サンプリングのための理論的基礎*

（小野勇一と共著）

1. サンプリングとは

たとえば，あるミカン園で病害虫の発生量を正確に知りたい場合，病害虫の全数を勘定することは労力的，時間的に不可能であるのが普通である。したがって，その一部分（サンプル＝標本）をとり出して全数を正しく推定する方法，すなわちサンプリング技術は必要不可欠となる。サンプリングではこのように‘推定する”のであるから，推定の結果は全数に対して当然，あるあたりはずれの範囲内に位置することになる。この範囲が推定値の誤差である。

サンプルをとる場合には抽出単位を定めることが必要である。果樹病害虫の場合には，葉とか枝とか樹とかがこれに当る。園全体はこれらの単位をすべて合せたものによって構成されている訳であるが，このような全体を母集団という。母集団には有限，無限の別があるが，本書でのべられる対象はすべて原理的には有限である。しかしながらとり出すサンプル数（これを“サンプルの大きさ”という）にくらべて母集団の大きさが非常に大きいときには計算上，無限母集団の仮定をおくことが多い。

推定値の誤差はサンプルのとり方，とり扱い方に影響されるとともに，とり出したサンプル間の異質性にも影響される。前者は非標本誤差とよばれ，数え間違いや調査者のくせ，集計の際の脱落などによるものである。後者は標本誤差とよばれ，本書**の大部分はこの誤差の検討結果の記述に費やされている。

以上によって，サンプリング技術の目的とは，まず非標本誤差を最小限にするための正しい標本とり扱い方法を定めることである。つぎに正しい方法で抽出した標本から母集団値の推定を行ない，かつその推定誤差の程度を判定することである。またこれと逆に推定値の誤差を一定の幅に入れたいとき（これを期待精度という），必要な標本の大きさを定めることでもある。

2. 標本の抽出

抽出単位を定める場合に重要なことは，それが集計単位となるべく一致するようにすることである。葉とか枝とかの自然的単位を抽出単位とする場合にはほとんど問題ないが，葉の表面の一部など，人為的に区切った単位を使用する場合には，この点に注意するか否かによって費やされる労力量，得られる情報量が左右される。本書**にのべる病害虫ではポンカンのヤノネカイガラムシで枝を，サビダニで芽とか葉・果実の一部分を単位としたほかは主として葉を抽出単位とした。

標本を抽出する場合，病害虫の分布等にかんして予備的な知識があるときには，それを利用する方がより高い精度の推定値をうることができる。たとえば樹の部位別方位別に密度差が予想されるときには，1本の樹を4方位にわけ，さらに各方位を上中下位部にわけるとともにこれを内外にわけ，つごう24部位にわける操作を行ない，各部位から一定の大きさのサンプルをとるというのがその1つである。このような操作を階層わけといい，この抽出方式を層別抽出という。本書にのべられる調査では，ときに層を減ずることはあっても大部分この層別抽出方式を採用している。層別抽出に対して，園全体から一定数の標本をとりだす方式を単純抽出という。

一部位からの標本抽出方法にも2通りある。その1つはその部位から無作為に一定の大きさのサンプルをとりだす方法で，ランダム抽出という。ランダム抽出では各単位に（たとえば葉に）一連番号をつけ，乱数表を用いて抽出するといった操作が最も望ましい方法である。これに対して単位の一連番号にある出発点をもうけ，それから一定間隔の番号の単位を抽出するといった方法を系統抽出という。果樹病害虫では実際上このいずれも実用的でないので，調査者が特別の選択を行なわないで，でたらめにその部位内のあちこちから葉をとるといった方法がふつう行なわれる。この方法でも実際上はランダム抽出とほぼ同一の結果をうることができる場合が多い。

3. 分布の型

発生量推定の精度はその病害虫の分布の型に強く影響される。たとえば，葉

単位抽出の場合，葉に着生した虫数が各葉とも全く同一であれば1枚の葉を抽出するだけで誤差なしの密度推定ができる。また逆に数葉にのみ集中して着生しているときはほとんどすべての葉を抽出しなければ，満足な推定値は得られないであろう。実際の分布は通常この両極端の中間である。このような葉毎の虫数のバラツキの度合を示す統計量として分散（s^2）が用いられる。

$$s^2=\frac{1}{q-1}\left(\sum_{i=1}^{q} x_i{}^2-q\bar{x}^2\right) \tag{1}$$

x_i は各葉上の虫数（$i=1,\ 2,\ \cdots,\ q$），q はサンプル葉数，$\bar{x}$ は x_i の平均値（$\sum x_i/q=\bar{x}$）。もし頻度分布表を用いれば各虫数(x)に対する頻度を f_x として

$$s^2=\frac{1}{q-1}\left(\sum f_x x^2-q\bar{x}^2\right) \tag{2}$$

分布の型は一般に葉毎に虫のいる確率が全く等しいような分布——これを機会分布またはランダム分布という——を考え，これから前述した両極端のいずれに傾くかによって，一様分布と集中分布とにわける。

（1） ポアソン分布

機会分布を数学的に表現する式としてよく用いられるのはポアソン分布である。ポアソン分布では1枚の葉につく虫数が，0，1，2，…，r 頭である確率は1枚当りの母平均を m としたとき，それぞれ

$$e^{-m},\quad me^{-m},\quad \frac{m^2}{2!}e^{-m},\quad \cdots,\quad \frac{m^r}{r!}e^{-m} \qquad (r!=r(r-1)(r-2)\cdots\cdots 2.1)$$

