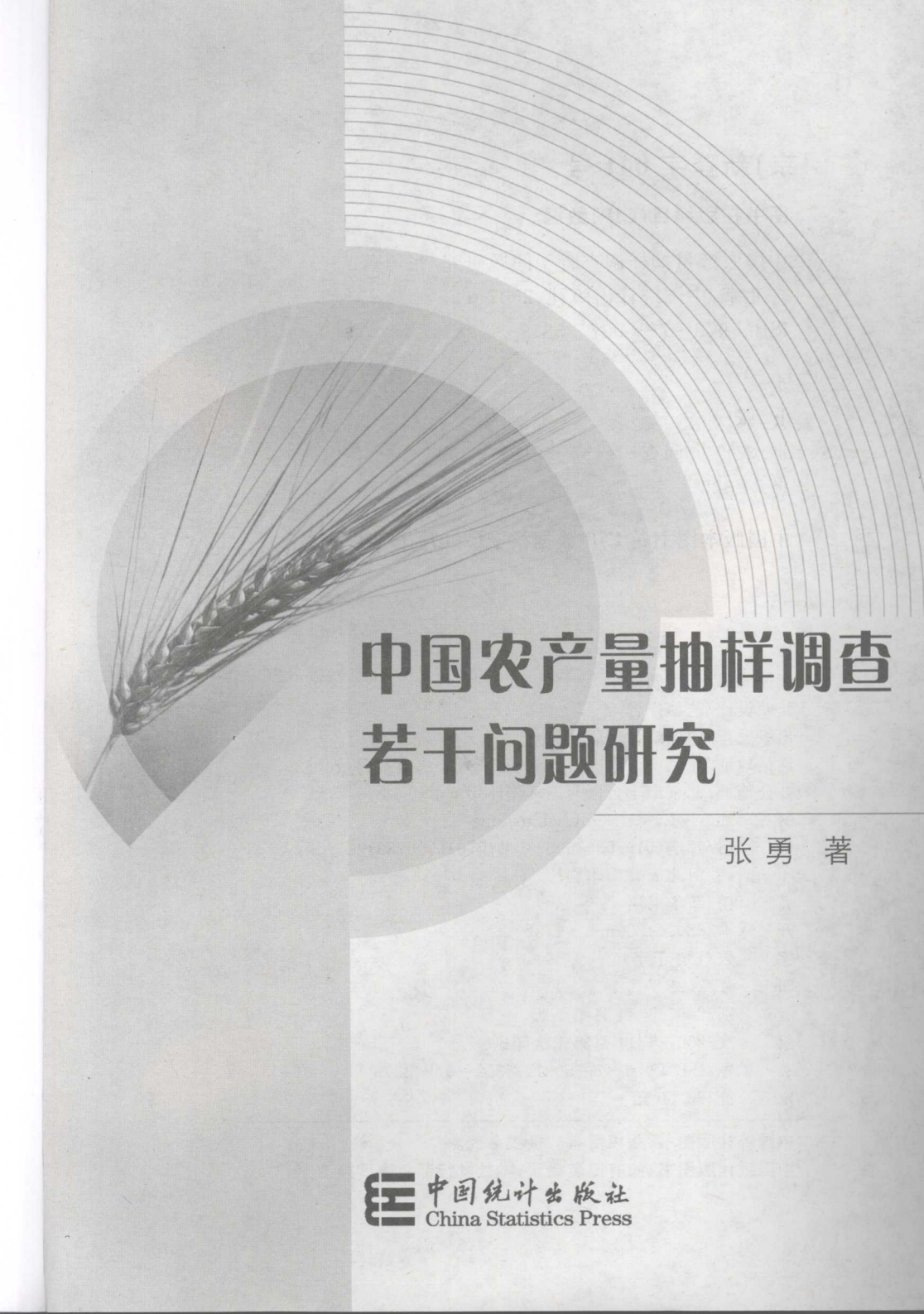


The cover features a large, stylized graphic on the left side. It consists of a wheat stalk with a dense head of grain, positioned diagonally. The stalk is set against a background of overlapping circles in shades of pink, purple, and blue. To the right of these circles, there are several thin, concentric, semi-circular lines that curve upwards and to the right, creating a sense of motion or a stylized horizon. The overall color palette is a mix of cool blues and purples, and warm pinks and yellows.

