


2005
农村统计制度
和方法研究

国家统计局农村社会经济调查司 / 编

Research on
Rural Statistical
System
and Methods


 中国统计出版社
China Statistics Press

2005 农村统计制度 和方法研究

国家统计局农村社会经济调查司 / 编

Research on
Rural Statistical
System
and Methods

江苏工业学院图书馆
藏书章

 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

农村统计制度和 method 研究. 2005 / 国家统计局农村社会经济调查司编.

—北京: 中国统计出版社, 2005. 8

ISBN 7-5037-4768-4

I. 农...

II. 国...

III. 农业统计-研究-中国

IV. F322

中国版本图书馆 CIP 数据核字(2005)第 077493 号

农村统计制度和 method 研究 - 2005

作 者 / 国家统计局农村社会经济调查司

责任编辑 / 姚立

封面设计 / 智道工作室·黄俊杰

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 75 号 邮政编码 / 100826

办公地址 / 北京市丰台区西三环南路甲 6 号

电 话 / (010) 63459084 63266600-22500 (发行部)

印 刷 / 河北天普润印刷厂

经 销 / 新华书店

开 本 / 787 × 1092mm 1/18

字 数 / 450 千字

印 张 / 27.5

版 别 / 2005 年 8 月第 1 版

版 次 / 2005 年 8 月第 1 次印刷

书 号 / ISBN 7-5037-4768-4/F·2145

定 价 / 58.00 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。

《农村统计制度和研究方法研究 - 2005》

编辑委员会

顾 问：李德水 邱晓华
主 任：鲜祖德
副 主 任：张淑英 曾玉平 盛来运 徐志全
编 委：（以姓氏笔画为序）
王明华 王萍萍 毛 峰 李仁元
李永强 阳俊雄 陈培成 赵建华
郝安民 黄秉信 郑云阳 侯 锐
编辑人员：黄秉信 周 巍 钱春林 白 康
蔡 旭 徐海霞

编者的话

为贯彻全国统计工作会议和全国农村统计工作会议会议精神,不断改进和完善现行农村统计制度和办法,深化农村统计制度方法改革,进一步提高农村统计服务水平,国家统计局农调总队2004年在全国农调系统和省(区、市)统计局农村处组织开展了农村统计调查制度方法的招标课题研究。参与中标课题研究的既有农村统计工作者,部分课题还有从事农村统计研究的专家学者共同参与。课题研究大多结合该地区实际,有些课题研究还着眼于全国范围带有共性的特点,在农村统计制度方法的诸多方面,就理论与实际应用的现状、成果、问题作了深层次的探讨,有难度、有新意、有前瞻性,对于推动农村统计方法制度改革,提高农村统计调查数据质量和效率有重要参考价值。现将研究成果汇编成《农村统计制度和办法研究(2004-2005)》一书,供广大农村统计工作者、有关专家、学者交流与参考。

全书共分成三个部分,第一部分是统计调查制度方法研究,包括农产品价格、第二次农业

普查方案、农村住户、农产品成本、农村抽样调查方法研究等；第二部分是统计调查及分析研究方法研究，包括县域统计指标体系、农业经济统计核算、农村可持续发展统计、小城镇综合实力评价、农业发展速度研究等；第三部分是统计调查推断方法研究，主题是农村抽样调查。

本书的编辑主要是对研究报告的文字和公式进行了订正、核实，基本保持了研究报告作者的自身观点和看法。因此报告中的有些观点并非代表编者，请读者在阅读时注意。由于时间仓促、报告内容的专业性比较强，难免出现一些差错，望读者见谅。

编者

2005年6月

目 录

■ 第一部分 统计调查制度方法研究

- 多主题抽样设计在农产量调查中的应用 国家统计局农调总队 3
- 完善农产品生产价格调查体系的研究与探索 浙江省农调队 15
- 完善农产品价格调查体系研究 重庆市农调队 24
- 对第二次全国农业普查方案设计的思考 江西省农调队 36
- 农村入户调查操作规程研究 江苏省农调队 49
- 农村入户调查操作规程研究 内蒙古自治区农调队 92
- 关于农产品成本调查方案的研究 黑龙江省农调队 100
- 农产品生产成本调查方法研究 广西壮族自治区农调队 128
- 农产品成本调查方法研究 湖南省农调队 143
- 农村抽样调查网点抽选及现场调查方法研究 山西省农调队 156

■ 第二部分 统计调查及分析评价方法研究

- 全面实现小康社会县域统计(监测)指标体系研究 ... 安徽省农调队 165
- 农林牧渔业及其相关产业经济统计核算 福建省统计局 185
- 农村生态环境与可持续发展统计指标体系和
统计调查方法体系的研究 山东省农调队 217
-

