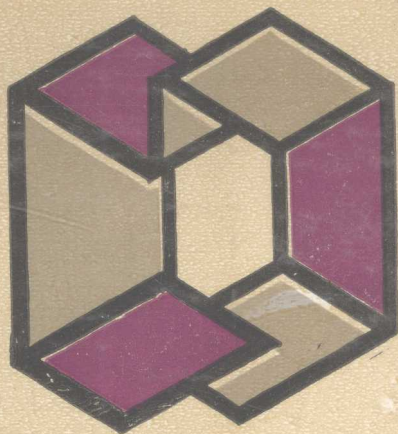


抽 样 调 查 方 法 研 究

THE RESEARCHES
OF SAMPLING
SURVEY



抽样调查方法研究

● 吴国培 著

厦门大学出版社

抽样调查方法研究

The Researches of Sampling Survey

吴国培 著

厦门大学出版社

[闽] 新登字 09 号

闽 0 第 1001 号

福建省出版集团中，《福建省出版集团中》，《福建省出版集团中》

。第 1001 号

立福建省出版集团中，《福建省出版集团中》，《福建省出版集团中》

。第 1001 号

福建省出版集团中，《福建省出版集团中》，《福建省出版集团中》

。第 1001 号

抽样调查方法研究

吴国培 著

*

厦门大学出版社出版

福建省新华书店发行

福建省尤溪县印刷厂印刷

*

开本 850×1168 1/32 6.125 印张 150 千字

1993 年 9 月 第 1 版 1993 年 9 月 第 1 次印刷

印数 1—2000 册

ISBN 7-5615-0775-5/F·122

定价：3.80 元

序

抽样调查和推断作为一种科学的认识方法，为世界各国推广应用已经有几十年的历史了。现在不论社会经济调查还是自然技术试验都离不开抽样方法。它为人类对客观事物的观察、推理、决策提供一条十分广阔又极具实用价值的途径。

抽样调查技术在我国统计工作中也获得成功的运用，特别是通过农产量调查和城乡住户调查取得丰富的社会经济信息，同时也积累了可贵的经验。我们应该充分认识到，随着我国政治经济体制改革深化，社会经济条件发生了深刻的变化，社会主义市场经济不断发展，对于统计调查也提出了新的需求。一方面，企业转换经营机制，经济愈搞活，对外愈开放，愈需要国家做好政策导向和宏观调控，对于统计信息必然提出更新更高的要求，不论在内容上、时效上、质量上、数量上都需要有一个飞跃。另一方面，市场体系的发育、发展，竞争日趋激烈，市场纷繁复杂，变化万端，而且多种经济成分并存，利益主体多元化，经营分散多样。这些都大大增加了信息采集的难度，也说明仅靠传统的全面报表调查方式已完全不能适应新形势的要求。抽样调查恰好填补这一不足，它既可以收到非全面调查的好处，又可以达到全面调查的效果，在经济上、时效上、可靠性、准确度等方面的优越性是显然的。我国正在建立以经常性的抽样调查为主体，多种方式结合运用的统计调查体系，作为大力推进统计信息技术革命的重要组成部分。我国统计部门领导号召：“今后的统计调查工作必须更加广泛地研究和应用现代抽样调查技术，从根本上改变过分

依赖全面统计报表的状况，逐步缩小其活动范围。”可以预期，抽样调查在我国将进入崭新的阶段。

随着抽样实践的发展，抽样调查、推断、决策理论和方法也成为统计学的中心课题。抽样理论成了统计学中发展最快最为活跃的一分支，它不但大大充实了统计学的内容，甚至推动着哲学认识方法和归纳逻辑的变革。正如哲学家自己所说的那样：“哲学观点不同和归纳逻辑学派首先来源于数理统计上的不同传统，今天的哲学家，无论从事于自然科学的哲学还是社会科学的哲学，都再也不能够忽视数理统计和决策论了。”（参阅江天骥《归纳逻辑导论》，湖南人民出版社）。统计抽样和推断所以具有如此重要的认识方法论上的意义，在于它坚持以概率论为基础的随机原则，探索到表面上似乎偶然性支配一切的社会自然现象，在其定期的、反复的百分比和平均数中可以显示出必然的规律性。随机原则的实质意义在于排除人的主观意识作用，包括排除调查者的有意选择，也包括被调查者的合作态度。这个原则从提出到被人们接受经过了漫长的历史时间，它一开始受到怀疑、嘲弄、拒绝，由于统计学家、数学家的不懈努力，也由于统计实践的检验，才使抽样的随机原则逐步为人们所认识接受，并普遍加以推广。而统计科学也因而飞跃到一个新的阶段。

