

抽样调查 原理与 方法

马祖谟 编著
中国统计出版社

SAMPLE SURVEY PRINCIPLES AND METHODS

抽样调查原理与方法

马祖谟 编著

中国统计出版社

(京)新登字041号

抽样调查原理与方法

CHOU YANG DIAO CHA YUAN LI YU FANG FA

马祖谟 编著

中国统计出版社出版

(北京三里河月坛南街38号 100826)

新华书店北京发行所发行

北京市通县永东印刷厂排版

大厂回族自治县印刷厂印刷

850×1168毫米 32开本 9.375印张 3插页 22万字

1992年2月第1版 1992年2月北京第1次印刷

印数: 1—3000

ISBN7-5037-0696-1/C·417

定价: 5.90元

前 言

抽样调查是一种科学的统计调查方法，是统计学的重要组成部分。抽样调查的理论和方法，早已为世界各国经济发达国家普遍采用。近十几年来，由于在经济、社会等许多领域得到广泛应用，它又获得了迅速发展，成为统计学中最活跃的分支之一。

在我国，党的十一届三中全会以来，抽样调查方法的研究和应用，愈来愈受到有关部门的重视，并在农村经济调查、城乡住户调查和人口调查等方面取得显著成绩。1981年9月15日国务院在批转国家统计局《关于加强和改革统计工作的报告》中强调指出：

“积极推广抽样调查，做到统计调查方法灵活适用。……今后布置新的统计任务，增加新的统计指标，凡是适合用抽样调查的，就不用全面报表，现在已经建立的全面报表制度，适合改用抽样调查的，要逐步用抽样调查代替。”在社会主义初级阶段，随着有计划商品经济的发展，抽样调查方法在统计工作改革和现代化过程中的作用，将会显得更加重要。

本书是为满足目前高等财经院校抽样调查课的教学和广大统计工作者的实际需要编写的，目的在于介绍抽样调查的基本理论与方法及其在实际中的应用。全书注重理论联系实际，充分考虑到不同对象、不同水平读者的实际状况，不过分追求纯数学的推导，但又适当地对一些重要方法和公式给予必要的论证，以加深读者的理解；同时，又通过大量实例说明方法和公式的应用，并结合原理介绍了一些实际经验。每章之末附有小结，提出复习指导，列有公式系统表，以揭示本章重点、主线以及各部分内容、各公式

之间的内在联系与区别。这样，有助于读者牢固掌握所学知识。

1989年11月10日，全国统计教材编审委员会数理统计专业组在给统著者的信中说：“经全国统计教材编审委员会数理统计专业组委员讨论认为，《抽样调查原理与方法》一书较系统地介绍了抽样调查的常用方法和基本原理，其难易程度也适合目前高校财经统计专业本科学生的水平，可以在这基础上加以修改定稿”。根据专业组的意见和近几年的使用情况，又作了修改。在出版中还得到中国统计出版社卢精诚等同志的指导和帮助，在此一并表示衷心感谢！本书不足之处，欢迎读者指正。

编著者

1991年7月4日

内 容 简 介

本书比较系统地介绍了抽样调查的常用方法与基本原理。主要内容有：抽样调查方案的设计和实施、简单随机抽样、分层随机抽样、整群随机抽样、等距抽样、多阶段抽样及不等概率抽样和比率估计方法等。书中对各种抽样方法的特点、应用场合、估计效率以及综合运用都作了比较深入的分析 and 讨论。

