

新编统计学

—描述与推断

XINBIAN

TONGJIXUE

陈湛匀 著

上海科学技术文献出版社

新编统计学

——描述统计与推断统计

陈湛匀 著

上海科学技术文献出版社

(沪)新登字801号

新编统计学
——描述统计与推断统计
陈湛匀 著

上海科学技术文献出版社出版发行
(上海市武康路2号 邮政编码 200031)

全国新华书店经销
上海科技文献出版社昆山联营厂印刷

开本 787×1092 1/32 印张 11 字数 266,000
1994年12月第1版 1994年12月第1次印刷
印数: 1-1100

ISBN 7-5439-0560-4/O·100

定价: 17.20元

《科技新书目》337-277

前 言

八年前,当时我在讲授统计学课程时,学校希望我进行统计学教学内容改革,经过几次实践和研讨,自己越来越想写成这样一本书——《新编统计学——描述统计与推断统计》,希望能适应和满足高等财经院校教学和统计工作者自学的需要。

统计学被誉为 20 世纪富有显著成就的学科之一。笔者认为,就统计学而言,不论社会经济统计学或数理统计学,以及各种分科应用统计学,它们的原理和基本方法是共同的,而各种统计方法,按其性质和作用,可以归纳为两类:一类是描述统计方法。它包括:根据经济事物性质确定经济指标及其计算方法;经济指标的具体数据的收集、整理和分析方法,如各种统计调查方法,分组和数据汇总方法,计算平均数、离差和指数的方法,时间序列分析法等。另一类是推断统计方法。它包括估计和检验两个组成部分。推断统计是以概率论作为基础的,需要更多地借助于数理统计。

本书按照上述体系展开撰写的,在书中努力体现统计学的社会经济内容,突出概率与推断统计的地位,反映现代统计学分支与应用性强的特色。具体地说,它有以下几点:

1. 以简练手法介绍描述统计,删除“统计图”等章节内容,详细介绍了推断统计和现代统计学分支应用;
2. 反映统计工作顺序和数理逻辑;
3. 吸收国内外统计科学新动向和新成就,例证和取材尽量结合管理和经济应用的实际;

4. 为避免繁琐、复杂的数理推导和论证,强调应用性,概念的引进,在保障科学性的前提下力求直观;方法的讲述,在保障理论严谨性的前提下,不拘泥于形式逻辑上的严格性。

全书共分十一章,可作为高等财经院校统计学课程的教材或自学用书,亦可作为统计实务工作者的参考用书。

本书承蒙郑德如教授审阅,在此表示感谢。

限于作者水平,书中错误或不妥之处在所难免,希望读者批评指正。

作 者
1994年2月

引 论

统计(Statistics)一词在不同的场合有不同的含义:1. 统计工作,即统计实践活动,早在古代奴隶制国家,由于赋税、徭役和征兵等需要,就开展了人口、财产、领地等统计调查活动;2. 统计资料,是指在统计工作中所取得的一系列数字资料,它与统计工作紧密相联,因此也很早就有了,据历史记载我国夏禹时代就开始有人口统计数字;3. 统计学,它是一门认识社会与自然的方法论科学,是统计工作的经验总结与理论指导,其内涵与外延随着统计实践的发展而发展。

一、统计学发展史略

统计学产生于资本主义萌芽时期的欧洲,迄今已有三百多年的历史。在其发展过程中先后出现了许多学派,下面以表格形式对几个主要学派进行评述与比较。

学派名称	产生时间、 国 度	代 表 人 物	主 要 特 点
政治算术 学派	17世纪中 叶英国	①威廉·配第(W. Petty), 其名著《政治算术》以数字资 料为基础,用计算与对比的 方法,比较英、法、荷三国实 力。 ②约翰·格朗特(J. Gra- unt)发表《关于死亡表的自 然和政治的观察》,创造性地 编制了初具规模的生命表。	①研究目的在于探讨社会 经济现象的规律性,为制定 政策提供依据。 ②研究对象是国家的人 口、土地、资本等社会经济现 象。 ③研究方法是对社会经 济现象以数字、重量和尺度 进行表述、推算与比较。

(续表)