中国农产量抽样调查 若干问题研究

张勇 著

 中国统计出版社
China Statistics Press



中国农产量抽样调查 若干问题研究

张 勇 著

 中国统计出版社
China Statistics Press

(京)新登字 041 号

图书在版编目(CIP)数据

中国农产量抽样调查若干问题研究/张勇著.

—北京:中国统计出版社,2007.11

ISBN 978—7—5037—5319—0

I. 中…

II. 张…

III. 农产量调查—研究—中国

IV. F326

中国版本图书馆 CIP 数据核字(2007)第 170577 号

作 者/张 勇

责任编辑/陈悟朝

装帧设计/艺编广告

出版发行/中国统计出版社

通信地址/北京市西城区月坛南街 57 号 邮政编码/100826

办公地址/北京市丰台区西三环南路甲 6 号

网 址/www.stats.gov.cn/tjshujia

电 话/邮购(010)63376907 书店(010)68783172

印 刷/河北天普润印刷厂

经 销/新华书店

开 本/880×1230mm 1/32

字 数/180 千字

印 张/7.25

版 别/2007 年 11 月第 1 版

版 次/2007 年 11 月第 1 次印刷

书 号/ISBN 978—7—5037—5319—0/F.2575

定 价/20.00 元

中国统计版图书,版权所有。侵权必究。

中国统计版图书,如有印装错误,本社发行部负责调换。

前言

本书内容主要来自于本人的博士论文,是专门研究的成果,阅读对象是熟悉抽样调查理论、方法和应用的专家和学者。为了保持论文的本来面目,本次出版没有对论文的整体结构和内容作大的变动,只对少数文字和数字进行了调整。作者曾将论文的部分内容,在有关杂志上发表和在研讨会上进行了交流。在交流中,相关人士对这部分内容表现了极大的关注,并向作者索要相关材料。在此次出版专著之际,考虑到满足更多读者的需要,特别增加了两个附录。附录1主要是对农业抽样调查一些基本的概念和应用进行介绍,并以韩国相关的农产量调查作为实例。附录2主要是对抽样设计的基本知识进行介绍。这样,有利于读者进一步了解抽样调查的理论、方法和应用背景,从而更好地在实际中使用抽样调查方法。当然,它的介绍方法与一般教科书介绍的方法不同。它是针对相关的农业抽样调查的基本概念、方法和应用进行的介绍,是可以直接服务于正文的。读者如果要了解或掌握抽样调查的一般理论和方法,还应该通过专门的教科书或专著来学习和研究。读者如果先阅读附录,再阅读正文,则更容易理解正文。

论文是在本人的导师金勇进教授的认真指导下完成的。在论文的选题、框架结构、内容安排以及论文文字的使用等方面,金老师都提出了很多具体、宝贵的意见。在此特表示衷心的感谢。金老师严谨的治学态度,渊博的知识,诚恳待人的优秀品德,都是本人终身学习的榜样。本人永远忘不了他对本人在学习、工作等方面的严格

要求,也忘不了他对本人在学习、工作和生活等方面的热情鼓励和帮助。在中国人民大学学习的这三年,本人经历了角色转变、学习、工作和生活等方面的巨大压力。没有金勇进等老师与同学们的鼓励和帮助,本人能够坚持学习下来是件很困难的事。

还要衷心地感谢中国科学院数学与系统科学研究院的冯士雍研究员。多年来,冯老师对本人的工作和学习一直给予了很大帮助。在本人遇到问题向他请教时,他总是热情、认真地进行指导,并对本人的一些想法给予鼓励和肯定。与冯老师在一起的时光,是愉快和有收获的时光。

