计学原理的讲述,淡化统计方法本身的数学背景,写作的过程中尽量避免复杂的数学推导,具有高中水平的读者就可以读懂。

(2)通过案例来驱动统计学原理与方法的学习和应用。本书各章开头都安排了“案例引子”,通过实际生活中的各种实例引导读者进行学习,既可以激发读者学习的兴趣,又可以将复杂的事物简单化、具体化,帮助读者更好地理解和学习统计学的原理和方法。同时,在各章理论内容结束后都安排了“应用案例”模块,这些案例与实践联系极为紧密,在对应用案例进行分析的过程中,我们力求统计实践的严格性与现实性。例如,一般而言,每种统计学方法和模型都是有其假定条件的,这些假定在我们介绍理论内容时都阐述过,但限于本书内容的深度,我们没有过多地从理论上去探讨,而是将其放在了应用案例中,直接使用高深一些的统计学方法验证这些假定是否满足。这么做,既可以保证统计应用的严谨性与真实性,培养读者良好的科学素养,同时还可以开拓读者的视野,激发其学习的热情,为其统计学后续内容的学习提前埋下伏笔,作好铺垫。

(3)结合统计软件,实现统计方法的应用。一方面,现代统计涉及大量的计算,不借助于统计软件,是很难完成数据的整理和分析的;另一方面,统计软件的学习,必须结合相应的统计理

论,要在应用中进行学习。为此,我们在本书中增加了对统计软件操作的介绍。所有的统计软件可以分为两类,一种是界面菜单式,如 Excel 和 SPSS 等;另一类是语言编程型,如 SAS、R 等。本书在这两类中分别选择了 Excel 和 R 进行了介绍。当然,本书的内容较为基础,大部分都可以用 Excel 去实现,因此,本书主要以 Excel 为主,以 R 软件为辅,来介绍统计操作的实现。而本书之所以选择 R 软件作为辅助的统计软件,一方面是为了弥补 Excel 等界面操作型软件的不足;另一方面是由于该软件具有免费性、开源性、编程简单直观等诸多优势,在国外十分流行,在国内近几年也逐渐流行和火热起来,然而与统计基础理论相结合介绍 R 软件操作的书籍却比较少见。为此,我们把 R 软件如何实现基础的统计方法纳入到本书的编写中,并在附录中增加了对 R 软件基本操作的介绍。

(4)统计的学习离不开练习,因此为了帮助读者理解和巩固所学的知识,提高实际动手的能力,我们在各章后面安排了类型和数量较多的思考与练习题。这些习题中,一部分是基础理论型,是为巩固知识所设立的;另一部分属于实训操作型,通过这部分习题的练习,不仅能够提高动手能力,激发学习兴趣,还能够促使读者从实验结果的分析中更好地去理解所学的统计学内容。

本书可以作为理、工、经济、管理、人文社会科学专业以及其他领域的统计学教材,也可以作为其他人员学习统计方法的参考书。

本书由汪朋担任主编,苟凌滨和苏婕担任副主编,这三位老师负责全书编写大纲的设计以及编写组的管理工作。编写内容的具体分工为:汪朋负责各章应用案例的选取和撰写,并负责第一章,第四章的第四、五节和第六章的编写;苟凌滨负责第一章至第五章的情景导入的选取和撰写,并负责第三章和第五章的编写;苏婕负责第六章到第九章的情景导入的选取和撰写,并负责了第二章和第八章的编写;侯淑静负责第九章的编写;陶青负责第七章的编写;刘颖负责第四章的第一、二、三节的编写。最后,由汪朋、苟凌滨和苏婕三位老师对全书进行了最后的修改、总纂和技术处理,并提供了配套的思考与练习题。

本书的编写过程中,我们参考了国内外出版的大量本学科教材和专著,敬列于参考文献中。对编写这些著作的学界前辈、专家和同行们,我们表示崇高的敬意和衷心的感谢!