如果说，统计抽样的随机原则，排除人的主观意识作用，曾经作为统计科学的重大课题耗费科学家们大量精力，那么它的另一命题即统计抽样又必须发挥人的主观能动性，充分利用一切可供利用的辅助信息，则留给人们更大的科学研究空间。从表面上看，统计抽样的这两个原则似乎是相互矛盾无法并存的。实则不然，抽样的随机原则不能理解为盲目性，更不是天然就能够实现的，且不说保证随机原则的实现需要发挥主观能动性去创造这一条件，更何况统计抽样还要追求另一原则，实现最佳效果的原则。它表现为在一定费用条件下抽样误差最小或是在一定误差

要求下抽样费用最省。可以说，所有抽样调查组织的改善都是和人们对调查对象有所了解，充分运用各种辅助信息有关。我们感谢统计学家长期的努力，为改进抽样调查和推断方法，完善抽样组织形式，提高抽样效果提供了大量极有价值的研究成果。但是许多重大课题仍需人们继续努力去探索开发，寻求更好的答案。

吴国培同志长期从事抽样理论与方法的研究，参加国家、省市抽样调查改革的设计，并承担国家抽样科研项目，发表了一系列抽样论文，并出版《应用抽样方法》一书，成果累累。本书（《抽样调查方法研究》）凝集了作者研究成果的精粹，得到国内同行专家的高度评价。这本书的主要特点，一是站在世界科学发展的前沿，继续探讨理论上尚未解决的问题。例如等距抽样和分层抽样的效果比较问题，抽样调查组织形式的改进问题，以及不同组织形式的抽样误差特别是等距抽样的误差估计问题等等。二是密切联系实际，紧紧抓住实践中的新问题开展研究并取得可喜的进展。例如多主题抽样、样本轮换等等都是我国抽样实践亟需解决的问题。三是不满足于理论的是非之争，而是开拓应用问题的研究。例如对我国农产量抽样调查方案和设计问题，抽样方法在普查中的应用问题，我国统计调查体系的改革问题，乃至抽样调查中非抽样误差的分析问题都作了大量的研究。

趁这本书出版的机会，我乐意把它介绍给读者。

黄良文

1993年4月于厦门大学白城

前言

抽样调查是一门科学，也是一门艺术，它有着坚实的理论基础和科学的方法论，被广泛地、灵活多样地运用于各种社会经济的统计调查。

抽样调查理论与方法的建立发展，曾被人们称为世界统计发展史上的一次重大变革。最早的抽样，是19世纪由著名统计学家拉普拉斯(P. S. Laplace)设计和主持的法国人口抽样调查；1895年，在第5届国际统计学会上，挪威的统计学者凯尔(A. N. Kiaer)首先提出了“代表性调查”的卓越见解，为抽样调查奠定了思想理论基础。尔后，许多统计学家，如尼曼(J. D. Neyman)、汉森(M. H. Hansen)、赫维茨(W. N. Hurwitz)、马哈拉诺比斯(P. C. Mahalanobis)和科克伦(W. G. Cochran)等，在社会经济抽样调查的理论、方法和实践方面，作出突出的贡献。

在当今世界各国，无论是进行社会经济基本情况的调查，还是某些一时性专题性的统计调查，抽样方法都是一种十分重要的手段，在纪念国际统计学会成立100周年之时(1985年)，一些著名统计学家撰文讨论和交流，达到如下的共识：广泛应用抽样方法，与应用计算机信息技术，建立国民经济核算体系和扩大统计服务职能这四个方面，构成了当今国际统计的四大发展趋势。

我国在50年代开始了全国性城市职工和乡村农民家庭的家计抽样调查。60年代初，开展了大规模的农产量抽样。80年代以来，我国重新组建了城市和农村二支规模庞大的抽样调查队伍，它