と表わされる。ポアソン分布では分散は平均値と等しい，すなわち $s^2=m$ または $s^2/m=1$ である。虫の分布がこの分布に合致するかどうかの判定法は2通りがある。その一つは，x_i に対する確率を求め，それに q をかけて期待頻度を算出し，実際に得た頻度と比較する方法である。第1表の計算例にも示したが，このときの確率はいちいち計算しなくてもポアソン分布表（北川，1951）などを用いれば，m の推定値として $\bar{x}$ さえわかれば，各確率は数値として与えられている。あと一つの方法は，先述の分散の性質を利用する。それには虫数の分散を計算して $\bar{x}$ との比をとり，

$s^2/\bar{x}<1$………一様分布（平均値に比して分散が小）
$s^2/\bar{x}=1$………機会分布≒ポアソン分布
$s^2/\bar{x}>1$………集中分布（平均値に比して分散が大）

表1　ミカンハダニの分布型判定のための計算例〔小野〕
（久留米，高良内の昭和40年9月2日の新葉濃密調査野帳からの1例）

x	f_x	xf_x	x^2f_x	POISSON 分布 $m=1.18$		N B 分布 $k=2.202$	
				F	$\chi_0{}^2=\dfrac{(f-F)^2}{F}$	F	$\chi_0{}^2$
0	119	0	0	95.9	5.56	121.6	0.06
1	104	104	104	113.1	0.73	93.2	1.25
2	38	76	152	66.7	12.35	51.9	3.75
3	30	90	270	26.3	0.52	25.3	0.86
4	12	48	192			11.5	0.02
5	7	35	175				
6	1　} 21	6	36	10.0	12.10	8.5	0.03
7	0	0	0				
8	1	8	64				
	312 = q	367 = T	993	312.0	31.26	312.0	5.97
				$P(\chi_0{}^2)<0.005$		$0.10<P(\chi_0{}^2)<0.25$	
	$\bar{x}=T/q=1.176$			$N=5-2=3$		$N=6-3=3$	
	$s^2=\dfrac{\sum f_x x_i{}^2-T^2/q}{q-1}=1.805$						
	$s^2/\bar{x}=1.534$						

と判定する。この示数の検定をするには，$n_1=q-1$，$n_2=\infty$ として，危険率 α を 0.05 としたときの F の値を統計数値表の F 分布表から求める（この表は大ていの統計教科書にのっている)。もし計算された $s^2/\bar{x}$ がこの F の値より大であれば1より有意の差があると判定する（鳥居，1960)。ただし，$s^2/\bar{x}<1$ のときは $\bar{x}/s^2$ として検定する。しかしながら，この示数は集中度が同じでも $\bar{x}$ が大になるにつれ，大きくなる傾向があるので，いくつかの資料の分布型の相互比較のためには平均値を等しくしておく必要がある（MORISITA 1964)。

（2） 負の二項分布

集中分布の一つの型として最もよく用いられるのは負の二項分布である。この分布は英名（Negative Binomial Distribution）の頭文字をとって NB 分布ともよばれる。一般式は

$$P_{(x)}=\frac{(k+x-1)!}{(k-1)!x!}\cdot\frac{p^x}{(1+p)^{k+x}} \tag{3}$$

で表わされ，1葉当り虫数（x）に対応する確率がこの式によって計算される。なおこの式での k は集中の度合いを示す数値であって，これが無限大のときは

負の二項分布はポアソン分布と一致するし，0の時は対数級数分布という分布になる。このように k の値は m と共に NB 分布の母数[1)]であるので，この k を推定することが分布型あてはめ計算のうえでまず必要である。k の計算法には3通りがある。

a）I法. NB 分布では，

$$\sigma^2=m+m^2/k \tag{4}$$

となる。σ^2 は母集団の分散。あてはめ計算をしようとする資料から s^2 を計算する。この s^2 を σ^2 の代わりに，また $\bar{x}$ を m の代わりに用いて，

$$k=\frac{\bar{x}^2}{s^2-\bar{x}} \tag{5}$$

として k を推定できる。p は $\bar{x}/k$ とする。

b）II法. NB 分布の虫数 0 の期待確率は，

$$P_{(0)}=\left(1+\frac{m}{k}\right)^{-k} \tag{6}$$

である。資料から $P_{(0)}$ を 0 の頻度 q_0 と全サンプル数 q の比（q_0/q）として計算する。$\bar{x}$ を m の代わりに用いれば，

$$\log\left(\frac{q}{q_0}\right)=k\ \log\left(1+\frac{\bar{x}}{k}\right) \tag{7}$$

k の値をいろいろ変えて試行錯誤的に計算し，両辺が等しくなったときの k が求める値である。

c）III法. 展開項の性質から

$$N\log\left(1+\frac{\bar{x}}{k}\right)=0.4343\cdot\sum\left(\frac{A_x}{k+x}\right)$$

A_x は個体数（$x+1$）以上の頻度の合計。計算はやはり試行錯誤法による。この方法は正確な k を与えるが計算が面倒である。

以上のようにして k, p を推定し，個体数 0 の項の確率 $p_{(0)}$ を前述の式より計算すれば，以下

個体数	1	の項の	確率は	$P_{(1)}=k\dfrac{p}{(1+p)}P_{(0)}$
〃	2	〃	〃	$P_{(2)}=\dfrac{k+1}{2}\cdot\dfrac{p}{1+p}P_{(1)}$
⋮	⋮	⋮	⋮	⋮
〃	r	〃	〃	$P_{(r)}=\dfrac{k+r-1}{r}\cdot\dfrac{p}{1+p}P_{(r-1)}$