本书课件、例题、案例和习题及有关的 R 代码等相关资料可以向西安交通大学出版社索取,也可以通过电子邮件向作者索取,邮件地址为:pwang@xzmy.edu.cn。

本书是西藏民族大学“特色教材建设”项目、“经济统计学专业综合改革”项目和“统计学专业实践教学模式创新研究”项目的阶段性成果,本书的出版得到了西藏民族大学特色教材建设项目的经费支持,在此表示感谢。本书的出版也得到了西安交通大学出版社的大力支持和帮助,特别是出版社的李逢国老师为本书的组稿、编辑做了大量的工作,在此一并表示衷心的感谢。

最后,由于作者水平有限,加之时间紧迫,书中不妥甚至错误之处在所难免,恳请使用本教材的老师、同学和其他读者批评指正。

编者

2016 年 1 月

目 录

Contents

第一章 总论	(1)
情景导入	(1)
第一节 统计的产生和发展	(2)
第二节 统计的研究对象和方法	(5)
第三节 统计学中的一些基本概念	(10)
第四节 统计软件介绍	(18)
应用案例	(19)
思考与练习	(20)
第二章 统计数据的搜集与整理	(22)
情景导入	(22)
第一节 统计数据的搜集	(22)
第二节 统计数据的整理	(40)
第三节 次数分布数列	(47)
第四节 统计表和统计图	(52)
应用案例 1	(64)
应用案例 2	(65)
思考与练习	(68)
第三章 统计数据分布特征的描述	(71)
情景导入	(71)
第一节 分布集中趋势的描述	(72)
第二节 分布离散程度的描述	(87)
第三节 分布形状的描述	(94)
第四节 应用 Excel 计算描述性统计指标	(96)
应用案例	(99)
思考与练习	(102)
第四章 时间数列	(107)
情景导入	(107)
第一节 时间数列概述	(108)
第二节 时间数列的水平分析	(111)
第三节 时间数列的速度分析	(119)
第四节 时间数列因素分解与测定	(126)

第五节 时间数列预测	(141)
应用案例	(148)
思考与练习	(153)
第五章 统计指数	(159)
情景导入	(159)
第一节 统计指数概述	(159)
第二节 综合指数	(162)
第三节 平均指数	(168)
第四节 指数体系与因素分析	(171)
第五节 几种常见的经济指数介绍	(179)
应用案例	(186)
思考与练习	(187)
第六章 抽样分布与参数估计	(192)
情景导入	(192)
第一节 概率基础	(192)
第二节 抽样分布	(203)
第三节 参数估计	(209)
第四节 样本容量的确定	(225)
应用案例	(227)
思考与练习	(229)
第七章 假设检验	(234)
情景导入	(234)
第一节 假设检验概述	(235)
第二节 单个正态总体的检验	(242)
第三节 两个正态总体的检验	(248)
第四节 假设检验的 P 值	(255)
应用案例	(258)
思考与练习	(260)
第八章 方差分析	(264)
情景导入	(264)
第一节 方差分析的基本原理	(265)
第二节 单因子方差分析	(268)
第三节 双因子方差分析	(273)
应用案例	(283)
思考与练习	(288)
第九章 相关与回归分析	(292)
情景导入	(292)

第一节	变量间的关系及其度量	(292)
第二节	一元线性回归分析	(300)
第三节	多元线性回归分析	(313)
第四节	非线性回归分析介绍	(318)
第五节	用于回归分析的 R 函数介绍	(323)
应用案例		(326)
思考与练习		(332)
附录一	R 软件的基本操作简介	(341)
附录二	常用统计表	(367)
附表 1	标准正态分布函数表	(367)
附表 2	标准正态分布分位数表(下侧)	(369)
附表 3	t 分布分位数表(上侧)	(371)
附表 4	χ^2 分布分位数表(上侧)	(373)
附表 5	F 分布分位数表(上侧, $\alpha = 0.05$)	(374)
参考文献		(376)

第一章

总论



情景导入

无处不在的统计

下面的现象或问题你能够解释或解决吗？想一想为什么？

1. 当你买了一台电脑时，被告知三年内可以免费保修。那么，厂家凭什么这样说？说多了，厂家会损失，说少了，会失去竞争力，也是损失。到底这个保修期是怎样决定的呢？

2. 电视上经常播放一些比赛，比如“全国青年歌手大赛”，在比赛评分中，我们总会听到诸如“去掉最高分、去掉最低分，选手最后得分 95.23 分”之类的话语，为什么要这么做呢？

