

万宝瑞 刘秀印 王治国著

电子计算机 在农产品产量 预测中的应用

农业出版社

电子计算机在农产品 产量预测中的应用

万宝瑞 刘秀印 王治国 著

农业出版社

2-617/13

电子计算机在农产品产量预测中的应用

万宝瑞 刘秀印 王治国 著

• • •

责任编辑 肖毅为

农业出版社出版 (北京朝内大街130号)

新华书店北京发行所发行 农业出版社印刷厂印刷

787×1092 毫米 32 开本 4 印张 1 插页 81 千字

1986 年 12 月第 1 版 1986 年 12 月北京第 1 次印刷

印数 1—1,800 册

统一书号 4144·621 定价 0.76 元

前 言

党的十一届三中全会以后，中国农村普遍实行了家庭联产承包责任制，原来的“三级所有，队为基础”的经营管理模式、传统的农产品产量统计方法和计算手段已经不适应新形势发展的要求。当前，统计方法已由过去主要依靠全面统计转向抽样调查、重点调查、典型调查等多种方法；统计内容，除考察以农户为经营单位的生产、收支和经济效益外，同时还要统计市场上对农副产品的需求量，尽快地提供产前、产中、产后的农业经济信息，为国家制定政策，为农业行政部门采取经济决策和为农村专业户发展生产服务。

因此，以电子计算机为计算手段，对农产品采用新的预测方法实为当务之急。我们在研究、分析国内外抽样调查和预测方法的基础上，根据我国地域辽阔，自然经济条件复杂，农业统计数据不够完善和电子计算机应用基础较差等因素，本着操作简单、适于推广的原则，建立了“农产品产量预测方法及电子计算机程序设计”课题，先后用了两年时间完成课题任务。本书则是在本课题研究的基础上，同时参考国内外大量资料，经过归纳、概括编写而成。

农产品抽样调查是世界上应用较广的一种方法。解放前，我国曾有个别人采用。五十年代，六十年代，由于“左”

的思想影响，抽样调查一直没有很好地开展，党的十一届三中全会以后，这种方法已引起人们的极大重视。1981年已把抽样调查方法用到统计资料搜集工作上。但是由于世界各国社会制度不同，自然经济条件，经营管理水平的差异，加之农业科学研究的复杂性，到目前为止，还没有一套符合我国国情的抽样调查、产量预测方法及其软件包。本课题设计的NCD软件包在国内还是第一次，并得到专家鉴定和验证。其优点如下：

1. 借助电子计算机的优势，通过多种方法，以抽样率、样本代表性检查值和抽样误差为约束条件，从多组样本中筛选出最优的一组，从而使样本代表性最好，使误差限制在最小的范围。例如，研制的三种分层抽样方法，在给定抽样率的情况下，可以选择代表性最好的样本。象这种选优的办法，手工是无法进行的。

2. 采用双圈式等距循环抽样法，并采用公式： $NM = N + K - 1$ 来求按圆圈排列的最末单位序号。一些专家认为，这是本课题的创新。这种方法的抽样起点是总体中全部单位，电子计算机可以从多组中选择最优的一组，这样就能够克服在直线等距抽样中，由于抽样起点所引起的系统误差。

3. 本软件包提供六种抽样方法，三种预产方法可供各地区不同条件选用，这样就能够缩小抽样、预产误差。例如，自然、生产条件差异不大的地区，可选用单产等距循环抽样；在经济条件复杂，亩产差异大的地区，可选用按单产分层抽样；在自然条件复杂，数据间差异较大的地区，可选用按自然条件分层随机抽样等等。

4. 农产品产量预测软件包 NCD, 包括 16 个程序, 适合于各种条件, 操作简单, 使用方便, 只要把基础数据转入机内, 就可以及时提供可靠的农产品产量信息, 这对我国农业现代化管理有着重要意义。

5. 抽样调查方法, 预产方法的理论基础是概率论。以单产分层随机抽样为例, 在有足够多的样本组合情况下, 样本代表性的分布是正态分布。又如, 按总产等距循环抽样, 抽样误差和抽样率的关系, 与计算大样本必要单位数的公式是符合的。

为了尽快地实现农业现代化管理, 农牧渔业部为全国各省市提供一批 IBM-pc 微型机, 这就为全国建立一个完整的计统网络创造条件。为了验证 NCD 软件包可行性, 我们受农牧渔业部计划司委托, 曾对 1983 年全国夏粮产量和 1984 年全国粮食产量进行预测, 结果表明, 1983 年预测产量与核实上报产量相差约为 2%, 1984 年相差为 1.1%, 为国家进行产量预报提供重要依据。

参加课题研究和本书执笔者有: 中国农业科学院农业经济研究所万宝瑞同志、刘秀印同志。农牧渔业部计划司统计处王治国同志。

全书共分六章: 第一、二章侧重介绍农产品产量抽样调查的基础知识; 第三、四、五章从计算机应用角度阐述六种抽样调查方法、三种产量预测(推算)方法及其程序设计; 第六章系统地介绍了农产品产量抽样调查和预测程序的使用方法等。

在课题研究和本书撰写过程中, 曾多次得到刘宗鹤教授

和张象枢教授的关心和指导，在此表示衷心感谢。

对这样一个新问题的研究，对于我们来说还是第一次，由于时间仓促，我们水平有限，书中一定会有很多缺点和错误，敬请广大读者批评指正。

1985年2月

目 录

1	第一章 抽样调查设计
1	一、抽样调查及其作用
3	二、抽样调查中几个基本概念
7	三、农业上常用抽样调查方法
12	第二章 抽样误差
12	一、几个误差的基本概念
15	二、抽样误差的计算
	第三章 农产品产量抽样调查和预产
32	(NCD) 程序包概述
32	一、NCD程序包的功能
33	二、NCD程序包的结构
37	第四章 抽样调查方法与程序设计
37	一、按总产排队进行等距循环抽样
50	二、按单产排队进行等距循环抽样
50	三、按面积累加数进行等距抽样
57	四、单产分层随机抽样
66	五、单产分层等距循环抽样
69	六、按自然条件分层随机抽样
73	七、抽样方法分析
77	第五章 预产方法及程序
77	一、预产方法
80	二、预产程序介绍

83	三、计算结果举例
86	四、预产（推算）方法分析
	第六章 NCD农产品产量抽样调查程序
87	包的使用和操作方法
87	一、工作环境
87	二、程序包的启动和各功能模块的工作方式
91	三、原始数据格式及其输入方法
100	四、参数选择和错误信息
106	附录：NCD程序包应用举例

第一章 抽样调查设计

一、抽样调查及其作用

(一) 什么叫做抽样调查

按统计调查的范围来分，统计调查可分为全面调查和非全面调查。抽样调查属于非全面调查，是非全面调查中用途最广的一种调查方法。所谓抽样调查，就是按照随机原则从总体中抽取一部分单位进行观察，并以抽中部分的结论来推断全部总体的方法。在数理统计中，抽样调查也叫统计抽样法，或叫抽样检验。

例如对收割前的农作物进行亩产调查时，抽取一小部分农作物地块进行调查，根据抽样的资料来测算农作物的平均亩产，并用以推算总产。

(二) 抽样调查的作用

抽样调查是一种非全面调查，它在社会经济调查中，具有重要作用。它与全面调查相比，有如下优点：

1. 可行性 全面调查的工作量较大，在实践中往往不能达到目的。采用抽样调查，则可以抽选一部分单位进行深入细致的研究，并从数量上推算全部总体的参数。例如农作物产量调查，农民家计调查等。另外还有很多情况，对调查的对象不能进行全面调查，例如对电灯泡寿命的检验，棉纱产

品拉力的检验，我们只能抽取一部分产品进行试验。只能采用抽样法调查。

2. 及时性 运用抽样方法调查，比全面调查涉及单位少，机动灵活，可以快速收集资料进行汇总分析，在短时间内就可以取得可靠资料，为政府及有关部门及时制订政策和经济决策提供依据。

3. 准确性 运用抽样法进行调查，调查单位少，调查人员对调查单位可以进行深入细致地了解。可以利用概率论原理，在一定概率保证的要求下，确定抽样推断的误差范围。从部分推断全体，保证必要的准确性。

为了保证抽样调查的科学性，运用抽样法调查时，要按照一定的方法和程序进行。抽选单位时，必须遵守随机原则，排除人的主观影响，使各单位都有同等被抽中的机会，抽中与否全凭偶然机会。同时要抽选足够的单位数，以提高抽中单位的代表性。

（三）抽样调查的应用范围

抽样调查在统计工作中应用很广泛，主要有：

1. 对一些事实上不可能或不必要进行全面调查时，采用抽样法处理比较妥当。例如农产量预测调查，只能抽取部分地块进行抽样观察，以此推算全部。对工业带有破坏性产品检查，只能检查或破坏一部分产品，而不能检查或破坏全部产品。对手工业生产和农村集市贸易，以及对城乡居民家庭生活情况的研究，只能采用抽样调查，而不可能进行全面调查。此外，为了争取时间，节约人力、物力和调查费用，常常进行抽样调查。

2. 利用抽样调查资料可以检查和核对全面调查资料的质量，并加以修正补充。全面调查由于范围广，工作量大，参加人员多，在登记时容易发生遗漏或重复，因而在全面调查后，马上进行抽查，运用两种资料进行核对和修正，以提高全面调查资料的准确性。例如在人口普查中，经常采用这种办法。

3. 抽样调查有利于对一些观察项目多，或者需要深入研究的总体进行调查和检验。例如对产品和商品质量检查，限于人力、物力和时间，往往不可能采用全面调查，即使采用，由于数量大，也容易发生差错。因此，运用抽样调查，可以提高统计工作的质量。

但是，抽样调查也和其他统计方法一样，不是万能的，有局限性。因为它是运用抽样调查所得的资料去推断全部，所以总是会出现一定误差的。另外，在研究社会经济现象时，抽样调查只能适用于随机变量部分，而不适用于非随机变量。同时，抽样推断，只能做纯数量的推断，而不能提供经济分析。

二、抽样调查中几个基本概念

（一）随机现象

数理统计是一门研究现实世界中随机现象的规律性的现代化学科。什么是随机现象，先看几个例子。

例1：某个地区在下月15日是否下雨。

那个地区由于受到当时的气温、气压、风向、风力、湿度等气象因素的影响（有些因素在目前是无法控制的），我们

不能完全准确地预测该地区下月15日是否下雨。

例2：一位顾客到商店购买衣服、鞋帽，所需要的号码可能是大号的、中号的，也可能是小号的。因此，售货员在顾客购货之前不能准确预言他所需要的型号。

例3：某生产队某个小麦品种在某种管理条件下，亩产可能低于400斤，可能在400斤与500斤之间，也可能高于500斤，在未收割以前也是不能准确预言的。^①

上述现象的共同特点是：在一定条件下，一种事物可能出现这种结果，也可能出现另一种结果，呈现出一种偶然性。换句话说，我们在事前不能准确地预言它的结果。这种现象称为随机现象。上面举的各种现象都是随机现象。

（二）概率

概率是一个数学名词，其简单的含义是：在一定的条件下，一个事件发生或不发生的可能性的的大小。例如，某机械制造厂，生产某种机械产品。虽然以同样的原料同样的工艺，同样的技术水平等，在这些条件下，而生产的某种机械产品这一事件，并不一定都是合格的，其中有成品、次品和废品等。但是，究竟出成品的可能性多大，出次品或废品的可能性多大，对这些可能性大小的计算就是通过概率进行计算和表示的。

由此可见，对某一事件在一定的条件之下，这一事件发生或不发生，发生的可能性有多大，不发生的可能性有多

^① 例1、2、3，引用《数理统计学》，宋元村、黄玉喜编著。

大，用数学方法预测出它的发生或不发生的可能性的_{大小}，就是概率。

(三) 随机变量

在一定条件下，以一定的概率采取的某一个可能数值的变量，就叫做随机变量。先举几个例子。

例 1：某百货公司在某一天内将接待顾客的人数。

例 2：秋后某块土地上粮食的产量。

例 3：某电话总机在一小时内接到呼唤的次数。

上述例子的共同特点是：这些随机现象的结果是直接用量来表示的。但这些数量的取值，不能事先确定，而是受偶然因素的影响改变的。这种随着偶然因素而改变的变量，就叫做随机变量。

随机变量与普通变量的区别（如 $y = f(x)$ 中的变量 x ， y ）在于：前者是一个变量，但这种变量是随着许多偶然的因素变化而改变着；后者也是变量，但其中一个变量确定后，另一个变量不受偶然因素的影响随之而确定。

随机变量在试验之前是一个不确定的量，它有许多可能值，具有多种多样的可能性，但在试验之后，它只取许多可能值中的一个，只有其中的一种可能性成为现实，这个值叫做随机变量的观察值。观察值是一个确定的数，必须和随机变量本身区别开来。例如，某电话总机某日接到呼唤的次数，在某日之前，是一个不肯定的数字，可是某日一过去，就是一个完全确定的数字了。

(四) 数学期望

随机变量 ξ 的平均值叫做数学期望。对于离散型随机变

量来说，它的数学期望等于随机变量 ξ 各个可能值分别乘以这个值的概率，然后再相加。

例如：如果离散型随机变量的分布列是

$$p_i \begin{array}{c} X_1 \quad X_2 \quad X_3 \cdots \cdots X_n \\ p_1 \quad p_2 \quad p_3 \cdots \cdots p_n \end{array}$$

那么随机变量 ξ 的平均值（数学期望） $M\xi$ 就是：

$$M\xi = X_1 p_1 + X_2 p_2 + \cdots \cdots + X_n p_n +$$

$$= \sum_{i=1}^n X_i p_i$$

（五）总体和样本

在抽样调查中，我们把所要研究的调查单位的全体称为全及总体，或称一般总体，简称总体；把由全及总体中抽取出来作为样本的部分称为样本总体，简称样本。

（六）总体平均数与样本平均数

在抽样调查中，全及总体各单位某一个数量标志值的算数平均数，叫做全及总体平均数，简称总体平均数。样本总体各单位某一数量标志值的算数平均数，叫做抽样平均数，也叫样本平均数。全及总体是由总体全部标志值决定的，所以总体平均数就是唯一确定的量。而样本指标，则是根据抽取部分单位确定的，因此，它随着所抽选的单位的不同而变化。所以抽样平均数随着抽选样本的不同而变化，它仍然是一个随机变量。

（七）简单算术平均值和加权算数平均值

简单算术平均数是把一组数值相加后，再用数值的个数

去除而得的平均值。例如有 n 个观测值 $X_1, X_2, \dots, X_n$, 其算术平均值便等于各观测值之和除以观测值的个数, 即:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{\sum_{i=1}^n X_i}{n}$$

如果在 n 个数值中, 彼此有相同的 X_i , 并用 f_i 表示具有相同数值 X_i 的个数。则 X_i 的加权算术平均值 $\bar{X}$ 可用下式计算:

$$\bar{X} = \frac{\sum_{i=1}^k f_i X_i}{n} = \sum_{i=1}^R X_i \cdot \frac{f_i}{n} \quad (k < n)$$

这里, n 个数值中有 k 个不同的数值。由上式可见, 加权算术平均值等于各个数值 X_i ($i = 1, 2, \dots, k$) 乘以各个数值的权数 $\frac{f_i}{n}$ (即频率) 之和。

(八) 标志值

一个总体可以通过一个或 n 个数字指标来反映它的特征。比如某省的粮食产量, 可以通过各县的单产、播种面积、总产等来进行统计。我们把用来表示总体某方面特征的数值称为标志值。比如上例中各县的单产、播种面积、总产都可以作为计算全省综合指标的标志值。

三、农业上常用抽样调查方法

在抽样调查中, 根据工作目的和工作条件的不同, 可以采用不同的抽样方法。角度不同抽样方法分类也不同。

(一) 抽样调查按组织形式可分为

1. 纯随机抽样 平常所说的随机抽样，实际上就是单纯随机抽样，也叫纯随机抽样。它是对总体不加任何分组排队，完全凭着偶然的机会从中抽取一定数目的单位进行调查。从理论上讲，纯随机抽样最符合抽样调查的随机原则，因而，它也是抽样调查的基本形式。在不使用计算机的情况下，纯随机抽样通常是采用抽签的办法或利用随机数表进行抽取的。在利用计算机进行抽样调查时，计算机自动对各总体单位进行编号，通过随机数发生程序产生随机数，进行抽样，既省去了人工对总体单位编号，也省去了人工抽签或编制随机数表的繁琐劳动。

2. 分层抽样 分层抽样也叫类型抽样，就是事先将总体单位按照某种条件划分成若干层（或组），然后在各层中再按随机的原则抽取一定数目的调查单位组成样本。如果在各层中，按各层在总体中所占单位数的比重，依照比例抽取单位数，这样的抽样就叫做按比例分层抽样。在各层中抽取样本时，可采用纯随机抽样方式，也可以采用等距离抽样方式。

由于采用分层抽样方式所取得的样本，一般具有较好的代表性，所以在抽样调查实践中，这种方法得到了广泛的利用。例如，在一个县内进行农作物产量调查时，可以事先把全县的生产大队（行政村，以下简称村）按地势分为山区、丘陵和平原三个类型区，然后再按各类型生产大队（村）所占的比例确定调查的生产大队（村）数，用等距抽样法抽选。由此可见，分层抽样法是科学分组和抽样原则相结合的方法，可以抽选出具有较好代表性的生产大队（村）。分层抽