感谢中国人民大学统计学系的袁卫教授、易丹辉教授、赵彦云教授、顾岚教授、何晓群教授、吴喜之教授、高敏雪教授、彭非教授、王晓军教授、杜子芳教授等,他们对本人的学业和知识获得都给予过鼓励和帮助。感谢统计学系的刘爱萍老师、林秋池老师、阎华老师、王汀老师、于军老师、陈虹老师、刘畅老师、张云老师等,他们在本人的学习过程中都曾给予过鼓励和帮助。本人的同学也给予了很多帮助。在此,向所有提供过帮助的老师 and 同学表示感谢。

感谢国家统计局农村社会经济调查总队的领导和有关同志们,他们对于本人完成博士论文给予了充分的鼓励,并在材料、数据等方面给予了很大帮助。特别是曾玉平和周巍二位同志,他们与本人一同研究相关的问题,并整理了许多材料、处理了大量的数据。感谢国家统计局统计教育中心的领导和有关同志们,他们对本人的工作、生活和学习也非常关心,并在不同方面减轻本人的压力,为本人坚持学习提供了条件。

最后,本人还要感谢自己的家人和朋友,他们对本人的学习给予了支持和帮助。

作 者

2007 年 10 月



内容提要

本书采用综合解决问题的研究方法,借鉴国内外的先进经验,结合中国的国情,采用新的思路,对中国农产量调查进行整体抽样设计改革的研究。在对抽样理论和方法进行研究的基础上,尝试提出新的理论和方法,解决具体的实际问题,使中国农产量抽样调查更好地适应实际的发展。全书分为6章,后面还有两个附录。

第1章是引论,主要是对研究背景的介绍。

这里分析了选题的理论和实践意义。掌握主要农产品产量的数据,是政府管理国家的需要,也是许多企业发展必须考虑的问题,获得准确、及时的农产品产量的数据,对中国的经济发展和社会稳定具有重要的意义。本项研究能够找出现行抽样设计中的主要问题,并且对存在的主要问题探寻出解决方法,为实际部门进一步认识调查工作中存在的问题提供依据,并在理论和方法上提供解决实际问题的参考和帮助。因此,这项研究对抽样的理论和实际具有重要意义。

这里还介绍了本书的主要特色和研究中要突破的难题。本书的特点是理论联系实际,尽量用简单适用的方法来解决复杂的实际问题。

第2章是对中国农产量调查抽样设计中需要讨论的问题进行的研究,主要研究了两个问题。一个是对中国农产量调查样本抽选方式的研究,另一个是对中国农产量调查对称等距抽样方法的研究。

在对中国农产量调查样本抽选方式的研究中,笔者对部分专家提出的现行抽样方式是多相抽样,提出了疑问。通过分析,笔者认为现行抽样方式不能当作多相抽样,而应该还其本来面目,就是多阶抽样。笔者对有关专家的用意表示理解,但是如果错误地理解样本抽选方式,就会错误地组织样本抽选工作,并使用不正确的估计公式来处理数据,因此会对调查结果带来负面的影响。

在对中国农产量调查对称等距抽样方法的研究中,笔者指出:现行的对称等距抽样方法不是一般意义上的对称等距抽样方法,而是一种 PPS 抽样方法。相关管理部门对这种抽样方法,使用简单随机抽样的估计,对指标就可能会出现有偏估计的问题,也不能得到合适的精度。

第 3 章是对中国农产量调查可行抽样设计进行的研究。

根据中国的国情,书中提出了农产量调查中平均亩产、总产量、播种面积调查的几种可行抽样设计。可行的抽样设计以 PPS 系统抽样为基础,辅助变量可分别选为户数、播种面积、耕地面积、切块数。从稳定、准确、及时、方便等性质上考虑,辅助变量以切块数为最优,耕地面积次之,播种面积再次之,户数为最差的。这里提出的可行的抽样设计,原理和估计方法都不复杂。因此,便于理解和实施,也便于解决分级管理需要的问题。