3. 1936 年，美国《文学文摘》杂志根据收回的 237.6 万张关于总统选举的民意测验问卷预言，阿尔夫·兰登(Alf Landon)将以 57% 对 43% 的优势超过富兰克林·罗斯福(Franklin Roosevelt)。但结果是：罗斯福以 62% 对 38% 的一边倒优势赢得了 1936 年的选举。样本量如此之大，为什么还会产生这么大的误差呢？

4. 某班在同一门课程的两次考试中平均分数分别 80 分和 70 分，而小红在这两次考试中的分数分别为 90 分和 80 分，你认为小红哪一次考试成绩更理想？为什么？

5. 战争是残酷的，没有人喜欢战争，当战争真的来临时，如何确定应征入伍、走上前线的人员名单是一件很重要的事。在越南战争中，美国政府制定了一个“抓阄”的征兵计划：把 1~366 的号码随机分配给一年中的每一天，然后由军事部门按分配的号码顺序把生日与之对应的年轻人分批征召入伍。具体办法是把 366 个写上日期的乒乓球放进一个大容器中，然后随机抽取，第一个拿出来的乒乓球上的日期是 9 月 14 日，那么这一天出生的适龄青年就将第一批入伍。抽取的结果发现，在较小的号码(1~183)中，有 73 个分配给了前半年的日子，而 110 个分配给了后半年的日子，这意味着后半年出生的人有更大的可能性先上前线，这在一定程度上背离了保证大家机会均等的初衷，那么是什么原因导致这种结果呢？

6. 很多人都读过莎士比亚的作品，人们为莎翁的才华所折服，但是也有人根据历史背景分析怀疑这些作品是否真的属于莎士比亚。统计学家利用统计侦探方法帮助莎士比亚捍卫了尊严，他们是怎么做的？你能想到更好的方法吗？同样道理，对于一个从来没有研究过《红楼梦》的统计学家来说，如何根据比较写作习惯就能找出从哪一段落开始就不是曹雪芹的手笔呢？

回答以上问题或多或少都需要借助于统计学的知识和方法。随着人们认识自我、认识自然、认识社会的要求不断提高，仅仅进行定性的分析和描述已远不能满足需要，定量研究已成必然趋势。如今，作为定量研究重要手段的统计学已经几乎应用于所有领域：战略管理、人力资源管理、市场营销学、生产管理、质量控制、工业工程、金融、精算、审计学、经济分析与预测、教育学、医学、气象学、工业、农业、动物学、生态学、人口统计学、地质学、遗传学、政治学、心理

学、文学、体育学等。

随着科技日趋发达,尤其是互联网这一令人惊叹的新生事物的出现,催生着大数据时代的来临,人们有越来越多的途径接触到越来越多的数据信息,以至于我们的大脑在很多时候处于超载状态——在海量的信息和数据面前难辨真伪、不知所措。而统计学知识则可以借给我们一双慧眼,帮助我们去伪存真,从浩瀚的信息中提取出真正有用的数据,得到有价值的结论。因此,统计学绝不是摆弄数字的乏味游戏,而是真正帮助我们生活、工作、学习的重要方法和工具。下面我们通过第一章介绍统计的含义、它的历史渊源、研究对象和方法及一些重要的概念和要素,为后面各章的学习打好基础。

第一节 统计的产生和发展

一、统计活动的产生和发展

统计实践活动先于统计学的产生。从历史上看,统计实践活动自人类社会初期,即还没有文字的原始社会起就有了。最初的统计,是社会统计,即只是反映社会基本情况的简单的计数工作。在原始社会,人们按氏族、部落居住在一起打猎、捕鱼,分配食物时就要算算有多少人、多少食物才能进行分配。所以,从结绳记事开始,就有了对自然社会现象的简单的计量活动,有了统计的萌芽。

在奴隶社会,奴隶主国家为了对内统治和对外战争的需要,进行征兵、征税,开始了人口、土地和财产统计。我们现在能够看到的我国最早的统计资料,就是关于公元前 21 世纪(夏朝)人口和土地数字的记载:夏朝时分中国为九州,人口约 1355 万人,土地约 2438 万顷。另据历史记载,在秦穆公时期,商鞅变法,在其调查研究中明确提出:“强国知十三数,欲强国,不知国十三数,地虽利,民虽众,国愈弱至削。”这说明我国古代的一些清醒的政治家、军事家早就意识到统计的重要性。在国外,古希腊、罗马时代,已开始了人口数和居民财产的统计工作。公元前 3050 年,埃及为建造“金字塔”,在全国进行人口和财产的调查。

