

第十章 简单随机抽样

第一节 简单随机抽样

一、简单随机抽样概念和计算公式

从含有 N 个单元的总体中，随机等概率地抽取 n 个单元组成样本，用来 n 个单元的总体的方法，叫简单随机抽样。

简单随机抽样应用的计算公式如下：

1. 样本单元数的确定

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

式中： n 样本单元数

t 可靠性指标，当可靠性规定为 95% 时， $t = 1.96$ ，95-4% 时， $t = 2$ ； σ_x^2 总体方差； Δ 绝对误差。如用相对误差

$$E = \frac{\Delta}{x} \times 100 \text{ 时，则}$$

$$n = \frac{t^2 C^2}{E^2}$$

如果抽样比 $f \geq 0.1$ 时，则

$$n = \frac{t^2 C^2 N}{E^2 N + t^2 C^2}$$

式中： C 总体变动系数； N 总体单元数；

E 相对误差， $E = 1 - p\%$ ； $p\%$ 调查要求达到的精度。

2. 样本平均值

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

3. 总体方差的估计值

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

4. 标准误 $s_{\bar{x}}$

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}}$$

5. 相对误差限 (相对误差)

$$E\% = \frac{\Delta \bar{x}}{\bar{x}} \times 100$$

6. 估计精度

$$P\% = 1 - E\%$$

7. 估计值误差 (抽样绝对误差)

$$\Delta_{\bar{x}} = t \cdot s_{\bar{x}}$$

8. 总体平均值估测区间

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

9. 总体总值的估测区间

$$N(\bar{x} \pm t s_{\bar{x}})$$

二 简单随机抽样的工作步骤

简单随机抽样是森林抽样调查的最基本方法，可视为各种抽样调查的基础。这里试以简单随机抽样调查为例，着重说明进行森林调查设计时各各环节所涉及的问题。以后各章与之相同之处从略。

（一）确定抽样总体

抽样调查是用样本估计总体。在设计抽样调查方案时，应先确知调查对象的大小和划定其范围界限，即确定总体和总体单元。

正规的抽样调查，先确定总体各单元的位置，而后进行抽样。在森林调查实践中，往往因为总体单元数太大（例如，一个200公顷的森林总体，以0.06公顷为一个单元， $n = 3333$ 个），只要保证组织样本时不致产生重叠或遗漏现象，就必须把每个单元位置事先一一确定出来。至于具体布点方法将在后面讲述。

从抽样精度来看，样本单元数的多少与总体大小无关，在变动系数相同时，总体越大，抽样效率越高。

从生产要求来看，抽样调查的目的不同，其总体大小也应不一样。例如，森林资源清查（Ⅰ类调查）可用一个省或更大区域做为总体；规划设计调查（Ⅱ类调查）可以作业区、林班等为总体。

（二）确定样地大小和形状

样地的形状和大小（面积）的选择是森林抽样调查设计一项重要内容。凡样地形状和大小选择适当者，都可在一定程度上提高效率。

1. 样地形状 一般采用圆形或方形。

（1）圆形样地：适用于地形平坦，林内通视良好，样地面积不大的情况下。

圆形样地的设置，可用样地中心向八个方向用测绳或测竿按半径确定样地的边界（图10—1），并在外围林木标以记号。

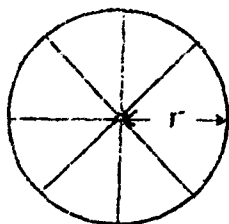


图 10—1 园形样地设置

样地位置坡度起伏明显时，应按八个实测方向的坡度，本着斜距改水平距的原理进行改正，以确保园形样地面积的准确。实测时不需计算，有相应数表供查用。

坡度不明显地段设置样地，样地面积的改正可用下式表达：

$$r'^2 \pi = r^2 \pi \frac{1}{\cos \theta}$$

其垂直坡面最大坡度的样地半径为：

$$r' = r \sqrt{1 / \cos \theta} \quad (\text{参看表 } 10-1)$$

设置样地时，要求测定该坡面垂直方向最大坡度（ θ ），用以确定样地半径，其水平方向半径应沿等高线量测，不需进行改正。

(2) 方形样地：具有设置简便、边界明确、灵活性大等特点应用较为普遍。

方形样地一般是沿测线菱形设置（如图 10—2），将即由中心点向外引出两条对角线，定出样地的边界。在坡地上测设时，要根据坡度对对线进行改正（可直接查森林调查常用表）

