第二篇　森林抽样调查

第九章　森林抽样调查基本知识

第一节　总体样本和样本组织方法

一、什么是森林随机抽样调查

从调查的对象中，按照予定的精度，随机地抽取一定数量的样地或样木进行实测，用以估测全林。这种方法简称森林抽样调查。

森林抽样调查与标准地调查法（典型选样）主要不同之处，是前者能较好的排除主观意识的影响，抽到尽可能客观地反映调查总体特征的样本，且对总体反映的准确程度，可以根据概率理论加以估计。我国五十代代　始，延续至今的标准地调查法，实践证明，对于已积累相当丰富调查经验的人说来，能够获得很高精度的森林资料，但就一般人，受经验的局限性和主观意识的影响，所得资料偏高、偏低有之，且往往偏高。尤其是事先有一定精度要求，并能按精度科学地安排调查工作，事后又能给出调查精度的可靠估计。在这方面，标准地调查法很难做到，而且也缺乏理论根据。森林抽样调查是以数理统计理论方法为依据。调查前，按照予定的精度去计划安排工作；之后，能对森林调查结果所达到的精度、可靠性做出准确　的估计。面对全国、全省大面积森林，调查任务十分繁重，抽样调查所具有的优点就能显得重要。它是一种科学的森林调查方法；它能够有计划地节约人力、物力和时间。这也是目前存在着的标准地调查被抽样调查逐渐代替趋势的关键。笔者认为，对此要作具体分析，千篇一律效果不见得好，所以，二者不可偏　，更何况，典型选样本身有很多优点，特别是易掌易普及，这实际上还涉及到一个现有基础理论水平问题。

二、总体和总体单元

1. 概念

人为划定调查研究的全部对象叫总体；而构成总体的一个个单位叫做总体单元。在统计概念上，全部单元构成总体，总体是全部单元的集合体。关于总体和总体单元的确定，有如下两点基本要求：

一个总体各单元属于同一类；

一个总体内各各单元在统计意义上只有一项特征不同。

如：在森林抽样调查中，为测算森林蓄蓄积量，而把总体划分为许多面积相等的(如面积均为 0.06 公顷）单元。只能划为相等的，不能划为不等的单元，因为在统计意义生上不能存在由面积大小不等的单元所组成的总体。至于各各单元的蓄积量，则是不相等的，是变量，是所谓只有一项特征不同。

又如：调查某一林分平均高时，林分内每株树高都是一个单元，树高标志着各各单元属于同一类，不能使树高与胸径或其它因为子掺杂并立地单元。而且每个单元树高值是一随机变量，是不相等的，也即只有一项特征不同。这样一来，就把森林划为具有予定变量的总体，以便统计估测。由此，也可以进一步理解到，对于同一片森林（同一调查对象），因调查要求的具体项目不同，可以确定为许多不同的总体，而且相互之间不能混淆。

2. 有限总体与无限总体

组成总体的单元数是有限的，称之为有限总体。如在有 5 0 0 株林木的标准地内，无论测其平均高或平均直径，这 5 0 0 株树构成的总体为有限总体。假如欲测算的不是林木平均高或平均直直径，而是测算平均叶面积，这时如以每个树叶面积为单元，可以想象得出，总体单元数目是难以查清的，这样即构成了无限总体。在实际工作中，

只要单元数目相当多，就可当无限总体对待。

3. 总体参数

描述总体特征的数值称之为总体参数。参数必须"充分"。参数便于不同总体之间进行比较。总体最重要的参数为单元观测值、算术平均值、标准差及变动系数等等。在"数理统计"书中已详细论述，不再重复。

三 样本和样本单元

森林抽样调查是一种以样本调查所获得的数据进而推断全体的一种调查方法。

抽样就是从总体中随机地抽取一部分单元进行观测。被抽中的这部分单元的总体称之为样本。样本单元也是总体单元，是被抽中的总体单元。而总体单元却不一定是样本单元。

样本和样本单元的统计概念无须深述，这里仅就实际工作中需要明确的问题进行讨论。

1. 抽样必须遵守的条件，是其抽取过程必须是随机的；决不能以调查方便的单元代替调查不方便的单元。

所谓随机的，主要指消除和摆脱主观意识的影响。如属于典型选样的标准地调查法就难以做到这一点。诸如标准地、标准木等都是人为的选取林分或林木的代表，"林分缩影"和平均木等。这种人为典型选样法，尽管也极力遵重客观实际，但选取的结果总是有意无意地带有偏见，至于有偏或无偏乃至偏差多大却无法确定，至少不能象随机抽样那样能够进行有根据的估计。这里所述，可一 概括，即抽样要严格遵守随机原理并消除任何偏见。在实际抽样调查方案设计中，将总体划分为相等大小的单元，使总体中每个单元都有一个已知机会可作样本；采取随机抽样方法，以保证总体每个单元被抽中的机会均