本书取材广泛，理论密切联系实际，思路清晰，重点突出，结构严谨，文笔流畅易读；章末的学习指导和公式系统表具有特色，为读者掌握和应用所学知识提供了有利条件。

本书适用于高等财经院校师生、广大统计实际工作者和统计理论工作者及中等财经专业学校教师等。

29.911
C5
29.911
抽样调查原理与方法 / 1
国家统计局出版社, 1992.2
266页; 大32开

目 录

第一章 引言	(1)
第一节 抽样调查的意义和作用	(1)
第二节 抽样调查中的基本概念	(9)
第三节 抽样调查方案的设计和实施	(14)
复习思考题	(23)
习题一	(24)
第二章 简单随机抽样	(26)
第一节 简单随机抽样的概念和方法	(26)
第二节 简单重复抽样	(32)
第三节 简单不重复抽样	(47)
第四节 总体成数的估计	(56)
第五节 比率估计量	(66)
本章小结	(75)
习题二	(79)
第三章 分层随机抽样	(84)
第一节 分层随机抽样的概念	(84)
第二节 总体平均数的分层抽样估计	(87)
第三节 样本容量在各层的最优分配	(90)
第四节 样本容量的确定与分层抽样的 估计效率	(109)
第五节 总体成数的分层抽样估计	(114)
第六节 总体总量和总体中具有某特征单元总数的 估计	(123)

本章小结.....	(126)
习题三.....	(129)
第四章 整群随机抽样	(135)
第一节 整群随机抽样的概念.....	(135)
第二节 总体平均数的等群抽样估计.....	(138)
第三节 总体成数的等群抽样估计.....	(143)
第四节 不等群抽样估计方法.....	(151)
第五节 不等概抽样估计方法.....	(160)
第六节 整群抽样的比率估计方法.....	(169)
本章小结.....	(173)
习题四.....	(174)
第五章 等距抽样	(178)
第一节 等距抽样的概念.....	(178)
第二节 有序等距抽样的方法.....	(183)
第三节 总体平均数的等距抽样估计.....	(194)
第四节 分层等距抽样.....	(206)
第五节 二维等距抽样.....	(211)
本章小结.....	(214)
习题五.....	(214)
第六章 多阶抽样	(215)
第一节 多阶抽样的概念.....	(215)
第二节 总体平均数一阶单元等大小的 两阶抽样估计.....	(218)
第三节 总体成数一阶单元等大小的 两阶抽样估计.....	(231)
第四节 一阶单元及二阶单元等大小的 三阶抽样估计.....	(236)
第五节 对一阶单元分层的两阶抽样估计.....	(238)
第六节 一阶单元大小不等的两阶抽样估计.....	(241)

本章小结	(245)
习题六	(246)
习题答案	(253)
附录1	(257)
附录2	(259)
附表1 随机数字表	(260)
附表2 正态分布的双侧分位数表	(264)
附表3 泊松 (Poisson) 分布参数 λ 的置信区间表	(265)
参考书目	(266)

第一章 引 言

第一节 抽样调查的意义和作用

一、抽样调查的概念和特点

(一) 全面调查与非全面调查

统计调查中，把所研究的大量同类事物或现象的全体称为总体，构成总体的每个基本单位称为总体单元或个体。统计调查的目的是，取得尽可能准确的综合数字和资料，用以说明总体的性质和规律。统计调查的方法，按其所包含的范围来说，可概括地分为全面调查和非全面调查两大类。

全面调查是对总体的全部单元，按某一时点或某一时期逐个进行观察、记录，说明总体在某一方面或某些方面的性质和规律。例如，我国建国以来，在社会经济领域中普遍实行的统计报表制度，就是一种全面调查。有时，一些机构在一段时间里，就某一特定问题组织的大规模普查，也是一种全面调查。如人口普查，工业普查，科技普查，耕地普查等。

在我国，中央和各省（自治区、直辖市）、县都设有统计机构，各企、事业单位也都设有统计部门或统计人员，这是开展全面调查的良好条件。但是，全面调查有很多局限性，有时必须或者最好进行非全面调查，即抽取总体的一部分单元进行调查。实际工作中，在下述情况通常进行非全面调查：

1. 有些总体本身的性质决定了不可能对它全面调查，而只

能进行非全面调查。

总体中所含单元的个数，称为总体的容量或大小，常用 N 表示。含有无限多个单元的总体，称为无限总体。对这样的总体，显然不能全面调查，而只能调查它的一部分单元。例如，检验连续不断生产过程中产品的质量，调查商店中顾客的购买情况，观察每天公共汽车上的乘客人数等。这里，产品的个数，顾客与乘客的人数，尽管在某一段时间内是有限的，可以全面调查；但是从长远来说它们又都处于连续不断的过程之中，在这个过程中，它们都是无限多的，因而不可能全面调查。

