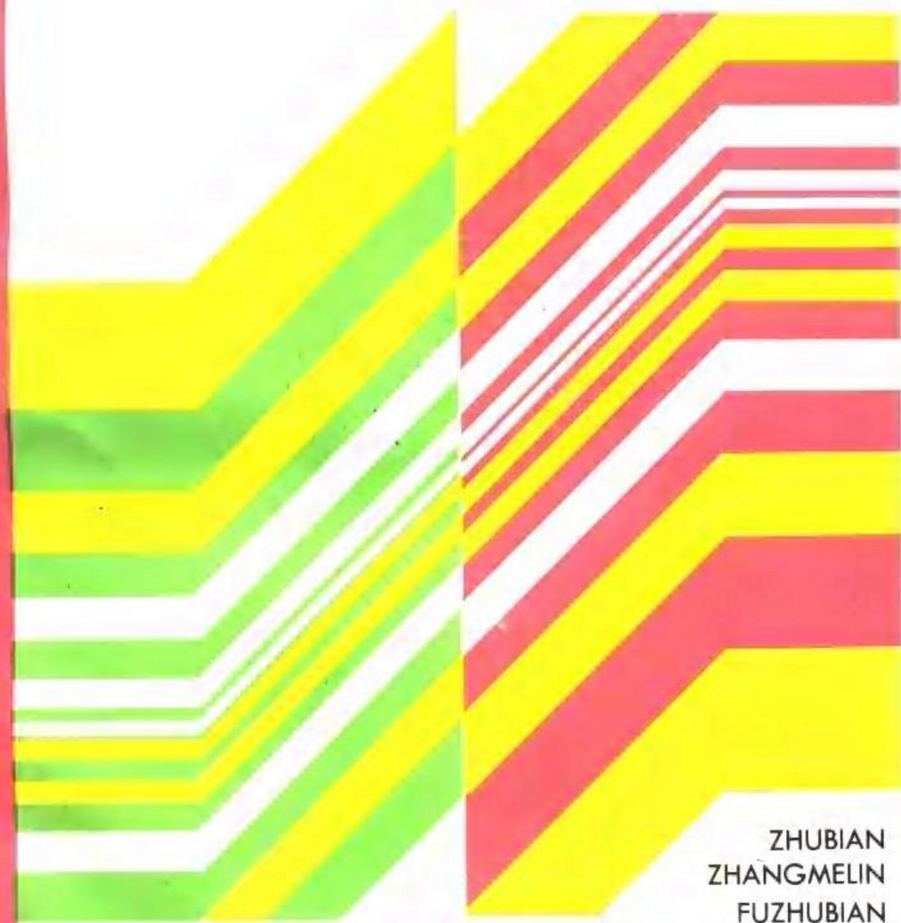


# 新编统计学

- 主 编 张梅琳
- 副主编 高启中 沈根荣

LIXIN KUAIJI CHUBANSHE  
立信会计出版社



XINBIAN TONGJIXUE

ZHUBIAN  
ZHANGMELIN  
FUZHUBIAN  
GAOQIZHONG  
SHENGENRONG

01  
08  
67

# 新编统计学

主 编 张梅琳  
副主编 高启中 沈根荣  
主 审 卞祖武

立信会计出版社

(沪)新登字 304 号

**新编统计学**

张梅琳 主编

立信会计出版社出版发行

(上海中山西路 2230 号)

邮政编码 200233

新华书店经销

上海联合科教文印刷厂印刷

开本 850×1168 毫米 1/32 印张 9.5 字数 230,000

1994 年 12 月第 1 版 1996 年 4 月第 4 次印刷

印数 27,001—32,000

ISBN7-5429-0207-5/F·0199

定价: 12.00 元

## 前 言

统计学是我国高等院校财经类专业通开的核心课程。本书是在该课程教改的基础上编写的,它突破了传统统计学教材的框架,强调教材内容要适应市场经济的需要,适应统计工作现代化的要求,并尝试性地将计算机在统计中应用的内容编入教材。本书可以作为高等院校财经类专业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。

本书共十章,内容有导论;描述统计,包括统计图表法和数据汇总;推断统计,包括抽样推断、显著性检验和统计变量、统计函数等内容;统计预测,包括回归分析和时间数列;其它统计信息处理,包括统计比率和统计指数;以及计算机在统计中的应用等内容。

参加本书编写的有:张梅琳(第一、四、五、六章)、刁德成(第三章)、倪加薇(第二章)、卞祖武(第七章)、高启中(第八、九章)、沈根荣(第十章)。张梅琳任全书主编,高启中、沈根荣任副主编,卞祖武任主审。本书编写与出版过程中得到了朱祖萱、周锡桥、洪家敏、许德因、孙远欣等许多同志的热情支持和帮助,在此一并致谢。