として各項の確率が計算できる。これに全サンプル数 q を乗ずれば各項期待

頻度をうることができる。

(3) χ^2-検　定

資料から得られた頻度分布が，上述のようなモデルから計算された期待頻度分布と合致するか否かを判定する方法としてカイ二乗検定（χ^2-検定）がある。期待頻度を F_x，実際に得た頻度を f_x（個体数 x のサンプル数）項数を h とすれば χ^2 は，

$$\chi_0^2=\sum_{x=0}^{h}(f_x-F_x)^2/F_x \tag{8}$$

として計算される。一方，統計教科書の χ^2 分布表から危険率が 5 %($\alpha=0.05$) のときの χ^2 値をよむ。$\chi_0^2>\chi^2$ なれば有意の差ありと認定する。χ^2 表において自由度 n を定めねばならないが，これは標本値によって代用される母数の数に影響される。

$n=$項数$-$(母数の数$+1$)

とする。ポアソン分布では $n=h-2$, NB 分布では $n=h-3$ となる。なお，χ_0^2 計算のさいに F が 5 以下の項は，その前後の項と合計して 5 以上とする（この場合項数はそれだけ減少する）。

(4) 計　算　例

分布型あてはめ計算の実例を示す。久留米市高良内で行なわれた，成木園におけるハダニ調査結果の一部分を例とした。この調査では一樹約 300 葉を各樹について行なったが，ここではそのうちの 1 本をとりだして第 1 表に示した。サンプル葉数 $\sum f_x=q=312$ を抽出して，総虫数 $T=367$ 頭，$\bar{x}=1.176$ 頭，$s^2=1.805$ であった。

まずポアソン分布への適合度をテストした。$s^2/\bar{x}=1.534$ である。F 分布表では n_1 が 100 以上のときはほとんど 1 であるのでこの分布は問題なく，ポアソン分布から有意にはずれているといえる。また $s^2/\bar{x}>1$ であるので集中型である。次に $m=1.18$ としてポアソン分布を用いて x_i に対する確率を求め，312 を乗じて期待頻度を求めた（第 1 表ポアソン分布の項）。χ_0^2 は 31.26 となり，χ^2 表より得た $n=5-2$ としたときの $\chi^2(\alpha=0.05)=7.81$ よりはるかに大である。すなわち有意の差がある。

つぎに負の二項分布への適合性テストを試みる。I 法で k を求めると，(5)

式から

$$k=1.176^2/(1.805-1.176)=2.202.$$

Ⅱ法では（7）式から，

$$\log\left(\frac{312}{119}\right)=k\ \log\left(1+\frac{1.176}{k}\right)$$

$\log(312/119)=0.4186$ である。$k=2$ とすると右辺$=0.4018$，$k=3$ で 0.4310 となった。そこで $k=2.5$ とすると右辺$=0.4187$ となったので，さらに 0.01 きざみで k を変化させ $k=2.498$ のとき右辺$=0.4186=$左辺となり，求める k を得た。I 法の k を用い，各項の確率を計算して，312 を乗じて得たのが第1表 NB の項の期待頻度である。χ_0^2 は 5.97 となり，$n=6-3$ のときの $\chi_0^2(\alpha=0.05)=7.81$ であるので有意の差はみられない。すなわち NB 分布にあてはまるといえる。

ハダニが NB 分布をすることは，大串（未発表）によっても認められている。彼は長崎において，6～7年生の若木30本から各100葉を抽出し，各樹毎にポアソン分布と，NB 分布とのあてはめを試みた結果，ポアソンを満足する例は皆無であったが，NB には30例中26例もの適合をみた。他の病害虫についても NB 分布が適合する例は多くあり（たとえばヤノネカイガラムシについては伊藤 1967参照），集中型分布は病害虫の一般的性質であるようである。

集中分布をする傾向がある対象について，よく使用されている分散分析法などを用いる場合には変数変換を行なって正規分布に近似させ，かつ平均値による分散への影響を除去しておく必要がある。負の二項分布は x_i を $\sqrt{x_i+0.5}$ または $\log(x_i+1)$ と変換することによってこの目的を達することができる。ヤノネカイガラムシについて $\sqrt{x_i+0.5}$ 変換を行ない，分散分析を行なった結果は後章の（ヤノネカイガラムシの樹内分布）の項に示した。

（5） I_δ-指　数

ポアソン分布の $s^2/\bar{x}$ 判定は $\bar{x}$ に影響されることはすでに述べた。NB 分布においても k がたとえ計算されたとしてもあてはめ計算をしてみないことにはその適合性は云々できない。森下（1959）は $\bar{x}$ もしくは q に影響されない分布型判定指数として I_δ-指数を提唱した。

$$I_\delta=q\delta=q\frac{\sum x_i(x_i-1)}{T(T-1)}\ \text{または}\ =q\frac{\sum x_i^2-T}{T(T-1)} \tag{9}$$

頻度分布からは

$$I_\delta = q\frac{\sum f_x x^2 - T}{T(T-1)} \tag{10}$$

$\sum x_i$ または $\sum f_x x = T$ である。

I_δ は分布型がランダムであれば常に 1 となる。また集中型では 1 より大となり，一様型では小となる。ランダムであるかどうかの検定には，

$$F_0 = \frac{I_\delta(T-1)+q-T}{q-1} \tag{11}$$

として F 表の $n_1=q-1$，$n_2=\infty$ とおいたときの $F(\alpha=0.05)$ の値と比較する（$F_0>F$ ならば有意）。または

$$\chi_0^2 = I_\delta(T-1)+q-T \tag{12}$$

とおいて $n=q-1$，$\alpha=0.05$ の χ^2 と比較する。さきのハダニの例では $I_\delta=1.4541$,

$$\chi_0^2 = 1.4541(367-1)+312-367=477.201$$

となる。$n=311$ のときの $\chi^2(\alpha=0.05)=412.388$ であるのであきらかに有意の差があるといえる。

I_δ-指数の検定は q が大なるときにはそれほど重要でないが，q が小さいとき，1 に近い I_δ 値で分布型の性質を云々するときには検定しておくことが必要である。