第 4 章是对中国农产量多目标调查及相关问题进行的研究。

这里重点研究了解决多目标调查问题的方法,就是使用比估计和 MPPS 抽样设计。在笔者提出的可行的抽样设计中,使用的都是 PPS 系统抽样。对 PPS 系统抽样设计的比估计,在国内外还少见系统的讨论。此外,MPPS 抽样设计的本质是比估计和 PPS 抽样设计的结合,也需要 PPS 系统抽样设计的比估计知识。这里对 PPS 系统抽样的比估计进行了系统的研究。

MPPS 抽样设计是上世纪末才出现的新方法,中国相关管理部门在广东省进行了这种方法的试点,以解决多目标调查的问题。这种方法本身还在不断完善之中,笔者通过研究,对估计方差的计算提出了改进的想法。在美国专家的帮助下,搞清了一些理论和方法

问题。笔者使用第一次农业普查的数据对这种方法进行了模拟,得到的结果很令人满意。

第5章是对中国农产量连续性调查样本问题进行的研究。这里研究了国家抽样调查县的代表性问题,经常性的样本轮换的问题和区域框中面积样本的获取。

中国从1984年确定国家抽样调查县至今,已经20多年了,除了行政区域调整外,没有变动国家调查县。人们对国家调查县的代表性提出了疑问,是可以理解的。本书使用实际中常用的系数调整法对此问题进行了研究。

中国农产量抽样调查是经常性定期进行的。但是,如果长期固定样本,也会出现一些问题。比如,当调查对象是人、家庭或单位时,如果样本使用时间过长,就会引起调查对象的厌倦情绪。他们不愿意或不积极配合,从而使调查质量下降。这就是样本老化的问题。为解决样本老化问题,本书结合提出的可行抽样设计,提出采用复合样本的方法。这种方法简单易行,具有可操作性。

中国一直使用的是名录框抽样设计,因为农业的许多指标都与地域有着紧密的相关关系,所以对农业调查采用区域框设计有很多有利条件。这里对中国如何获取面积样本进行了初步研究。笔者认为,中国农业调查应该借鉴外国的先进经验、根据中国的国情进行科学合理的抽样设计,改名录框抽样设计为多样框抽样设计,以提高农产量抽样调查的效率。

第6章是对全书内容进行的总结。

这里首先总结了本书主要的研究内容和解决的问题,并指出了需要进一步研究的一些问题。然后对农产量调查设计作了一些说明。指出,抽样设计只是调查设计的一部分,调查者还要重视调查设计的其他部分,才能保证调查的质量。