等。这两点是抽样调查调查必须严格遵守的先决条件。

2. 样本的组织方法。先决条件既已明确，而据此提出的基本抽样方法主要有三种：一是纯随机抽样（包括抽签法、随机数字法和经验数字法）；二是机械抽样，又叫系统抽样；三是整群抽样。这三种基本抽样方法，在实际应用中又派生多种形式的组织样本的方法，在后面章节中再进行讨论。

四 样本大小

样本单元个数 n≥50，可称之为大样本；n＜50为小样本。

在总体平均数估计中，决定样本大小的理论根据，涉及如下定理：

1. 如果总体频率分布是正态分布，则无论样本单元数多少，样本平均数的概律分布必定是正态分布；

2. 如果总体频率分布不是正态分布，则在一般情况下，只要样本单元数充分大（n≥50），样本平均数的概率分布接近正态分布。

在森林抽样调查中，如果予先不了解总体的分布律，一般多采用大样本进行总体平均数估计，而正态分布是其主要根据。当确知总体频率分布为正态分布或近似正态分部时，应采用小样本资料，根据学生氏（t分布）分布，以90%以上的可靠性对总体平均数进行估计。

第二节 森林抽样调查的特征数

一、最佳估计值

样本特征数称之为统计量，它是抽样总体参数的估计值。

无偏性、有效性和一致性是对总体参数估计值的基本要求。一般情况下，总体参数是未知数，用样本特征数（$\bar{x}$、S^2）做为相应总体参数的估计值，这里就存在一个估计值好坏问题。以无偏性、有效性和一致性为标准，如总体参数与它的估计值为最佳估计值。

1. 无偏：如样本平均数的平均数为 $\bar{x}_{\bar{x}}$，总体平均数为 u，总体频率分布为正态分布。

因 x_i，$\bar{x}_j$ 属同分布，

所以 $\bar{x}_{\bar{x}} = u$

即样本平均数的平均数是总体平均数的无偏估计值（证明从略）。

当随机变量 x_i 的精确分布难以求出时，n次调查结果所得几个（$\bar{x}$）平均值，当 $n\to\infty$ 时

则 $$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i = u$$

则 $$E(\bar{x}) = u$$

所以样本平均值的数学期望值是总体平均值的无偏估计值。无偏是表示平均的看没有偏差的意思。

非无偏的估计值可举总体方差 σ^2 的估计值 S^2 为例：

$$ES^2 = \frac{n-1}{n} б^2 \neq б^2$$ 是有偏的

如进行修改，使 S^2 成为 S^{2*}

$$S^{2*} = \frac{n-1}{n} S^2$$

则 S^{2*} 是无偏的。

2. 有效性：如果随机抽取一个大样本，并用予定的统计量（如算术平均值）进行估计，这个总体总数的估计值将有特定的方差。如统计量是"有效"的，则随着样本的大小的增加，误差的分布将趋于正态分布。因此，统计量可以依据其平均值的标准差越小，则估计越有。所以统计量的有效性可由方差大小来说明。

3. 一致性：如果样本逐步增大，则参数的估计值将趋于与总体参数相等的一个固定值。能符合这种要求的估计值称为一致估计值。

上述最佳估计值的标准说明如下问题：

① 进行森林抽样调查时，以样本特征数作为总体参数估计值要求满足一致、无偏所要求的条件，即予知总体的分布为正态分布。如在难以确定总体分布情况下，必须𠃋大样本单元数，否则无法进行估计。

② 在满足一致、无偏的条件下，样本特征数（如平均值、方差）是相应总体参数一致、无偏的估计值，业已严格证明，无需过虑。

③ 样本平均数的数学期望值（ $\bar{x}_{\bar{x}}$ ）是总体平均数的 u 的一致、无偏估计值，样本平均数 $\bar{x}$ 的分布，只要总体单元数 N 和样本元数 n 都足够大，而 n / N 不太大，则可以证明它是平均值为 $\bar{x}$ ，方差为 $б_{\bar{x}}^2$ 的分布。样本平均数的这两个性质是抽样误差估计的理论依据。

显然，如要设计出一个符合要求的调查方法，对类似上述一些数理统计的基本知识，必须熟知。

二 森林抽样调查的几个特征数

(一) 样本平均数

抽样调查中是用样本平均数 $\bar{x}$ 作为总体平均数 u 的估计值。其计算公式：

$$\bar{x}=\frac{\sum_{i=1}^{n}x_i}{n}$$

式中：$\bar{x}$ 样本平均数；

x_i 第 i 个单元的观测值；

n 样本单元数。

如按分组数据统计时，则

$$\bar{x}=\frac{1}{\sum_{i=1}^{h}f_i}\sum_{i=1}^{n}f_ix_i=\frac{1}{n}\sum_{i=1}^{n}f_ix_i$$

式中：h 划分的组数；

x_i 第 i 组组中值；

f_i 第 i 组频数；

n 样本单元数， $n=\sum_{i=1}^{h}f_i$。

(二) 标准差及方差

标准差及方差，均表明单元观测值对平均数的离散程度。

总体方差定义为：

$$\sigma^2=\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$$

总体方差及标准差一般是未知数，只能通过样本资料进行估计。通常样本方差小于总体方差，经修改后方为总体方差无偏估计值（S^2）。

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

标准差：

$$S=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$S=\frac{1}{n-1}\left[\sum_{i=1}^{n}x_i^2-\frac{1}{n}\left(\sum_{i=1}^{n}x_i\right)^2\right]$$

变动系数C：标准差的相对值即变动系数。其估计值为：

$$C=\frac{S}{\bar{x}}\times 100$$

(三) 标准误：用样本（一个平均值去估计总体平均值所产生的差异，这个差异用标准误表示：

$$S_{\bar{x}}=\frac{S}{\sqrt{n}}$$

式中：S 总体方差估计值

n 样本单元数

$S_{\bar{x}}$ 标准误

四）估计误差限及估计区间

假定总体频率分布为正态分布，总体平均值为u，总体标准差为σ时，样本平均值（$\bar{x}$）落在

$(u-\sigma, u+\sigma)$ 的概率为68.3%

$(u-2\sigma, u+2\sigma)$ 的概率为95.4%

$(u-3\sigma, u+3\sigma)$ 的概率为99.7%

$(u-1.96\sigma, u+1.96\sigma)$ 的概率为95%

总体方差σ在抽样调查中用$S_{\bar{x}}$代替，这样：

$\bar{x} \pm S_{\bar{x}}$

$\bar{x} \pm 1.96S_{\bar{x}}$

$\bar{x} \pm 2S_{\bar{x}}$

$\bar{x} \pm 3S_{\bar{x}}$

$\bar{x} \pm t\Delta_{\bar{x}}$

所示范围，叫做估计区间。

1、1、1.96、2、3……等数值用t表示，称之为可靠性指标。而$\Delta_{\bar{x}}$为绝对误差限，以某衡量估测误差的大小。

$$\Delta_{\bar{x}} = tS_{\bar{x}} = t\frac{S}{\sqrt{n}}$$

并记 $E = \frac{\Delta_{\bar{x}}}{\bar{x}} 100\% = \frac{tS_{\bar{x}}}{\bar{x}} 100\%$ 为相对误差限。

从误差计算公式可知，抽样误差的大小，取决于可靠性、样本标准差及样本单元数量因子。当可靠性要求一定，若样本标准差越大而样本单元数越少，则估测精度越低，反之则高。

用样本估计总体值时，不仅要求算出估测误差限，同时要求给出这一误差限的可靠程度，即总体平均值（u）落在$\bar{x} \pm tS_{\bar{x}}$区间的把握性大小。把握性大小用可靠性指标t值表示。t采用小样本时，根据要求的可靠性，按自由度（$n-1$）由“t分布表”查得。

当$S_{\bar{x}}$不变时，要求可靠性越大，则估测区间越大，对估测的意义越小；要求可靠性越小，则估测区间也越小，虽估测精度较高，但这种精度的可靠性小，因此估测意义也就不大。