4. 密度推定の精度

（1） 単純ランダム抽出の場合

大きさ Q の母集団から，大きさ q の標本をランダム抽出して，標本平均 $\bar{x}$，分散 s^2 を得たとする。$\bar{x}$ は母集団の平均値 m の推定値であるので $\varepsilon'=|\bar{x}-m|$ を推定値の誤差という。くり返し q をとれば $\bar{x}$ はそのたびに少しずつ異なるが，q が大きければ全体的には m のまわりに正規分布をする。このときの $\bar{x}$ の分散（$s^2\bar{x}$）は，ただ一回の q サンプルからでも

$$s^2\bar{x} = \frac{Q-q}{Q}\cdot\frac{s^2}{q} \tag{13}$$

として推定することができる。平均値の標準偏差 $s\bar{x}$ は標準誤差（*S.E.*）とよばれる。正規分布の性質から誤差は

$$\varepsilon' = |\bar{x} - m| = ts_{\bar{x}} \tag{14}$$

で与えられ，普通 95% 以上の信頼度をうるために q が大きければ $t=2$ として計算する。平均値の幾%の誤差であるという表現のためには相対誤差

$$\varepsilon = \left|\frac{\bar{x} - m}{\bar{x}}\right| = t\frac{s_{\bar{x}}}{\bar{x}} \tag{15}$$

が用いられる。

第1表のハダニの例では，$\bar{x}=1.176$，$q=312$，$s^2=1.805$ であった。この例では Q は q よりはるかに大きい，すなわち $Q \gg q$ であるので

$$s^2_{\bar{x}} \fallingdotseq \frac{1}{q}s^2 = \frac{1.805}{312} = 0.0058$$

$$s_{\bar{x}} \fallingdotseq \pm 0.077,\quad \varepsilon' = \pm 0.154$$

と計算される。つまり一葉当りハダニ数は 1.0～1.3 頭であると推定される。また，この例で ε は 13.0 %である。

(15) 式はまた I_δ によっても表わすことができる。(9) 式と (1) 式から，分散は

$$s^2 \fallingdotseq (I_\delta - 1)\bar{x}^2 + \bar{x}$$

と表わされる。いま，q のサンプルをくり返しとったとき，各回のサンプル個体数間の I_δ を $\overline{I_{\varDelta s}}$ (これをサンプル間 I_δ とよぶ) とすれば，(15) 式および後述の (17) 式から相対誤差は

$$\varepsilon = ts_{\bar{x}}/\bar{x} \fallingdotseq t\sqrt{(\overline{I_{\varDelta s}} - 1) + \frac{1}{T}} = t\sqrt{(\bar{I}_{\varDelta s} - 1) + \frac{1}{q\bar{x}}} \tag{16}$$

と表わされる。

サンプル間 I_δ は推定されるべき値である。そこでもとの集団 (大きさ Q) の I_δ を $I_\varDelta$ とすれば，

$$\overline{I_{\varDelta s}} = \frac{Q - q}{Q - 1} \cdot \frac{1}{q}(I_\varDelta - 1) + 1 \tag{17}$$

$I_\varDelta$ もまた通常不明であるが，1回の抽出から得た I_δ ($I_{\delta 0}$ とする) を用いてサンプル間 I_δ を

$$\overline{I_{\delta s}} \fallingdotseq \frac{Q - q}{Q - 1} \cdot \frac{1}{q - 1}(I_{\delta_0} - 1) + 1 \tag{18}$$

として推定する (ただし $q>2$)。

$Q \gg q$ ならば，

表2 ミカンハダニを用いての抽出率と精度に関する実験結果〔小野〕

$\bar{x}$ による発生密度グループ	m	Q	I_Δ		抽出率 (%) 2	5	10	15	20	25	30	35	40
	卵の場合												
0.26～1.90 (4本)	0.825	286	12.41	q		14	29	43	57	72	86	100	114
				$I_{\delta 0}$		5.754	11.96	11.50	8.45	9.70	13.72	11.39	11.86
3.92～9.67 (5本)	6.42	282	4.790	q	6	14	28	42	56	71	85	99	113
				$I_{\delta 0}$	3.210	5.472	3.823	3.870	4.870	4.475	4.190	5.463	6.608
11.9～22.12 (6本)	16.48	334	2.960	q	7	17	33	50	67	84	100	117	134
				$I_{\delta 0}$	2.038	2.338	2.795	2.059	2.546	2.875	3.192	2.926	3.473
34.5～66.7 (5本)	46.37	221	2.321	q	4	11	22	33	44	55	66	77	88
				$I_{\delta 0}$	2.280	2.123	2.456	2.223	2.904	2.219	2.007	2.406	2.439
	成虫の場合												
0.03～0.20 (7本)	0.090	277	9.05	q			28	42	55	69	83	97	111
				$I_{\delta 0}$			2.01	7.49	8.02	0.58	5.97	11.08	8.84
0.38～0.87 (7本)	0.517	326	4.818	q			33	49	65	82	98	114	130
				$I_{\delta 0}$			4.375	3.481	4.864	5.235	4.146	4.335	3.685
1.39～3.24 (6本)	1.945	238	3.886	q			24	36	48	59	71	83	95
				$I_{\delta 0}$			2.943	3.131	3.349	3.711	2.690	4.059	4.195