农村社会、经济、环境全面可持续发展统计指标	
体系和方法研究	四川省农调队 254
农村社会、经济、环境全面可持续发展统计指标	
体系和方法研究	甘肃省农调队 271
山东小城镇综合实力评价课题研究	山东省农调队 288
小城镇综合实力评价指标体系与方法研究	天津市农调队 300
小城镇综合实力评价方法研究	广西壮族自治区农调队 309
缩减法计算地县农业发展速度的实证研究	四川省统计局 321
市、县农业发展速度计算方法改革研究	江苏省农调队 340
地、县计算农业发展速度的方法研究	山西省农调队 357
新一轮县以下农村样本轮换后调查工作的组织与管理 ...	河北省农调队 369

■ 第三部分 统计调查推断方法研究

农村抽样调查中非抽样误差的控制与测算研究	河南省农调队 385
农村抽样调查非抽样误差研究	福建省农调队 400
降低和控制农村抽样调查现场登记误差的研究	江西省农调队 420

第一部分

统计调查制度方法研究

多主题抽样设计 在农产量调查中的应用

国家统计局农调总队

本文在对与规模成比例的概率抽样(PPS)方法,以及由此发展的多主题与规模成比例的概率抽样(MPPS)方法,其特点和适用性进行分析的基础上,结合我国农产量调查的实际,探讨了农产量调查中多主题调查的实现方式。

一、研究背景

多主题抽样调查,或称多目标抽样调查,在农村统计调查实际工作中,就是使用一套样本或尽可能使用较少个数的样本,来准确、及时地获得使用者需要的多个调查指标,其特点是能够省时、省力、省物和提高抽样调查的效率。这个问题对实际调查部门来说具有非常现实的意义。通常一个调查可以满足多目标需要的方法有两种,一种是通过普查的方式,另一种就是利用抽样调查。从理论上讲,只要对总体目标单位进行普查,就能得到所有的指标。但是对大型调查而言,普查需要的资源很多,它的使用场合是非常有限的。因此,利用抽样设计来满足多主题调查是现实当中经常遇到的问题。

我国农产量抽样调查,从本质上说就是要利用一套样本同时对若干目标变量做出统计估计,也即多目标抽样调查。它对提高农产量调查的利用效率,节约调查财力和人力,减少调查误差等具有重要作用。目前我国的农产量调查内容包括,粮食综合播种面积、粮食分品种播种面积(稻谷、小麦、玉米等)和相应的单产调查、棉花播种面积和单产调查,今后还将逐步对油料、糖料、水果、蔬菜、水产等开展抽样调查。

20世纪90年代中后期,美国农业部国家农业统计署(简称NASS)根据发展

中国的需要,开发出了利用名录框进行的多变量与规模成比例的概率(MPPS)抽样方法。近年来,国家统计局农村社会经济调查总队与美国农业部国家农业统计局(NASS)合作,在广东省进行了MPPS抽样设计的试点研究和推广工作,已取得了初步成果。当然,MPPS抽样设计不仅仅是适用于发展中国家,也不仅仅是适用于农业调查,它对解决多目标调查问题能起到很好的作用。

二、PPS 抽样方法

我们知道,按照传统的抽样调查理论,在进行抽样设计时,一般只针对一个目标变量(主题)来进行,实施调查后可以用于推算与之相关的目标变量(主题)的若干观测值,但推断结果的准确性取决于这些变量与设计变量相关性的强弱。怎样将多个目标变量(主题)同时纳入抽样设计过程,并且满足一系列目标变量(主题)推算的需要,恰恰是实际工作中经常遇到的具体问题。

在我们日常进行的社会经济抽样调查中,为解决多主题调查的问题,通常先对设计变量按照有关属性进行分层,分层后再利用抽样框的相关资料进行多阶段抽样。其中与规模成比例的概率抽样(PPS方法)是经常用到的。

在PPS抽样设计中,初级抽样单位第*i*个个体的抽样概率 $\pi_i = nz_i$ (*n*是样本量),它与设计变量 y_i 的规模 z_i 成正比即 $z_i \propto y_i$,因此 $z_i = y_i/Y$ 。对于一阶段的抽样设计,其总体估计量和方差如下:

$$\hat{Y} = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{z_i} \quad (1)$$

$$V(\hat{Y}) = \frac{1}{n} \sum_{i=1}^n z_i \left(\frac{y_i}{z_i} - Y \right)^2 \quad (2)$$