限于水平,本书舛误之处一定不少,尚望读者不吝批评指正。

编 者

# 目 录

<b>第一章 导论</b> .....	(1)
第一节 数据信息和统计学.....	(1)
第二节 统计数据信息搜集的方法.....	(4)
第三节 统计数据信息处理的方法.....	(6)
第四节 计算机在统计中的应用.....	(9)
思考题 .....	(11)
<b>第二章 描述统计 I:统计图表法</b> .....	(12)
第一节 频数表 .....	(12)
第二节 统计数列和统计表 .....	(16)
第三节 频数图 .....	(20)
思考题 习题.....	(26)
<b>第三章 描述统计 II:数据的汇总</b> .....	(28)
第一节 统计变量集中趋势的测定 .....	(29)
第二节 统计变量离散程度的测定 .....	(39)
思考题 习题.....	(49)
<b>第四章 推断统计 I:抽样推断法</b> .....	(52)
第一节 抽样的必要性 .....	(52)
第二节 概率基础 .....	(56)
第三节 随机变量和随机变量的概率分布 .....	(60)
第四节 三种常用的概率分布 .....	(68)
第五节 抽样推断 .....	(75)
思考题 习题.....	(88)

<b>第五章 推断统计 I:显著性检验</b> .....	(92)
第一节 显著性检验的基本要素 .....	(92)
第二节 母体均值的显著性检验 .....	(98)
第三节 母体比例的显著性检验 .....	(103)
第四节 母体方差的显著性检验 .....	(109)
思考题 习题 .....	(117)
<b>第六章 推断统计 II:统计变量和统计函数</b> .....	(120)
第一节 变量和函数 .....	(120)
第二节 用统计表、统计图和方程式表示统计函数 .....	(123)
第三节 统计函数的估计 .....	(125)
第四节 统计函数的设计 .....	(133)
思考题 习题 .....	(138)
<b>第七章 统计预测 I:回归分析预测</b> .....	(140)
第一节 一元线性回归分析预测 .....	(140)
第二节 多元线性回归分析预测 .....	(155)
第三节 非线性回归分析预测 .....	(161)
思考题 习题 .....	(172)
<b>第八章 统计预测 II:时间数列</b> .....	(175)
第一节 时间数列概述 .....	(175)
第二节 时间数列的分析指标 .....	(178)
第三节 时间数列分析与预测 .....	(188)
思考题 习题 .....	(207)
<b>第九章 其它统计信息处理法 I:统计比率与统计指数</b> .....	(210)
第一节 统计比率 .....	(210)
第二节 统计指数 .....	(215)
第三节 统计指数在因素分析中的应用 .....	(222)
第四节 指数数列 .....	(225)
第五节 价格指数的编制与应用 .....	(227)

思考题 习题 .....	(235)
<b>第十章 其它统计信息处理法 II: 计算机在统计中的应用...</b>	<b>(239)</b>
第一节 电子计算机概述 .....	(239)
第二节 统计软件包概述 .....	(246)
第三节 SPSS/PC+统计软件包 .....	(255)
思考题 ... .....	(270)
<b>部分习题答案.....</b>	<b>(271)</b>
<b>附表.....</b>	<b>(277)</b>
1. 二项分布表.....	(277)
2. 泊松分布表.....	(282)
3. 正态分布表.....	(284)
4. $T$ 分布表.....	(285)
5. $F$ 分布表 .....	(286)
6. $X^2$ 分布表 .....	(288)
7. SPSS/PC+命令分类总表 .....	(289)
<b>参考书目.....</b>	<b>(293)</b>

# 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)1991年我国国内生产总值为20188.3亿元人民币<sup>①</sup>,其中第一产业为5288.6亿元;第二产业为9102.2亿元;第三产业为5797.5亿元。(2)1990年城镇总人口为30191万人,乡村总人口为84142万人,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。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

---

<sup>①</sup> 资料来源:1990、1991、1993年统计年鉴。



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。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”。英语将这三者区别是，当(Statistics)一词，作为复数词应用时，指的是统计资料；作为单数词(Statistic)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，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，但仅从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存在着差异、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——定性的科学的认识规律。

## 第二节 统计数据信息搜集的方法

统计数据的搜集是统计整理和分析,以及统计推断和预测的基础。统计数据的搜集就是根据统计研究的目的和要求,有组织、有计划地向调查对象搜集原始资料和次级资料的过程。原始资料又称为初级资料,是指为了某个研究目的而进行实地调查所获得的,未经任何加工整理的的第一手资料。次级资料是指那些已经加工过的,往往是公开发表的数据。如从定期公开发表的“统计年鉴”、“会计报表”、报刊杂志上取得数据。由于次级资料需经原始资料的加工、整理而得到,因此原始资料的搜集的准确、及时、完整尤为重要,它是保证统计数据信息质量的关键。

原始资料搜集的主要组织方式为:专门调查和统计报表制度。

专门调查是指为了某些特定目的而专门进行的调查。这种调查常是一次性调查。一般是指普查、抽样调查、重点调查和典型调查。

普查是为了某一特定的目的、专门组织的一次性全面调查。如全国的人口普查、能源普查、工业普查等。普查对时间的要求很高,这是因为通过普查所搜集的资料,往往是表明某一现象在某一时点的情况。所以必须严格遵守普查所规定的时间要求。开展