$$\overline{I_{\delta s}} \fallingdotseq \frac{1}{q-1}(I_{\delta 0}-1)+1.$$

したがって (2) 式は

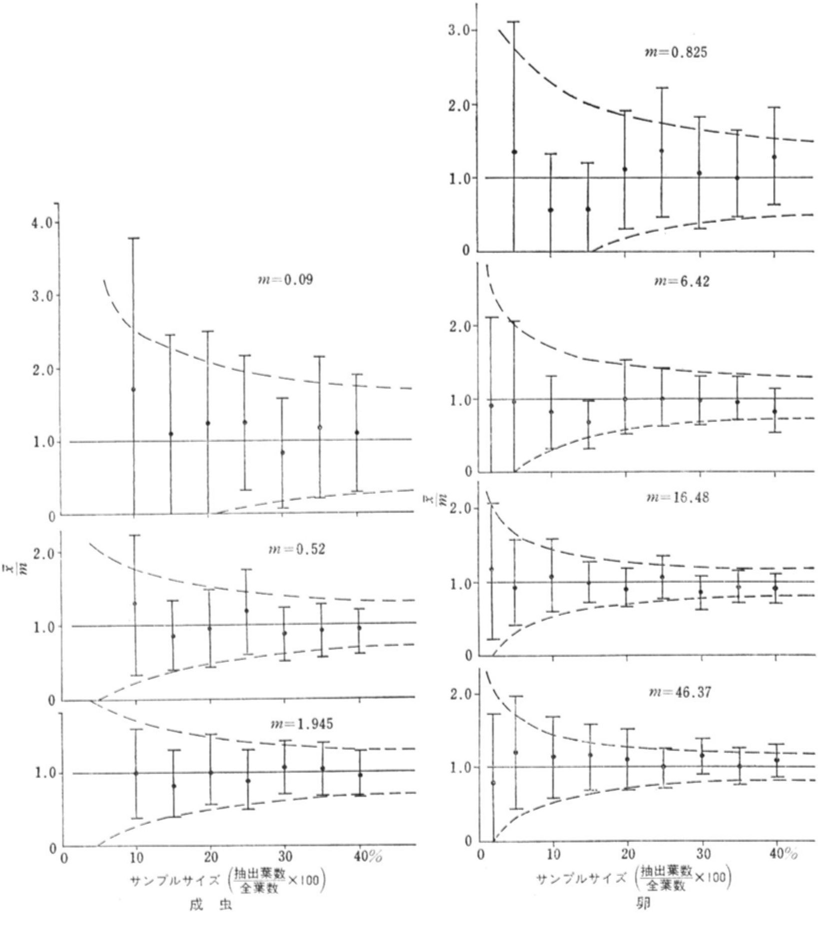
$$\varepsilon \fallingdotseq t\sqrt{\frac{Q-q}{Q-1}\cdot\frac{1}{q-1}(I_{\delta 0}-1)+\frac{1}{T}} \tag{19}$$

無限母集団または $Q \gg q$ ならば

$$\varepsilon \fallingdotseq t\sqrt{\frac{1}{q-1}(I_{\delta 0}-1)+\frac{1}{T}} \tag{20}$$

となる。

(15) 式も (20) 式も同一の母集団について ランダム抽出をする場合，サイズ q が大になれば T もまた大となり，結果的に ε は小さくなることを示している。このことはサンプルサイズと精度との基本的な性質であるが，これはつぎのようなモデル実験によって明瞭に示されている。実験は長崎農林センターの西野敏勝技師によって行なわれた。方法は3年生普通温州20本を用い，1966年3月せん定後全葉につき一葉ずつ別々にミカンハダニの卵，幼虫，成虫数を記録した。1本の樹の全葉を母集団とみなせば，母集団の大きさ (Q),



第1図　ミカンハダニに関するサンプリング実験の結果〔小野・西野原図〕

母集団平均値 (m), 母集団 $I_\delta(I_\Delta)$ を (16), (17) 式に入れることによって，サンプリングした場合の誤差の範囲を計算することができる。第1図の点線にかこまれた部分は q をかえた場合のその範囲を示すものである。ただし今回の例では20本の樹をハダニの発生密度の多寡に応じてグループわけし，各グループの平均値をもって母集団の各値とした（第2表）。この母集団から抽出率 (q/Q) を 2, 5, 10, 15, 20, 25, 30, 35, 40%と変えて標本抽出（実際は1葉ずつの記録からランダム抽出）を行なった。もし正しくランダム抽出されれば,

この結果から計算される誤差の推定範囲は上述の期待範囲をこえないはずである。抽出率の各々の場合につき，$\bar{x}$，I_δ を求め，(4) 式によって ε を計算した結果を第1図に示した。この場合 $q<30$ では t は 2 より大となるので t 表により値を求めた。図には比較の便宜上 $\bar{x}$ の代わりに $\bar{x}/m$ として示した。卵，成虫いずれの場合も期待される誤差の範囲と推定範囲はかなりよい一致を示している。しかもその一致の程度は m が大きいほどよいことが明瞭であろう。

このように同一の母集団については ε と q とは互いに関連するものであるから，観点をかえればある誤差範囲内に推定値 $\bar{x}$ を入れたい場合に，必要なサンプルの大きさを定めることもまた可能である。$q \gg I_{\delta_0}$ とすれば (20) 式から近似的に

$$q>\left(\frac{t}{\varepsilon}\right)^2\left(I_{\delta_0}-1+\frac{1}{\bar{x}}\right) \tag{21}$$

