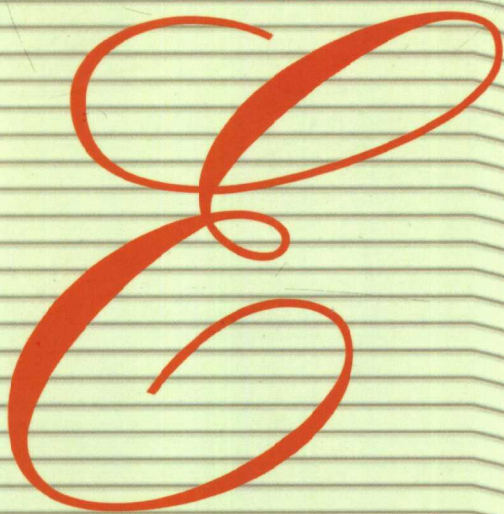




复旦卓越·经济学系列

应用统计学

张梅琳 主编



复旦大学出版社

复旦卓越·经济学系列

应用统计学

主 编 张梅琳

復旦大學 出版社

图书在版编目(CIP)数据

应用统计学/张梅琳主编. —上海:复旦大学出版社,2004.11
(复旦卓越·经济学系列)
ISBN 7-309-04224-7

I. 应… II. 张… III. 应用统计学 IV. C8

中国版本图书馆CIP数据核字(2004)第106212号

应用统计学

张梅琳 主编

出版发行 复旦大学出版社

上海市国权路579号 邮编 200433

86-21-65118853(发行部) 86-21-65109143(邮购)

fupnet@fudanpress.com <http://www.fudanpress.com>

责任编辑 李 华

装帧设计 陈 萍

总 编 辑 高若海

出 品 人 贺圣遂

印 刷 上海复旦四维印刷有限公司

开 本 787×960 1/16

印 张 20.25

字 数 385千

版 次 2004年11月第一版第一次印刷

印 数 1—6 000

书 号 ISBN 7-309-04224-7/F·932

定 价 29.00元

如有印装质量问题,请向复旦大学出版社发行部调换。

版权所有 侵权必究

简要介绍

本书秉持“打好基础,反复操练,注重应用”的理念,由浅入深地介绍统计学中最重要、最实用的数量分析方法,力图用相对较精简的篇幅介绍统计学的精髓;而且可以边学习,边着手用统计软件处理数据。无论你是数量分析初学者,还是高手,花些时间阅读本书,都会有意想不到的收获。

本书特色:

1. 丰富的范例引用,帮助读者从统计学的概念、初步方法开始奠定良好的基础;
2. 浅显易懂的文字并佐以图表说明统计学的各数量分析方法;
3. 与我国经济改革紧密结合的案例分析,将统计学的数量分析方法融会贯通,便于理解;
4. 适合作为非统计专业的统计学教材和经济管理人员的参考书。

前 言

21 世纪是信息经济时代,它的显著特点就是信息爆炸、泛滥。在这些杂乱无章的信息中谁能看出端倪,谁就能掌握新经济时代的制胜先机。虽然,信息经济离不开计算机的普及和应用,但是,从本质上讲,信息经济所依赖的不只是信息处理手段的先进,更重要的是信息收集、整理、分析的准确性和可靠性,而准确、可靠的信息收集、整理和分析离不开统计学学科的发展。因为统计学是助你把握复杂信息现象的工具。也可以说,统计学在一定程度上,对信息经济的发展具有重要的支撑作用。对此国家教育部也十分重视,将统计学列为财经类专业 10 门核心课程之一。由于本书注重应用,将统计指标及统计分析方法通过范例演练,降低了统计学学习的进入障碍,使之能较易体会统计学的应用精髓,因此,本书适合高等院校非统计专业的财经类学生作为学习统计学的必读教材。本书也可作为人口学、医学、法学、新闻学、档案学等非财经类专业学生学习统计学的选读教材和作为广大经济管理人士的参考书。

全书共 12 章,分为 3 个层次,以利于读者和教师分段学习及授课。

第一层次从第一章到第四章,属于基础篇。

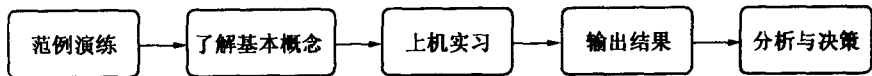
主要介绍描述统计的内容,包括统计数据搜集、萃取、整理及基本统计分析方法的应用。经过这一层次的学习,将了解统计调查的组织方案,运用和掌握 Excel 及 Eviews 软件对调查结果进行整理和制作图表,通过总量指标、相对数指标和绝对数指标等进行数据的初步分析。

