

抽样调查的 理论与方法

主编 佟哲晖



中国统计出版社

抽样调查的理论与方法

主编 佟哲晖

中国统计出版社

(京)新登字 041 号

图书在版编目 (CIP) 数据

抽样调查的理论与方法/佟哲晖主编. —北京: 中国统计出版社, 1994. 10

ISBN 7-5037-1782-3

I. 抽… II. 佟… III. 抽样调查-高等教育: 函授教育-教材 IV. C811

中国统计出版社出版

(北京三里河月坛南街 38 号 100826)

新华书店经销

新乐市印刷厂印刷

*

787×1092 毫米 32 开本 12.25 印张 26.5 万字

1994 年 10 月第 1 版 1994 年 10 月新乐第 1 次印刷

印数: 1—5, 500 册

*

ISBN 7-5037-1782-3/C·1014

定价: 11.5 元

(版权所有, 不得翻印)

前 言

在统计理论和实践中,抽样调查占有很重要的地位。新中国成立以来,统计部门在应用抽样调查进行城乡居民家计调查、农产量调查、人口普查后的抽样复查等方面,都取得了显著的成果,有关抽样理论与方法的研究也有很大进展,与世界各国的学术交流正日益扩大。当前,我国正在经历由计划经济体制向社会主义市场经济体制转变的时期,统计体制和统计方法也必须进行相应的改革。1994年全国统计会议提出,在统计调查方法改革的目标模式中,应将经常性的抽样调查放在主体的地位;除现有的两个抽样调查队而外,另组建企业调查队,并在工业、建筑业,批发零售贸易、餐饮、社会消费品零售、固定资产投资等行业,开展进度性抽样调查。这是我国统计调查方法体系中一项重要的改革。为了适应新形势的要求,我们受国家统计局培训中心、中央广播电视大学的委托,编写了这本《抽样调查的理论与方法》,作为经济信息管理专业的教材使用。本书还可作为高等学校统计学专业教学的参考读物。

全书共十二章,从结构看,可分成三大部分。第一部分(第一、二、三章)概要地介绍了抽样调查的基本概念、概率与概率分布、抽样分布等内容。其中有些是属于数理统计的内容,但为了能从理论来源上了解抽样调查的数理根据,介绍这一部分内容仍然是十分必要的,还因为在学习本书后面各章内容

时,也要经常用到这些东西。当然,学习时可根据具体情况,有选择地阅读。

第二部分(从第四章至第十一章)以较多篇幅分别叙述了统计实践中常用的抽样调查方法,并介绍了比估计、回归估计、双重抽样等内容。在写法上,除了着重说明每种调查方法的特点和适用范围外,也注意从理论上进行适当论证,力求使读者既了解每种调查方法各自的特点,又能掌握它的理论依据,把理论、方法、应用三者有机地融合起来,以便在实际应用时能得心应手。这部分内容是全书的重点。

第三部分(第十二章)介绍了我国统计实践中抽样调查方法的应用情况。

本书在编写过程中曾参考了国内外出版的许多专著、论文及其它参考资料,从中获得不少教益;辽宁省农调队、城调队向我们提供了大量第一手资料,给以大力支持帮助;国家统计局石占前、夏荣坡同志对本书的编写提出了宝贵意见并做了大量组织工作,在此一并表示衷心的感谢!

本书的全部稿子是由孙慧钧讲师(博士研究生)在搜集、阅读大量资料的基础上,经过综合整理研究执笔编写的;主审是姚志学教授;最后由佟晖教授负责总纂定稿。

由于水平所限,时间匆促,书中难免有错误和不妥之处,诚恳希望广大读者给予批评指正。

编 者

1994年9月于大连

目 录

第一章	绪论	(1)
第一节	抽样调查的概念和作用.....	(1)
第二节	抽样调查中的几个基本概念.....	(8)
第三节	抽样调查的组织形式及调查方法	(12)
第四节	抽样调查的设计	(17)
第二章	概率与概率分布	(22)
第一节	概率的概念及其基本运算规则	(22)
第二节	随机变量和概率分布	(31)
第三节	二种重要的概率分布	(39)
第四节	随机变量的数量特征	(44)
第五节	大数定律与中心极限定理	(50)
第三章	抽样分布	(54)
第一节	统计量与抽样分布的概念	(54)
第二节	样本平均数的抽样分布	(58)
第三节	样本成数的抽样分布	(62)
第四节	抽样误差	(65)
第五节	总体参数的估计	(75)
第四章	单纯随机抽样	(88)
第一节	单纯随机抽样方法及其特点	(88)
第二节	样本指标及其抽样误差	(93)
第三节	样本容量的确定.....	(100)
第五章	系统随机抽样	(109)
第一节	系统随机抽样方法及其特点.....	(109)