2. 有些总体，理论上虽可全面调查，但实际上却不能或不必要全面调查。

含有有限个单元的总体，称为有限总体。对这样的总体，虽说可以逐个调查，但实际中有些调查是一种破坏性或消耗性试验，因而实际上不允许全面调查。例如，检验一批药品的效力，棉纱的强力，轮胎的耐磨力，子弹的杀伤力以及农作物种子的发芽率等等，只能进行非全面调查。有些调查虽不是破坏性试验，但实际却难以进行或不必要全面调查。如城乡住户调查，面对如此众多的住户，要调查他的几十个甚至上百个项目，显然是难以办到的。再者，由于同类居民的家计情况比较接近，因而也无必要全面调查。又如集贸市场调查，森林木材积蓄量调查等，也属这种情况。

3. 为节省人力、财力，或受其条件限制而只好进行非全面调查。

例如，我国在1982年7月1日进行的第三次人口普查，虽然只调查19个项目，由于要调查10亿多人口，工作量巨大，仅动员普查员和普查指导员就有518万。此外，还有570万人参加了户籍整理等准备工作，50万人参加了试点工作，所花经费上亿元。因此，人口普查大约每10年才搞一次，在两次普查之间，一般只做抽样调查。

4. 为提高时效性, 或受时间限制, 而只能进行非全面调查。

例如, 我国第四次人口普查大致分为三个阶段: (1) 准备阶段。从1989年5月至1990年6月底, 为时14个月。(2) 普查登记、手工汇总和编码阶段。从1990年7月1日至1990年10月底, 为时4个月。(3) 电子计算机汇总、资料的整理分析和工作总结阶段。从1990年11月至1993年, 为时38个月。全部工作历时要4年8个月。如果需要经常或及时掌握人口变化情况, 只有进行非全面调查。

(二) 抽样调查的概念和特点

非全面调查, 可以说是广义的抽样调查, 按其抽样的原则, 可分为意识抽样与随机抽样两类。

意识抽样是, 调查者主观地、有意识地从总体中抽取部分单元进行调查。如典型调查和重点调查, 都属于意识抽样。典型调查在国外称为判定抽样或目的抽样, 是调查者在对总体已有一定了解和分析的基础上, 根据自己的调查目的, 有计划地从总体中抽取个别典型单位, 通过解剖麻雀、系统分析来观察趋势, 指导一般。如毛泽东同志的《兴国调查》和《湖南农民运动考察报告》, 社会学家费孝通教授的《江村调查》等。典型调查针对性强, 对于制定政策, 指导工作具有重要意义。但是, 由于调查者判断的依据是他自己认可的单位, 因此, 调查结果在很大程度上依赖于调查者本人的主观意识, 以及他们的知识和经验。重点调查是, 抽取总体中在数量或比重上占有重要地位的少数单位进行调查。通过对这些单位的调查, 掌握总体的主要情况。例如, 通过对一些大型油田的调查, 掌握国家石油的生产情况; 通过对几家大型商店的调查, 了解市场物价、购买动向等情况; 通过对几所重点大学的调查, 了解高等教育和大学生的状况等。

随机抽样是, 按照随机原则, 即保证总体中每个单元都有一定概率被抽取的原则, 抽取部分总体单元进行调查。所抽部分单元称为随机样本, 简称样本。所以, 随机抽样也称概率抽样, 而

意识抽样也称非概率抽样。本书所称抽样调查，即指随机抽样。与全面调查和意识抽样比较，随机抽样具有以下特点：

1. 按照随机原则抽取样本，这是抽样调查（即概率抽样）与其它非全面调查的根本区别。

按照随机原则抽取样本，总体中各个单元都有一定的概率被抽取，排除了意识抽样中，由于人们主观因素的作用而产生系统偏差的可能性。因而，能以极大的概率保证被抽取的单元在总体中的分布比较均匀，使得样本中单元的分布与总体更加接近，增强了样本的代表性。

