

新编统计学

(修订本)

主编 张梅琳

ZHUBIAN
ZHANGMEILIN

LIXIN KUALIJI
CHUBANSHE

立信会计出版社

XINBIAN TONGJIXUE

图书在版编目 (CIP)数据

新编统计学/张梅琳主编. —2版(修订本). —上海:立信会计出版社, 1999.9
ISBN 7-5429-0476-0

I. 新... II. 张... III. 统计学 IV. C8

中国版本图书馆 CIP 数据核字(1999)第 44301 号

出版发行 立信会计出版社
电 话 (021)64695050 × 215
(021)64391885(传真)
(021)64388409
地 址 上海市中山西路 2230 号
邮 编 200233
E-mail *lxaph@sh163c.sta.net.cn*
出 版 人 陈惠丽

印 刷 上海申松立信印刷厂
开 本 850 × 1168 毫米 1/32
印 张 9.875
字 数 239 千字
版 次 1997 年 5 月第 2 版
印 次 2000 年 8 月第 11 次
印 数 3 000
书 号 ISBN 7-5429-0476-0/F·0444
定 价 15.00 元

如有印订差错 请与本社联系

前 言

本书是在1994年版的基础上修订的。它突破了传统统计学教材的框架,注重教材内容适应市场经济的需要,适应统计工作现代化的要求,并将计算机在统计中应用的内容编入教材。本书可作为高等院校财经类统计学课程的必读教材,以及社会学、人口学、新闻学、法学、档案学等非财经专业学习统计学的选读教材,也可以作为广大经济管理人士的参考书。

本书分四部分:描述统计部分(第一章至第四章),为读者提供统计数据搜集、整理及基本统计分析方法;推断统计部分(第五章至第七章),主要讲述了参数估计和显著性检验等内容;统计预测(第八章至第九章),介绍了回归分析和时间数列的预测方法;计算机在统计中应用的内容放在第十章。

经过几年的教学实践,积累了经验,发现了本书第一版中的一些问题,并吸收了广大读者的意见,修订本就是在这一基础上编写的。修订本修改了第一版中存在的不足之处,并致力于提高教材的质量,因此在选材和叙述上尽量做到联系财经类专业的实践,注重应用,力图将概念写得清晰易懂,便于教学。为适应经济发展的需要,加强了推断和计算机在统计中应用的内容。例如,根据近年来计算机发展迅速、统计软件更新快这一新情况,对计算机部分的内容作了大量的修订和补充。

本书由张梅琳主编,卞祖武教授主审。各章编写分工如下:第一、五、六、七章张梅琳,第二章丁又双,第三章刁德成,第八章卞祖武,第四、九章高启中,第十章沈根荣。本书修订过程中,周锡樵、许

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	9
思考题	10
第二章 描述统计 I : 数据搜集与整理	12
第一节 数据的搜集	12
第二节 数据的整理	17
第三节 频数图	25
思考题 习题	30
第三章 描述统计 II : 数据的汇总	33
第一节 统计变量集中趋势的测定	34
第二节 统计变量离散程度的测定	44
思考题 习题	54
第四章 描述统计 III : 统计比率与统计指数	58
第一节 统计比率	58
第二节 统计指数	63
第三节 统计指数在因素分析中的应用	70
第四节 指数数列	74

第五节	价格指数的编制与应用	76
思考题	习题	83
第五章	推断统计 I: 抽样推断法	88
第一节	抽样的必要性	88
第二节	概率基础	92
第三节	随机变量和随机变量的概率分布	96
第四节	三种常用的概率分布	106
第五节	大数定律及中心极限定理	112
第六节	抽样推断	115
思考题	习题	129
第六章	推断统计 II: 显著性检验	133
第一节	显著性检验的基本要素	133
第二节	总体均值的显著性检验	140
第三节	总体比例的显著性检验	145
第四节	总体方差的显著性检验	151
思考题	习题	160
第七章	推断统计 III: 统计变量和统计函数	163
第一节	变量与函数	163
第二节	用统计表、统计图和方程式表示统计函数	166
第三节	统计函数的估计	168
第四节	统计函数的设计	176
思考题	习题	182
第八章	统计预测 I: 回归分析预测	184
第一节	一元线性回归分析预测	184

第二节 多元线性回归分析预测·····	199
第三节 非线性回归分析预测·····	205
思考题 习题·····	215
第九章 统计预测 I: 时间数列 ·····	219
第一节 时间数列概述·····	219
第二节 时间数列的分析指标·····	223
第三节 时间数列分析与预测·····	233
思考题 习题·····	252
第十章 计算机在统计中的应用 ·····	255
第一节 电子计算机概述·····	255
第二节 统计软件概述·····	265
第三节 SPSS 统计软件包 ·····	274
思考题·····	285
部分习题答案 ·····	286
附表 ·····	292
1. 二项分布表 ·····	292
2. 泊松分布表 ·····	298
3. 正态分布表 ·····	300
4. t 分布表·····	301
5. F 分布表 ·····	302
6. X^2 分布表 ·····	304
参考书目 ·····	305

第一章 导 论

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 1996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 1990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 1993年我国工会会员1亿多人,基层工会达61万多个^①。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

① 资料来源:1990、1993、1996年有关统计资料。

计工作,推动统计工作的不断发展和提高。由于在现代汉语中“统计资料”、“统计工作”和“统计学”是三个涵义不同的名词,意义明确、不会造成混淆,因此我们口语常将这三者简称为“统计”。英语的区别是:当 statistics 一词作为复数词应用时,指的是统计资料;statistic 作为单数词应用时,指的是统计学。

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。这就规定了统计学的性质是通过大量数据信息,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。用这种方法论来指导统计工作,能帮助统计工作不断提高。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。数量性是统计的首要特点,但仅凭这一点不足以将统计学与其他的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,以掌握全体的综合特征。

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指客观事物的具体的数量方面。统计的数据都是具体的,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把定性和定量结合起来,遵循定性——定量——定性的科学认识规律。

统计学又是一门多学科渗透、融合的科学。统计学历经几百年的发展,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。以研究领域为标志对其进行划分,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。包括社会统计学、经济统计学、科技统计学、环境统计学等。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,即描述统计法和推断统计法。

一、描述统计方法

描述统计(descriptive statistic)主要包括对已收集到的统计

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随机性变量。由确定性因素影响所形成的变量为确定性变量,由随机性因素影响所形成的变量为随机性变量。

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作出解释,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(一个月或一年)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。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。

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。如价格指标体系、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

统计推断(statistical inference)也称为归纳统计学(inductive statistic),是统计学的一个重要组成部分,其主要内容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,由于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。

这里要加以说明的是,总体、总体单位以及前述的标志和指标都是相对而言的,即不是固定不变的,它们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。故检验问题则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

近年来,由于各国社会、经济高速发展,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表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。电子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 16M 的内存,而目前 Pentium(译名奔腾)则能直接存取 4096M(4G)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。计算机的高速运算,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。

对于计算机在统计中的应用问题,我们将在本书的第十章作较详细的阐述。

思 考 题

1. 试说明统计资料与统计学的关系。
2. 试述统计学的性质及其特点。

3. 解释并举例说明下列概念：统计总体、总体单位、标志、变异、指标、变量。

4. 试说明标志与指标的区别和联系。

5. 统计科学研究的基本方法是什么？

6. 试说明计算机对统计工作的作用。