第二节	几中常用的系统抽样方法·····	(113)
第三节	样本指标及其抽样误差的测定·····	(116)
第四节	系统抽样的精度·····	(124)
第五节	总体结构问题·····	(129)
第六章	分层随机抽样 ·····	(137)
第一节	分层随机抽样方法及其性质·····	(137)
第二节	样本指标及其抽样误差的测定·····	(140)
第三节	分层抽样的组织形式·····	(154)
第四节	样本容量的确定·····	(167)
第五节	分层抽样的精度·····	(179)
第七章	整群随机抽样 ·····	(189)
第一节	整群抽样方法及其性质·····	(189)
第二节	等群随机抽样·····	(192)
第三节	不等群随机抽样·····	(200)
第四节	整群抽样的精度·····	(212)
第五节	样本容量的确定·····	(214)
第八章	随机抽样中的比估计 ·····	(217)
第一节	比估计的概念及其性质·····	(217)
第二节	单纯随机抽样下的比估计·····	(220)
第三节	系统抽样下的比估计·····	(239)
第四节	分层抽样下的比估计·····	(248)
第五节	整群抽样下的比估计·····	(262)
第九章	随机抽样中的回归估计 ·····	(271)
第一节	回归估计的概念及其性质·····	(271)
第二节	单纯随机抽样中的回归估计·····	(274)

第三节	系统抽样中的回归估计·····	(286)
第四节	分层抽样中的回归估计·····	(292)
第十章	双重随机抽样 ·····	(305)
第一节	双重随机抽样的概念及其性质·····	(305)
第二节	为分层而进行的双重抽样·····	(307)
第三节	为进行比估计而进行的双重抽样·····	(317)
第四节	为进行回归估计而进行的双重抽样·····	(324)
第十一章	二阶(级)抽样 ·····	(330)
第一节	二阶抽样的概念及其性质·····	(330)
第二节	一阶单位大小相等的二阶随机抽样·····	(333)
第三节	一阶单位大小不等的二阶随机抽样·····	(339)
第十二章	抽样调查方法在我国的应用 ·····	(350)
第一节	住户调查·····	(350)
第二节	农产量调查·····	(357)
第三节	系统抽样调查的应用·····	(365)
第四节	样本轮换·····	(368)
第五节	人口与工业产品质量抽样调查·····	(374)

第一章 绪论

第一节 抽样调查的概念和作用

一、抽样调查的概念

抽样调查是建立在随机原则基础上,从总体中抽取部分单位进行调查,并依据概率估计原理,应用所得到的资料对总体的数量特征进行推断的一种调查方法。例如,从某地区全部职工当中随机抽取部分职工,以家庭为单位,按月调查取得有关收入、支出等方面的资料,并依据这些资料推断全区职工的收支状况,这就是一种抽样调查。从调查方法上看,它是属于一种非全面调查,但又与一般调查方法不同,它不只停留于搜集资料和整理资料,而且还要对资料进行分析,并据以推断总体的数量特征,从而提高统计的认识能力。因此,抽样调查的理论和方法在统计中占有很重要的地位。

抽样调查具有以下主要特点:

1. 从调查总体中抽取的部分单位是依据随机原则进行的,总体中每个单位都有同等被抽中的机会。被抽中的单位在统计中称为样本,它相当于一个代表团,这个代表团也就是总体的缩影。

2. 抽样调查的目的在于通过样本获得资料,并进一步从数量上推断总体。因抽样而产生的误差可以根据抽样原理加

以测定并控制在一定范围之内,从而使抽取的“代表团”能反映出总体的数量特征。

二、抽样调查的主要作用

一般地说,抽样调查主要用于下列一些现象的调查:

1. 实际工作不可能进行全面调查,而又需要了解其全面情况的现象。例如,具有破坏性或消耗性产品的质量检验,如灯泡的使用寿命,轮胎的行程,人体的血液检验以及糖果、烟、酒等食品和调味品等的质量检查,都不可能把所有产品一一进行检查和试验,只能采用抽样调查的方法。