2. 样地大小 样地大小的确定，实质上就划分总体单元。总体单元划分的越大（单元面积）各单元变动越小。变动系数随着单元面积的增大而减小，而当单元增大到一定程度时，变动系数趋于稳定（图 10—3）。最适宜的样地面积是变动系数趋于稳定处，如图

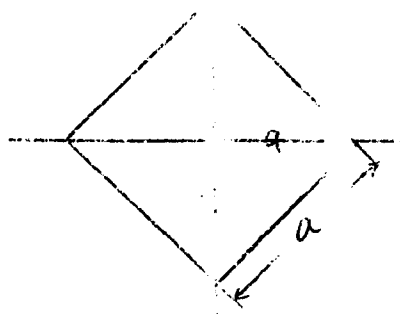


图 10—2 方形样地设置意图

表 10—1 不同坡度时圆形样地半径

面积 (公顷)	半径 (米)	不同坡度时的半径 (米)							
		5°	10°	15°	20°	25°	30°	35°	40°
0.005	3.99	4.00	4.03	4.06	4.12	4.19	4.29	4.41	4.56
0.01	5.64	5.65	5.69	5.74	5.82	5.93	6.06	6.23	6.45
0.02	7.98	7.99	8.04	8.12	8.23	8.38	8.57	8.82	9.12
0.04	11.28	11.31	11.37	11.48	11.64	11.85	12.13	12.47	12.89
0.06	13.82	13.85	13.93	14.06	14.25	14.52	14.85	15.27	15.79
0.08	15.96	15.99	16.08	16.24	16.46	16.76	17.15	17.63	18.23
0.10	17.84	17.88	17.98	18.15	18.40	18.74	19.17	19.71	20.38

，样地面积以 0.06 公顷最为适宜。

样地的大小决定于调查方法、林分年令、变动系数及交通条件等因素。例如，幼龄林的蓄积量很少，株数很多，样地面积过大，检尺工作量很大，所以在森林调查中可根据林分的年令及特点确定样地的

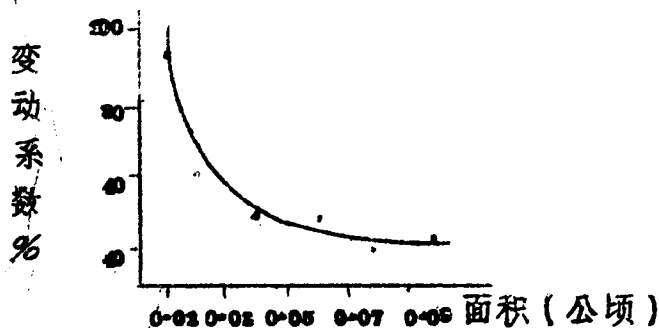


图 10-3 变动系数随面积变化系曲线

的大小。在一般情况下多采用 0.06 公顷的样地，林分变化大时 0.1 公顷；而幼龄林采用 0.01 公顷较为适宜。

3. 确定样本单元数

有限总体样本单元数按下式计算：

$$n = \frac{t^2 \sigma^2 N}{E^2 N + t^2 \sigma^2}$$

若为无限总体 ($f = \frac{n}{N} \leq 0.05$ 时)，则按下式计算：

$$n = \frac{t^2 \sigma^2}{E^2}$$

森林抽样调查的估计误差，一般情况要求。

为保证估计精度，常把由公式计算的样本单元数 n ，增加 10—20% 的安全系数。

森林抽样调查的估计误差，一般情况下要求不超过 5—10%，并以 95% 以上的可靠性进行估计，所以上式中 E 、 t 均视调查要求的可靠性和精度而估计的。其中可靠性指标 t 值，如为大样本，则据要