第二层次从第五章到第十章,属于数据分析篇。

主要介绍推断统计、预测和决策的内容,包括参数估计、显著性检验、方差分析、回归分析和预测、时间数列分析和预测、决策分析方法等。通过这一部分的学习,可掌握绝大部分统计分析方法,并应用 Excel 及 Eviews 软件进行复杂的数据处理,应用这些处理结果来帮助判断与决策。

第三层次从第十一章到第十二章,属于综合篇。

主要介绍国民经济核算体系及案例分析的内容。全书的构思用图来表示为:



本书的特点是从范例、实用性和易掌握性出发,重点突出、内容丰富,可运用和操作性强。全书运用大量的图表来阐述概念,通过上机实习来掌握统计学的应用精髓,掌握精髓后,再来进一步理解较复杂的数学结构,以达到举一反三的效果。

本书的编者积多年统计教学和使用各种软件进行统计分析的经验及体会,集思广益编写此书。本书由张梅琳主编和主审。本书各章的编写分工如下:第一、二、三、四、九、十章张梅琳;第五、六、十一章施方;第七、八、十二章王惠萍;部分范例、习题收集由金戈、龚水燕、刘忠勋、刘霖云完成;部分计算机软件应用与开发由沈根荣完成。

尽管各位编者竭尽了自己的所能,但是限于时间仓促,疏漏之处在所难免,恳请读者不吝赐教,以期再版时予以修正。

本书的出版,得到了复旦大学出版社的支持、帮助,对为了本书出版付出心血的编辑们及有关同事、朋友的关心,致以衷心的感谢。

编者

2004年8月

目 录

第一章 导论	1
第一节 数据信息和统计学.....	1
第二节 数据信息处理的方法.....	4
第三节 计算机在统计中的应用.....	7
思考题.....	9
第二章 数据搜集与整理	10
第一节 数据的搜集	10
第二节 调查设计	17
第三节 数据的整理	21
第四节 频数图	26
第五节 计算机软件的应用	30
思考题	37
习题	37
第三章 数据的汇总	39
第一节 统计变量集中趋势的测定	39
第二节 统计变量离散程度的测定	50
第三节 计算机软件的应用	58
思考题	63
习题	63
第四章 统计比率与统计指数	66
第一节 统计比率	66
第二节 统计指数	70
第三节 统计指数在因素分析中的应用	76
第四节 价格指数的编制与应用	80
思考题	87
习题	88
第五章 抽样基础	91
第一节 概率概述	91

第二节	随机变量和随机变量的概率分布	94
第三节	三种常用的概率分布	101
第四节	大数定律及中心极限定理	106
第五节	抽样设计	109
第六节	抽样分布	114
第七节	Excel 的应用	118
思考题		125
习题		126
第六章	参数估计和假设检验	128
第一节	点估计	128
第二节	参数的区间估计	129
第三节	假设检验的一般方法	139
第四节	总体均值的假设检验	143
第五节	总体比例的假设检验	146
第六节	总体方差的假设检验	148
第七节	置信区间与假设检验之间的关系	149
第八节	Excel 分析	151
思考题		154
习题		155
第七章	方差分析	157
第一节	单因素方差分析	157
第二节	无交互作用的双因素试验方差分析	166
第三节	有交互作用的双因素试验方差分析	170
第四节	Excel 在方差分析中的应用	175
思考题		180
习题		180
第八章	相关分析与回归分析	183
第一节	相关分析	183
第二节	一元线性回归分析	189
第三节	多元线性回归分析	201
第四节	计算机软件的应用	205
思考题		211
习题		211

第九章 时间数列	214
第一节 时间数列概述	214
第二节 时间数列的分析指标	216
第三节 时间数列分析与预测	225
第四节 计算机软件的应用	238
思考题	246
习题	246
第十章 决策分析	248
第一节 决策概述	248
第二节 随机型决策	252
第三节 非随机型决策	255
第四节 风险分析	258
思考题	262
习题	262
第十一章 国民经济核算体系及主要指标	264
第一节 国民经济核算体系	264
第二节 国民经济核算体系的基本统计指标	266
思考题	271
习题	271
第十二章 统计学案例分析	273
案例一 橡胶制品加工工艺质量的统计控制	273
案例二 城市特征和 FDI 的空间经济计量分析	282
案例三 时间数列灰色预测模型	290
部分习题答案	298
附表 统计表	303
参考文献	312