2. 在数量上以样本推断总体，这是抽样调查不同于其它调查方法的主要特点。

总体单元之间虽然存在着不同程度的差异，但抽样调查按随机原则抽取充分多的单元组成样本，大量单元平均的结果随机性相互抵消而趋于稳定，这符合大数定律的基本条件，因而可以用样本推断总体。并且，无论总体的分布如何，根据中心极限定理，此时所产生的误差都趋于正态分布，因而可以利用正态分布计算这种误差。由此可见，抽样调查的另一个重要特点是：

3. 可以在一定的概率保证下，计算抽样误差，并把它控制在允许的范围之内。

需要说明的是：抽样调查所研究的对象是有限总体，如果需要的话，是可以对它进行全面调查的，这与古典概率主要研究无限总体是有所不同的。再者，抽样调查的理论基础是大数定律与中心极限定理，所以要求必须按照随机原则抽取样本，并且，它只适用于对大量现象的统计调查。但这并非说抽样调查只适用于随机现象，即总体单元之间的差异是单纯由于随机因素影响而形成时，才能运用抽样调查。我们要把总体的特点与抽样的随机原则区别开。

二、抽样调查的优越性

抽样调查不仅能使统计方法由单纯性的数量描述,提高为由样本推断总体,而且可以进行预测。它扩展了统计研究的领域,增强了人类的认识能力。与其它调查方法比较,它的优越性可以概括为:省、快、广、准。

第一,省。与全面调查比较,抽样调查可以节省人力和财力,提高调查的经济效益。

抽样调查中,由于调查的单位相对很少,调查中的组织、登记、汇总、分析等工作量大为减少,因而可以节省人力、财力和物力。特别是,当总体规模庞大,单元众多,分布很广,情况复杂时,更可以显示出它的优越性。例如住户调查,国家统计局1984年3月制定的《城镇居民家庭收入和基本情况一次性调查方案》中规定:

调查户数最多不得超过以下规定的比例,但也不得少于最少的调查户数。

	抽选比例 (%)	各市、县 最少调查户数(户)
城市(按非农业人口数分组)		
100万人以上的城市	6	2000
50~100万人的城市	8	1200
20~50万人的城市	10	800
不足20万人的城市	—	500
县城	—	500

调查户数=本市(县)非农业居民户数×抽选比例(%)

例如,某城市非农业居民家庭有22.5万户,抽选比例为千分之八,则应调查1800户:

$$22.5(\text{万户}) \times 8\% = 1800\text{户}$$

第二,快。与全面调查比较,抽样调查可以节省时间,加快

工作进程，提高调查的时效性。

由于抽样调查涉及的单位少，工作量小，各项工作都由专业人员承担，因而可以节省调查时间，加快工作进程，及时获得所需要的数据和资料。特别是在急需信息或要随时掌握情况变化时这一点尤为重要。例如，瞬息变化的市场情况调查，收购季节的农产量调查，工业生产过程中的质量控制，甚至民意测验等等。

第三，广。与全面调查比较，抽样调查可以增加调查项目，扩大调查范围。

全面调查中，由于需要调查全部总体单元，所以每增加一个调查项目，就要增加很多人力、财力和时间。有些调查项目需要有经过专业训练的调查人员和使用专门的设备才能取得数据和资料，此时更不能轻易增加调查项目，而只能限于一些最基本、最重要的项目进行调查，并且不能经常进行。抽样调查由于只调查相对很少的总体单元，便于组织领导，容易培训调查人员和配置调查设备，因此不仅可以在同一范围内增加调查项目，进行深入细致的调查，而且在同样的人力、经费等条件下，还可以扩大调查范围，在更广泛的领域中获得资料和信息。

第四，准。科学设计和严格实施的抽样调查，能够获得精确、可靠的调查结果。

如国家统计局、国家计划生育委员会、国务院第三次人口普查领导小组《关于认真做好1983年人口变动情况抽样调查工作的意见》中指出：“出生、死亡等人口变动资料，是制定国民经济和社会发展规划，掌握人口政策，推行计划生育工作所必要的基本数据之一。但是，在目前依靠全面布置统计报表搜集人口变动数字，往往不够准确。按照上报的出生、死亡数字计算的全国人口自然增长数，与全国总人口的增长数相差较大。以1982年为例，按各地上报的人口自然增长数字为1247万人，自然增长率为12.42%。与全国总人口增加数1495万人相差248万人。而根据1982