抽样调查要求在确定最合适的可靠性下获得最大估测精度，所以进行森林抽样设计时，不要把样本单元（n）数、抽样精度（$P_{(\%)}=1-E_{(\%)}$）、可靠性指标（t）三者分割，应同时考虑。

(五) 有限总体改正项

所谓估测值误差限，是指以样本实测值去估测总体时所存在的差异。如不考虑样本实测值本身的系统误差时，估测误差是指没有进行实测的部分，即$N-n$的部分。而我们所计算的误差$\triangle_{\bar{x}}=t \cdot S_{\bar{x}}$包括了样本所这一实测部分，即扩大了误差，因此在计算时应该减去实测部分，这一部分叫有限总体修正项，用$\sqrt{1-\frac{n}{N}}$来表示。加修正项的误差计算公式为：

$$\triangle_{\bar{x}}=t \cdot S_{\bar{x}} \cdot \sqrt{1-\frac{n}{N}}$$

标准误的计算式为

$$S_{\bar{x}}=\sqrt{\frac{S^2}{n}\left(1-\frac{n}{N}\right)}$$

式中：　$1-\frac{n}{N}$有限总体修正项；

$\frac{n}{N}$为抽样比，用"f"表示，$f=\frac{n}{N}$。

一般总体抽样较大，则N很大，而样本单元不多，抽样比f常是很小，所以修正部分的误差值也是很小的，此种情况可不必修正。

至于抽样比f多大才可不进行修正？这是人为规定的，一般说：

当抽样比$f=\frac{n}{N}\geq 0.05$时，作有限修正；

抽样比$f\leq 0.05$时，可不必作有限修正。

第三节 有关森林抽样调查的若干问题

一、抽样调查必须予先确定总体范围，否则将影响精度和效率。总体范围大小，应根据调查目的、经济条件、经营水平，结合行政界或自然境界等进行区划。一般说来，总体范围越大，抽样调查工作效率越高。

二、样本单元的抽取，必须遵守随机的原则，这是抽样调查不同于典型选样的关键所在。抽样调查要求按予先设计的概率抽取，所设计的概率可以是相等概率，也可以是不等概率的。一般调查要求是等概的，即各单元被抽中的机会相等。在实际工作中，等概率是相对的，绝对等概是难以做到的。有时因照片比例尺和 影差影响，抽取抽样地和规定地点时会有误差；有时样地落在不能 登的 坡上而被舍去以及补点时会所遭成的不等概等，都多少的影响了等概原则。这些影响除设法使其减少到最小限度外，重要的是不能带有主观因素去选择样地。

三、已予知总体为正态分布或趋于正态分布，抽取大样本或小样本均可。采用小样本可减少工作量。如不知总体频率分布，应采用大样本。

四、抽样调查的误差，前者所讲的只是抽样误差，不包括样地测定误差、面积误差，前者所讲的只是抽样误差，不包括样地测定误差、面积误差、材积表误差……等等。因为这些误差不是抽样本身带来的，用其它方法调查也同样存在。

第三节 有关森林抽样调查的若干问题

一、 抽样调查必须予先确定总体范围，否则将影响精度和效率。总体范围大小，应根据调查目的、经营条件、经营水平，结合行政界或自然境界来进行区划。一般说来，总体范围越大，抽样调查工作效率越高。

二、 样本单元的抽取，必须遵守随机的原则，这是抽样调查不同于典型选样的关键所在。抽样调查要求按予先设计的概率抽取，所设计的概率可以是等概率，也可以是不等概率。一般调查要求是等概的，即各单元被抽中的机会相等。在实际工作中，等概率是相对的，绝对等概是难以做到的。有时因照片比例尺和投影差影响，抽取抽样地和现地定点时会有误差；有时样地落在不能攀登的陡坡上而被迫舍去以及补点时所造成的不等概等，都多少的影响了等概原则。这些影响除设法使其减少到最小限度外，重要的是不能带有主观因素去选择样地。

三、 已予知总体为正态分布或趋于正态分布，抽取大样本或小样本均可。采用小样本可减少工作量。如不知总体频率分布，应采用大样本。

四、 抽样调查的误差，前者所讲的只是抽样误差，不包括样地测定误差、面积误差、材积表误差……等。因为这些误差不是抽样本身带来的，用其它方法调查也同样存在。