

统计学新编

● 李晶·孙英 编著

四川大学出版社

统计学新编

TONGJIXUE XIN BIAN

李 晶 孙 英 编著

四川大学出版社

1995年·成都

(川)新登字014号

责任编辑:李勇军

封面设计:冯先洁

版式设计:孙 瑛

统 计 学 新 编

李 晶 孙 英 编 著

四川大学出版社发行 (成都市望江路29号)

四川省新华书店经销 郫县犀浦印刷厂印刷

850×1168mm 1/32开本 9.125印张 2插页 220千字

1995年9月第1版 1995年9月第1次印刷

印数:0001—5000册

ISBN7-5614-1293-7/F·184 定价:9.50元

目 录

第一章 总论	(1)
第一节 什么是统计	(1)
第二节 统计学的几个基本概念	(5)
第三节 统计的基本方法	(9)
第二章 统计调查	(14)
第一节 统计调查的意义和种类	(14)
第二节 统计调查方案	(17)
第三节 统计调查的组织形式	(21)
第四节 搜集统计资料的方法和调查误差	(25)
第三章 次数分布	(32)
第一节 次数分布	(33)
第二节 次数分布的图示	(41)
第三节 频率和累计次数	(43)
第四章 代表值	(49)
第一节 算术平均数	(50)
第二节 中位数	(56)
第三节 众数	(61)
第四节 几何平均数	(64)
第五章 离散的代表值	(68)
第一节 全距	(68)
第二节 平均差	(70)
第三节 标准差	(72)

第六章 指数	(81)
第一节 指数的意义.....	(81)
第二节 指数的编制.....	(83)
第三节 指数体系与因素分析法.....	(95)
第七章 时间数列(I)	(107)
第一节 时间数列概述.....	(107)
第二节 时间数列发展水平指标和比较指标.....	(111)
第三节 时间数列的平均指标.....	(115)
第八章 时间数列(II)	(123)
第一节 总论.....	(123)
第二节 长期趋势.....	(130)
第三节 季节变动.....	(136)
第四节 周期波动.....	(147)
第五节 经济时间数列分析的实例.....	(150)
第九章 相关分析与回归分析	(157)
第一节 相关分析的意义和内容.....	(157)
第二节 相关分析.....	(161)
第三节 直线回归.....	(174)
第四节 曲线回归.....	(181)
第五节 相关与回归分析的一些问题.....	(184)
第十章 抽样推断	(190)
第一节 抽样推断的意义和作用.....	(190)
第二节 概率分布.....	(193)
第三节 抽样分布.....	(206)
第十一章 参数估计与假设检验	(225)
第一节 参数估计的基本原理.....	(225)
第二节 总体参数的估计.....	(233)
第三节 抽样组织方式.....	(242)

第四节	假设检验的概念和步骤·····	(249)
第五节	平均数的假设检验·····	(255)
第六节	成数的假设检验·····	(261)
第十二章	计算机在统计中的应用·····	(266)
第一节	电子计算机概述·····	(266)
第二节	计算机在统计工作中的应用·····	(271)
第三节	统计软件包概述·····	(273)
第四节	SPSS/PC+统计软件包·····	(278)
附录	·····	(283)
后记	·····	(287)

第一章 总论

统计是人们在社会经济生活中必不可少的工具,是人们认识世界、探索现象数量差异的本质及其规律的方法,是人们进行明智决策的一门艺术。随着人类社会进入信息时代,统计作为一种方法和工具就变得越来越重要。本章主要介绍:(一)什么是统计;(二)统计学的几个基本概念;(三)统计的基本方法。

第一节 什么是统计

一、统计的概念

在日常的政治、经济和社会文化活动中,人们经常要接触各种数字化的信息,例如:(1)我国国土面积960万平方公里,其中山地约320万平方公里;高原约250万平方公里;平原约115万平方公里;丘陵约95万平方公里。(2)1994年我国国内生产总值为43800亿元人民币^①,其中第一产业为8231亿元;第二产业为21259亿元;第三产业为14310亿元。(3)1992年全社会固定资产投资为720亿元,比1991年增长30%以上;1993年全社会固定资产投资8800亿元,比1992年增长22.2%。在这些大量的信息之中,如果存在着

^① 资料来源:国家统计局关于1994年国民经济和社会发展的统计公报。

说明社会现象与自然现象的规律性,那么对于社会现象来说就可以根据了解到的规律性为制定政策提供依据,而对于自然现象来说则可以发现控制的手段。因此,初次处理大量信息就显得十分必要。

那么,由社会现象和自然现象构成的大量信息中,是否存在规律性呢?我们认为,确实存在。不仅如此,还可以利用某种方法,整理和掌握作为这种现象的规律性。这种方法就是统计方法。有关统计的理论就是统计学。因此,统计学在社会现象和自然现象的分析方面就成了极为重要的分析工具。

什么是统计学呢?