为及时准确地掌握我国城乡社会经济多方面的统计资料，发挥了重要的作用。我国各专业统计部门也正积极地应用和推广抽样技术，比如，工业、商业和运输业等统计中的抽样调查。与此同时，我国抽样理论方法和应用的研究，也在不断深入和提高。可以断言，在我国实行社会主义市场经济下，在建立健全新的国民经济核算体系下，抽样方法将会发挥更大更广泛的作用。

伴随着抽样调查的广泛应用，以及人类社会对抽样方法所提出新的更高的要求，研究和发展抽样理论方法及其应用，是国内外统计学界所关注的热门课题之一。在这种背景下，作者拟写了本书，对抽样理论、方法和应用中的若干重要问题进行了比较全面系统的讨论研究，在学习和吸收国内外有关的新近科研成果的基础上，密切联系我国具体抽样实践，重点研讨了等距抽样、分层抽样、多主题抽样、样本轮换和非抽样误差分析等问题，提出了一些自己的见解和方法。本书也对我国应用抽样方法和统计调查体系改革等问题，进行了专门的分析和研究。

本书共有十章，大体可分为三大部分。第一部分关于等距抽样和分层抽样。这二种抽样是国内外常用的方法。本部分侧重于分析它们的抽样效果和抽样误差与偏差的估计方法问题。第二部分关于抽样调查中的新近课题，主要包括多主题抽样、样本轮换和非抽样误差等问题，这些问题在国内外研究的时间不太长，但却是目前国内外抽样调查工作中急待研讨和解决的问题。第三部分关于应用抽样方法的研究，主要包括我国农产量抽样方案的研究，抽样方法在普查中应用的研究，以及我国统计调查体系改革的研究。

下面列出各章的主要内容和研究问题：

第一章 等距抽样问题研究（一）

本章主要从理论和实践两方面，分析等距抽样的抽样效果。本章进行了等距抽样与纯随机抽样的比较；线性等距抽样与一般等

距抽样的比较;线性趋势总体中对称等距抽样效果分析;不等概抽样与等概抽样的比较,最后进行不等概等距抽样与不等概纯随机抽样的比较。通过这些比较和分析,一方面深化了对等距抽样效果的认识,另一方面也对我国的不等概对称等距抽样的具体实践,进行了论证分析。

第二章 等距抽样问题研究(二)

等距抽样误差的估计问题,尤其是线性趋势总体下等距抽样误差的估计方法,是国内外抽样理论和实践中的一个难点。本章首先讨论了抽样调查中估计等距抽样误差的困难所在,接着讨论以纯随机抽样方法估计等距抽样误差的条件及其关系,以一个样本估计等距抽样误差的估计方法及其局限性,以多个样本估计等距抽样误差的估计方法及其局限性。最后,本章对估计等距抽样误差这一难点问题,经过努力探索,试图提出了一种在线性趋势总体下估计等距抽样误差理论和方法,即以一等距样本,进行交叉子样本估计。本章给出这种估计的方法和模型,并证明这种估计的无偏性,还讨论了这种估计方法在抽样实践中具体应用及其优越性。

第三章 分层抽样问题研究(一)

本章主要分析分层抽样的抽样效果。本章讨论了总体分层与降低抽样误差的关系;进行了分层抽样与纯随机抽样的比较;比例分层抽样与内曼分层抽样的比较;分层抽样与等距抽样的比较;分层抽样与整群抽样的比较;分层抽样与二阶抽样的比较。通过这些比较分析,从理论和实践上论证了分层抽样比其他抽样方法,一般会具有更高的估计精度。

第四章 分层抽样问题研究(二)

本章主要讨论合理分层问题和层权偏差问题。本章讨论了分层标志的选择,分层层数的确定等对分层抽样效果所产生的影响。在考虑多个标志进行总体分层下,可以结合运用统计聚类分析方

法,这里,讨论统计聚类分析方法在分层抽样中的应用问题。在事先难以或无法对总体进行分层的情况下,可以考虑采用抽样后分层技术,这里,讨论了事后分层的估计量、估计效果和估计偏差等问题。接着,本章讨论了层权偏差产生的原因及其所带来的影响和纠正方法。研讨了如何应用二重抽样方法来纠正层权偏差问题。