显然当 $z_i = y_i/Y$ 时,方差为零。在实际情况中,我们一般不能用 y_i 计算 z_i ,因为调查的目的正是要取得这些观测值,所以通常用另外一个与 y_i 高度相关的“规模度量”值,或者采用 y_i 的前期指标值计算 z_i 。如果第*i*个个体的“规模度量”的大小表示为 M_i , $M_0 = \sum_{i=1}^N M_i$,那么 $z_i = M_i/M_0$ 。

当我们扩展到两阶段抽样时,第一阶段抽选初级抽样单元(PSU),比如说我们抽选行政村,第二阶段抽选次级抽样单元(SSU),比如说在行政村内抽选调查

户。在忽略有限总体修正系数时,两阶段PPS的总体估计量 $\hat{Y} = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{z_i}$,这里 \hat{Y}_i

是由第二阶段样本对第 i 个初级抽样单位的估计值 \hat{Y}_i 的方差为:

$$V(\hat{Y}) = \frac{1}{n} \sum_{i=1}^N z_i \left(\frac{y_i}{z_i} - Y \right)^2 + \frac{1}{n} \sum_{i=1}^N \frac{M_i^2 S_{2i}^2}{m_i z_i} \quad (3)$$

这里 m_i 是从第 i 个 PSU 中 M_i 个 SSU 中抽选出的第二阶段的样本量, S_{2i}^2 是第 i 个 PSU 内部 SSUs 间的方差。请注意(2)式与(3)式不同之处,是(3)式多了一项 PSU 内的方差。如果通过仔细地选择 z_i ,使 PSU 间的方差最小化,我们就可以减少总方差。通常情况下,PSU 间的方差较 PSU 内的方差要大。

通常针对农村以户为对象开展的抽样调查工作,两阶段 PPS 抽样设计的优点在于:当对省一级总体进行估计时,第一阶段我们用每村总户数作规模度量,采用放回的 PPS 抽样方法抽选样本村。第二阶段在样本村内按简单随机抽样抽选农户,当每村抽选农户数为常数时,便构造成了一个两阶段自加权的样本。这种自加权设计在计算总体总量和方差时,有比较简便的形式。

但是,当调查主题与户数这一规模指标相关性不强时,就需要提供针对调查指标相关性更强的设计变量。如果我们有每个村的农作物产量和家畜总量数据,可以作为抽样框进行 PPS 抽样。但问题是当我们对多主题进行调查时,很难保证针对一种主题的设计方案对其他主题也是有效的。比如说,按牛的饲养数量作为 PPS 规模度量而抽选的样本,对生猪调查来说,效果显然是要打折扣的。在多主题调查情况下,利用农户总数可以作为一种折衷的办法,但前提是需要每村中牛的饲养数量及生猪的饲养数量与农户总数是高度相关的,否则一项针对性调查很难扩展到其他主题,而要提高对其他调查主题的代表性,或者要单独抽选一套样本,或者要大大增加样本量。所以,问题的关键是发现一种适宜的规模度量方法,使其对多主题调查有良好的适用性,同时样本量尽可能地减少。

三、MPPS 抽样方法

多变量与规模成比例的概率抽样(MPPS)方法,是单变量与规模成比例的概率抽样(PPS)技术的进一步推广,其特点是入样概率的确定由多项主题同时决定。由美国统计学家 Sarndal 和 Brewer 等人最先提出有关理论,美国农业部农业统计署的专家 Jeffrey T. Bailey 和 Phillip Kott 等率先在农业抽样调查领域进行了实践。

我们以初级抽样单位(行政村)的抽选为例,每个初级抽样单位(PSU)的入样概率是由多个变量的规模度量共同决定的,即每个村唯一的入样概率由设计

变量的辅助指标和每一主题期望的样本量共同决定。每个村的人样概率 π_i 是前面提到的相对规模即 z_i 乘以样本量 n_i 得到的。即：

$$\pi_i = \min \left[1, \max \left[n_1 \frac{x_{1,i}^{3/4}}{\sum_{i=1}^N x_{1,i}^{3/4}}, \dots, n_k \frac{x_{k,i}^{3/4}}{\sum_{i=1}^N x_{k,i}^{3/4}} \right] \right] \quad (4)$$

这里, $x_{k,i}$ 是第 i 村第 k 主题的规模数量, n_k 是第 k 个主题期望的样本量。 k 是调查中涉及的全部主题(多目标变量)的个数, N 是抽样总体中全部行政村的个数。对目标变量取 $3/4$ 次幂* 是因为通常抽样框资料用的是上一年度的相关数据, 我们不希望多主题(多目标变量)中规模数量大的村有很大的人样概率 π_i , 这样在 π_i “取大取小”的过程中可以适当减少抽中村的个数。设定 π_i 的上限为 1, 是因为在实际调查时我们不希望同一样本被抽中两次以上。