2. 对某些现象虽然可以进行全面调查,但是由于总体范围大、单位数目多,又缺乏原始记录可作依据,实际进行全面调查有许多困难。例如,城乡居民的家庭收支调查,虽然可以对城乡中每户居民逐一进行调查,但工作量太大,耗费人力、物力和财力过多,而且常常缺乏原始记录。这时可以采用抽样调查,掌握部分家庭的收支状况,也就足以说明总体的特征。

3. 对普查或全面调查资料的质量检查和修正。例如,为了检查人口普查资料的正确性,在普查完毕后要抽5%—10%的居民户对一些重要的指标进行详细的复查,用复查结果修正普查资料。比如,我国建国以来的四次人口普查所公布的人口资料,都是经过抽样调查修正普查数据后的结果。

4. 当被调查总体中的单位数无限多时,事实上不可能进行全面调查,只能进行抽样调查。例如,江河、湖泊、海洋中的鱼尾数,大气或海洋的污染程度等就属于这种情况。

从调查方法上看,抽样调查是非全面调查的一种,它比全面调查有许多优点。

首先,抽样调查节约费用,快速及时。由于抽样调查需要

调查的单位少,搜集、整理和汇总工作量小,所以同全面调查相比较,能大大节约人、财、物力,能快速地进行汇总和分析,尤其对于急需的有关信息可以及时得出所需结果。

其次,抽样调查能够提高资料的准确性和可靠性。由于抽样误差的大小取决于样本容量的大小,也取决于抽样的组织方式,因此,在抽样时可以通过抽取单位数的多少和抽样组织方式的变化来调节和控制抽样误差,同时可以用较少的费用,对所需要的工作人员进行严格的业务培训,提高业务人员的业务素质,减少调查登记误差,更进一步提高资料的准确性和可靠性。

第三、抽样调查有广泛的应用领域。目前世界上许多国家在以下一些方面广泛采用抽样调查法:①农产品产量调查;②土地资源利用调查;③城乡居民和职工家庭收支调查;④工业产品质量检查;⑤劳动就业调查;⑥市场、物价和购买力调查;⑦饮水、住宅、人民健康和社会福利调查;⑧科学实验效果调查;⑨环境污染调查;⑩人口、工业、农业等各种普查后的复查;⑪民意测验等。

在我国,抽样法已被广泛应用于生产技术及社会生活的各个领域。目前,国家统计局调查制度中所包括的统计指标,依靠抽样方法取得的资料已达三分之一左右。在城乡住户调查、农产量调查、价格统计、市场调查等领域,应用抽样调查已取得了很好的成果,在人口统计社会统计、交通统计、商业统计等领域,抽样调查也正在发挥越来越重要的作用。从1995年起,还将在工业、固定资产投资和第三产业中的生产经营活动组织抽样调查。随着我国社会主义市场经济的发展,抽样调查的应用范围将逐步扩大,所发挥的作用也将越来越大。

但是抽样调查也有其自身的局限性。首先,由于总体构成的复杂性,通常它无法提供总体中各个组成部分的资料。其次,组织抽样调查要遵守某些条件,被调查的对象也有一定限制。一些重要的反映国情、国力的统计资料,仍然必须组织全面调查。因此,既要大力推广和应用抽样调查,又不能用抽样调查取代其它调查。应该根据资料的性质和调查对象之不同,采用不同的统计调查方法。

三、抽样调查理论与方法的产生和发展。

抽样调查是一门较新的学科,但是它的形成却经历了一个漫长的历史过程。

最初人们对自然界和社会经济现象的数量观察采用逐个清点的办法,仅是对具体事物的记录,以反映当时社会的一些重要事项。例如,我国历代的户口登记,欧洲中世纪各国的财产登记等都是现代统计学所说的全面调查和普查的雏形,是原始形态的全面调查。

经过长期的实践,人们发现在社会和自然界中,如果对个别的偶然现象进行大量的观察和汇总,能够发现某种共同的规律。这就是大数定律。大数定律的发现为抽样理论的诞生奠定了数理基础。随着现代数学和现代统计学的发展,特别是数理统计学的发展,人们着手研究采用实地调查和数学推断相结合的办法,根据对部分事物的观察结果来推断大范围 and 全体事物。各国科学家们在这一方面作出了巨大的努力。例如,法国著名数学家拉普拉斯(P. S. Laplas)在他的一篇关于1786年巴黎人口出生、死亡和婚姻状况的论文里,建议用某些地区的出生率来推算整个法国人口,并对推算数字的误差进行了研究,同高斯(C. F. Gauss)共同建立了误差分布规律。