普查的主要目的是为了摸清和掌握有关国情、国力的基本情况。其涉及面广、调查单位多、需花费较多的人力、物力和财力。但它可获得较准确的全面信息。

抽样调查是一种非全面调查。它根据随机的原则从调查对象的全体中抽取部分单位进行观察,并依据其结果推断全体数量特征的一种调查方法。由于抽样调查能用较少的时间、花较少的财力,达到近似全面调查的效果,因此可应用的范围较广。有关抽样调查的理论和方法,将在第四章中专门介绍。

重点调查也是一种非全面调查。它不同于抽样调查,不是随机地抽取部分单位组织调查,而是有意识地在全部单位中选择部分重点单位进行调查。所谓重点单位是指在调查的量的方面,占有较大的比重。例如要了解全国钢铁企业的生产情况,只要调查鞍钢、太钢、包钢、攀钢、首钢、宝钢等大型钢铁企业,就能达到调查的目的。这是因为这些钢铁企业虽在全国只占少数,但它们的产量在全国钢产量中占的比重很大。显然,重点调查可节省人力,物力和时间,能通过调查,掌握所需的基本情况,但它不能从数量上推断全体,这是重点调查的主要局限。

典型调查是根据调查的目的和要求,在对调查对象进行全面分析的基础上,有意识地选择部分有代表性的单位进行调查,因此它也是一种非全面调查。典型的选择一般根据调查单位之间差异的大小分别采用不同的方式,如调查单位之间差异小,波动不大,可适当选择 1 至 2 个典型单位,作“解剖麻雀”式的调查;如调查单位之间差异大,波动幅度大,可采取“划类选典”的方式,即将调查的全部单位分成若干类型,然后在每一类型中选择典型进行调查。这是一种深入了解活情况的搜集统计资料的方法。它的局限性与重点调查类似,不能由若干典型单位的数值对全体作出推断。

尽管有以上几种调查组织方式,但我国目前搜集统计资料的

主要方式还是统计报表制度,统计报表制度有定期、临时、全面、非全面之分。它是按照国家有关法规的规定,自上而下地统一布置、自下而上地逐级提供统计资料的一种统计调查方法。所谓全面统计报表制度,要求调查对象中的每一个单位都要填报,这种调查方式有时可以结合普查来应用;而非全面统计报表制度,只要求调查对象中的部分单位填报,有时它需要与重点调查、典型调查、抽样调查相结合。例如现在由国家统计局定期组织的城乡居民生活和农产量抽样调查,就是通过统计报表与抽样调查相结合,来搜集数据的。

这里要加以说明的是政府统计部门搜集统计信息与利用统计方法开展某项研究搜集数据,这二者在目的、方法上是有区别的。政府统计部门搜集统计数据的目的主要是掌握国民经济基本统计数字,供党政领导编制国民经济计划、检查计划执行情况以及制订方针、政策应用。由于国民经济的基本数字,涉及面广,取得这些数字的工作任务较繁重,因此只有设置专门统计机构、经常开展工作,才能取得。例如在国家统计局下,在全国的各省、市、县都建有统计局、处等。它主要通过统计报表制度来搜集统计资料的。

至于利用统计方法开展某项研究与政府的经常性统计工作的要求不同,它是随着研究任务不同而确定的。一项具体的研究任务完成了,围绕着这项研究所需要搜集的那些统计资料也就丧失了作用。因此利用统计方法进行某项研究时,首先要根据研究目的确定搜集那些统计资料。其搜集资料的方法,往往是开展非全面调查,如抽样调查、典型调查和重点调查。

### 第三节 统计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方法。

## 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、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。标志如以品质属性来表示,即以文字来表达如工人的性别,分为男、女;文化程度可分为小学、初中、高中等,称为品质标志。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的不同的具体表述,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的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。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。数量指标说明事物总规模和水平的各种总量指标,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的各组标志总量和全部标志总量。质量指标指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的对比求得。

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。

## 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,以样本来推论母体。

所谓母体(Population)又称总体是指特定的统计研究对象的全部个体的总和。将构成母体的各个单位,称为总体单位或母体单位。母体可分为无限母体和有限母体,如果母体中所包含的单位数是无限的,则称该母体为无限母体。如连续大量生产的零部件,其产量可以是无限的;如果母体中所包含的单位数是有限的,则称该母体为有限母体。对无限母体只能抽取部分单位进行非全面调查,据以推断母体。对有限母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母体,并且是有限母体。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 4000

户)进行调查,这若干户(如 4000 户)的家庭收入就是样本(Sample),通过样本调查的结果来推测母体的某种特征。

这里要加以说明的是,母体或总体,母体单位或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为总体,而该企业每一职工为总体单位,这时该企业的产值就是指标,则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。

估计是根据母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,以这些样本指标数值作为母体的参数估计值,来推断母体的某个数量特征。

检验是先对母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。

#### 第四节 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 and 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,目前 486 机的内存可达 4M—64M 之间,并配有一台或两台软盘驱动器及二台硬盘。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,可以存储 600 兆字节信息,信息的存储量相当于 500 张高密度软盘或 30 万页文献,使大量统计信息的存贮已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。仍以微型机为例,目前 486 机传输速度可达每秒 100 多 M 字节,使所需统计信息可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 and 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