第五章 多主题抽样问题研究

目前国内外的社会经济抽样调查,一般是多目标多主题的,但如何保证这种抽样取得良好的多主题估计效果,还有待于进一步研究。本章首先讨论了多主题抽样的基本条件——代表性和相关性,并重点研讨了我国农村抽样中,多主题结合到村级的可能性和必要性。接着讨论多主题抽样的样本容量问题,提出了按相关系数和变异系数设计多主题抽样样本容量的方法。最后,研讨和设计了多主题抽样的抽样框编制形式——混合分层形式、混合系数形式和混合主分量形式。

第六章 样本轮换与动态抽样估计

本章研究样本轮换和不同时间抽样估计的问题,这两个问题有着一定内在联系。本章首先分析了我国抽样实践中,实施样本轮换制度的必要性,接着研讨了国内外样本轮换的轮换层次、轮换方案和确定轮换点形式等问题。确定轮换比例是样本轮换中的一个重要问题,本章认为合适的轮换比例主要取决于前后二期轮换总体的相关性,并由此得到计算轮换比例的公式模型。根据这种思想和模型,论证分析了目前我国农村抽样中采用 $1/4$ 轮换比例是合理的。本章还具体设计了我国农村抽样和城乡居民住户抽样的样本轮换流程图。在部分轮换样本条件下,产生了动态抽样估计问题。本部分研讨了动态抽样中,关于总体均值、总体总值、总体均值变动率、总体成数变动率和动态平均成数的估计量及其估计量方差,并给出了动态抽样估计下,确定最佳轮换比例的方法。

第七章 我国农产量抽样方案的问题研究

我国的农产量抽样方案在理论上和实践上都有许多优点，但也存在一些问题。本章分析了该方案所存在的四个主要问题，进而提出对该方案加以改进的思想，并设计了“改进方案”——分层、不等概、二阶段和等距抽样的方案。这个“改进方案”既保持了现行方案的各种优点，又克服了现行方案中的主要缺点，还能进一步提高抽样估计精度。“改进方案”中研讨了关于农产量播种面积和平均产量的抽样调查问题，给出它们的抽样估计方法。最后，讨论了有关“改进方案”的若干问题，主要是设计样本容量问题和满足各级管理部门需要的问题。

第八章 非抽样误差分析

非抽样误差是一种在抽样调查过程中出现的，但又不是抽样方法本身所引起的统计误差。这种误差日益引起国内外统计界的重视。这种误差产生于整个统计调查过程中，从方案设计到实施调查到资料处理的各个环节。本章具体分析了每一环节中可能产生这种误差的各种因素。抽样总误差包括变量误差和偏差。本章具体分析了它们的性质、关系、影响作用以及处置方法。本章还进一步提出变量误差和偏差中都包含非抽样误差因素，并以几何关系式和数理关系式，描述了它们间的关系。对如何减小或避免非抽样误差，本章提出几种检验方法，并着重研讨了无回答误差的抽样估计问题、观测误差的估计问题、回归抽样中的误差估计问题、偏差的区间估计问题。

第九章 抽样方法在普查中的应用

在各种普查中结合应用抽样方法，这是国内外普遍采用的统计调查方式。这种方式对于提高统计调查的时效性，节省人财物力和时间，增加调查信息和提高调查结果准确性等，具有重要作用。本章讨论了普查中应用抽样方法的估计效果问题，调查经费开支问题和抽样框设计问题。进而，研讨了抽样方法在普查中的几种主要应用——普查方法和内容的检验，普查资料结果的检验，

普查资料的超前汇总，以及普查中的专题抽样调查。目录抽样是一种新近发展起来的全面调查与抽样调查相结合的形式，它对于偏态且内部差异大的总体，具有比其他抽样方法更好的估计效果。本章研讨了目录抽样的理论、方法、模型及其具体应用，尤其是在我国工业企业调查中的具体应用。