书后的两个附录是新增加的。附录1对农业抽样调查一些基本的概念和应用进行介绍,并以韩国相关的农产量调查作为实例。附录2对抽样设计的基本知识进行介绍。

目 录

第 1 章 引论 1

第 1 节 选题的依据 2

1.1.1 选题的理论和实践意义 2

1.1.2 国内外研究现状 3

第 2 节 研究的整体框架 13

1.2.1 主要研究内容和研究目标 13

1.2.2 研究中要突破的难题 14

1.2.3 特色及创新点 16

1.2.4 研究路线 17

第 2 章 中国农产量调查抽样设计中需要讨论的问题研究 23

第 1 节 中国农产量调查样本抽选方式的研究 24

2.1.1 中国农产量调查的抽样设计 24

2.1.2 关于现行抽样方式为多相抽样的讨论 31

2.1.3 样本抽选方式定位解决的问题 39

第 2 节 中国农产量调查对称等距抽样方法的研究 42

2.2.1 中国农产量调查对称等距抽样方法 42

2.2.2 中国农产量调查对称等距抽样方法的实质 45

2.2.3	中国农产量调查对称等距抽样方法的模拟和分析	47
2.2.4	小结	58
第3章	中国农产量调查可行抽样设计的研究	59
第1节	现行播种面积调查方法	60
3.1.1	调查户的抽选	60
3.1.2	估计方法	60
第2节	现行播种面积调查估计中的主要问题	64
3.2.1	利用乡村户数作推算权数的问题	64
3.2.2	其他原因的分析	66
第3节	几种可行抽样设计的研究	69
3.3.1	以户数为辅助变量的三阶抽样	69
3.3.2	以播种面积为辅助变量的三阶抽样	73
3.3.3	以耕地面积为辅助变量的三阶抽样	75
3.3.4	以切块数为辅助变量的三阶抽样	77
3.3.5	以户数为辅助变量的二阶抽样	81
3.3.6	以切块数为辅助变量的二阶抽样	83
第4节	几种可行设计的数据模拟和比较分析	85
3.4.1	省直接抽村的数据模拟和结果比较	85
3.4.2	分阶段抽样的数据模拟和结果比较	88
3.4.3	关于几种可行抽样设计的总结	91
第4章	中国农产量多目标调查及相关问题的研究	95
第1节	关于PPS系统抽样设计比估计的研究	96
4.1.1	简单随机抽样的比估计	96
4.1.2	PPS系统抽样设计的比估计	98
4.1.3	比估计的数据模拟和分析	102

第2节 关于 MPPS 抽样设计的研究	108
4.2.1 MPPS 抽样设计的研究背景和思路	108
4.2.2 MPPS 抽样设计	111
4.2.3 关于 MPPS 抽样设计的其他讨论	113
4.2.4 关于 MPPS 抽样设计估计的研究	116
4.2.5 关于几种抽样设计估计的总结	119
4.2.6 MPPS 抽样设计的数据模拟及分析	122
第3节 中国农产量多目标调查的开展	130
4.3.1 中国农产量多目标调查的形势	130
4.3.2 中国农产量多目标调查的改进	131
第5章 中国农产量连续性调查样本问题的研究	134
第1节 国家抽样调查县的代表性问题的研究	135
5.1.1 广东省抽样调查县的代表性问题的研究	135
5.1.2 进一步的研究	136
第2节 经常性的样本轮换问题的研究	144
5.2.1 中国农产量调查样本轮换的分析	144
5.2.2 复合样本抽选方法	146
第3节 面积样本获取的研究	149
5.3.1 相关管理部门的准备和打算	149
5.3.2 分层和切块的划分	151
5.3.3 中国农产量调查样本讨论的总结	159
第6章 总结	160
第1节 研究内容总结	161
6.1.1 本书的主要研究内容	161
6.1.2 需要进一步研究的问题	162
第2节 关于农产量调查设计的说明	163

6.2.1	关于抽样设计方面的说明	163
6.2.2	调查设计的其他方面的说明	164
附录 1	农业抽样调查的基本概念及应用实例	165
附录 2	抽样设计的基本知识	189
参考文献	215	
后记	218	

第1章

引 论

从新中国成立至今,中国开展农产量抽样调查已经有 50 年的历史了。国内关于农产量调查有关问题的讨论一直没有停止,特别是在抽样设计方面还存在着不同的认识。国内外很多专家都对中
国农产量抽样调查提出了要求和建议,有关调查部门也在不断地探讨和解决调查实施过程中抽样设计存在的问题。实际的要求和探讨,表现出来的形式是对指标的质疑,实质上是要求对抽样理论、方法和应用的完善。因此,这项研究对抽样调查理论、方法和实际工作的发展都具有参考和促进作用。