求的可靠性查概率积分表()得,如为小样本,则按自由度 $n-1$,由 t 分布表中查出。

前式中总体变动系数较难确定。如无以往调查资料可供参考,可利用航测照片判读,并结合踏查对照估计 σ 值。而较准确的方法是通过予备调查确定 σ 值。

总体变动系数 σ 值,常采用极差法,其计算公式为:

$$\sigma = \frac{x_{\text{大}} - x_{\text{小}}}{6\bar{x}} \times 100\%$$

式中: $x_{\text{大}}$ 总体单元中最大观测值
 $x_{\text{小}}$ 总体单元中最小观测值
 $\bar{x}$ 总体平均数估计值

上式是根据正态分布推导出来的。在正态分布的总体内,落在 $(\bar{x} \pm 3\sigma)$ 区间的总体单元内为 99.7%,落在区外 0.3% 的单元数是个别的极大、世极小的单元。

令 $(\bar{x} + 3\sigma)$ 的单元值为 $x_{\text{大}}$, $(\bar{x} - 3\sigma)$ 的单元值为 $x_{\text{小}}$ 。

则 $(\bar{x} + 3\sigma) - (\bar{x} - 3\sigma) = x_{\text{大}} - x_{\text{小}}$

$$\therefore \sigma = \frac{x_{\text{大}} - x_{\text{小}}}{6}$$

$$\sigma = \frac{\sigma}{\bar{x}} \times 100\% = \frac{x_{\text{大}} - x_{\text{小}}}{6\bar{x}} \times 100\%$$

由于在予备调查中, $x_{\text{大}}$ 和 $x_{\text{小}}$ 常常不一定正好选出最大、最小值,有人主张不除以 6,而除以 4.6,即

$$= \frac{x_{\text{大}} - x_{\text{小}}}{4.5} \times 100\%$$

此式可在预备调查中试用。

(四) 布点

根据样地大小确定点间距，若样地面积为 a ，样地形状为正方形，则点间距 L 为：

$$L = \sqrt{a}$$

用一块较密的网点板，使肉眼的面积按布点图的比例计算等于样本单元的面积。将网点板复在总图上，在总图范围内将每个网点进行编号。

用随机数字表，按计算的样本单元数抽取样本，并将被抽中的样本单元刺在布点图上或再刺到在航空相片上。

(五) 外业调查

在地形图或航空相片上，确定各样地中心点与某一明显地物的方向和距离。在实地找到该地物后，按既定的向和距离引点，以此点为样地中心点设置样地，然后进行测树调查。

(六) 内业分析

内业分析，主要是应用本节“一”所列公式对外业取得的材料进行整理和计算，估计出总体平均值的区间值，这也是调查的最后成果，至于估计误差、可靠性和精度，与区间值不可分割。缺此得不出区间值，而且也会使成果无意义。为说明内业分析计算步骤，兹举一例“假设以某林为总体，其林地总面积 235 方公顷，随机抽取 35 个 0.06 公顷的样地组成样本，其调查结果见表。

1. 总体平均数的估计值

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{35} \times 285.24 = 8.15$$

简单随机抽样计算表

样地号	样地蓄积	样地号	样地蓄积	样地号	样地蓄积	样地号	样地蓄积	样地号	样地蓄积
1	7.12	9	9.29	17	9.22	25	7.01	33	10.36
2	9.83	10	9.39	18	11.02	26	8.34	34	10.14
3	7.70	11	5.09	19	8.20	27	9.18	35	11.15
4	7.65	12	7.33	20	5.17	28	6.13		
5	8.76	13	6.92	21	5.20	29	10.40		
6	9.96	14	7.18	22	6.20	30	11.43	Σ	285.24
7	12.35	15	8.23	23	7.45	31	8.49	Σx_i	2450.3296
8	8.02	16	6.37	24	6.04	32	4.72		

2. 总体方差的估计值:

$$s_x^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i \right)^2}{n} \right) = 3.6972$$

3. 标准误:

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \sqrt{\frac{3.6972}{35}} = \sqrt{0.1056} = 0.325$$

此例是按无限总体计算标准误的, 因为抽样比 $f = \frac{n}{N} = 0.01 < 0.05$

4. 估计误差 (可靠性为 95%)

$$\Delta_{\bar{x}} = t \cdot S_{\bar{x}} = 1.96 \times 0.325 = 0.64$$

可靠性指标是：样本查概率积分表得 z 值（一般以 $n \leq 30$ 作为大小样本的划分。本例 $n = 35$ ，看作为大样本）。

5. 相对误差

$$E_{\%} = \frac{\Delta_{\bar{x}}}{\bar{x}} \times 100\% = \frac{0.64}{8.15} \times 100\% = 7.85\%$$