1891年,挪威统计局长凯尔(A. N. Kiaer)进行了有关挪威国民收入和财富的统计抽样调查。在这次抽样调查中,可能代表全国国民收入和财富情况的许多城市、乡区和村落被挑选出来。在这些区域内按某种年龄(如17,23,28岁等)各种社会阶层(五种),各种婚姻状况以及某种字母起首姓名等标志,把居民的有关情况详细记录下来,从这样的记录中推断出挪威全国国民收入和财富。类似的抽样法也由凯尔用于推算全国人民的年龄、收入和职业间的情况。随后也被推广到农业统计方面。比如1901年丹麦农产品的收获量调查和以后德国巴登地方的牲畜调查等。

早期对抽样理论做出杰出贡献的有俄国数学家切比雪夫(YeGbiweB)、瑞士数学家贝努里(J·Bernoulli),他们阐明了抽样论数理基础中著名的大数定律;法国数学家德姆瓦(DeMoivre)在1773年推导出常态曲线公式,对平均误差的问题进行了研究;比利时统计学家魁特莱(A·Quetelet)提出了统计结果的稳定性,把数理统计方法应用到自然和社会的许多方面,获得了重大突破;英国数学家葛尔登(F·Galton)和皮尔逊(Karl·Pearson)建立了变异和相关的广泛概念。

以后经过戈塞特(W·S·Gossett)、费喧(R·A·Fisher)等的共同努力,逐步发展完善了抽样理论,特别是戈塞特构造的七分布——小样本的统计方法,克服了大样本的限制,把抽样理论和实践不断推向前进。

我国统计学家、北京大学教授许宝禄(1910——1970)为抽样理论和方法的创建做出了举世瞩目的贡献。他的《抽样论》系统地论述了抽样法原理,为开展大规模的社会经济调查、人口调查,草原、农田估产,森林木材调查、能源调查等提

供了方法依据。

随着抽样调查理论的日臻完善,本世纪 30 年代以后,特别是第二次世界大战以后,抽样法得到了广泛的应用。美国在二战期间利用当时发明的极少量电子计算机,首次制订了成批产品抽样检验方案来评定大量军事物资的质量,效果显著。这些方案在战后也推广到民用产品的质量保证方面。印度从三十年代开始就采用随机抽样方法组织劳动就业、土地利用、农产量、城乡居民家计等一系列大规模抽样调查,把科研和实际调查结合起来,在抽样理论和实际工作方面取得很大成效。1957 年,印度著名统计学家马哈拉诺比斯(P·C MaHaLanobis)曾来我国讲学,介绍有关抽样理论和方法。

我国在解放以前,也曾在工业产品质量检验,医学、生物学等方面采用过抽样调查,但是大规模应用抽样调查是在新中国成立以后。1955 年 1 月,国家统计局颁布了《1954 年农民家计调查方案》,规定以系统抽样方法抽取调查乡,再按类型比例与系统抽样相结合的方法抽取调查户,全国抽取了 15000—20000 户农户。当时观察的主要对象是个体农户,合作化后转变为社员户的生产情况调查。我国的职工家计抽样调查是在 1956 年开始的,国家统计局制定了《全国职工家计调查方案》,采取类型比例与系统抽样相结合的方法,全国抽取工业部门国营和公私合营职工 6000 人。1957 年职工家计调查扩大到商业、教育、机关团体等部门。经过五十年代、六十年代的艰苦努力,我国的抽样调查工作取得了不少成绩,积累了丰富的经验。但在文化大革命的十年动乱期间,我国的统计调查工作一度被迫中断。党的十一届三中全会以来,我国的抽样调查工作跨入新阶段,建立了农村和城市两支抽样调查队。

农村调查队主要从事农产量调查、农村住户调查和农村社会发展、经济生活、精神文明建设等调查。城市调查队主要从事城市住户调查、职工生活费用价格指数、零售物价指数调查等。此外，抽样调查在工业、商业、交通运输等生产管理中也得到了广泛应用。