年抽样调查结果，人口自然增长率为14.49%，据此推算，自然增长的人口数为1454万人，与总人口增加数基本一致。这说明抽样调查是当前取得准确的人口变动数字的较好方法。”

抽样调查只抽取总体的部分单元进行调查，为何可能比全面调查获得更精确、更可靠的调查结果呢？回答这个问题，需对两种调查方法中的误差及其产生原因进行分析。

误差可以分为两大类，即调查误差和抽样误差。

第一类，调查误差，也称工作误差或登记误差，是在统计调查过程中，由于工作上的种种原因造成的。常见的主要原因可归纳如下：

(1) 调查方案存在缺点。此时，即使严格实施调查方案也会产生误差。调查方案中的概念与调查目的不相符或者含义不清，以及计算口径不一致或者公式使用不当都会产生系统性误差。如在住户调查中，“家庭人口”是一个十分重要的项目，它必须与住户家庭收支和消费的计算范围完全一致；否则，将产生较大的系统性误差。

(2) 实际调查范围与总体真实情况有出入。由于受到客观条件限制，在实际调查时所涉及的范围，即被抽样总体 (the sampled population) 与所要研究的总体，即目标总体 (the target population) 的真实情况之间存在差异。即因为测量不准、情况不明或发生变化等原因，调查时所使用的总体单元的名单或图表，即抽样框可能有重复或遗漏现象。例如，在与人口有关的调查中，预先制定的名单在使用时已经发生了变化；又如，在农产量调查中，所测得的播种面积与实际播种面积有出入，等等。

(3) 回答误差。有些情况下，由于调查者或被调查者的原因，回答前后不一致，或者作出不适当的回答。有的是因为对问题理解不透，记忆不清而无意造成的；有的则是问题涉及当事人的利害关系而有意夸大或缩小，甚至弄虚作假造成的。如住户调查中，居民可能有意少报收入，高报学历；在农产量调查中，为

了得到表扬、奖励而多报，或者为了取得贷款而少报；在个体经营者调查中，少报营业额更是常见的。在调查中，有时还会遇到被调查者拒绝回答，或者回答不完整，甚者找不到被调查者的情况。

(4) 测量、登记、整理、汇总、计算乃至印刷过程中的差错造成的误差。

上述调查误差，是全面调查与抽样调查中都存在的。并且，一般地说，由于抽样调查涉及单元少，填报层次少，调查人员专业素质好，工作质量高，可以比全面调查大大减少这种调查误差。

第二类，抽样误差，也称代表性误差，是按照随机原则抽取样本时，在不含上述调查误差的情况下，单纯由于样本与总体之间存在着差异而产生的误差。因而，抽样误差是抽样调查固有的，是不能消除和避免的。但是，根据概率论的知识，在一定可靠程度下，可以计算抽样误差可能的最大范围。并且，通过抽样方案的设计，在一定条件下，把它降到最低或控制在允许的范围之内，所以抽样误差又称可控制误差。

由此可见，我们科学地设计并严格实施抽样调查方案，把抽样误差控制在允许的范围之内，并使抽样调查中的调查误差远远小于全面调查中的调查误差，从而，当抽样调查中的两种误差之和，小于全面调查中的一种误差时，便可以使抽样调查的结果比全面调查更精确、更可靠。在实际中，特别当总体规模很大，单元很多，情况复杂时，这是容易做到的。因此抽样调查方法愈来愈多地为世界各国所采用，而且把抽样调查与全面调查结合起来，用全面调查后进行的抽样调查检验它的质量，确定它的误差，并对它进行修正与补充。例如，世界各国在人口普查后，一般都要抽取5%到10%的居民户进行调查，并对普查资料做出补充。我国1982年第三次人口普查后，根据《人口普查登记质量抽样检查细则》规定，又在全国抽取了972个生产队和居民小组的187362人重新进行登记