在初级抽样单位(PSU)的人样概率确定以后, 样本的抽选方法通常按不等概率系统抽样或者泊松抽样方法确定。

(一) MPPS 不等概率系统抽样

MPPS 情况下的不等概率系统抽样方法与通常 PPS 情况下的不等概率系统抽样方法是完全类似的。

令 $\{\pi_i\}$ 是每一个村的人样概率, $i = 1, 2, \dots, N$, 且 $\sum_{i=1}^N \pi_i = \text{int}[m]$, r 是 $[0, 1)$ 内的随机数, 则当, $\sum_{j=1}^{i-1} \pi_j < r + k$, $\sum_{j=1}^i \pi_j \geq r + k$, $k = 0, 1, 2, \dots, m-1$ 时, 总体中第 j 个单位被抽中。

(二) MPPS 泊松抽样方法

我们知道, 不等概率系统抽样其样本量恰好等于人样概率之和(取整后的值), 而泊松抽样的样本量不是固定的, 其样本量等于人样概率之和的期望值。泊松抽样过程中要使用永久随机数(PRN), 进一步如果对总体中每一单位对应的永久随机数进行修匀变换, 就可以有效地减少实际样本量与期望样本量之间的差异, 也就是说使实际抽中的样本数量与理论上的期望样本量更加接近。

1. 对抽样总体赋予永久随机数

利用均匀分布给抽样总体中的每一个行政村配上一个永久随机数, 在随后

* 后面一节给出理论解释, 当然也可以对原变量值不做任何数字变换。

进行的抽样过程中该随机数将不再变化。

永久随机数(PRN)的赋值方法:PRN来自 $[0,1)$ 上的均匀分布。在抽样过程中,为了使得总体中每个村对应的PRN分布在县一级更加均匀,利用下面公式对所有县包含的村进行计算:

$$PRN = (e + k) / N \quad (5)$$

$k=0,1,2,3,\dots,N-1$, e 是来自 $[0,1)$ 上的均匀分布, N 是县一级总体行政村的总个数。

对总体PRN进行修匀的作用是减少泊松抽样过程中实际样本量与期望样本量之间的差异。

2. 计算村的入样概率 π_i

设每个村的入样概率为 π_i ,它取决于多个目标变量组成的抽样框资料以及每个主题期望的样本量。公式如下:

$$\pi_i = \min \left[1, \max \left[n_1 \frac{x_{1,i}^{3/4}}{\sum_{i=1}^N x_{1,i}^{3/4}}, \dots, n_k \frac{x_{k,i}^{3/4}}{\sum_{i=1}^N x_{k,i}^{3/4}} \right] \right] \quad (6)$$

这里, $x_{k,i}$ 是第 i 村第 k 主题的指标标志值(即规模大小), n_k 是第 k 个主题期望的样本量。 k 是调查中涉及的全部主题(多目标变量)的个数, N 是抽样总体中全部行政村的个数。设定 π_i 的上限为 1,是因为在实际调查时我们不希望同一样本被抽中两次以上。

显然每个村入样概率的确定与各主题(变量)对应的样本容量有关,而样本容量如何确定呢?在实际操作中,一般先根据以往调查的经验样本量,给各主题赋予一个初始值,借助计算机程序计算出各主题的标准差和变异系数(C.V),再按照各单项调查主题对调查精度的要求,即各调查主题期望达到的变异系数,对样本量的初始值做出调整,即抽样精度高时(变异系数低),可以适当减少对对应主题的样本量,而抽样精度低时(变异系数高),可以适当增加对应主题的样本量,直到各项样本量的大小,可以达到期望的抽样精度。采取这种不断逼近的方法,可以确定一个理论上满足要求的最小样本量,但在实际工作中,为了保证实际调查时的抽样精度,该最小样本量还可适当扩大一些。

3. 利用泊松方法抽选样本

若 $\pi_i > prn$,则抽中第 i 村 (7)

其中 π_i 是第 i 村的入样概率, prn 是与其对应的永久随机数。

(三) 总体总值的估计

$$\hat{Y}_k = \sum_{j=1}^N x_{k,j} \frac{\sum_{i=1}^n w_i \hat{Y}_{k,i}}{\sum_{i=1}^n w_i x_{k,i}} \quad (8)$$