第一章 导 论

随着我国经济体制改革的快速推进以及经济全球化趋势的日益显现,人们逐步认识到统计在信息时代中的重要作用。站在 21 世纪的起跑线上,统计学的发展将紧跟时代前进的步伐,以系统、科学的统计理论及信息技术为基础,以改革和创新的思维寻求突破,逐步实现统计现代化。让我们以此为主线来阐述本书的基本概念、核心内容和主要框架。

第一节 数据信息和统计学

一、数据信息和统计学的关系

人们在日常政治、经济和社会文化活动中,经常要接触各种数字化的信息。例如:①某年我国国内生产总值(GDP)为 67 800 亿元,按可比价格计算比去年增长 9.7%。其中第一产业增长 5.1%,第二产业增长 12.3%,第三产业增长 8%。②某年城镇总人口为 30 191 万人,乡村总人口为 84 142 万人,城市人均居住面积为 6.7 平方米,乡村为 17.8 平方米。③某年我国工会会员 1 亿多人,基层工会达 61 万多个。这些数字化的信息统称为统计资料。换言之,统计资料是反映客观事物数量特征的数字信息。平时我们编制国民经济计划,进行市场调研,作经营决策,企业管理以及从事科学研究时,都需要运用某些统计资料。

统计资料的质量好坏、数量的多少取决于统计工作,统计资料是统计工作的成果。统计工作是包括对社会、经济以及自然现象的数量方面进行搜集、整理、分析的活动过程。

统计工作的发展和深化需要统计理论的指导。有关统计的理论就是统计学。总起来说,统计学是一门关于大量数据资料的搜集、整理、描述、推断和分析的学科,是一门方法论的科学。统计学来源于统计工作,是统计工作经验的理论概括,它反过来又指导统计工作,推动统计工作的不断发展和提高。

二、统计学的性质及其特点

上面我们介绍了统计学的定义,即统计学是一门关于数据资料的收集、整理、描述、推断和分析的方法论学科。它是用来研究数据现象的,这种现象既包括社会

现象又包括自然现象。由此,统计学不得不面对两个问题,第一个问题是这些数据从哪里来的,就是说,所展示的数据现象是真的吗?怎样找出或采集数据?第二个问题是这堆数据在说明什么?它对我们的生活、工作等有什么特别意义呢?统计学就是在探索和回答这两类问题中得到了发展。统计学的整个结构也在探索这两类问题的解答途径中形成。这就规定了统计学的性质是通过大量数据信息搜集和挖掘,对事物的数量方面进行研究,进而达到探索数据信息内在数量规律的目的。因此,统计学与统计工作的关系是理论与实践的关系,表现在性质上则是:统计工作是对客观事物开展实质性调查研究的工作,而统计学则是对这种调查研究的理论概括。可见,从统计工作总结出来并使之上升为理论的,是如何做好统计工作的原理、原则和方法。统计方法的运用效果取决于基础数据,而收集何种基础数据,怎样节省收集数据的成本,如何降低数据收集过程中的误差,都需要一定的理论指导,统计学为回答这些问题提供了许多很有效的解决方案。用这种方法论来指导统计工作,能帮助统计工作不断提高和拓展。因此,统计理论是一种方法论科学,不是实质性科学。对客观事物进行统计研究,探讨其本质和规律性,是统计工作的任务,不是统计学的任务。

显然,统计学的最突出的特点是数量性和大量性。数量性是统计学的首要特点,但仅凭这一点不足以将统计学与其他研究事物数量方面的学科相区别,如会计学、审计学等。统计学与之相区别的主要特征是大量性。统计学是运用大量的数据资料来综合反映事物的数量方面的。这主要是由于客观事物中个体与个体之间的差异、各自之间的差异是普遍存在的。例如在同一民族、同一性别、同一年龄的人群中,其身高、体重各不相同。这种差异是各种偶然因素影响的结果。因此,少量的数据并不能代表事物的一般特征,不能说明事物发展的规律性。必须运用足够大量的数据资料加以综合分析,才能使事物中非本质的偶然因素的影响互相抵消或削弱,事物的一般特征,也就是其共性才能显示出来。这是马克思主义哲学的共性与个性的对立统一规律的具体表现。