在封建社会,由于经济十分落后,统计发展缓慢。统计广泛迅速地发展是在资本主义社会。资本主义社会取代封建社会后,经济文化有了很大的发展,社会分工日益发达,引起对情报、信息和统计的新的需要。统计已不限于人口、土地、财产等内容,它逐步扩展到了更为广泛的领域,产生了诸如工业、农业、商业、银行、保险、交通、邮电、外贸、劳动、就业等各个方面形成的各种专业的社会经济统计。1830—1849 年,欧洲出现“统计狂热”时期,各国相继成立了统计机关和统计研究机构,统计成为社会分工中的一种专门的行业。

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。统计学作为一门科学,便在这样的背景下应运而生。

二、统计学的产生和发展

从上面的叙述可以看到,统计活动虽有几千年的历史,但在学术上作为一门学科的统计学的历史却没有这么长。一般认为,统计学产生于 17 世纪中叶的欧洲,距现在只有 300 多年。其发展主要可分为三个阶段:

(一)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了一定的学术派别,主要有国势学派和政治算术学派。

1. 国势学派

国势学派又称记述学派,产生于 17 世纪的德国。由于该学派主要以文字记述国家的显著事项,故称记述学派。其主要代表人物是海尔曼·康令(Hermann Conring,1606—1681)和阿亨华尔(Gottfried Achenwall,1719—1772)。康令于 1660 年把国势学从法学、史学和地理学等学科中独立出来,在大学中讲授“实际政治家所必需的知识”;阿亨华尔在哥廷根大学开设“国家学”课程,其主要著作是《近代欧洲各国国势学纲要》,书中讲述“一国或多数国家的显著事项”,主要用对比分析的方法研究了国家组织、领土、人口、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。因在外文中“国势”与“统计”词义相通,后来阿亨瓦尔正式将它命名为“统计学”。国势学派只是对国情的记述,偏重事物性质的解释,未能进一步揭示社会经济现象的规律,也不研究事物的计量分析方法,不注重数量对比和数量计算,只是用比较级和最高级的词汇对事物的状态进行描述。所以,人们也把它叫做记述学派(旧学派或德国学派),并认为国势学派有统计学之名而无统计学之实。

2. 政治算术学派

政治算术学派产生于 17 世纪资本主义的英国,代表人物是威廉·配第(William Petty,1623—1687),他在 1671—1676 年间写成《政治算术》一书。在撰写该部代表作时,正值第三次英荷战争,国内经济困难,国外面临着荷、法两国的威胁。威廉·配第为了让人们知道和确信“英国的事业和各种问题,并非处于可悲的状态”,在《政治算术》中用数字比较分析了英、荷、法三国的经济实力和造成这种实力差距的原因,并从贸易、税制、分工、资本和利用闲散劳动力等多方面提出了英国的强盛之道。这种用数字来表述,用数字、重量和尺度来计量,并配以朴素的图表,是前所未有的,是现代描述统计学广为采用的方法和内容。由于威廉·配第对于统计学的形成有着巨大的功绩,因此马克思评价道:“威廉·配第——政治经济学之父,在某种程度上也是统计学的创始人。”

政治算术学派的另一个代表人物是约翰·格朗特(John Graunt,1620—1674),他以 1604 年伦敦教会每周一次发表的“死亡公报”为研究资料,在 1662 年发表了《关于死亡公报的自然和政治观察》的论著。书中通过大量观察发现了人口各年龄组的死亡率、性别比例等重要的数量规律,并对人口总数进行了较为科学的估计;并且第一次编制了“生命表”,对死亡率与人口寿命作了分析,从而引起了普遍的关注。因此,他被认为是人口统计学的创始人。