当前我国正在经历由计划经济体制向社会主义市场经济体制转变的时期。在新形势下，各方面对统计信息的需求不是减少，而是增加了，统计工作不仅不能削弱，而且必须加强，但统计体制和统计方法都必须继续进行改革，而统计调查方法的改革又是其中的重要一环。1994年全国统计工作会议上提出了《统计调查方法改革总体方案》，其中规定我国统计调查方法改革的目标模式是：建立以必要的周期性普查为基础，以经常性的抽样调查为主体，同时辅之以全面统计报表，重点调查和科学推算综合运用的统计调查方法体系。显然，在这个方案中，抽样调查被置于主体的地位。相应的主要改革措施包括：进一步改革和完善城、乡两个经济调查队，并另外组建企业调查队，作为开展企业调查的专门组织机构，各地方要尽快建立和完善基本统计单位名录库，为今后有计划地推广应用目录抽样作好准备。从1995年起，在工业统计、批发零售贸易业、餐饮业及社会消费品零售统计、固定资产投资统计中，开始采用目录抽样法搜集资料。根据即将在全国广泛开展抽样调查的客观形势，进一步学习和应用抽样调查的理论和方法，是我国统计工作者面临的一项重要任务。

第二节 抽样调查中的几个基本概念

一、总体和样本

总体是指由调查对象的全部单位构成的集合体,也称为全及总体。总体的范围可以是一个企业、村庄、街道、机关、学校,也可以是一个县、地区、省、市,甚至一个国家。例如,要调查、观测某个村庄居民的年龄构成,这个村庄的全体居民就构成一个总体;如果要对全市所有职工家庭生活状况进行调查,则该市全部职工家庭便构成一个总体。

总体按所包含的单位数目的多少,可分为有限总体和无限总体。有限总体中包含有限个单位,是可以计算出来的。有限总体包含的单位数称为总体容量,记为 N 。无限总体包含无限个单位,不能用计数方法取得。例如,一个地区的人口、一个企业一年的产量、一批灯泡都构成有限总体,但是一个湖泊、海洋有多少尾鱼,一个森林区有多少株树等,这些都无法确切知道,它们构成无限总体。抽样调查所研究的总体绝大部分是有限总体,如果有限总体中包含的单位数目相当多,也可以近似地看作无限总体。

抽样总体是按随机原则从总体中抽选出来的部分单位组成的集合体,在抽样调查中通称为样本。例如,从一批灯泡中抽取 20 只进行测试,这 20 只灯泡构成一个样本;从某市全部职工中抽选 200 户进行调查,这 200 户职工家庭就构成一个样本。构成样本的单位数目,称为样本容量,以 n 表示。

样本容量与总体容量 N 相比是个很小的数,最大的样本

容量 n 一般也不超过总体容量 N 的 $\frac{1}{3}$, 往往是百分之几或者千分之几。但是当总体容量 N 很大时, 样本容量 n 本身也必须足够大, 这样才能保证样本对总体的代表性。根据 n 的大小, 样本分为大样本和小样本。在统计学上, 一般把抽取 30 个以上单位, 即 $n \geq 30$, 叫做大样本, 而把抽取 30 个以下单位, 即 $n < 30$ 的样本, 称为小样本。

统计总体与总体单位通常都是以实物形态表现的实体, 但在某些场合, 总体的内涵也有时变化。当总体单位及所研究的标志一经确定, 为了便于叙述, 有时把标志值抽象化, 把原总体与反映总体单位特征的变量等同起来, 而把变量值的集合看成总体。在抽样推断中研究的是随机变量, 因而本书在以后的叙述中, 就是把随机变量可能取值的全体组成的集合称为总体。但是也不应该忘记, 在最后回归到具体问题的研究时, 仍然要还原为原来的实际存在的总体。

二、总体指标与样本指标

根据总体中各单位的标志值计算的用于反映总体数量特征的指标, 称为总体指标, 也称为全及指标或总体参数。例如, 一批灯泡的平均使用寿命, 某市全部职工家庭平均每月收入, 一批产品的合格率和废品率等, 都是总体指标。由于总体是唯一确定的, 因此, 总体指标是个确定的数值。

总体指标主要有总体平均数、总体成数、总体方差和总体总量指标等。

设总体中 N 个总体单位某项标志的标志值为 Y_1, Y_2, \dots, Y_N , 其中具有某种属性的有 N_1 个总体单位, 不具有某种属性的有 N_0 个总体单位, 则总体指标的计算公式为: