

农业抽样调查 技术及原理



山西科学教育出版社

农业抽样调查技术及原理

张淑莲 吴国定 编著
刘执鲁 张祖信

*

山西科学教育出版社出版 (太原并州北路十一号)

山西省新华书店发行 山西省七二五厂印刷

*

开本: 787 × 1092 1/32 印张: 7.125 字数: 170 千字

1988年5月第1版 1988年5月山西第1次印刷

印数: 1—3500册

*

ISBN 7—5377—0035—4

S·6

定价: 1.85元

前 言

农业动态监测，是为了解掌握资源利用、生产情况、科技实施效应和农村经济变化等必不可少的一门先进的综合性技术。动态监测需要通过抽样调查和统计计算来实现。这是急待普及的一项科学技术。

山西省农业区划委员会，自1981年迄今在全省11个地（市）、33个县（市）建立了农业动态地面监测网络，覆盖面积达23000平方公里。这项工作由山西省农业科学院综考所负责技术指导。

我们在总结实践经验的基础上，为给工作在第一线的科技人员提供农业抽样技术和统计方法，撰写了这本书。介绍了有关概率论、数理统计的原理及常用的抽样调查方法，并通过实例加以说明，力图通俗易懂。附录部分还针对本书中介绍的各种抽样方法专门编制了适合PC——1500袖珍计算机使用的BASIC语言程序及使用说明。本书可供专业工作者、生产管理人员和有关院校师生阅读。

本书在撰写过程中，得到山西省农业区划委员会及有关单位的大力支持，山西农业大学姜藏珍副教授曾帮助审稿和

指导，在此一并致谢。

由于作者水平有限，书中错误在所难免，望广大读者提出宝贵意见。

作者 1988. 4. 于太原

目 录

第一章 抽样调查	(1)
第一节 统计调查基础知识.....	(1)
第二节 抽样调查的基础知识.....	(2)
一、抽样调查的种类.....	(2)
二、抽样调查的概念.....	(2)
三、随机变量.....	(3)
四、误差和偏差.....	(4)
五、样本的大小及样本容量的计算.....	(5)
六、标志的符号和下标.....	(6)
七、总和符号及运算公式.....	(7)
八、样本平均数.....	(8)
九、标准差及方差.....	(9)
十、样本平均数的方差及其标准差.....	(11)
第二章 概率论知识简介	(13)
第一节 基本概念.....	(13)
第二节 事件的概率.....	(14)
一、频率.....	(14)
二、频率的稳定性.....	(14)

三、概率.....	(15)
四、概率的性质.....	(15)
五、古典概型(等可能概型).....	(16)
第三节 概率分布.....	(18)
一、离散型随机变量的定义及其概率分布.....	(18)
二、连续型随机变量的概率密度.....	(21)
第四节 置信概率(即可靠性).....	(29)
第五节 概率论应用举例.....	(32)
第三章 简单随机抽样	(35)
第一节 简单随机抽样的概念.....	(35)
一、简单随机抽样的概念.....	(35)
二、简单随机抽样的符号和定义.....	(35)
三、举例说明简单随机抽样公式、符号的应 用.....	(38)
第二节 简单随机抽样的工作步骤.....	(39)
一、确定抽样总体.....	(39)
二、确定样地的大小和形状.....	(40)
三、确定样本单元数.....	(40)
四、用简单随机抽样对玉米测产.....	(41)
五、例题3—3计算结果的分析与探讨.....	(43)
第四章 成数抽样	(46)
第一节 成数抽样的基本概念.....	(46)
一、总体成数与样本成数.....	(46)
二、确定样本单元数.....	(48)

三、各因子总体成数的估计·····	(49)
四、确定估计值的误差·····	(50)
第二节 应用举例·····	(51)
第五章 分层抽样 ·····	(60)
第一节 分层抽样简介·····	(60)
一、分层抽样的概念·····	(60)
二、分层抽样的原则·····	(60)
三、分层抽样的有关符号及定义·····	(61)
四、分层抽样的有关计算公式·····	(62)
五、分层抽样的效率·····	(63)
第二节 分层抽样的工作步骤·····	(64)
一、确定抽样方法·····	(64)
二、选择分层因子·····	(64)
三、确定各层的总体单元数·····	(64)
四、确定要求精度及可靠性, 计算总体样本 容量·····	(65)
五、分层整理资料·····	(65)
六、样本容量的计算和样本单元的分配·····	(65)
(一) 样本容量的计算·····	(65)
(二) 样本单元的分配·····	(67)
第三节 分层抽样应用举例·····	(69)
第六章 两阶抽样 ·····	(74)
第一节 两阶抽样简介·····	(74)
一、两阶抽样的概念·····	(74)

二、两阶抽样所用的符号和定义·····	(75)
第二节 一阶单元大小相同的总体特征数的 估计法·····	(75)
第三节 一阶单元大小不等的两阶抽样法·····	(77)
第四节 应用举例·····	(79)
第七章 整群抽样 ·····	(88)
第一节 整群抽样方法简介·····	(88)
一、等群抽样·····	(88)
二、不等群抽样·····	(88)
三、整群抽样的三个原则·····	(89)
四、整群抽样与分层抽样的区别·····	(89)
第二节 等群抽样·····	(89)
一、等群抽样的意义·····	(89)
二、等群抽样的几个特征数及计算·····	(89)
第三节 不等群抽样方法·····	(93)
一、不等群抽样的基本概念·····	(93)
二、不等群整群抽样的标志值、符号及算 式·····	(95)
三、总体方差公式·····	(96)
第四节 应用举例·····	(97)
第八章 大样本、小样本抽样估计方法简介 ·····	(101)
第一节 总体平均数的抽样估计及重复抽样 估计方法·····	(101)
第二节 总体频率的大样本抽样估计方法·····	(109)

第三节 平均数与方差的简捷计算公式介绍…… (119)

第四节 小样本法估计总体平均数及总体频率…… (124)

附录 (亦适用于Pc—1500袖珍机的BASIC语言程

序) …… (131)

一、随机抽样计算程序…… (131)

二、简单随机抽样调查结果计算程序…… (134)

三、分层抽样程序…… (139)

四、按属性分层因子分层时的抽样程序…… (145)

五、分层抽样调查结果计算程序…… (149)

六、等群抽样计算程序…… (151)

七、不等群抽样的计算程序…… (152)

八、一阶单元大小相等的两阶抽样 (调查结果计算程序) …… (154)

九、一阶单元大小相等两阶抽样结果计算程序…… (157)

十、一阶单元大小不等的两阶抽样结果计算程序…… (160)

附表 …… (165)

1. 正态分布的密度函数表…… (165)

2. 正态分布表…… (168)

3. 正态分布的双侧分位数 (U_{α}) 表…… (174)

4. 二项分布表…… (175)

5. 二项分布参数P的置信区间表…… (180)

6. 泊松 (Poisson) 分布表..... (192)
7. 泊松 (Poisson) 分布参数 λ 的置信区间
表..... (209)
8. 学生氏 t 分布的双侧分位数 (t_α) 表..... (211)
9. 随机数表..... (213)

第一章 抽样调查

抽样调查属于统计调查范畴内的一种方法。

第一节 统计调查基础知识

统计调查是有计划、有组织地向调查单位搜集材料的过程。由于统计研究的目的和对象的特点不同，一般统计调查分三类。

一是按搜集资料的组织方式，分为统计报表制度和专门调查。专门调查是为了某种特定目的而组织的一种搜集资料的调查组织形式。例如各种普查、重点调查、典型调查、抽样调查等。

二是按调查的范围，可分为全面调查和非全面调查。

三是按调查对象在时间上进行观察记录的连续性不同，可分为经常性调查和一次性调查。

几年来，在农业动态监测一系列实践中，我们采用了统计调查第一类专门调查中的抽样调查方法，进行有关监测计算。我们认为，抽样调查能以较少的人力、物力，

在较短的时间内，推算出说明整个总体的综合指标。在只需说明整个总体情况，而不需说明各级的详细状况的前提下，都可采用抽样调查。

第二节 抽样调查的基础知识

抽样调查，是按随机原则从总体中抽取一定数目的单元，组成样本，根据对样本的各项指标计算的结果，从数量上推断出总体的一种非全面调查。在抽样过程中要遵循随机原则。

一、抽样调查的种类

常用的抽样调查有：简单随机抽样、成数抽样、分层抽样、两阶抽样、分层成数两阶抽样等。根据具体情况，我们将几种常用的抽样方法分章详述于后。

二、抽样调查的概念

抽样调查，首先确定调查范围，在被调查的对象中，按要求的精度，遵循随机原则，抽取一定数量的单元组成样本，进行测量和调查，并用其指标数值去推断总体相应指标。为了便于掌握抽样调查的方法、公式、步骤及原理，下面介绍几个基本名词并加以解释。

1. 单元 按预先规定的样本容量的大小，将被调查总体划为若干个观测和调查的单位，叫做单元。

2. 总体及总体单元数 单元的集合叫总体。用 N 表示总体单元的个数。

3. 样本及样本容量 在由 N 个单元组成的总体中，随机抽取 n 个单元，组成一个新的单位叫样本。样本中含有单

元的总数，叫做样本容量，记作n。

4. 抽样调查的一般方法和步骤 按工作方案中的要求，抽取一定数量的样本单元，并对预定数量的样本单元，进行定位、定量调查。根据所取得的资料进行分析研究、统计计算，进而推断出总体的各项相应指标。

三、随机变量

在抽样调查中，对每个单元可以测定许多调查因子。而这些调查因子分别属于两种标志：一种是属性标志，例如作物种类、监测样点所属地类等情况；另一类是数量标志，如作物的亩产量、总产量、种植面积等。在随机抽样调查过程中，有许多随机现象的结果，表现为一系列不同数值，其中每个数值都有可能出现，而且具有一定的概率，即随机现象以量的形式出现，就形成随机变量。

例1—1：有6包不同品种的蔬菜种子，用1、2、3、4、5、6编号放在一个盒子里，每次从中摸出1包，可能出现

$$C_6^1 = \frac{6!}{1!(6-1)!} = 6 \text{ (种) 结果 (每个品种的菜籽被摸中的}$$

的概率皆为1/6)。这时，我们将“可能出现6种结果”这一随机事件发生的各种可能，用变量X取不同值来表示，则X就是随机现象各种结果的变量，叫做随机变量。本文中皆取X的随机变量。

随机变量可分为连续变量的离散变量两种。一般说来，能用有限数值标定被调查体的长度、面积、容积、重量等标志值的变量，叫连续变量。用正整数计量的标志值变量，叫离散变量。如植株数/米²、发芽种子的粒数等，采用离散变量描述。

四、误差和偏差

真值：观测对象客观存在的数量，称之为真值。

观测值：对观测对象每次进行观测所得的值，称为观测值。

误差：真值与观测值之差，称为误差。若以 X 表示真值， X_i 表示第 i 次观测值，误差 $a_i = x - x_i$ 。

在实际工作中，真值是无法得知的，但可以采取抽样的方法，通过多次对观测值分析研究，找出其规律性，从而推算出真值的估计值（极限）。

至于抽样误差，是指样本均值与总体均值之差。在抽样调查中，我们是用样本均值估计总体均值的，其差值的大小，直接影响对总体特征数估计的精度。

例1—2：一个总体是由15个单元组成的，其标志值分别为222.8、116、511.9、327.8、452.3、189.9、408.1、217.9、350.8、346、332、163.6、97、433.3、84.6（单位：公斤/亩）。若随机抽取5个单元组成样本时：①以222.8、346、511.9、163.6、84.6组成一个样本其平均值为265.8；②以166、452.3、408.1、97、217.9组成样本时，其平均值为268.3。已知总体均值为286.9。样本①均值与总体均值之差为21.1；样本②的均值与总体均值之差为18.6。显然用样本②的均值去估计总体时，误差较小，精度就会相对较高。

偏差则往往产生在调查过程的各项工作中，如用工具、仪器测量时视觉的错觉，样本抽取不正确，估计方法错误等，都是产生偏差的原因。如果采取相应的办法，这种偏差是可以缩小直至纠正的。例如经过对仪器的校正和处理，通