第1节 选题的依据

1.1.1 选题的理论和实践意义

中国在经历了 20 多年的改革开放以后,经济发展正在进入一个新的阶段。与改革开放前的计划经济相比,国家的整个管理和发展的理念与方法都在发生巨大的变化。中国是一个农业大国,农业人口占全国人口的比重仍很大,全国人民的吃饭问题不是一个小问题。此外,农产品还是许多其他产业的重要原材料。农产品产量一直是党和政府关注的重点,它不仅关系到国家的安全和稳定,而且还关系到农业乃至整个国家的可持续发展。掌握主要农产品产量的数据,是政府管理国家的需要,也是许多企业发展必须考虑的问题。正确决策的做出,需要建立在数据准确、及时的基础上。因此,获得准确、及时的农产品产量的数据,对中国经济的可持续发展和社会稳定具有重要的意义。

中国负责提供全国农产量数据的部门是国家统计局农村社会经济调查总队(现改为农村社会经济调查司),本书称其为相关管理部门。为了向决策者提供准确、及时的农产量数据,相关管理部门做了大量的工作。但由于种种原因,部分研究者、决策者乃至普通公民对中国农产量调查数据的质量存在疑问,认为存在数据不准的问题。解除人们的疑问,解决数据不准的问题,需要人、财、物的投入,其中进行科学的调查是必不可少的因素。在科学的调查中,我们要借鉴外国的先进经验,结合中国的国情。

农业调查的基本类型可分为两类,即农业普查和农业抽样调查。抽样设计是对抽样调查来进行的。抽样调查又分为概率抽样调查和非概率抽样调查。我们主要讨论概率抽样调查,因为国际上大型调查的事实已经证明:一个国家农业部门及时的、可靠的全国统计数据只有通过建立恰当的、周期性的以概率抽样方法为基础

的全国农业调查来提供。经常性的农业调查是建立在概率抽样和估计方法基础之上的,它是能够提供一个国家有关农业部门的及时、可靠的基础数据的唯一方法。这种调查还可以计算估计的统计精度。见联合国粮农组织(2000年)的《多样框农业调查》。中国农产量调查就是采用概率抽样调查来进行的。

实际中存在的主要理论和方法问题有:由于对基本抽样方式认识的不同,可能导致抽样的组织实施不同和估计的误用,从而影响调查数据的准确性和精度;现行对称等距抽样设计不是一种规范的抽样设计,因此可能造成对农产量数据的高估或低估,并且不能得到估计的精度。中国农产量调查抽样设计现在采用的是名录框对称等距抽样方法。这种方法在实施中存在一些理论和实际的问题,已经很难适应实际需要,因此需要用新的方法和思路来进行改革。需要解决的主要问题有:在目前情况下,中国有哪些抽样设计是可行的,这些抽样设计的结果如何;中国农产量多目标调查是调查中必须面对的一个问题,如何解决多目标调查中存在的问题;中国农产量调查是一种连续性调查,在这种调查中如何解决样本代表性的问题等。通过对这些问题的研究,我们能够找出现行抽样设计中的主要问题,并且对存在的主要问题探寻出解决方法,为实际部门进一步认识调查工作中存在的问题提供依据,并在理论和方法上提供解决实际问题的参考和帮助。因此,这项研究具有重要的理论和实际意义。

1.1.2 国内外研究现状

中国农产量调查的发展历史也是研究的历史。新中国建国以后,中国的农产量抽样调查就没有停止过。因此,研究也就没有停止过。但对现在仍有重大影响的,可以从1963年的农产量抽样调查算起。外国对中国农产量调查的问题的研究,多是针对某些具体实际问题开展的。

1. 1963年的农产量抽样调查方案

这个方案中的一些设计方法对现在仍有相当大的影响,其特点



如下。

(1)比较合理地划分了抽样阶段、确定了抽样比率。

《方案》规定:对主要农作物的播种面积和产量,采取四阶抽样的方式。四阶为省抽县、县抽大队、大队抽地块和地块抽样本点。每省抽本省总县数的 $1/6$ 为调查县。每调查县抽本县大队总数的 $2\% \sim 3\%$ 。大队抽地块数,因地区而异:平原和地块较集中、面积较大的地区,抽的地块数为 $7 \sim 15$ 块;山区和地块分散、面积较小的地区,抽的地块数为 $20 \sim 30$ 块。地块抽样本点,要求在 7 个以上,地块面积较大者,相应增加样本点数。