学派名称	产生时间、 国 度	代 表 人 物	主 要 特 点
国势学派 (记述学 派)	17~18世 纪 德国	① 康令(H. Conring) 最早在德国赫尔斯达德大 学讲授“欧洲最近国势学”, 奠定了国势学的基础。 ② 阿亨瓦尔(G. Achen- wall) 著有《欧洲各国国势学概 论》,把国势学称为 Statist- ik 即统计学,被一直延用 至今(但含义已经不同)。	① 研究目的是为当权者 提供管理国家的方法,熟悉 各国的财富。 ② 研究对象是叙述国家 最必要的显著事项以及最高 政策关系的总体。 ③ 研究方法是各国政治、 社会情况以文字记述为主。
数理统计 学派	19世纪中 叶 比利时、英国	① 凯特勒(A. Quetelet) 著有《社会物理学》,把大 数定律,正态分布应用于统 计之中,较好地解决了统计 数据处理和计算的依据。 ② 英国戈塞特(N. S. Gosset)与费歇(R. A. Fish- er)、波兰尼曼(J. Neyman) 与小皮尔逊(E. S. Pearson)、 美国瓦尔德(A. Wasld)和 科克兰(W. G Cochran) 为后期代表人物。	① 最初是统计学与概率 论的结合,并认为统计学是 一门独立的方法论科学。 ② 早期多属描述统计学派, 后期多属推断统计学派。 ③ 当代数理统计越来越 广泛应用数学方法和电脑技 术,形成许多新的分支与边 缘学科。
社会统计 学派	19世纪后 半叶 德国	① 克尼斯(K. G. A. Kni- es)。 著有《作为独立学科的统 计学》,认为统计学是一门独 立的社会科学,是用数值解 说大量现象的一门科学。 ② 恩格尔(C. L. E. Eng- el)、梅尔(G. V. Mayr)均 为早期代表人物。 ③ 美国恰笃克(R. E. Chaddock)、日本麓川虎山、 德国弗拉斯科波(P. Flask- amper)等为后期代表人物。	① 从渊源上看,融合了政 治算术、国势学的观点,继承 和发扬了凯特勒强调研究社 会现象的传统。 ② 早期坚持“实质性科 学”的立场,采用大量观察法 研究社会总体。后期不再坚 持纯粹的实质性科学的主张, 但仍强调以事物的质为前提。 ③ 当代社会统计已过渡 到方法论,吸收了以概率论 为基础的数理统计方法。

在我国,解放前也有社会统计学派和数理统计学派,解放后从前苏联大量引进社会经济统计学,并一度全盘否定数理统计学,直至改革开放后,后者才得到迅速发展,对社会经济统计

学也进行了必要调整，但仍难以适应社会主义市场经济的发展，因此迫切需要建立新的统计学体系，为此本书也试图在这方面做些努力与尝试。

二、描述统计与推断统计

综观当前世界各国统计研究成果，基本上可以认为统计学是一门研究现象总体数量关系的方法论科学，而统计方法可以概括为两种：描述统计与推断统计。

描述统计(Descriptive Statistics)侧重于对客观现象进行数字的计量、概括、描绘和表述。具体工作包括对统计资料的收集、整理、列表图示以及测度集中趋势、离散趋势和偏度等，目的是反映现象的数量特征和数量关系，而不对现象的未来发展作进一步的推论。

推断统计(Inferential Statistics)也称归纳统计(Inductive Statistics)，着重研究如何根据部分数据去推论总体的情况，因为对客观现象收集统计资料时，往往不可能对全部单位进行调查，即使可能也要付出太大的人力、财力、物力，故通常采用抽取部分样本对总体进行推断的方法。当然样本只是总体的一部分，包含的信息不完备，而且抽样是随机进行的，必然会出现误差。但推断统计可以根据概率论的原理，使推断所产生的不确定性得到度量，使样本对总体作出更为可靠的推断。

在20世纪之前，统计学基本上处于描述统计阶段，自20世纪初以后，社会领域和自然领域都向统计学提出更高的要求，仅靠描述统计方法不能解决现象间复杂的数量关系，而且不能对未来做出估计与预测，所以统计学迅速向推断统计发展，以戈塞特、费歇、尼曼为主要代表，使推断统计的理论不断丰富并成为统计学的主流。

推断统计的内容主要包括两个方面：1. 估计，即根据样本

资料计算某一指标数值(如平均数、方差等),并作为总体参数的估计值;2. 检验,即以样本取得数据作出某种假设,并根据样本资料计算结果来推断所作假设能否接受。

由上述可知,描述统计是统计学的基础,而推断统计是现代统计学的核心。本书基本上照顾到这两方面的内容,第一章至第四章主要介绍描述统计,其余各章主要介绍推断统计。但有时候描述统计与推断统计是密不可分的,虽然有不少描述统计内容,但我们出于对篇章结构合理性的考虑,还是将其归入推断统计一列。至于第十一章决策分析则是推断统计学中的估计与假设理论的最新发展结果,我们着重对其中应用较广的方法进行评述。

三、统计学的基本概念

和其它学科一样,统计学也有自己特定的概念和范畴,下面是一些常用的基本概念,其它具体的概念则将在各有关章节中加以介绍。

1. 总体与总体单位

总体是指客观存在的、在同一性质基础上结合起来的许多个别单位的整体。总体中每个个别单位必须具备某些共同性质,即同质性,这是构成总体的必要条件。

总体单位就是构成总体的各个个别单位。如果总体中包括的总体单位数是无限的,则称为无限总体;如果总体中包括的总体单位数是有限的,则称为有限总体;如果总体单位数有限,但数量很大,也可视为无限总体。