大量性运用于统计调查就是大量观察法,即必须观察事物的全部或足够数量的单位,才能使事物的个别的、偶然的差异相互抵消,从而显示事物的集体的、必然的规律性。运用大量观察法,我们可以对调查对象的所有单位都进行调查,如统计报表、普查;亦可只对能够反映全体特征的部分单位进行非全面调查,如抽样调查、重点调查。大量观察法并不否定个别典型单位的作用,它可以与典型调查相结合,取得典型单位的数据,以加深对事物的认识。

统计学的第二大特点是差异性和具体性。客观事物的差异是普遍存在的。例如商业职工的工龄、工资各不相同,各个商业企业的效益也有差异。正因为这些个体之间存在着差异,才需要综合大量个体的差异,以掌握集体的综合特征。

差异可以表现为数量方面(如工龄、工资、产量、销售额等),也可以表现为非数量方面(如性别、民族、所有制等)。非数量方面的差异,只有将之转化为数量方面的差异,才能成为统计学研究的对象。

客观事物的数量方面将随着时间的变化,表现为一定的差别和差异,因此统计学对客观事物时间上的差异也展开研究。例如对事物发展的动态分析等。

统计学研究的数量方面是客观事物具体的数量。统计学的数据都是具体的,反映一定研究范围的数字。这与数学不同。数学完全撇开研究对象的具体内容和质的特征而研究抽象的数量关系。统计学是在一定质的规定下研究事物的数量方面,它需要在质和量的辩证统一中进行研究。因此必须对事物质的规定性有了正确的认识后,才能统计它们的数量。

统计学认识事物是从定性认识开始再进入定量认识的,定性认识是定量认识的前提和基础,明确这点尤为重要。例如要统计商业增加值,首先要明确商业增加值的概念。商业增加值是用货币表示的商业部门劳动者在一定时期内经营活动的最终成果,它是从总产出中扣除中间消耗而得到的。如果不明确商业增加值的概念,不把总产出中转移的中间消耗的价值与新增价值区分开来,那么就无从正确统计商业增加值的数量。由此可见统计研究必须将定性和定量结合起来,遵循定性一定量一定性的科学认识规律。

统计学是一门融合自然科学和社会科学,多学科渗透、融合的科学。统计学的起源可追溯至 18 世纪甚至更早,然而统计学主要的发展却迟至 19 世纪末叶 20 世纪初期才真正开始,到了 20 世纪 40 年代才逐渐成熟。统计学和概率论的关系密切,可以说任何统计问题的研究都或多或少地牵涉到概率论的运用,因为后者实际上已成为前者的主要工具。随着芯片制造技术和软件工程的迅猛发展,计算机学科的不断渗入,统计方法得到了大规模推广。统计学依赖于计算机软件的计算和分析能力的不断增强或者说计算成本的大幅下降,其应用领域不断开拓,计算机让统计学变成了真正的实用学科。统计学历经百余年的发展,其学科内各学派之间既有融合,又有分化,其学科外受到自然科学和社会科学的相互影响和渗透,特别是当今现代边缘科学即交叉科学不断涌现,统计学与它们相结合,产生了一系列专门的或专业的统计学,从而使统计学成为一门多科性的学科体系。以研究领域为标志划分,统计学学科体系大体上可分为三个部分:① 数理统计学。它以概率论为基础,以抽样为核心,研究随机变量的变化规律,是在社会经济统计和自然科学及计量技术统计实践基础上抽象和概括出来的纯方法论学科。② 应用技术统计学。包括生物统计学、心理统计学、气象统计学、统计物理学、医药与卫生统计学等。这些专门的统计学是数理统计学同自然科学和应用技术科学相结合的交叉学科,是具有自然科学和技术科学性质的学科。③ 社会经济统计学。它是以人口、