比较遗憾的是,政治算术学派的学者没有使用“统计学”这个名称,他们的著作有统计学之实,却没有统计学之名,存在着名不副实的缺陷。

国势学派和政治算术学派共存了将近 200 年,两派互相影响互相争论,但总的来说,政治算术学派的影响要大得多。

历史上,曾就“国势学”和“政治算术”哪一个才是统计学的真正起源问题产生过持续长达一个多世纪的争论。1850 年,德国人克尼斯(K. G. Kniex,1821—1898)根据当时统计学的发展实践,概括了大多数人的意见,写了《独立科学的统计学——关于统计学的理论和实际上的纠纷的解决——同时即是关于阿亨瓦尔以来的统计学的批判的历史的一篇论文》一书,提出了

国家论和统计学的分工,主张将政治算术称为统计学,而国势学派所称的统计学仍称为国势学。这个意见,逐渐为大多数人所接受,从而使得“国势学”与“政治算术”的争论告一段落。

(二)近代统计学时期

这个时期大致是从18世纪末到19世纪末。著名的大数法则、最小平方法、相关与回归分析、指数分析法、时间数列分析法以及正态分布等理论都是这个时期建立和发展起来的。代表学派主要有数理统计学派和社会经济统计学派。

1. 数理统计学派

统计学的另一个重要起源是概率论。14世纪,在工商业比较繁荣的意大利以及地中海沿岸其他地区,由于赌博游戏盛行和保险活动的萌起,人们已经对“机会”问题产生了兴趣。不过真正意义上的概率论,是从17世纪开始的。帕斯卡(B. Pascal)和费马(P. Fermat)关于“得点问题”的讨论,奠定了概率论的基础。在早期概率论的研究中,做过重要贡献的数学家有:莱布尼茨(G. Leibniz)、贝努利(J. Bernoulli)、棣莫佛(A. de Moivre)、贝叶斯(T. Bayes)、拉普拉斯(F. Laplace)、高斯(C. Gauss)、勒让德(A. Legendre)、辛普森(T. Simpson)、布丰(C. de Buffon)、泊松(S. Poisson)等。其中,拉普拉斯是古典概率的极大成者,他给出了概率的“古典”解释,并把数学分析方法系统地引进概率论,建立了较为严格的概率数学体系。高斯和勒让德在误差研究过程中提出了最小二乘法,高斯还导出了正态分布曲线。

“政治算术”研究的是简单的、确定的数量关系,而概率则研究复杂的、随机性现象,因此将概率论引入到统计学中,极大地充实和深化了数量问题研究的内容,而作出这一开创性工作,同时也在统计学史上起着承前启后作用的是19世纪比利时的凯特勒(A. Quetelet, 1796—1874),其著有《社会物理学》等著作。他最先提出,用数学中的大数定律——平均数定律,作为分析社会经济现象的一种工具。他提出,社会现象的发展并非偶然,而是具有其内在规律性的。但他在解释社会规律时,不能正确地把社会规律与自然规律区分开,提出社会规律与自然规律一样永恒不变的错误观点。凯特勒写过不少运用概率论的著作,到19世纪60年代,他又进一步将国势学、政治算术、概率论的科学方法结合起来,使之形成近代应用数理统计学。

其后,经过多方面的研究,特别是数理统计学吸取生物学研究中的有益成果,由高尔顿(F. Galton, 1822—1911)、皮尔逊(K. Pearson, 1857—1936)、戈塞特(W. S. Gosset, 1876—1937)和费希尔(R. A. Fisher, 1890—1963)等统计学家,提出并发展了回归和相关、假设检验、 χ^2 分布和 t 分布等理论,数理统计学逐渐发展成为一门完整的学科。

2. 社会统计学派

社会统计学派产生于19世纪后半叶,创始人是德国经济学家、统计学家克尼斯,主要代表人物有厄恩斯特·恩格尔(Ernst Engel)、乔治·冯·梅尔(Georg von Mayr, 1841—1925)等人。他们融合了国势学派与政治算术学派的观点,沿着凯特勒的“基本统计理论”向前发展,但在学科性质上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。

由于数理统计学的产生和发展,在一些根本性的问题上,与社会统计学派有了分歧。社会统计学专门研究社会现象,而数理统计学既研究社会现象又研究自然现象,这就发生了统计学研究领域的争论。另外,社会统计学原是一门实质性科学,而数理统计学是一门方法论科学,这就又发生了统计学到底是一门什么性质的科学的争论。时至今日,这两派仍在争论,在争论中两派又互相渗透。一方面,由于数理统计方法在社会实践中的广泛应用,对社会统计学发生

