 DOI: https://doi.org/10.12349/ecin.v3i3.9665

Application of Sampling Estimation and Hypothesis Testing in the Sample Survey of Grain Production

Dedong Liu

Statistics Station of Manzhuang Town, Daiyue District, Tai'an, Shandong, 271024, China

Abstract

This study investigates the application of equal-distance sampling and hypothesis testing in corn yield estimation. Using the queueing method with independent markers and half-distance starting point sampling, we calculated the average yield per mu and total yield range of the plots. The paper addresses issues of inverted null and alternative hypotheses and incorrect rejection domain judgment in hypothesis testing. The two-tailed Z-test results (|Z|=1.78<1.96) indicate no significant difference between the sampled yield and actual yield. In conclusion, the 3-meter-long-row equal-distance sampling method demonstrates controllable error and reliable results, making it suitable for widespread application.

Keywords

equidistant sampling; sampling inference; maize yield estimation; hypothesis testing

抽样估计和假设检验在粮食产量抽样调查中的应用

刘德东

泰安市岱岳区满庄镇统计站，中国 · 山东 泰安 271024

摘　要

本文围绕玉米地产量调查，阐述等距抽样及假设检验的应用。采用无关标志排队、半距起点等距抽样，抽取样本推算地块亩均及总产量区间。修正假设检验中原假设与备择假设颠倒、拒绝域判断错误的问题，经双侧 Z 检验，|Z|=1.78 < 1.96，不拒绝原假设，表明抽样测产与实际产量无显著差异。综上，该 3 米长行等距抽样测产方法误差可控、结果可靠，可推广应用

关键词

等距抽样；抽样推断；玉米测产；假设检验

抽样估计一般指抽样推断。抽样推断是基于抽样调查的统计分析方法，通过计算样本指标推断总体数量特征，包含参数估计和假设检验两类核心问题。其以随机抽样为基础，利用样本数据估算总体参数（如均值、总量），并量化抽样误差与非抽样误差的差异。该方法广泛地被统计系统应用于各类调查工作中，例如住户调查，人口抽样调查，粮食产量调查等。在国家调查队系统的粮食产量调查中常采用的是“等距抽样”，这种方法便于操作，实用性强。

等距抽样：等距抽样又称机械抽样，先将总体中各个单位按一定顺序排列好，然后根据总体单位数和抽取单位数计算出抽选间隔 (N/n)，再按照这个距离去抽取调查单位。总体各单位排列的方法有两种：一种是不编号排列；另一种是有编号排列。所谓无关标识排列就是按照与标志值不相关的标识对总体单位进行排列，如我们要掌握职工工资情况就按工作时间长短进行排队。

间隔和排队确定后，确定抽样起点就显得尤为重要。为了便于抽取样本单位再把所有（N）单位分成若干（n）相等的小组，每一组中都有 N/n=f 个单位。等距抽样确定起点常用的有三种方法：随机起点等距抽样，半距起点等距抽样和随机起点对称等距抽样。

随机起点等距抽样。先在第 1 组中任取一个单元开始选样，再每隔一定的单元数选取一个，直到达到 n 的总数为止。这一步体现的是选样的随机性，即起始样本单元 fl 的抽取确定了后面所有样本位置的定位。若是有关标志排队，这样抽到的样本单位其标志值可能系统地偏高或偏低情况。

半距起点等距抽样。把每一小组的中点单位定位第一组的中选单位，然后每隔 f 个单位再抽选一个单位直至抽选够 n 个单位。如果是有关标志排队，这种抽样方法虽然从一定程度上避免了上述哪种随机起点等距抽样带来的不足，但也具有破坏随机原则的可能性。这种方法可操作性比较强，粮食产量检测中常被采用。

随机起点对称等距抽样。即从第一个小组中随机抽取

【作者简介】刘德东（1966–），男，中国山东泰安人，统计师，从事基础统计研究。

一个单位作为第一个样本单位而后每两个小组合并成一个大组，在每个大组中对称抽出两个样本单位，使每对样本单位与其相近的上限或下限的距离相等。若样本单位 n 为奇数时经过合并成大组将剩下一个小组，则应将这个组放在中心位置，先在这个组随机确定一个单位作为样本单位，然后在该组两边并成大组，仍用对称抽样的方法抽取样本。随机起点对称等距抽样方法，保留了前两种方法的优点，避免了他们的局限性。

关于等距抽样的抽样平均误差，采用无关标志排队的等距抽样，近似于简单随机抽样，采用的标志与调查项目无关，可以看作把总体不加任何限制的简单排列，因此，每隔抽样单位的位置都是随机的，所以计算抽样平均误差时可以按照简单随机抽样来处理。

泰安市岱岳区满庄镇华家岭村南，村民王士国耕种一块玉米地（岱岳调查队抽样点），长 60 米，宽 47.6 米，共有 73 垅玉米。现采取等距抽样，抽取 30 个 3 米长垅为样本，实割实测，推算这块玉米地的产量。

步骤和计算如下：样本间隔 = 总垅长 / 样本单位数 =60×73/30=146（米），为了便于操作取样本间隔 140 米。

长势如何是自然形成的，可认为是无关标志排队，拟采用半距起点等距抽样方法（也可随机起点），第一个样本间隔的中点选取第一个样本单位。即从 140/2=70 米处（实际在地边第 2 垅的 10 米处），前后各取 1.5 米为第一个样本单位。以后在每隔 140 米前后各取 1.5 米为一个样本单位，一直取够 30 个样本单位为止，最后一个样本位于（70+140*29）倒数第 5 垄（第 68 垄）50 米处。实割实测后得各样本的产量，样本点实测产量及分组如下表：