6. 估计精度

$$P_{\%} = 1 - E_{\%} = (1 - 0.0785) \times 100\% = 92.15\%$$

7. 置信区间

$$\bar{x} \pm \Delta_{\bar{x}} = 8.15 \pm 0.64 = 7.51 \sim 8.79$$

8. 林地蓄积量估计量

每公顷蓄积量：

$$M/h_a = \frac{\bar{x}}{0.06} = 135.8 \text{ (m}^3\text{)}$$

总蓄积量

$$M = 135.8 \times 235 = 2507 \text{ (m}^3\text{)}$$

总蓄积量的估计误差

$$\Delta_M = \frac{\Delta_{\bar{x}}}{0.06} \times 235 = \frac{0.64}{0.06} \times 235 = 2507 \text{ (m}^3\text{)}$$

总蓄积量置信区间

$$M \pm \Delta_M = 29406 \sim 34420 \text{ (m}^3\text{)}$$

此例如以 $n \leq 50$ 时视为小样本，则 z 值用自由度 $n-1$ ，由 z 分布查得 z 值为2.03。此时估计误差：

$$\Delta_{\bar{x}} = 0.10 \times 0.325 = 0.0325$$

$$E_{\%} = 8.0\%$$

$$E_{\%} = 9.0\%$$

置信区间

$$\bar{x} \pm \Delta_{\bar{x}} = 19.8 \pm 0.0325$$

$$\Delta_{\bar{x}} = 2.0\%$$

总体蓄积量的置信区间

$$N \pm \Delta_N = 2930 \pm 34498$$

由此可以看出，当 $n > 30$ 时，视为大样本划分标准是可以的，本例两者计算结果相近，也说明了这一点。

三、对简单随机抽样方法的评价

简单随机抽样是基本的抽样方法，其他抽样方法实质上是它的各种变形。这种抽样方法要求总体的分布是正态分布，当总体不是正态分布，只在样本单元数充分大时（ $n \geq 50$ ），样本平均数的分布近于正态分布，可以对总体用此法进行估计。当样本单元数较小，采用此法应注意总体的分布，当总体不是正态分布时，就会使估计有偏。

简单随机抽样在样地或样点抽取方面，准备工作与设计工作比较麻烦。另外，样地或样点分布分散且定位比较困难。

第二节 成数抽样

在森林调查工作中，全林总面积一般是已知的（利用地形图或进行面积测量），而各地类所占面积是未知数（即使测量也是比较困难的），可以用成数抽样法估测地类面积。成数抽样法还可以用于造林成活率、种子发芽率及林木病腐率等等方面。

一、成数抽样概念和计算公式

1.1 概念

总体中具有其特点的单元数与总体单元数的比值，叫做具有某特点单元的总体的成数。具有某特点的单元数为 M ，则具有某特点的总体的成数为：

$$p = \frac{M}{N}$$

在 N 个单元的总体的中，随机抽取 n 个单元组成样本，样本中具有某特点的单元数为 m 与样本单元数 n 的比值，叫做具有某特点的单元的样本成数：

$$p = \frac{m}{n}$$

用样本成数估计总体成数的方法叫做成数抽样

总体成数的抽样估计，是总体平均数抽样估计的特别，若总体或样本中任意单元 x_i 具有某特点定值时为“1”，否则为“0”，

$$\bar{x} = \sum_{i=1}^N x_i = 1 \times M + 0 \times (N - M) = M$$

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

$$m = x_1 + x_2 + \dots + x_n = \sum_{i=1}^n x_i$$

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

式中:

$\bar{x}$ 总体平均数

$\bar{x}$ 样本平均数

M 具某点 S 特点总体单元数

m 具某特点样本单元数

n 样本单元数

N 总体单元数

P, p 分别为总体和样本成数

2. 计算式

样本成数

$$p = \frac{m}{n} \quad \text{是总体成数无偏估计值。}$$

总体估计值方差

$$\begin{aligned} s_x^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \\ &= \frac{np - np^2}{n-1} = \frac{p}{n-1} (1-p) \end{aligned}$$

标准误

$$S_{p重} = \sqrt{\frac{p(1-p)}{n}} \quad (\text{重复抽样时})$$

$$S_{p非} = \sqrt{\frac{p(1-p)}{n} \left(1 - \frac{n}{N}\right)} \quad (\text{非重复抽样时})$$

误差限

$$\Delta_{p重} = t \cdot S_{p重} = t \sqrt{\frac{p(1-p)}{n}}$$

$$\Delta_{p非} = t \cdot S_{p非} = t \sqrt{\frac{p(1-p)}{n} \left(1 - \frac{n}{N}\right)}$$

样本单元数的确定

$$n_{重} = \frac{t^2}{\Delta_{p重}^2} \{p(1-p)\} \quad (\text{重复抽样样本单元数})$$

$$n_{非} = \frac{n t^2 p(1-p)}{\Delta_{p非}^2 N + t^2 p(1-p)} \quad (\text{非重复抽样样本单元数})$$

用相对误差表示时

$$E_{p重} = \frac{\Delta_{p重}}{p} = \frac{t}{p} \sqrt{\frac{p(1-p)}{n}}$$

$$n_{重} = \frac{t^2 (1-p)}{E_{p重}^2 p} \quad (\text{重复抽样样本单元数})$$

$$n_{非} = \frac{t^2 c_p (n-N)}{t^2 c_p^2 + E_{p非}^2 N}$$

式中：

E_p 相对误差限

C_p 变动系数

下表给出 $p = 0.1-0.9$ 范围内变动系数和相对误差限为 10% 时所需样本单元数。

表

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
C_p	0.09	0.16	0.21	0.24	0.25	0.24	0.21	0.16	0.09
σ	0.30	0.4	0.46	0.49	0.50	0.49	0.46	0.40	0.30
C_p	300	200	154	123	100	82	66	50	33
n	3600	1600	933	600	400	267	171	100	44

上表表明： $p = 0.5$ 时方差最大，由它向两边的方差对称；在 $p = 0.3-0.7$ 范围内方差和标准差的差异很小，故在此范围里，样本大小相等时，用成数估计值将有近似的标准误；变动系数的变化幅度较大，尤其当 p 小时，表中还说明成数抽样要达到一个满意的精度需要一个很大的样本。

三、用成数抽样确定各类型面积的方法

用成数抽样确定各类型面积的方法有三种：样点法、地块法和截距法。这里仅介绍前二种。

1. 样点法：在总体内随机或系统抽取样点，用以估计总体里各类型面积成数叫做点法。

(1) 确定样点数。用样点法确定各类型面积，一般所需样点数很多，在非重复抽样条件下亦可视为重复抽样。所以，可用下式计算样本单元数：

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

式中： E_p 相对误差限
 t 可靠性指标

式中 p 为第 1, 2, 3, ..., k 种和地类中占面积最小的地类总体成数估计值。采用最小地类的总体成数来确定样本单元数，可保证各地类的估计精度都不低于要求精度。 p 越小，所需样本单元数越大。

(2) 布点：根据确定的样点数，在图上随机或系统布点（（参看简单随机抽样）。若有航摄照片，可转判。

(3) 样点设置和调查：根据布点图或航摄照片进行现地观测，观测的方法依情况而定，可实测，对坡观测等。实测时，样点定位同简单随机抽样。不管用什么方法，均需保证观测的结果确切无误，即观测后该样点所属地类的判断与实际所属地类完全一致。林、插花地小面积林中空地不构成单独地类的样点，列入附近地类中。

(4) 内业统计和精度精度计算

确定各类型面积的估计值：

根据观测结果的记载，在第 1, 2, ..., k 等各该地类中所落的样点数分别为 $n_1, n_2, \dots, n_k$ ， $\sum_{i=1}^k n_i = n$ ，则 该地类所占面

积的估计为：

$$\hat{A}_1 = A \frac{n_1}{n} = A p_1, \quad \hat{A}_2 = A \frac{n_2}{n} = A p_2, \dots, \quad \hat{A}_k = A \frac{n_k}{n} = A p_k$$

式中： $p_1, p_2, \dots, p_k$ ，分别为各地类总体成数估计值，

$$\sum_{i=1}^k p_i = 1 ;$$