了深刻的影响,由此,社会统计学逐渐由原来的实质性科学向方法论科学转变;另一方面,数理统计学中的“应用统计”,则逐渐向社会统计学靠拢。

(三)现代统计学时期

20 世纪至今为现代统计学时期,这一时期的主要特征是描述统计学已转向推断统计学,1907 年,英国人戈塞特提出了小样本 t 统计量理论,丰富了抽样分布理论,为统计推断奠定了基础。英国的费希尔提出了极大似然估计量的概念,迅速成为了估计参数的重要方法,他还提出样本相关系数的分布、实验设计和方差分析等方法。英国科学家高尔顿提出了相关与回归思想,并给出计算相关系数的明确公式。英国统计学者 K. 皮尔逊发展了拟合优度检验,还给出了卡方统计量及其极限分布,波兰学者奈曼(J. Neyman, 1894—1981)创立了区间估计理论,并和 E. 皮尔逊发展了假设理论。美国学者瓦尔德提出决策理论和序贯抽样方法。美国化学家威尔科克松(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。由马哈拉诺比斯领导的印度统计研究所和 20 世纪 30 年代后期奈曼发表的两篇论文,使抽样的数学理论在 20 世纪 30 年代得到了迅速发展。

统计学大致经过以上三个阶段发展到今天,随着统计学理论知识的发展与健全,统计学的应用领域得到了极大地扩展,出现了许多新型的交叉学科,比如统计应用到法律、文学等学科。同时,伴随着计算机技术的飞速发展,统计学还在模糊现象、突变现象及混沌现象等方面开辟新的研究领域。

第二节 统计的研究对象和方法

一、“统计”一词的含义

统计的英文词为“Statistics”,其语源最早出自中世纪拉丁语的“Status”(各种现象的状态和状况),由这一词根组成的意大利语“State”,表示国家的概念及关于国家结构和国情这方面知识的总称。

统计一词作为学科名称最早使用的是 18 世纪德国哥丁根大学政治学教授阿亨瓦尔,他把国势学称为“Statistik”,即统计学。在英国,早在 17 世纪就出现用数字来说明社会的科学,但使用的是另一个完全不同的名称:“政治算术”(Political arithmetic),直到 18 世纪末,英语“Statistics”才作为德语“Statistik”的译文传入英国,即用数字表示事实。

随着社会经济和统计学自身的发展,“统计”的含义已经起了变化,它包含有统计工作(活动)、统计资料和统计学等三种含义。比如,“统计一下学生人数”,是指统计工作;“看一下天气的统计”,是指统计资料;“今天我们开始学习统计”,是指统计科学。

统计工作:即统计实践或统计活动。它是指对社会、政治、经济、文化等现象的数量方面进行搜集、整理、分析的工作活动过程的总称,即一种调查研究活动。

统计资料:是统计工作的成果。它是指统计部门或单位进行工作所搜集、整理、编制的各种统计数据资料的总称,包括数据资料和统计分析资料。如统计资料汇编、统计年鉴、统计手册、统计图表、统计分析报告等。

统计学:是有关数据的学科,是一门搜集、整理和分析统计数据的方法论科学,即统计

理论。

统计的这三种含义具有密切的联系:统计工作是人们的统计实践,是主观反映客观的认识过程;统计资料是统计工作的结果。统计工作与统计资料是过程与成果的关系,通过统计工作可以取得统计资料,反过来进行统计工作,往往需要一定的统计资料作为工作的依据。统计学是统计工作经验的总结和概括,反之,统计学所阐述的理论和方法又是指导统计工作的原则和方法。因此,统计学和统计工作之间存在着理论和实践的辩证关系。

从“统计”一词的三种含义中可以看到,处于中心地位的是“统计工作”这一层含义,因为统计工作是否科学合理完全决定了统计资料的合理性,而“统计学”的建立和完善则是为了更好地指导统计工作。一个完整的统计工作大致会经历统计设计、数据的搜集、数据的整理、数据的统计分析和统计资料的积累与应用等五个基本环节。