样本产量 Xi	样本单位数 Ni	Xi×Ni	离差 Xi-X	离差平方 Ni
2.9	4	11.6	-0.2	0.16
3.0	8	24	-0.1	0.08
3.1	9	27.9	0	0
3.2	5	16	0.1	0.05
3.3	3	9.9	0.2	0.12
3.6	1	3.6	0.5	0.25
合计	30	93		0.66

平均每个样本产量 X=93/30=3.1（斤）

样本标准差 $S=\sqrt{\frac{1}{n-1}\sum (x_i-\bar{x})^2 Ni} = \sqrt{\frac{0.66}{29}} = 0.15$(斤)

抽样平均误差 $Ux= \frac{S}{\sqrt{n}} =\sqrt{\frac{0.15^2}{30}} = 0.027$(斤)

极限误差 $Z_{a/2}* \frac{S}{\sqrt{n}}=1.96*0.027=0.053$

在 95% 的概率保证下，每 3 米长垅的平均产量的可能区间是：3.1+-Δx, 即 3.1+-0.053 在 3.047 斤至 3.153 斤之间。

每亩地平均产量 = 样本单位平均产量 * 每亩地样本单位数 =X*666.67 平方米 /（样本长度 * 平均垅宽）

=3.1*666.67/3.0*（47.6/73）=1060.34（斤）

平均亩产量的抽样平均误差 = 每亩地样本单位数 * 样本单位抽样平均误差 =666.67/3*（46/73）*0.027（斤）

=9.23 斤

则平均亩产量的可能范围是：1060.34+-1.96*9.23，即在 1042.24 至 1078.44 斤之间。

整块玉米地的面积是：60*47.6/666.67=4.2 亩。

则整块玉米地的总产量区间范围是：4377 斤至 4529 斤之间，概率为 95%。

以上是“3 米长行”抽样调查法的调查结果，能否利用这一结果推断（代替）整个地块的产量情况，进而这种方法是否可以进一步推广，我们需要进行一下“假设检验”。

统计假设检验又称 假设 检验，显著性检验 是最常见也是最基本的 统计 推断方法，即先对总体提出某种假设，然后通过对样本的分析来判断这个假设应被拒绝还是接受的过程。常用的假设检验方法有 Z 检验法、t 检验法、卡方检验法、F 检验法等 [1]。

基本思想及原理是假设检验的基本———“小概率事件”，其是一种具有一定概率特征的反证法。小概率思想是指在一次试验中，小概率的事件基本上不发生。而反证法的思想就是先提出所要验证的假设 H0，然后使用适当的统计方法利用小概率原理验证该假设成立与否。即我们是要对假设 H0 进行检验看它是否正确，首先要假定假设 H0 成立，再根据样本观测值来接受还是拒绝该假设。如果样本的结果导致了“小概率事件”，拒绝接受 H0；否则就接受 H0【1】。

这里所谓的小概率事件，是“在一次试验中几乎不发生的事件”，一般以 α 表示这个概率 (0<α<1)，称为检验的显著性水准。对于不同课题、问题，α 水准可能不同，一般认为小于 0.1、0.05 或 0.01 等的事件可被判定为“小概率事件”[1]。

基本步骤

提出检验假设又称无效假设，符号是 H0；备择假设的符号是 H1【2】。

H0：样本与总体或样本与样本间的差异是由抽样误差引起的【2】；

H1：样本与总体或样本与样本间存在本质差异【2】；

预先设定检验水准为 0.05；将备择假设当作真的情况下将其拒绝的概率，称为第一类错误概率 α，通常取 α =0.05 或者 α=0.01【2】。

明确了统计方法之后，我们可以通过样本观察值得到对应公式的统计量估计值，如 x2 值、t 值等，按照所得信息类型及特征可以选用 z 检验、t 检验、秩和检验以及卡方检验等方式展开检验工作。

根据观测结果的数据量及其分布确定检验假设成立的概率 P 值，判断最后的结论。若 P > α，则认为在给定的 α 水平上两组之间差异不显著，即接受 H0，也可以说是由于抽样误差造成的，在统计上并无意义。但是，如果 P≤ α，

则认为我们得出的结果是在假设水平上存在显著性差异，并拒绝 H0 接受 H1，说明这种差异不大可能是由抽样的偶然性所致，而是由于实验条件不同所造成，在统计上是允许存在的。P 值的大小一般可通过查阅相应的界值表得到[2]

实际收割后，整个地块的实测产量是 4465 斤，亩产 1063 斤。

下面检验一下这种“3 米长行”测产方法是否可行：我们需要进行一下假设检验，原假设 H0：U=U0=1063

备择假设 H1：U ≠ U0=1063

统计检验量 $Z = \frac{X-U0}{S/\sqrt{n}} = \frac{1060-1063}{9.23/\sqrt{30}} = -1.78$

Z 的绝对值为 1.78 ， 而 Za/2= 1.96 (95% 的概率)

因此：1.78 < Za/2

Z 的统计量值没有落在拒绝域内，所以不能推翻原假设，即原假设成立，在一定概率保证下抽样测产 1060.34/ 亩，与实割产 1063/ 亩是一致的。

综上所述，这种方法是可以推广和应用的。

参考文献

[1] 沈南山著,数学教育测量与统计分析,中国科学技术大学出版社,2017.01,107-108

[2] 陆克斌，崔久波主编；万志峰，董西红，石丽，沈菊，钟妙副主编；薛强主审,市场调查与预测,教育科学出版社,2013.08,204

[3] 蔡鹏伦，关于粮食大县粮食播种面积和产量抽样调查方法的研究——以浙江省××市为例，现代经济信息，2015（12）,353