统计学就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。它通过对大量客观现象数量方面的分析研究来反映事物发展变化的规律性,并根据其规律性预测未来状态,为制定科学的决策服务。统计学不等于统计。统计包括统计工作、统计资料和统计学三层含义。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。统计资料是通过统计工作实践活动所获取的全部资料,它包括数字资料和文字描述资料两种形式。可见,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结果,它是从事统计研究和各种社会经济研究必不可少的依据。对统计实践活动的理论概括就是统计学。统计学是一种方法论的科学,它来源于实践活动,又反过来指导统计工作实践,从而得到不断完善和发展。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,因此我们口语中将这三者简称为“统计”。英语将这三者的区别是,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。

通常,统计学可分为描述统计和推断统计。

(一)描述统计

描述统计是表述、分析、解释全部总体资料的数量特征的统计方法。其目的是通过适用、简单、明确、易于理解的形式去揭示现象的数量规律性。在社会、经济和日常生活中，我们所搜集或遇到的数据资料大部分都是杂乱无章、无规律可循的，而人们却可以通过统计描述的过程，用更加有意义或更加适用的形式，遵循描述数据的某些规则和程序，使现象的数量规律性呈现出来。

(二)推断统计

推断统计是用总体中的部分数据的特征去推断全部总体的数量特征而进行的统计方法。推断统计在实践中具有相当重要的意义。它既可以节省费用支出，又可以节省时间，是非常有效、实用、科学的统计方法。例如，从某国的选民中抽出3000人，由3000人组成的样本推测全国选民这一总体，了解支持某个政党所占百分比有多少。近年来，这种推断统计有着特别惊人的发展。

描述统计与推断统计虽然方法不同，用途各异，但二者的目的是一致的，即揭示大量现象的数量规律性。在实际工作中，推断统计常常以描述统计为基础，使用描述统计方法的测定结果进行推断。因此，人们使用统计方法去探索现象的数量规律性，不仅要掌握描述统计方法，还要掌握推断统计方法。

二、统计的特点

统计作为一种方法、一种工具，在认识和揭示现象的数量规律性的过程中具有自己独立的思维形式和分析特点。

第一，统计研究的是现象的数量表现。

在客观事物中，任何现象都是质和量的统一体。例如学习状况和学习成绩、人口质量和人口构成、社会生产的规模和发展速度、人民的物质文化生活状况和物价指数、某地区的天气状况和气象资料等等，这些现象的数量表现正是统计认识的基础，统计正是从现象数量表现入手，进行搜集、整理、分析，以揭示现象的数量规

律。

第二, 统计研究的是具有差异的数量表现。

由于客观现象的复杂性, 使各种现象的数量表现不尽相同, 出现差异, 从而使我们无法认识现象的本质。例如: 考生的成绩各不相同, 各年的物价指数不同, 各月的气象资料不同等等。正是由于现象间存在着数量方面的差异, 才要求统计采取其特有的统计方法去揭示现象的数量本质。所以我们说差异是统计存在的前提, 换句话说, 只要存在差异, 便有统计的用武之地。

第三, 统计研究的是大量现象的数量差异。

统计着眼于现象的数量表现的目的并不是描述个体现象的数量特征, 而是描述由大量单个现象所构成的整体的数量规律。例如, 对于某球队来说, 一次比赛的成绩可能是 $1:0$, 可能是 $2:1$, 也可能是 $2:3$ 。由于这些比分不同, 我们不能根据一次的成绩来断定这支球队的水平。当我们对这支球队的比赛进行多次观察时就会发现, 就每次比赛来看, 他可能输, 也可能赢, 但从整个的国际比赛的整体看, 赢的场次多, 输的场次少, 因此, 我们可以得出这支球队是世界级水平的结论。这个结论正是基于对这支球队的多次比赛的观察而得出的。可见, 统计正是从大量个体现象入手进而上升到对整体的数量特征进行认识的一种方法, 是我们认识自然现象、社会现象的工具。