统计设计就是要根据研究问题的性质,在有关科学理论的指导下,制定统计指标、指标体系和统计分类,给出统一定义、标准;同时提出搜集、整理和分析数据的方案和工作进度等。统计设计是统计研究的前期工程,其完成质量关系到整个统计研究的质量。做好统计设计不仅要有统计学的一般理论和方法作为指导,还要求设计者对所研究的问题本身具有深刻的认识和相关学科的知识。

经过统计设计,形成方案之后,就可以进入统计数据的搜集阶段。统计数据的搜集就是根据统计研究任务的要求,有计划、有组织地向调查单位搜集原始资料的工作。通过数据的搜集取得丰富的资料,增强人们对研究对象的感性认识。统计数据的搜集是认识事物的起点,同时也是数据整理和统计分析的基础。

统计数据的整理就是根据一定的目的和任务,将搜集得到的大量原始资料进行科学分组和综合汇总。它处于统计工作过程的中间环节,起着承前启后的作用。

数据的统计分析是将加工整理好的统计资料加以分析研究,采用各种分析方法,计算各种分析指标,来揭示社会经济过程的本质及其发展变化的规律性。通过这一过程由感性认识上升到理性认识。

统计资料的积累与应用是统计工作过程的后续环节。通过统计整理和分析,可以得到有关的统计资料,但统计资料的提供并不意味着统计研究的终结。统计的目的在于认识客观世界的规律。对于已经公布的统计资料需要加以积累,还可以进一步地加工,同时结合相关的实质性学科的理论知识去进行分析和利用。如何更好地将统计数据和统计方法应用于各自的研究领域是应用统计学研究的一个重要方面。

二、统计研究的对象与特点

(一)统计的研究对象

关于统计到底研究什么?统计学究竟是属于方法论科学,还是属于实质性科学?这个问题在理论界至今没有一个统一的明确说法,回答这个问题对于全面认识这门学科和为以后的论述铺平道路具有重要意义。目前比较流行的有以下三种观点:

第一种,规律派:认为统计学是研究社会经济现象发展规律的,即通过研究在一定时间、地点条件下的社会经济现象的数量表现,来揭示社会经济发展规律的独立的社会科学,是一门实质性科学。

第二种,数理统计学派:认为统计学的研究对象是随机现象,是以概率论为基础的应用数

学,是一门通用的数理方法学科。认为并不存在独立的社会经济统计学,它只不过是数理统计方法在研究社会经济现象时的应用。

第三种,方法论派:这一派认为统计工作和统计科学是不同的。统计工作研究的是大量的具体现象的数量方面,包括数量特征、数量关系和数量界限。而统计学研究的则是大量具体现象具体数量规律的方法,即统计工作的方法论。

目前第三种观点为大多数人所接受,本教材采用的便是这种观点。即我们认为,统计学属于认识具体现象数量方面的方法论科学,或者说是对一定总体现象的定量认识的方法论。

(二)统计的研究特点

统计的研究特点可以从统计学的学科特点和工作方法的特点两个方面来看。从学科上看,统计学是一门既不属于数学也不属于其他相关实质性科学的一门具有“寄生性”的独立性的方法论科学。说统计学具有“寄生性”是因为统计学从来都不是靠解决自身的问题发展壮大起来的,而是靠解决其他领域的问题发展起来的一门学科,可以这么说没有其他实质性学科的发展,就不会有统计学的产生、发展和壮大。

从统计工作的方法上看,具有数量性、总体性、具体性、客观性、变异性和广泛性等六个方面的特点。

1. 数量性

统计的认识力首先表现在它以准确的和无可争辩的事实为基础,同时,这些事实用数字加以表现,具有简短性和明显性。数量性的特点,是统计工作方法的重要特点,这一特点也把统计学和其他实质性的社会科学(如政治经济学)区别开来。