总体和总体单位是相对而言的,可以随着研究目的的不同而变化。例如,在对某校所有班级进行统计研究时,则该校所有班级构成一个总体,每个班级是总体单位;而在对该校某个班级全体学生进行调查时,则该班所有学生构成一个总体,而每一位

学生就是总体单位。

2. 标志与指标

标志是说明总体单位属性或特征的名称。按其性质不同，标志可以分为品质标志和数量标志。品质标志表示事物品质属性的特征，如学生的性别、民族、籍贯等；数量标志表示事物的数量特征，如学生的年龄、身高、成绩等。值得注意的是，品质标志虽然一般不能用数值表示，但是也有一些用数值表示的，如工人的技术等级、产品的质量等级等。

标志还可分为不变标志与变异标志。标志如果在总体各单位之间的具体表现完全相同，则称为不变标志，任何总体的每个总体单位至少有一个共同的、使它们能结合在一起的不变标志构成总体同质性的基础。标志如果在总体各单位之间的具体表现不尽相同，则称为变异标志，如学生的年龄、成绩等，均为学生的变异标志。

指标是反映统计总体数量化的科学概念，是总体在具体时间、地点、条件下综合性的数量表现。指标与标志的区别在于：①标志是说明总体单位特征的，是依附于总体单位上的，而指标是说明总体特征的，是依附于总体上的；②标志既可以用数值表示也可以不用数值表示，而指标只能用数值表示；③标志值一般不经过汇总就可以得到，而指标值必须经过汇总才能得到。

当然，指标与标志也存在密切联系：①许多指标是标志值汇总的结果，例如，若以某校所有班级为总体，则该校学生总人数(指标值)是各班级学生人数(标志值)汇总的结果；②随着研究目的的改变，标志和指标可能发生转化，如在上例中，若重新确定总体为该校所属教育局的所有学校，则该校学生总人数就转变为标志值。

一个指标只能反映总体某方面的数量特征，若要了解总体

的全貌及其发展过程，就需要运用指标体系。指标体系就是一系列相互联系、相互制约的统计指标所构成的整体。

3. 变异和变量

变异标志的变化称为变异，包括品质标志的变异和数量标志的变异。总体变异性是统计存在的前提条件，如果总体没有变异性，就不会产生统计实践活动，也就不会有统计学。

变量是可变的数量标志。可变数量标志的标志值称为变量值。变量按变量值的性质可以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，如人数、企业数等，其变量值一般采用计数方法取得；连续型变量指在整数之间可取小数的变量，如工资、成本等，其变量值一般通过测量或计算取得。一个变量属于哪个类型不是绝对的，这和采用单位和精确度有关。

根据变量所受因素影响的不同，还可以将其分为确定性变量和随机性变量。受确定性因素影响的变量称为确定性因素，如推广良种能够使农作物增产，这是确定性因素的影响，受随机性因素影响的变量称为随机性变量，如上述农作物增产的数量并不完全决定于推广良种这一因素，而还要受到土质、施肥等其他随机因素的影响。

目 录

引论	
第一章 统计资料	(1)
第一节 统计调查.....	(1)
第二节 统计整理.....	(9)
第三节 频数分布.....	(16)
第二章 集中、离散趋势和偏度指标	(31)
第一节 集中趋势指标.....	(31)
第二节 离散趋势指标.....	(47)
第三节 偏度指标.....	(59)
第三章 统计指数	(66)
第一节 统计指数概述.....	(66)
第二节 综合指数的编制.....	(70)
第三节 平均数指数的编制.....	(77)
第四节 指数体系与因素分析.....	(82)
第四章 时间序列	(95)
第一节 时间序列概述.....	(95)
第二节 时间序列的水平指标.....	(99)
第三节 时间序列的速度指标.....	(103)
第四节 时间序列的因素分析.....	(107)
第五节 时间序列预测方法简介.....	(120)
第五章 概率论的基本概念和方法	(135)
第一节 概率及其定义.....	(135)