三、统计的作用

统计作为一种方法和工具广泛地应用于各个领域。只要现象存在着数量差异, 只要存在着未知现象, 就存在着统计。在科学研究领域中, 没有一门科学可以完全离开统计。天文学要研究天体运动规律, 物理学要研究现象物理变化的规律……, 对这些领域的规律性的研究都离不开统计。只有使用统计方法才能认识现象的数量规律性, 从而更好地认识现象的内在规律。

在经济生活和日常生活中,人们在不同的行业,从不同的角度,以不同的方式应用统计方法去认识世界,改造世界。例如:气象预报员利用大片地区的气象资料进行预报;企业根据大量产品信息决定产品的方向;教育部门根据人口年龄构成确定师资和校舍的数量;居民根据消费品价格指数决定是否存款;银行利用统计数据确定货币投放量;球迷利用统计去评价运动员的水平。总之,人们总是有意无意地使用统计去解决周围的问题,从而使统计成为认识世界的工具。正如美国统计学家H·G·韦尔斯说“统计思维有一天会象读写能力一样成为有能力的公民的身份需要。”可见,掌握和运用统计能够更自觉地、科学地处理周围所发生的事情。

第二节 统计学的几个基本概念

进行统计研究,势必要涉及一些统计概念。为了叙述方便,有利于以后各章学习,我们先介绍几个基本概念。

一、统计总体与总体单位

凡是客观存在的,在同一性质基础上结合起来的许多具有差异的个别事物构成的整体,称为统计总体,简称总体。构成总体的每一个个别事物称为总体单位,简称个体。例如,我们要研究某班学生的学习成绩,因为每个学生是客观存在的,都具有是这个班级的学生这一共同性质,而每个学生的学习成绩各不相同。所以,由这个同一班级的许多学习成绩不同的学生所构成的班级就是统计总体,每个学生就是总体单位。可见,作为一个统计总体,具备三个基本特征,即同质性、大量性、差异性。同质性是指各个总体单位在其数量方面或属性方面必须具有共同的表现,否则对众多总体单位的综合便无法描述或揭示总体的数量特征。如研究中国人口,则

每个人是构成总体的总体单位,他们都必须具有中华人民共和国国籍,否则所概括的人口特征就不能描述中国的人口情况。大量性具有两层含义:其一是指统计总体必须是由数量众多的总体单位所组成,而不能是某一个个体。这是因为统计的目的是对总体的研究,而不是对个体的研究。如上例中,中国人口这个总体就不能是张三或李四其人,而是由众多人数所构成的整体。另外,只有在对许多个体的分析、综合、研究的基础上,才能揭示出总体的数量特征。其二是指要观察全部总体中的足够多的总体单位。在推断统计中,可以不必观察全部总体单位,但必须在一定误差的前提下观察到足够多的总体单位,否则也无法对总体特征作出具有一定可靠性的描述。差异性是指总体中各总体单位的具体表现不同。如人口有性别方面的差别、年龄方面的差别等。正是这种差异的存在,才需要统计对其进行综合概括,抽象分析,从而揭示总体的数量特征。可见,差异是统计存在的前提条件,没有差异,便没有统计。

总体与总体单位的划分并不是固定不变的,两者随着研究目的的不同而变换,即总体可能会变成总体单位,总体单位也可能变成总体。例如,研究某校各个班的学习成绩,这里总体单位是每个班级,而总体则是由各个班级所构成的整体。但是,如果我们改变了研究目的,研究某班各个学生的学习状况,则班级便由总体单位变成了统计总体,而每一个学生在这里便是总体单位。

二、标志和指标

所谓标志是指构成事物的各个单位所具有的属性和特征。例如上例中的学生便是总体单位,它身上具有许多特征,如“性别”、“年龄”、“学习成绩”等。标志都具有其具体表现。如年龄有20、22、25岁等表现,学习成绩有80分、90分表现,性别有男、女的表现。

标志按其性质不同可分为数量标志和品质标志。表示总体单

位数量特征的标志称为数量标志,如工人的年龄、工资、身高、体重,企业的职工人数、总产值、利润等。数量标志的表现只能是数值形式的,如总产值有200万元、300万元等。表示总体单位属性特征的标志称为品质标志。如工人的性别、婚姻状况等。品质标志可分为有序品质标志和无序品质标志。有序品质标志如产品等级、文化程度等;无序品质标志如婚姻状况、性别等。无序品质标志只能采取文字的形式,有序品质标志则既可采取文字形式,有时为了分析方便也可采用数值的形式表现。如产品等级可分为一级、二级、三级,文化程度有小学、初中、高中、大学等,它们可用数值表现为1、2、3等。