としてある ε を定めたときの q を求めることができる。

第1表のハダニの資料について計算例を示そう。これでは $I_{\delta_0}=1.4541$，$q=312$，$T=367$，$\bar{x}=1.176$ であった。対象木は葉数の大きい成木であったので (20) 式を用い，

$$\varepsilon \fallingdotseq 2\sqrt{\frac{1}{311}(1.4541-1)+\frac{1}{367}}=0.129$$

すなわち 12.9% の誤差が得られた（当然であるがこれは $s_{\bar{x}}$ による計算結果と一致する)。もし ε を 20% まで許容するとすれば，(21) 式より，

$$q>\left(\frac{2}{0.2}\right)^2\left(1.4541-1+\frac{1}{1.176}\right)\fallingdotseq 130.5$$

すなわち，131 枚程度抽出すればよいことになる。

(2) 層別ランダム抽出の場合

次に層別抽出における I_δ 法による誤差の計算法をのべる。簡単な階層わけとして，園内の L 本の樹を L の階層とみて，一つの層が Z 枚の葉（Z 単位）で構成されている場合について考える。この母集団について葉単位病害虫の分布を I_δ で次のように表現する。すなわち，樹と樹の間の I_δ を $I_{\Delta Y}$（層間 I_δ），1本の樹内で葉毎の I_Δ を $I_{\Delta(Z)}$，$I_{\Delta(Z)}$ の平均値を $\overline{I}_{\Delta(Z)}$（層内 I_δ），層に関係なくすべての葉毎の I_δ を I_Δ とする。このとき

$$I_\Delta = I_{\Delta Y} \cdot \overline{I_{\Delta(Z)}} \tag{22}$$

なる関係が成立する。

上述の母集団から l 本（層），z 枚（単位）をランダム抽出したとき，サンプル間 I_δ は，

$$\overline{I_{\Delta s}} = \frac{L-l}{L-1}\frac{1}{l}(I_{\Delta Y}-1) + \frac{Z-z}{Z-1}\frac{1}{lz}(\overline{I_{\Delta(Z)}}-1)I_{\Delta Y}+1$$

$$= \frac{L-l}{L-1}\frac{1}{l}(I_{\Delta Y}-1) + \frac{Z-z}{Z-1}\frac{1}{lz}(I_\Delta - I_{\Delta Y})+1 \tag{23}$$

と表わされる。I_Δ，$I_{\Delta Y}$，$\overline{I_{\Delta(Z)}}$ は母集団値であり，ふつう不明であるが，サンプル I_δ は a）の場合と同様に 1 回の標本抽出から得た $I_{\delta 0}$（I_Δ に対応），$I_{\delta y}$（$I_{\Delta Y}$ に対応）を用いて次のように計算できる。

$$\overline{I_{\delta s}} \doteqdot \frac{L-l}{L-1}\frac{1}{l-1}(I_{\delta y}-1) + \frac{Z-z}{Z-1}\frac{1}{l(z-1)}(I_{\delta 0}-I_{\delta y})+1 \tag{24}$$

この $\overline{I_{\delta s}}$ を用いて（16）式から ε を計算する。ただしもし $lz<30$ であれば t 値はやはり t 表から求めなければならない。

計算例として鹿児島で行なわれたポンカン園におけるヤノネカイガラムシ成虫についての調査結果を示そう。全樹数 $L=243$，$l=16$，1 樹当り $z=239$ として調査した結果，$T=79$，$\bar{x}=0.021$，$I_{\delta 0}=33.857$，$I_{\delta y}=2.242$ を得た。1 樹の葉数 $Z \gg z$ であるとして，（24）式より

$$\overline{I_{\delta s}} \doteqdot \frac{243-16}{243-1}\frac{1}{16}(2.242-1) + \frac{1}{16(239-1)}(33.857-2.242)+1$$

$$=0.0777+0.0083+1=1.0860$$

となった。$t=2$ として求めた ε は 62.8% となった。もし全樹を調査すると（$L=l$），I_δ が変らねば，$\overline{I_{\delta s}}=1.0083$ となり $\varepsilon=0.291$ とぐっと小さくなる。このようにサンプル I_δ の大きさは（24）式右辺第 1 項に強く影響される。つまり誤差を小さくするためには調査樹数（層数）をふやすことがまず必要であることになる。この点については後章で，ミカンハダニ，コナカイガラ類の項で改めて検討を行なう。

さて，期待精度（ε）を定めたときの必要サンプル数は層別抽出ではどうなるであろうか。もし全層調査をするとすれば（9）式の第 I 項は 0 になるので，各層から抽出すべき単位数 z は近似的に

$$z > \left(\frac{t}{\varepsilon}\right)^2 \left(I_{\delta 0} - I_{\delta y} + \frac{1}{\bar{x}}\right) \tag{25}$$

で与えられる。ただし $Z \gg z \gg 1$ とする。

l 層，z 単位抽出の場合は，ε が l, z の両方に左右されるので（25）式のように簡単にはならない。ただし $L \gg l \gg 1$, $Z \gg z \gg 1$ という条件下では，(24), (16) 式から

$$l > \left(\frac{t}{\varepsilon}\right)^2 \left\{(I_{\delta y}-1)+\frac{1}{z}\left(I_{\delta 0}-I_{\delta y}+\frac{1}{\bar{x}}\right)\right\} \tag{26}$$