过多次观察与分析；抽取样本的范围一定要与总体范围相一致；估计数据或以往资料尽可能地准确；工作人员在工作中注意避免失误等，这样在抽样工作中，以上偏差是可以避免的。

五、样本的大小及样本容量的计算

所谓样本的大小，是指样本单元数的多少。在抽样调查前，应利用公式对其样本单元数（即样本容量）进行预算。在简单随机抽样调查中，样本容量的求法，可用公式：

$$n = \frac{t^2 \sigma_x^2}{\Delta^2_{\bar{x}}}$$

在成数抽样调查中，样本容量的求法，可用公式：

$$n = \frac{t^2(1-p)}{E^2 p}$$

在分层抽样调查中，样本容量的求法，可用公式：

$$n = \frac{t^2 \sum_{H=1}^L W_H \sigma_H^2}{\Delta^2_{\bar{x}}};$$

$$n = \frac{t^2 \sum_{H=1}^L W_H \sigma_H^2}{E^2 \bar{x}^2}$$

在运用两阶抽样调查时，当一阶单元大小相同时，对二阶样本单元的求法，可用公式：

$$m = \sqrt{\frac{S_2^2}{S_1^2 - \frac{S_2^2}{M}}} \cdot \sqrt{\frac{C_1}{C_2}}$$

一阶样本单元数的求法，可用公式：

$$n = \frac{t^2 \left(S_1^2 - \frac{S_2^2}{m} + \frac{S_2^2}{m} \right)}{\Delta^2_{\bar{x}} + t^2 \cdot \frac{S_1^2}{N}}$$

以上公式将在以后各有关章节分别详细讲述。

在抽样调查中，还常用到“大样本”和“小样本”这样的术语。所谓大样本，是指样本单元数超过50（即 $n \geq 50$ ）的样本（有些情况也称 $n \geq 30$ 的样本为大样本）。样本容量小于50（即 $n < 50$ ）的样本，称为小样本。

在具体抽样调查工作中，由于总体分布律不同，总体均值所处位置也不一样。在不了解总体分布律的情况下，应采取大样本原则确定样本单元数，进行抽样调查。因为当样本单元数充分大时，不论总体分布如何，样本平均数的分布趋于正态分布。当然，也有例外情况，那就是：若对于调查总体分布确知系正态分布时，则可用小样本原则确定样本单元数，进行抽样调查。以上对于样本单元数计算的理论根据是中心极限定理。

六、标志的符号和下标

在抽样调查中，参数和变量均采用相应的字母作为符号，一般用 x 或 y 做为变量的符号（本文中取 X 为变量符号），取 t 、 E 、 P_c 、 f_i 等作参数符号。

为了限定标志符号的范围，通常用下标来标明。如用 x 表示小麦亩产量，则 X_i 就表示第 i 个单元的小麦亩产量， i 称为 X 的下标。有时下标可以是两重或多重的形式。例如 X_{iB} 、 X_{iA} ，第二重下标 B 和 A 是对 i 限定范围的。若 B 和 A 分别表

示不同的调查时期，B表示初测，A表示复测，则 X_{iB} 表示第*i*个单元初测值（亩产），而 X_{iA} 表示第*i*个单元复测时亩产值。又如，在分层抽样中对于变量 X_{Hi} ，下标H是表示第H层，而第二重下标*i*表示H层中的第*i*个单元的序号。 X_{Hi} 表示第H层中第*i*个单元的观测值。

七、总和符号及运算公式

在抽样调查的估计计算过程中，经常用不同方式对各种变量求和，常用的总和符号 Σ （读“西格马”），表示逐项

相加的意思。例如， $\sum_{i=1}^5 x_i$ 表示由 x_1 到 x_5 依次累加。即

$$\sum_{i=1}^5 x_i = x_1 + x_2 + x_3 + x_4 + x_5。其中i = 1到5表示变量数的$$

下限和上界。

常用的“和号”表达式介绍如下：

$$1. \sum_{i=1}^n x_i = x_1 + x_2 + x_3 + \cdots + x_n \quad (n \in N, \text{以下同})$$

$$2. \sum_{i=1}^n x_i^2 = x_1^2 + x_2^2 + x_3^2 + \cdots + x_n^2$$

$$3. \sum_{i=1}^n x_i y_i = x_1 y_1 + x_2 y_2 + x_3 y_3 + \cdots + x_n y_n$$