社会、经济、科技和文化等人类自身及其活动为对象的统计方法论。20 世纪后半世纪,人文社会科学的发展与统计学的关系越来越紧密,统计学的发展已经渗透到人文社会科学的许多领域,并由此产生许多新的学科,如人口统计学、历史统计学、教育统计学、心理统计学、社会统计学、科技统计学、环境统计学,等等。统计学与人文社会科学的结合,改变了原有单一学科发展的思路、视野和应用功能,对人文社会科学的发展具有极大的支撑作用;反过来,这种结合又促进了统计学的发展。但是,最根本的是统计学对人文社会科学巨大的推动作用,这种推动作为一个大趋势将在 21 世纪得到更充分的体现。上述三种统计学构成现代统计学的主体部分。此外,还有统计史、比较统计学等。

第二节 数据信息处理的方法

统计数据信息处理的方法主要包括两个方面,即描述统计方法和推断统计方法。简单地说:任何对数据(即样本)的处理导致预测或推论总体的效果称为推断统计。反之,如果我们的兴趣只限于手头现有的数据,而不准备把结果用来推断总体则称为描述统计。举个例子来说,依据过去 10 年来的统计,每年来上海旅游人数,平均每人在上海停留的日数,平均每人每天在上海的花费,10 年内哪一年创最高纪录等等都是属于描述统计的范围;但是如果我们根据这些年的数据预测来年可能的旅游人数及消费额就是推断统计的问题了。

一、描述统计方法

描述统计(descriptive statistics)是主要对已收集到的统计数据信息进行加工、分组、编制统计表、绘制统计图及计算相对数、平均数、方差等,以反映事物的数量特征和数量关系的统计方法。

对所搜集的资料进行描述和比较的主要动机是使反映客观事物的统计数据可以一目了然、条理清晰、便于使用。因此必须对原始数据资料按某一标志分类、整理,如按地区分组、按数量分组、按时间先后顺序分组等。所谓标志是指构成事物的各个单位所具有的属性和特征。例如每一工人可以有性别、年龄、民族、文化程度、工种、工龄、工资等属性和特征。标志可分为品质标志和数量标志。标志如以品质属性表示,即以文字来表达的,称为品质标志,如工人的性别分为男、女;文化程度可分为小学、初中、高中等。标志如以数量特征表示,即以数字来表达的,称为数量标志,如生产工人的年龄、工资、工龄等。

我们一般将品质标志不同的具体表述,称为变异。如民族标志,其具体表述为汉、满、蒙、回、藏、维……有时也将数量标志的不同取值称为变异。但一般是将数

量标志不同的具体表现称为变量。变量的不同取值,称为变量值或标志值。如工人年龄变量,其取值可30岁、40岁、50岁不等;工资变量,其取值可75元、80元、100元……

变量按其取值是否连续,可分为离散型变量和连续型变量。离散型变量指只能取整数的变量。职工人数、商店数、设备数等都属于离散型变量。连续型变量是指在整数之间可插入小数的变量。身高、体重、增加值、成本、利润等都属连续型变量。

变量按其所受因素影响的不同,可分为确定性变量和不确定性变量。由确定性因素影响所形成的变量为确定性变量,由不确定性因素影响所形成的变量为不确定性变量。

将标志的具体表现(变异和变量)进行登记、汇总,最后形成说明事物综合特征的各种数字资料,称为统计指标。如按生产工人的性别标志汇总计算出男、女各多少人;按工资标志汇总,计算出工资总额。因此统计指标是综合反映统计总体数量特征的。指标由指标名称和指标数值两部分组成。指标名称反映总体某一方面质的规定性,是对总体本质特征的一种概括,必须以一定的理论为依据做出解释,同时又要界定其范围,即哪些该计算,哪些不该计算。如,企业职工工资总额,它是指企业在一定时期(1个月或1年)按全部职工提供劳动数量和质量所支付的劳动报酬总额,包括规定的生产性奖金和津贴,而不包括各种非生产性奖金,即职工从企业股份中分得的利息、红利以及各种福利性货币收入和实物收入。这种计算范围就是统计指标量的规定性。同一指标名称,因具体的对象也就是总体的不同,以及统计的时间、地点、条件的不同,而产生不同的指标数值。由此可见,统计指标具有两个重要的特点:① 具体性。它是说明总体在具体时间、地点、条件下的数量特征,是总体某一方面质和量的辩证统一。② 综合性。它说明的不是个别单位或部分单位的数量特征,而是构成总体的全部单位的综合结果。因为指标数值的取得,是通过对总体单位调查登记并加以汇总整理的结果。