として z を定めれば l を求めることができる。しかし，これは相当大まかな方法であるので，もし I_δ 値と病害虫との間にある一般的な関係が定まる場合には，標本 I_δ から後述の補遺にのべた方法で母集団 I_δ を推定して一般性の高い調査設計をする方がより望ましい。

〔補遺〕 母集団 I_δ の推定

標本誤差 ε の計算で，もし q が 30 より小さいときは t 値は t 表により求めることが必要であるとのべた。しかしもしこの計算に使用する I_δ 値がすべて母集団値であるならば $t=2$ とおいても差支えない。また既にのべたように期待精度をうるための必要標本数の決定方式を広域対象により一般化しようとする場合には，母集団 I_δ を求めることが必要である。ちなみに，もし母集団 I^δ が求められるならば，果樹病害虫調査では一般に $Z \gg z$ であるので，l と z との関係は

$$l \doteqdot \frac{L(I_{\Delta Y}-1)+\dfrac{(L-1)\left(I_\Delta-I_{\Delta Y}+\dfrac{1}{\bar{x}}\right)}{z}}{(L-1)\left(\dfrac{\varepsilon}{t}\right)^2+I_{\Delta Y}-1}$$

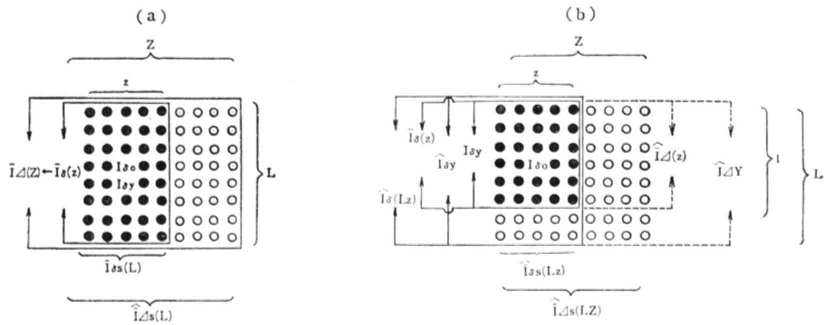
と表わすことができる。

さて，母集団 I_δ の推定法は，全層（L），z 単位抽出の場合と l 層 z 単位抽出の場合とにわけてのべる。第2図にその抽出状況を模式的に表現した。

i）$L \times z$ 単位抽出

第2図 a に示したように $I_{\delta 0}$, $I_{\delta Y}$ は標本から計算される値である。手続きとしてはじめに層内 I_δ の母集団値（$\overline{I_{\Delta(z)}}$）を推定し，その推定値を用いて層間 I_δ の母集団値（$I_{\Delta Y}$）を求め，最後に母全 I_δ（I_Δ）を求める。

層内 I_δ の加重平均は前節（7）の関係から，



第2図　層別抽出の模式図〔小野原図〕

$$\overline{I_{\delta(z)}}=I_{\delta 0}/I_{\delta Y}$$

として計算される。これから母層内 I_δ は,

$$\hat{\overline{I}}_{\Delta(Z)}=\frac{\left(1-\dfrac{Z-z}{Z-1}\cdot\dfrac{1}{z}\right)\overline{I_{\delta(z)}}}{1-\dfrac{Z-z}{Z-1}\cdot\dfrac{1}{z}\overline{I_{\delta(z)}}}$$

として推定される（∧は推定値の意味）。

次に各層より1単位ずつ抽出（各樹より1葉ずつ抽出）したと考えたときのサンプル I_δ は

$$\overline{I_{\delta s(L)}}=\frac{1}{L}(I_{\delta 0}-I_{\delta Y})+1$$

と計算できる。この値の母集団推定値は

$$\hat{\overline{I}}_{\Delta s(L)}=\frac{\left(1-\dfrac{Z-z}{Z-1}\cdot\dfrac{1}{z}\right)\overline{I_{\delta s(L)}}}{1-\dfrac{Z-z}{Z-1}\cdot\dfrac{1}{z}\overline{I_{\delta s(L)}}}$$

として求められる。

一方,

$$\hat{\overline{I}}_{\Delta s(L)}=\frac{1}{L}(\hat{\overline{I}}_{\Delta(Z)}-1)I_{\Delta Y}+1$$

なる関係から，母集団層間 I_δ は,

$$\hat{I}_{\Delta Y}=L\cdot\frac{\hat{\overline{I}}_{\Delta s(L)}-1}{\hat{\overline{I}}_{\Delta(Z)}-1}$$

となる。ただし $\hat{\overline{I}}_{\Delta(Z)}=1$ のときは $I_{\delta Y}$ を $\hat{I}_{\Delta Y}$ とする。そして再び (7) の関係から，

$$\hat{I}_{\Delta}==\hat{I}_{\Delta Y}\cdot\hat{\overline{I}}_{\Delta(Z)}$$

として母集団全 I_δ が推定される。

ii) $l\times z$ 単位抽出

i) では第2図aの横方向の母集団値が推定の主な対象であったが，ii) では同図bに示したように縦・横両方向での推定が必要となる。

まず縦方向について考える。いま L 層すべてから z ずつ抽出したと仮定すると，このときの層間 I_δ は，

$$\hat{I}_{\delta Y}=\frac{\left(1-\dfrac{L-l}{L-1}\cdot\dfrac{1}{l}\right)I_{\delta y}}{1-\dfrac{L-l}{L-1}\cdot\dfrac{1}{l}I_{\delta y}}$$