第二节	排列与组合	(138)
第三节	概率的基本运算规则	(144)
第六章	随机变量与概率分布	(153)
第一节	随机变量	(153)
第二节	离散型随机变量的概率分布	(154)
第三节	连续型随机变量的概率分布	(164)
第四节	大数定律与中心极限定理	(171)
第五节	统计量及其分布	(174)
第七章	参数估计和假设检验	(178)
第一节	估计理论概述	(178)
第二节	参数的区间估计	(181)
第三节	假设检验的一般方法	(194)
第四节	常用参数的假设检验	(198)
第八章	抽样推断	(208)
第一节	概述	(208)
第二节	抽样误差	(213)
第三节	抽样组织方式	(222)
第九章	回归和相关分析	(236)
第一节	概述	(236)
第二节	相关分析	(239)
第三节	回归分析	(249)
第十章	拟合优度检验和非参数方法	(267)
第一节	拟合优度检验	(267)
第二节	非参数方法	(272)
第十一章	决策分析	(281)
第一节	非确定型决策	(282)
第二节	风险型决策	(292)

第三节	竞争型决策.....	(311)
第四节	多目标决策.....	(319)

第一章 统计资料

统计资料(Statistical Data)是反映在一定空间、时间条件下客观事物某些特征的事实或数字,它是进行统计描述与统计推断的“原材料”。

第一节 统计调查

收集统计资料是整个统计研究工作的首要环节。统计资料的收集根据来源不同可分为两种:一种是对原始资料的收集,即直接对研究对象的情况进行登记或调查,通常称为统计调查;另一种是对次级资料的收集,如从有关部门、报纸杂志、统计公报等收集已经加工的现成统计资料。但归根到底,一切次级资料都是从原始资料过渡而来的,因此,原始资料的收集即统计调查是最根本的。

一、统计调查的分类

根据国内外的实践,可以从不同角度把统计调查分为不同的种类。

1. 按调查的范围大小划分

从调查是否包括所有的总体单位来看,可以分为全面调查和非全面调查。表 1-1 对两者进行了简明的比较,但在实际工作中,应根据具体研究目的与条件,选择或结合采用。

2. 按调查的进行方式划分

(1) 亲自调查 即由研究者亲临现场对调查对象进行直接

表 1-1

比较项目	全 面 调 查	非 全 面 调 查
调查范围	构成总体的所有个体	构成总体的部分个体
所获资料	比较全面、完整,但一般仅限于最重要、最基本的指标	间接推算总体资料,可以说明总体的状况与发展趋势
调查时间	旷日持久	可急可缓
调查费用	涉及人员多,费用大	一般较小,视具体要求而定
调查方式	普查	重点调查、典型调查、抽样调查

观察、计数和测量等活动,它只适用于规模较小的调查,因为要由少数人在短时间内采集全部资料。这种方法的最大优点是可以保证取得内容一致的统计资料。

(2) 委托调查 即派遣调查员进行访问调查。此法是将调查表交给受过专门训练的调查员,由调查员分别对调查单位进行访问,取得所需的资料。这种方法可以对比较复杂的问题获得比较确实的回答。

(3) 通讯调查 在调查范围较广,而时间、经费有限的情况下,可以采取通讯调查的方式,即研究者把调查表寄交调查单位,由调查单位按照表中的项目填写后寄还。通常收回的表格会少于发出的表格,因此实际得到的样本往往存在偏误。

3. 按调查的组织形式划分

在我国,从调查的组织形式来看,可以分为统计报表制度和专门调查。统计报表制度是国家统计系统和各业务部门为定期取得系统的全面的基本统计资料,而采用的一种自下而上的收集资料的组织形式。专门调查是指为了了解和研究某种情况或

某项问题而专门组织人力,深入基层收集原始资料的形式。

4. 按资料的连续程度划分

统计调查按登记资料连续程度不同,可分为经常性调查和一次性调查。经常性调查是一种连续性调查,它随着总体单位标志值的发展变化,连续不断地进行登记,获取标志值发展变化过程及其结果的统计资料。一次性调查是一种非连续性调查,它主要用于取得总体单位在某个时点上状况的统计资料。一次性调查可分为定期和不定期的两种。

二、统计调查方案

统计调查是非常复杂细致的工作,它由许多人把许多总体单位的资料收集起来,为了在调查过程中保持步调一致,需要有一个统一、完整的统计调查方案。通常可以根据下述步骤确定调查方案的具体内容。

1. 确定调查目的

制订统计调查方案的首要问题是确定调查目的。目的不同,调查的内容、范围、方式也会不同。目的不明,就无法确定向谁调查、调查什么、如何调查,由此产生的结果必然是:调查得来的资料可能并不都是需要的,而需要了解的情况却又得不到充分的反映,浪费了人力、物力和时间。

目的的确定要具体。任何调查都可以从不同角度,根据不同的目的来确定。例如人口可以从职业结构上来研究,也可以从文化程度上来研究。一般来说,确定调查目的要遵循两项原则:①从研究工作的需要出发,抓住最为重要的迫切问题;②从调查对象的实际出发,把需要和可能结合起来。

2. 确定调查对象和调查单位

调查对象就是统计总体,更准确地说,统计总体在调查阶段称为调查对象。确定调查对象,就是要确定调查的是什么样的