统计指标按其所反映的内容不同,可分为数量指标和质量指标。数量指标是说明事物总规模、总水平或工作总量的各种总量的指标,即绝对数指标,它可以按品质的变异(如所有制、性别)汇总得到各组单位数和全部单位总数的总量指标,也可以按数量标志值加总得到各组标志总量和全部标志总量,如销售额总量、GDP总量等。质量指标是反映事物的经济效益和工作质量的各种相对指标和平均指标。质量指标通常由两个数量指标,以绝对数对比求得,也可以由两个质量指标(即平均数或相对数)或一个数量指标一个质量指标对比求得。如恩格尔系数、基尼系数等。

统计指标按其所采用的计量单位的不同,可分为实物指标和价值指标。实物

指标采用实物计量,直接反映产品使用价值的数量;价值指标采用货币计量,反映产品的社会必要劳动量。

单个指标只能反映总体某一方面的数量特征。社会经济现象错综复杂,统计总体往往具有多重特征。为了全面反映总体的数量特征,就要用一整套统计指标。这种具有内在联系的一系列指标所构成的整体,称为统计指标体系。如为了反映企业财务状况,可建立固定资产、流动资金占用和周转,以及税金、利润等财务指标体系。

统计指标体系大体上可分为基本统计指标体系和专题统计指标体系两类。基本统计指标体系是反映国民经济和社会发展基本情况的指标体系。如国民经济核算指标体系、科技进步指标体系等。专题统计指标体系是针对某项社会经济问题而制定的专项指标体系。如价格指数体系、小康生活水平测定指标体系、企业经济效益指标体系等。指标体系可能会随着社会经济的发展变化而变化,但指标体系一经制定,就要力求保持相对稳定,以便积累历史资料,进行系统的比较分析。

二、推断统计方法

推断统计(inferential statistics)也称为归纳统计学(inductive statistics),是统计学的一个重要组成部分。推断统计分为三大类,就是估计、检验和选择。譬如说,有一家企业,想了解其产品的市场占有率,于是以随机抽样的方式,询问 100 位市民的意见,而后根据所得结果推断可能全市有多少人 would 购买其产品,这是估计问题。又如某机构想知道奥妙牌洗衣粉的洗净力是否比白猫牌洗衣粉强,首先假设奥妙牌比白猫牌好,然后经过试验来测定这种假说是否成立,在本例中,我们并不想估计任何参数,而只是想检验事先叙述的假设是否成立,其可靠性有多大,这就是检验问题。还有,新制造的三种药品中哪几种比目前所用的药品更有效?这是选择问题。它们都是以样本为基础,它们的主要内容都是研究如何应用概率理论,以样本来推断总体。

所谓总体(population)是指特定的统计研究对象的全部个体的总和。将构成总体的各个单位,称为总体单位。总体可分为无限总体和有限总体。如果总体中所包含的单位数是无限的,则称该总体为无限总体。如连续大量生产的零部件,其产量可以是无限的。如果总体中所包含的单位数是有限的,则称该总体为有限总体。对无限总体只能抽取部分单位进行非全面调查,据以推断总体。对有限总体既可以进行全面调查,也可以进行非全面调查。例如,我们研究上海市职工家庭收支,则上海市全部职工家庭收入就是总体,并且是有限总体。由于逐户调查上海市全部职工家庭的收支情况,在实际工作中是行不通的,而且也没有这种必要,因此只能抽取若干户职工家庭(例如抽取 4 000 户)进行调查。这若干户(如 4 000 户)

的家庭收入就是样本(sample),通过样本调查的结果来推测总体的某种特征。

当由样本来推断总体的特征时,统计总体有全级总体和抽样总体(即样本)两个不同范围的概念:用于说明全级总体的指标,我们称之为参数;根据样本资料汇总计算的指标,即样本指标,我们称之为统计量。如在参数估计时,以样本平均数在一定的概率保证下,估计总体均值。这里样本均值就是统计量。统计量是对全级总体参数进行推断的依据。

要说明的是,总体、总体单位以及前述的标志和指标都是相对而言的,即它们不是固定不变的,随着研究目的的不同,是可以转化的。当研究工业企业生产结构的情况时,全部工业企业就是总体,每一工业企业为总体单位,而全部工业企业的产值总量为指标,每一企业的产值为标志;当研究某企业内部生产结构情况时,则某工业企业的全部职工就是总体,该企业每一职工就是总体单位,这时该企业的产值就是指标,而该企业每一职工的生产量就是标志了。所以说指标是对总体而言的,而标志是对总体单位而言的。