只有两个标志表现的标志称为是非标志或交替标志。这种标志有些是客观存在的,如性别分为男和女,产品质量有合格和不合格;有些是主观划分的,如工资在100元以下和100元以上,参加工作30年以上和参加工作30年以下等。

标志按其表现是否相同分为不变标志和可变标志。不变标志是指在总体中各总体单位在某个标志上具体表现是相同的。如,在一个由8岁的学生所构成的统计总体中,对于年龄这个数量标志来说,各个总体单位的标志表现是相同的。因此,在这种情况下,年龄是一个不变标志。可见,不变标志是指总体各单位标志表现相同的标志,它是构成总体同质性的一个主要因素。可变标志是指总体中各总体单位在某个标志上的具体表现不同。如由50名学生所构成的总体中,学生的年龄有17岁、18岁,学生的身高有160公分、175公分,学生的性别有男、女。这些标志都称为可变标志,它是统计研究的主要对象。

指标是统计指标的简称,它是用来说明现象总体数量方面特征的。如研究50个学生的学习成绩,则反映总体特征的平均成绩就是指标。一个完整的统计指标应包括指标名称和指标数值两个部分。指标名称是说明现象总体特征的科学概念,它给统计总体某

一方面特征以质的规定。例如国内生产总值、社会商品零售额、人均国民收入、人口密度等，都是比较严格的概念。指标数值说明现象在具体时间、地点、条件下在该特征方面的具体表现，它是根据总体各单位的相应的品质标志表现和数量标志值汇总、计算而来。例如，研究某市各工业企业情况，则总体是该市全部企业，总体单位是每个企业，标志有所有制形式、职工人数、工业总产值、利润额等，它们都是由各个工业企业的相应数值汇总而来，而全市国有企业所占比重则是由全部的国有企业数除以全部工业企业数计算而来的。

统计指标按其反映总体现象的内容不同而分为质量指标和数量指标。前者反映现象各总体单位的水平、现象间的内在联系和比例关系，后者则用来说明现象的总量或规模。例如平均工资、单位产品成本、商品流转次数、利润率、劳动生产率等就是质量指标，而国内生产总值、工资总额、商品销售额、利润总额就是数量指标。

指标和标志虽然是两个不同的概念，但二者既有区别，又有联系。它们的主要区别是：①标志是说明总体单位特征的，指标是反映总体特征的。如研究某企业工人的工资水平，则工资是说明每个工人的特征的名称，即标志；而平均工资则是说明这个总体的特征，即指标。②标志既有能用数值表示的数量标志，也有能用文字表示的品质标志；而指标只能用数值表示。如上例中，工人的工资是用数值表示的数量标志，工人的工种则是用文字表示的品质标志，而全厂技术工人所占比重、全厂工人平均工资等指标都是由数值表现的。它们的主要联系是：①指标数值是由构成总体的各单位数量标志值或品质标志表现的次数汇总计算而来。如工资总额是由工人工资汇总而来，而技术工人所占比重则是由技术工人、非技术工人等品质标志表现的次数汇总、计算而来。②随着研究目的的变化，指标和标志二者之间可以转换。指标与标志的这种转换取决于总体和总体单位的转换。例如，研究各个小组所构成的班级的学

习成绩,则小组成绩是标志值,而班级平均成绩是指标数值。如果研究各位同学所构成的小组的学习成绩,则每个人的成绩是标志值,小组平均成绩是指标数值。可见,由于研究目的不同,二者关系可以转换。

三、变量

变量即可变的数量标志,它表明的是数量上的差异,所有上面列举的学习成绩、工人工资、年龄、身高等等,都称为变量。可变数量标志其具体表现是不尽相同的,亦即各个总体单位的变量取值不同,例如学习成绩有50分、60分、90分,年龄有21岁、25岁等。这些具体的数值称为变量值。变量值也就是标志值,是标志的数量表现,或数量标志的具体表现。