第十章 关于我国统计调查体系改革的目标模式

本章首先讨论了我国现有统计调查体系所存在的各种弊端，接着提出了建立“以全面报表和抽样调查为主，其他调查方式为辅”的新的统计调查体系的构想，认为这种新体系能基本适应我国社会主义市场经济管理和统计现代化的需要，并从理论和实践方面论证了建立这种新体系的必要性和可行性，也指出了建立这种新体系所需考虑的若干重要问题。本章最后，介绍了作者与厦门市统计局合作，进行厦门市工业企业统计调查体系改革的尝试。在这改革尝试中，突出的特点是，将全面报表与抽样调查相结合使用。这个尝试取得良好的成效。

本书的写作，得到厦门大学计划统计系各位师长和同事的全力帮助，尤其是博士生导师黄良文教授，他给予作者长期精心严格的指教和栽培，为作者在科研方面所取得的进步，付出了巨大的劳动。著名统计学者刘宗鹤教授，给予作者许多帮助和鼓励，并对此书稿给予充分的肯定。福建省统计局农调队和城调队，厦门市统计局的有关同志，为本书的资料准备给予了积极配合和大力支持。在此，一并表达作者对他们的衷心感谢。

本书是作者在教学科研和社会实践中不断积累和研究的基础上编著而成的。限于作者的水平，书中难免有不足之处，恳请各位同仁和读者给予批评指出。

作者 吴国培

1993年4月于厦大白城

目 录

(30)	 样本量估计与样本量计算	—
(32)	 样本量估计与样本量计算	—
(33)	 样本量估计与样本量计算	二
(35)	 样本量估计与样本量计算	三
(37)	 样本量估计与样本量计算	四
(38)	 样本量估计与样本量计算	五
序	 样本量估计与样本量计算	—
前言	 样本量估计与样本量计算	—
第一章 等距抽样问题研究 (一)	 样本量估计与样本量计算	—
(38)	——我国有关标志排队的等距抽样效果分析.....	(1)
(37)	一、等距抽样与纯随机抽样的比较.....	(2)
(30)	二、对称等距抽样与一般等距抽样的比较.....	(7)
(34)	三、不等概抽样与等概抽样的比较.....	(11)
(100)	四、不等概等距抽样与不等概纯随机抽样的比较.....	(14)
第二章 等距抽样问题研究 (二)	 样本量估计与样本量计算	—
(107)	——等距抽样的误差估计问题.....	(17)
(90)	一、测算等距抽样误差的困难所在.....	(17)
(111)	二、由纯随机抽样方法估计等距抽样误差.....	(21)
(31)	三、由单个样本估计等距抽样误差.....	(25)
(16)	四、由多个样本估计等距抽样误差.....	(28)
(10)	五、估计等距抽样误差的进一步讨论.....	(34)
第三章 分层抽样问题研究 (一)	 样本量估计与样本量计算	—
(133)	——分层抽样的估计效果分析.....	(38)
(132)	一、分层抽样的估计效果.....	(38)
(130)	二、分层抽样与纯随机抽样、等距抽样的比较.....	(40)
(132)	三、分层抽样与整群抽样的比较.....	(44)
(134)	四、分层抽样与二阶段抽样的比较.....	(46)
第四章 分层抽样问题研究 (二)	 样本量估计与样本量计算	—

——合理分层及层权偏差分析	(50)
一、分层标志的选择	(50)
二、总体分层与统计聚类分析	(53)
三、事后分层及其估计效果	(55)
四、层权偏差的影响与纠正	(61)
第五章 多主题抽样问题研究	(68)
一、多主题抽样的基本条件	(69)
二、多主题抽样的样本容量	(75)
三、多主题抽样框的排队问题	(81)
第六章 样本轮换与动态抽样估计	(86)
一、轮换样本的研究和实践	(87)
二、样本的轮换比例	(90)
三、轮换样本的流程设计	(94)
四、动态抽样估计问题	(100)
第七章 我国农产量抽样方案问题研究	(106)
一、“改进方案”的基本思想	(107)
二、播种面积的抽样估计	(109)
三、平均亩产的抽样估计	(111)
四、“改进方案”的若干问题	(113)
附录 (一)	(116)
附录 (二)	(116)
第八章 非抽样误差分析	(122)
一、非抽样误差的性质与原因	(123)
二、非抽样误差与总误差关系	(125)
三、无回答误差的抽样估计	(130)
四、观测误差的抽样估计	(132)
五、误差的其他估计问题	(134)
第九章 抽样方法在普查中的应用	(137)