推断统计方法主要是通过统计估计和检验两个方面来完成的。无论统计估计还是检验,都是概率论所揭示的随机变量的一般规律,是利用抽样调查获得样本信息,构造样本指标,即统计量,对总体的某些性质或数量特征展开的推断。

估计是以总体中抽出的样本为依据,根据样本资料计算出某一指标如样本平均数、比例、方差等统计量,以这些样本指标即统计量数值作为总体的参数估计值,来推断总体的某个数量特征。因此,估计问题侧重于用样本统计量估计总体的某一未知参数。

检验是先对总体的某个特征或属性是否存在作出假设,然后根据样本资料的计算结果(统计量)来推断所作的假设是否可以接受。例如在农业生产中可以检验新品种和对照品种是否存在显著性差异,以判断新品种是否有明显的增产效果,有无推广的价值。故检验问题侧重于用样本资料验证总体是否具有某种性质或数量特征。

第三节 计算机在统计中的应用

21世纪是信息的时代。现代统计工作就是一个基于现代信息技术,实施网络化运行和管理的庞大复杂的开环或闭环复合系统。随着统计方法制度改革的深化、信息技术的日益进步和统计组织机构的不断优化,统计的工作链将逐步缩短,整个统计系统的综合效率就可以大为提高,统计工作在社会中的地位和作用就更加显要了。我国加入WTO后,要求统计工作摆脱手工操作状况,实现统计现代化。这一任务是通过计算机的应用而实现的。计算机在统计中的应用主要体现在

统计信息的存贮和检索、统计资料的分析和检验等方面。

一、统计信息的存贮和检索

随着现代科学技术的迅猛发展,人类社会已开始进入到一个利用和开发信息资源的信息社会。信息数据数量之大,范围之广,变化之快,致使传统的手工与机械数据处理手段已无法适应社会、经济高速发展对统计所提出的要求,也难以提高数据处理的速度和精度。计算机在统计中的应用解决了统计信息的存贮和检索问题,从而解决了统计工作中的一大难题。计算机体积小,但可存贮的信息容量却极大。以微型计算机为例,80286 能够直接存取 16 M 的内存,而目前 Pentium(译名奔腾)则能直接存取 4 096 M(4 G)的内存,并配有一台或两台软盘驱动器及两台硬盘。由此可见,微型机的外存容量已相当可观。而新的光盘技术的发展,又使计算机的容量成倍增加。目前仅一张 5.25 英寸的磁光盘,可以存贮 600 兆字节信息,信息的存贮量相当于 500 张高密度软盘或 30 万页文献,使大量统计信息的存贮成为轻而易举的事情。而多媒体技术的发展,使计算机不仅可存贮文字、图形,而且可将声音、动画、视频以数据形式存贮在计算机内,使统计数据的存贮出现新的突破。计算机的高速运转,又使统计信息的检索大为便利,目前,在亿分之一秒的时间内执行一条指令已是常见的水平,使所需统计信息能在瞬息之间即从大量信息数据中检索出来,从而大大减少了人的劳动。尤其是计算机所配有的数据库,使统计信息得以有序存贮、自动加工处理和迅即取出。

二、统计资料的分析和检验

在今天的信息社会里,统计分析研究工作越来越引起经济、社会等各界的重视。经济工作者及管理决策部门,要想做到纵观全局,了解现状,推断未来,就需要对大量的信息资料进行分析处理。然而,今天要想通过传统的手工方法来获得及时准确的信息,是十分困难的。只有发挥计算机的强大功能,才能获得具有高度逻辑性、预见性的信息。

计算机在统计中的应用,可以及时、准确地将有关统计资料的统计参数如均值、方差、标准差、偏态等迅速地提供给人们,特别是对一些数据量较大,难以用手工完成的统计工作,计算机更体现了它的优越性。

计算机在统计中的应用,还表现在利用经过分析处理的次级统计资料进行统计检验、概率分析。通过对某种产品的实验研究,或对某种现象的抽样调查,有助于在实际工作中避免主观性和盲目性,并最终选择出一种最佳方案。过去由于统计数据量过大、过于复杂,因而给统计检验带来很大困难,而计算机的应用,使上述检验成为相当容易的工作。特别是采用计算机统计分析软件,像 Minitab, Excel,