变量按其取值是否连续可分为连续变量和离散变量。连续变量的数值从理论上说是连续不断的,在任意两个数值之间可能有无限个数值,如总产值、身高、体重。离散变量的数值只能以整数计量,如企业数、设备台数、职工人数等。

第三节 统计的基本方法

在统计学中,有许多统计方法被用来搜集、整理、分析现象的数据资料,如大量观察法、平均法、指数法、相关分析法等。其中最主要的方法是大量观察法和平均法。

一、大量观察法

大量观察法是统计学所特有的方法,它是指对客观现象的总体或总体单位的集合的观察。“大量”一词是外文翻译过来的,意思是指总体单位的集合,即统计总体。所以,大量观察所表明的并不

是单一数字或数量关系的观察,而是指各总体单位的某一数量标志值的集合。

我们知道,统计的目的是揭示现象总体的数量规律性。如果总体中的各个总体单位的标志值相同,我们就会立刻掌握总体的数量本质。例如,某班学习成绩统计都是80分,毫无疑问,班级的水平是80分。在这种情况下就没有进行统计分析研究的必要。实际上,类似情况是很少发生的。在我们所遇到的现象中,大多数总体单位的标志值是不同的,或者说他们的标志值之间存在着差异。这样,我们就不能马上认识、掌握总体的数量规律性,因此,需要采用一些特殊的方法去进行分析研究。

要认识现象的数量规律性,首先应该知道数量规律性是怎样发生的。我们知道,现象的发生和发展都是多种因素作用的结果。在这众多的因素中,有一些是主要的、基本的因素,它决定着现象的本质及其规律;有一些是次要的、伴随的因素,无论其是否存在,对现象的本质并无影响。前者使现象的数值出现规律性、一般性。后者使现象出现个别性、偶然性。如果分别地观察每一个总体单位,就会看到每一个总体单位的具体表现是不同的,如果将这些总体单位集合起来观察,就会发现次要的、伴随因素的作用相互抵消,相互减弱,而在每个总体单位中的影响微弱的主要因素便随之而扩大、集中、强化乃至呈现出来。这种呈现出来的主要因素作用的结果便是总体的数量规律性。可见,客观存在的各总体单位具体表现之间的差异是大量观察法的基础。换句话说,正是由于客观世界中存在着各总体单位的数量差异,才产生了观察、研究大量具有差异现象的大量观察法。

大量除指总体单位的集合以外,在一定意义上也可以理解为观察数目的大小多少。从实际应用上看,在作全面观察时,总体单位同被观察的单位是一样大的,这种观察的代表性是最充分的。然而,在非全面调查时,被观察单位必须达到足够的单位才能对总体

具有代表性。在这种意义上也可以把“大量”作为“数量大”来理解。

在统计学中,只有通过大量观察法才能认识总体的数量规律性,这一道理早已为人们所熟知。统计大师凯特勒曾对大量观察中的个体与总体的关系作过一个精彩的比喻。他说,“假定从最近的距离来注视画在平面上的大圆线上的一小部分,这时,无论圆线是怎样被用心画出来,但看起来总象是很多的点奇妙地、随便地排列着,并且好象是把他们胡乱地结合在一起似的。可是,如果把眼睛移开一些来看,那么,各个点就从眼界中消失了,它们便分而为一,形成一个规律分明的整体,于是也就可以认识到所画图线的性质了。”他认为当大量的个体构成总体时,个体性质即可消失,从而表现出总体的规律性。日本统计学家横山雅男也曾在他的《统计通论》中说:“一勺之水无色可辨,而大海则蔚然而紺碧;少量之气无迹可循,而太空则穹然而青苍。”这一简单的比喻道明了大量观察法的真谛。

大量观察法的对象是客观现象,而对客观现象进行大量观察乃是一种社会行为,它要根据统计任务的要求,事先确定观察的目的、特点等。

二、平均法

从统计研究来看,通过大量现象观察所得到的总体各单位的数据资料只是探索总体数量规律性的原料,要进行统计分析研究,还要采用各种统计分析方法对这些具有差异的数据资料进行分析,如分析总体的数量特征,揭示总体间的数量关系等。在这些分析方法中,最主要的是平均法。

所谓平均法是指测定集中趋势的方法。在大量观察情况下,总体中的各单位的数值之间虽然存在着差异,但经过统计整理,分组计数,会使杂乱无章的标志值消失而呈现出一种规律性的次数分布。在这个次数分布中,所有标志值规则而均匀地从中心值散布开