统计的特点是用大量数字资料说明事物的规模、水平、结构、比例关系、差别程度、普遍程度、发展速度、平均规模和水平、平均发展速度等。例如,国家统计局发布的关于2012年国民经济和社会发展公报时指出:全年国内生产总值519322亿元,比上年增长7.8%;全年居民消费价格比上年上涨2.6%,其中食品价格上涨4.8%;全年全国公共财政收入117210亿元,比上年增加13335亿元,增长12.8%;其中税收收入100601亿元,增加10862亿元,增长12.1%。年末国家外汇储备33116亿元,比上年年末增加1304亿元。还有其他领域的许多统计数字。这些统计数字都从各方面表明我国当前社会经济发展和深化改革的基本情况。

应当注意:统计不是单纯地研究社会现象的数量方面,而是在质与量的密切联系中研究现象的数量方面。唯物辩证法的质与量的辩证统一关系是:没有质量就没有数量,没有数量也就没有质量,量变引起质变,质变又能促进新的量变。这种质与量相互关系的哲学观点,是统计研究具体现象数量关系的准则。

2. 总体性

统计研究具体现象的数量方面指的是总体的数量方面。从总体上研究具体现象的数量方面,是统计学区别于其他实质性科学的一个主要特点。

统计对具体现象总体数量方面的调查研究,用的是综合研究方法,而不是对单个事物的研究,但其研究过程是从个体到总体,即必须对足够大量的个体(这些个体都表现为一定的差别、差异)进行登记、整理和综合,使它过渡到总体的数量方面,从而把握具体现象的总规模、总水平及其变化发展的总趋势。比如,了解市场物价情况,统计着眼于整个物价指数的变动,而不是某一种商品价格的变动,但物价统计必须从了解每种有关商品(即代表规格品)的价格变动

情况开始,才能经过一系列的统计工作过程,达到对于物价总体数量变动情况的认识。

3. 具体性

统计学研究的数量方面是指客观现象的具体的数量方面,而不是抽象的数量关系,这是它不同于数学的重要特点。

任何现象都是质量和数量的统一。一定的质规定一定的量,一定的量表现一定的质。因此,必须对现象质的规定性有了正确认识后,才能统计它们的数量。数学研究抽象的数量关系和空间形式,而统计则反映一定时间、地点条件下具体现象的数量特征,它是从定性认识开始,进行定量研究的。比如,只有对工资、利润的科学概念有确切的了解,才能正确地对工资、利润进行统计。

统计研究现象的具体性特点,把它和研究抽象数量关系的数学区别开来,但要注意,统计在研究数量关系时,也要遵守数学表明的客观现象量变的规律,并在许多方面运用数学方法。

4. 客观性

统计数量是客观事物的反映,表示客观现象在具体时间、空间,具体条件作用下,实际已经达到的水平和程度。它独立存在于外部世界,不是人的意志所能转移的。统计资料虽然是经过人们有意识的搜集、整理、汇总、加工,但都不能改变它的客观性。统计资料的客观性是统计质量的基础,基于此,统计资料不但确凿而且雄辩。统计工作只唯“实”,不唯“上”,维护统计资料客观性和真实性,是统计的基本传统。

然而,在统计实践中,“渗水”统计由来已久。最突出的是 20 世纪 50 年代后期“大跃进”那几年,为求“卫星上天”虚报产量,虚假的统计数字造成了严重恶果。改革开放几十年来,统计工作大有改进,但在新的历史条件下,由于利益格局变动等原因,有些地方或单位在统计数字里“渗水”的现象又有所露头。这种现象既违背了统计对象的客观性,也是一种不法行为,我们必须坚决抵制。

5. 变异性

统计研究的变异性是指构成统计对象的总体各单位,除了在某一方面必须是同质的以外,在其他方面又要有差异,而且这些差异并不是由某种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,高等院校这个统计对象,除了都是从事高等教育的教学活动这一共同性质之外,各高等院校在隶属主管部门、院校性质、招生规模、专业设置等各方面又有差异。工人作为统计数据资料对象,每个工人在性别、年龄、工龄、工作性质、工资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。

6. 广泛性

统计学几乎不同程度地渗透到所有人类活动的领域。统计研究对象的范围,既包括社会经济现象,也包括自然科技现象;既包括生产力,又包括生产关系;既有经济基础,又有上层建筑。此外,还要从社会经济与自然技术条件的联系中,研究技术条件对社会经济现象总体的影响。

(三)统计学的学科分类

根据统计学研究对象的内容可以将统计学分为描述统计学和推断统计学两大类。描述统