中国现行抽样设计基本上是 1984 年实施的抽样设计。在 1984 年的抽样设计中,也是采用多阶抽样设计,只是对县以下抽样阶段的单位进行了调整。而在 2000 年的样本轮换实施后,相关管理部门将多阶抽样设计解释成多相抽样设计,这种解释造成相关指标及其精度估计的变化。这种解释是否成立?如果成立,估计的变化将有利于提高调查的质量;如果不成立,估计的变化则会给调查质量造成负面的影响。这是一个需要研究的问题。

本设计方案中样本量的确定,不是按误差要求计算得出来的,而是根据经验和行政管理的方便。这也一直影响至今。

(2)确定了按有关标志排队、半距起点、等距抽样的方式。

《方案》要求:地块以上的各阶段样本单位的抽选,采取以总体各单位的预计平均单产由低到高顺序排队,以种植作物的播种面积作为辅助资料,进行累计。在累计播种面积中计算组距,以第一个组距半数为起点,等距抽取样本单位。为增加样本单位对总体的代表性,分季作物用各季预计单产进行排队,并以相应的播种面积作为辅助资料,编制抽样框。

设计中这种等距抽样的概念不同于国内外通常等距抽样的概念。通常等距抽样的概念是:等概率的系统抽样是等距抽样,见冯士雍等(1998年)的讨论。系统抽样也称机械抽样,是将总体中的单位按某种顺序排列,在规定的范围内随机抽取起始单位,然后按一套规则确定其他单位的一种抽样方法。等距抽样的特点是,等距是

相对总体单位的排列位置而言的,总体的每个单位被抽中的概率都是相同的。而方案中的等距是相对于总体单位的辅助指标播种面积而言的,总体的每个单位被抽中的概率是不同的。这实际上是一种 PPS 抽样。抽样设计包括抽取样本和利用样本对总体进行估计两个方面。抽取样本的方法,决定了估计的方法。方案中的估计针对当时的调查内容粮食产量,采用简单随机抽样的估计公式是不会出现太大的问题的。但对其他指标的估计,采用简单随机抽样的估计公式就可能出现問題。中国现行农产量调查的指标在不断增加,需要解决这个问题。但是,至今还没见过对这个问题的研究。

2. 1982 年的《农产量抽样调查试行方案》

1982 年国家统计局和农牧渔业部联合制发了《农产量抽样调查试行方案》。与 1963 年的《方案》比较,此方案继承了成功的作法,并进行了重大改进,主要体现在以下几个方面。

(1) 改原来的四阶抽样为五阶抽样。

1963 年方案中,采用的是省抽县、县抽大队、大队抽地块和地块抽点的四阶抽样,1982 年《方案》改为省抽县、县抽公社、公社抽生产队、生产队抽地块和地块抽样本点 5 个阶段。增加公社这一阶,虽然要扩大一些抽样误差,但一是能有效满足县、公社等基层领导对农产量数据的需求,便于调查工作的组织;二是改抽生产大队为生产队,范围缩小,调查工作量减小。

在当时的计划经济体制下,各级政府都要求获得农产量数据,对调查的效益并没有要求。而在中国市场经济发展的今天,如果要满足各级政府对农产量数据的需求,则必须考虑调查的效益。如何满足和在哪一级满足政府管理经济的需要,是一个值得研究和正在研究的问题。目前的研究主要集中于如何满足国家抽样调查县政府对农产量数据的需求。

(2) 重新安排了各阶段的样本量。

1982 年《方案》规定:省抽县,样本县应占总体的 20%,县数较多或较少的省、区、市可按此比例适当调整,但各省、区、市的调查县最少不得少于 7 个。县抽公社,样本公社应占总体的 30%,最少不