这里 $x_{k,j}$ 表示来自抽样框的控制数据, $\hat{Y}_{k,i}$ 表示对抽中村的实际调查数据, $w_i = 1/\pi_i$, 它是第 i 村入样概率的倒数即权重。

由于用上面多变量方法得到的权重用于推算时, 其含义不象单变量的 PPS 方法得到的权重那样易于理解, 所以在推算过程中, 一般用设计变量的比率估计

值进行修正, 提高估计量的精度。这里, 我们利用设计变量的比率估计值 $\frac{\sum_{j=1}^N x_{k,j}}{\sum_{i=1}^n w_i x_{k,i}}$

去修正加权推算的结果 $\sum_{i=1}^n w_i \hat{Y}_{k,i}$ 。特别地, 当 $x_{k,j} = 1$ 时, 相当于用总体个数 N 与权重之和 $\sum_{i=1}^n w_i$ 的比率来校准估计结果。

(四) 抽样误差的计算

总体总量估计值的方差和抽样标准差系数(估计量的变异系数 C. V)如下:

$$\hat{V}(\hat{Y}_k) = \frac{\left(\sum_{j=1}^N x_{k,j} \right)^2}{\left(\sum_{i=1}^n w_i x_{k,i} \right)^2} \sum_{i=1}^n (1 - \pi_i) w_i^2 \hat{e}_{k,i}^2 \quad (9)$$

$$\text{这里 } \hat{e}_{k,i} = \hat{Y}_{k,i} - x_{k,i} \frac{\sum_{i=1}^n w_i \hat{Y}_{k,i}}{\sum_{i=1}^n w_i x_{k,i}}, w_i = \frac{1}{\pi_i}$$

这里 $x_{k,i}$ 表示第 i 村第 k 个主题的抽样框数据, $\hat{Y}_{k,i}$ 表示第 i 村第 k 个主题的实际调查数据, w_i 表示第 i 村的入样概率之倒数即该村的权重。

抽样标准差系数(估计量的变异系数 C. V)的计算如下:

$$C.V = \sqrt{\hat{V}(\hat{Y}_k)/\hat{Y}_k} \quad (10)$$

在实际操作过程中, C.V 作为我们判断是否达到预期抽样精度要求的标准。如果达不到我们期望的 C.V, 则可以通过适当增加有关主题的样本量达到要求的抽样精度。通常, 我们经常使用的是多阶段抽样设计, 方差估计会变得更加复杂, 这样对实际调查数据也经常采用“刀切法”做出方差估计。

四、样本轮换在 MPPS 方法中的实现

通常在农村抽样调查过程中, 为防止样本老化和调查对象配合程度的下降, 年度间要考虑对样本进行一定程度的样本轮换, 这样做即可以使样本及时地反映变化了的客观现实, 也可以满足对样本进行连续观察的需要。使用多主题抽样设计 MPPS 方法, 结合永久随机数时 (PRNs) 的使用, 样本轮换的方法如下:

第一, 对抽样框中每一个初级抽样单位 (PSU) 指定永久随机数 PRNs (在以后年度不再改变的“固定”随机数), PRNs 来自 $[0, 1)$ 上的均匀分布。把所有的 PSU 按 PRNs 的大小以升序排列, 利用公式 (5)

$$PRN = (e + k)/N$$

$k = 0, 1, 2, 3, \dots, N - 1$, e 是来自 $[0, 1)$ 上的均匀分布, N 是县一级总体行政村的总个数。这样做的好处是使其在 $[0, 1]$ 上的分布更均匀, 减少抽中村样本个数的波动。

第二, 按 MPPS 方法即公式 (6) 计算每个 PSU 的抽样概率 π_i 。

第三, 抽选第一年的样本。利用泊松分布, 如果 $PRN \leq \pi_i$, 则抽中该 PSU, 那么 PRN 落在 $[0, \pi_i]$ 区间的 PSU 组成了第一年的样本。

第四, 以后年度的样本抽选。设与前一年样本的重复率为 O , 新样本的起点是重复率 O 的函数, 这里需要用到前期样本。定义新样本的终点如下:

$$l_i = u_i - (\pi_i \cdot O)$$

其中, l_i 下一年样本抽选区间的低限点,

u_i 上一年样本抽选区间的高限点,

π_i 上一年第 i 个 PSU 的抽样概率,

O = 前后两年期望的样本重复率。

$$u_i = l_i + \pi_i \text{ 或在 } l_i + \pi_i > 1 \text{ 时, } u_i = l_i + \pi_i - 1$$

其中, u_i 下一年样本抽样区间的高限点,