と推定される。またこの仮定のもとで，全 I_δ は

$$\hat{I}_{\delta(Lz)}=\hat{I}_{\delta Y}\cdot\overline{I_{\delta(z)}}$$

となる。ただし，$\overline{I_{\delta(z)}}=I_{\delta 0}/I_{\delta y}$

次に横方向について，i) と同様に各層から1単位ずつ抽出したと考えたときのサンプル I_δ は，

$$\hat{\overline{I}}_{\delta s(Lz)}=\frac{1}{L}(\hat{I}_{\delta(Lz)}-\hat{I}_{\delta Y})+1$$

と計算される。この母集団値は，

$$\hat{\overline{I}}_{\Delta s(LZ)}=\frac{\left(1-\dfrac{Z-z}{Z-1}\cdot\dfrac{1}{z}\right)\hat{\overline{I}}_{\delta s(Lz)}}{1-\dfrac{Z-z}{Z-1}\cdot\dfrac{1}{z}\hat{\overline{I}}_{\delta s(Lz)}}$$

として推定される。

以上により層内 I_δ，層間 I_δ，全 I_δ の母集団推定値は i) と同じように，

$$\hat{\overline{I}}_{\Delta(Z)}=\frac{\left(1-\dfrac{Z-z}{Z-1}\cdot\dfrac{1}{z}\right)\overline{I}_{\delta(z)}}{1-\dfrac{Z-z}{Z-1}\cdot\dfrac{1}{z}\overline{I}_{\delta(z)}}$$

$$\hat{I}_{\Delta Y}=L\cdot\frac{\hat{\overline{I}}_{\Delta s(LZ)}-1}{\hat{\overline{I}}_{\Delta(Z)}-1}$$

$$\hat{I}_{\Delta}=\hat{\overline{I}}_{\Delta(Z)}\cdot\hat{I}_{\Delta Y}$$

となる。$\hat{I}_{\Delta Y}$ の計算で，もし $\hat{\overline{I}}_{\Delta(Z)}=1$ のときは，やはり $\hat{I}_{\delta Y}$ を母集団推定値とする。

i) についての計算例

宮崎で行なわれたミカンハダニについての調査の結果，$L=468$，$z=32$ として，$T=1159$，$\bar{x}=0.078$，$I_{\delta Y}=62.806$，$I_{\delta 0}=283.772$，$\overline{I_{\delta(z)}}=283.772/62.806=4.518$を得た。$Z$は約 3,000 であったので $Z\gg z$ とみなして，

$$\hat{\overline{I}}_{\Delta(Z)}=\left(1-\frac{1}{32}\right)4.518\Big/\left(1-\frac{1}{32}4.518\right)=5.103$$

$$\overline{I}_{\delta s(L)}=\frac{1}{468}(283.772-62.806)+1=1.472$$

$$\hat{\overline{I}}_{\Delta s(L)}=\left(1-\frac{1}{32}\right)1.472\Big/\left(1-\frac{1}{32}1.472\right)=1.4949$$

よって，

$$\hat{I}_{\Delta Y}=468(1.495-1)/(5.103-1)=56.452$$

$$\hat{I}_{\Delta}=56.452\times 5.103=288.075$$

この値を用いればサンプル間 I_{δ} は，

$$\overline{I_{\Delta s}}=\frac{1}{468\times 32}(288.075-56.452)+1=1.0156$$

誤差は　$\varepsilon=2\sqrt{(1.0156-1)+\dfrac{1}{1159}}=0.2567$

と計算される。

ii) についての計算

前述のヤノネカイガラムシの例で改めて計算する。この例では

$$\overline{I_{\delta(z)}}=33.857/2.242=15.101$$

と計算される。この値および前述のデータから諸推定値は，

$$\hat{I}_{\delta Y}=\left(1-\frac{243-16}{243-1}\cdot\frac{1}{16}\right)2.242\Big/\left(1-\frac{243-16}{243-1}\cdot\frac{1}{16}\cdot 2.242\right)=2.430$$

$$\hat{I}_{\delta(Lz)}=2.430\times 15.101=36.696$$

$$\hat{\overline{I}}_{\delta s(Lz)}=\frac{1}{243}(36.696-2.430)+1=1.141$$

$$\hat{\overline{I}}_{\Delta s(LZ)}=\left(1-\frac{1}{239}\right)1.141\Big/\left(1-\frac{1}{239}\cdot 1.141\right)=1.142$$

$$\hat{\overline{I}}_{\Delta(Z)}=\left(1-\frac{1}{239}\right)15.101\Big/\left(1-\frac{1}{239}\cdot 15.101\right)=16.052$$

$$\hat{I}_{\Delta Y}=243(1.142-1)/(16.052-1)=2.150$$

$$\hat{I}_{\Delta}=2.150\times 16.052=34.506$$

となった。この値を用いれば

$$\overline{I_{\Delta s}}=\frac{243-16}{243-1}\cdot\frac{1}{16}(2.150-1)+\frac{1}{3808}(34.506-2.150)+1=1.0759$$

$$\varepsilon=2\sqrt{(1.0759-1)+\frac{1}{79}}=0.5951$$

となり，標本 I_δ をそのまま用いたときより誤差はやや小さくなる。

註

1）〔201頁〕 母数パラメーターとは母集団の特性を示すような量をいう．ポアソン分布では m がこれにあたる．

* 本篇は,「カンキツ病害虫の共同防除の合理化に関する研究」(九州果樹病害虫共同防除研究協議会編) (1969) に「II，サンプリングのための理論的基礎」として書かれたものである．

** 上記「カンキツ病害虫の共同防除の合理化に関する研究」を指す．