一、抽样方法在普查中应用的特点·····	(138)
二、抽样方法在普查中的几种应用·····	(143)
三、目录抽样方法及其应用·····	(150)
第十章 关于我国统计调查体系改革的目标模式·····	(156)
一、改革的必要性·····	(156)
二、新统计调查体系的构想·····	(159)
三、新统计调查体系的可行性·····	(163)
四、建立新统计调查体系的若干问题·····	(165)
五、厦门市工业统计调查体系改革的尝试·····	(167)
参考文献·····	(176)

第一章 等距抽样问题研究（一）

——我国有关标志排队的等距抽样效果分析

等距抽样是我国社会经济统计调查中应用最多的抽样方法之一。早在 50 年代中期，我国在城市职工和农村农户家计调查中，就开始使用这种抽样方法。当时，这种抽样方法称为机械抽样。60 年代初，在引进吸收国外先进的抽样技术和经验的基础上，我国对农产量的抽样调查，作出进一步改进和提高。1963 年国家统计局在《农产量抽样调查方案》中，规定使用“有关标志排队，累计辅助变量，多阶段，中点式的等距抽样”。这种抽样组织形式，构成我国社会经济抽样调查中的最基本最主要的模式。它广泛应用于我国的农业、交通运输、人口和城乡居民住户等方面的调查。

80 年代，我国的等距抽样方法在理论上、方法上和实践应用上，又有着新的改进和提高，这突出表现在农村社会经济的抽样调查方面。这种抽样调查组织形式改进为“有关标志排队，累计辅助变量，多阶段，随机起点的对称等距抽样。”这种抽样方法，在调查和推断农作物产量、农民住户等农村社会经济主要统计数字资料方面，取得重大成效，也引起了国际统计学界的极大兴趣和关注（见第 47 届国际统计学讨论会论文集，巴黎，1989 年）。这种抽样方法有下列三个主要特征：第一，有关标志排队。这是在编制总体抽样框时，采用某一有关标志，如粮食平均产量或农户人均收入，将总体各单位按它们标志值的大小，进行顺序排队。这种排队序列所构成的总体，实质是一种具有线性趋势的总体。第二，累计辅助变量。在农产量抽样调查中，这种辅助变量是取粮

食播种面积或农户有收入的人口数。通过累计辅助变量，并在这个序列中抽取样本单位，可以保证辅助变量越大的单位，被抽中的概率也越大，这实质是贯彻了不等概抽样的思想。第三，对称等距抽样。这是在有关标志排队总体中确定样本单位的科学方法。

下面，要对我国这种等距抽样方法的科学性、合理性和优越性，加以论证分析。具体地说，要讨论在有关标志排队的线性趋势总体下，等距抽样是否优于纯随机抽样？对称等距抽样是否优于一般等距抽样？不等概抽样是否优于等概抽样？以及不等概等距抽样是否优于不等概纯随机抽样？

一、等距抽样与纯随机抽样的比较

比较各种抽样方法的估计效果，其主要途径之一是比较它们的抽样误差大小。显然，抽样误差愈小者，其抽样效果愈好；而抽样误差较大者，其抽样效果较差。就等距抽样与纯随机抽样的效果比较，我们难于从一般意义上断言何者较好，而应根据现象总体的具体情况，分析它们的抽样误差及其估计效果。

一般地，等距抽样的方差形式，可表示为：

$$V(\bar{X}_{sy}) = \frac{N-1}{N} \sigma^2 - \frac{K(n-1)}{N} \sigma_{wy}^2 \quad (1.1)$$

式中， σ^2 为总体方差， σ_{wy}^2 为各等距样本方差的平均值，也可称为等距样本的平均方差。N 为总体单位数，K 为总体中等距样本的个数。

另一方面，纯随机抽样方差，可表示为（按不重复抽样）：

$$V(\bar{X}_{rn}) = \frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right) \quad (1.2)$$

式中，n 为样本容量。N 和 σ^2 分别为总体单位数和方差。

在下面的分析中，可以看出，等距抽样的误差是小于还是大于纯随机抽样误差，主要视现象总体的方差与等距样本的平均方