任意標本調査法（四）

關　彌　三　郎

層化抽出法は母集団の構成単位を、調査せんとする性質又はそれと密接な関連を有する他の性質に関して同質的な単位の層に分ち[1]、各層毎に単位を抽出して標本を構成する方法である。

1) 層化抽出法の特徴はすべての層より且各層毎に別々に抽出が行れる事にあるのであって、層化に際して同質的な単位の層に分つ事は、層化抽出法の必ずしも本質的な条件ではないのである。唯後述するように層内同質にする事が層化抽出法の精度を高める所以であるから、一般に層化抽出に於ては同質的な単位を集めて層を作るのである。

五　層化抽出法

(1) 層化一段抽出法に於ける推定式

今N個の調査単位より成る母集団πをL個の層に分ち、各層内の単位の数をそれぞれN_i個（$i=1, 2, \cdots, L$）

（但 $\sum_{i=1}^{L} N_i = N$）とし、i 番層に於ける j 番調査単位の変量を x_{ij} で表すと、i 番層の平均値 $\bar{x}_i$、分散 σ_i^2 は

$$\bar{x}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} x_{ij},\quad \sigma_i^2 = \frac{1}{N_i}\sum_{j=1}^{N_i}(x_{ij}-\bar{x}_i)^2$$

であり、母集団 π に於ける平均値 $\bar{x}$、分散 σ^2 は

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{L}\sum_{j=1}^{N_i} x_{ij} = \frac{1}{N}\sum_{i=1}^{L} X_i, \qquad (5\cdot1)$$

（X_i……i 番層の量的大いさ）

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{L}\sum_{j=1}^{N_i}(x_{ij}-\bar{x})^2$$

である。このL個の層の各より n_i 個（$i=1, 2, \dots, L$）宛の調査単位を抽出して、大いさ $n(=\sum_{i=1}^{L} n_i)$ の標本を構成した場合の母平均の推定式を求めよう。母平均は(5·1)より明らかな如く、各層の量的大いさを総計して単位総数で除した値であるから、先づ各層の量的大いさを推定し、それの総計を平均すればよいであろう。

i 番層に於ては N_i 個の単位より n_i 個の単位が抽出されるのであるから、その層の量的大いさ X_i の不偏推定値 X_i' は(2·34)より

$$X_i' = \frac{N_i}{n_i}\sum_{j=1}^{n_i} x_{ij}',\quad E(X_i') = X_i \qquad (5\cdot2)$$

であり、従って母平均 $\bar{x}$ の推定値 $\bar{x}'$ は

$$\bar{x}' = \frac{1}{N}\sum_{i=1}^{L} X_i' = \frac{1}{N}\sum_{i=1}^{L}\frac{N_i}{n_i}\sum_{j=1}^{n_i} x_{ij}', \qquad (5\cdot3)$$

となる。これの期望値を求めると

$$E(\bar{x}') = \frac{1}{N}\sum_{i=1}^{L} E(X_i') \overset{1)}{=} \frac{1}{N}\sum_{i=1}^{L} X_i = \bar{x} \qquad (5\cdot4)$$

であるため、(5·3)式は母平均の不偏推定値である。そして(5·3)の分散は

$$V(\bar{x}') = \frac{1}{N^2}\sum_{i=1}^{L} V(X_i')^{2)}$$

$$= \frac{1}{N^2}\sum_{i=1}^{L} N_i^2\,\frac{N_i-n_i}{N_i-1}\,\frac{\sigma_i^2}{n_i} \quad [(2\cdot35)] \qquad (5\cdot5)$$

若し N_i が n_i に比して可成り大であり有限母集団修正を無視し得る時は

$$V(\bar{x}') = \frac{1}{N^2}\sum_{i=1}^{L} N_i^2\,\frac{\sigma_i^2}{n_i} \qquad (5\cdot6)$$

である。

1) 2) 標本構成単位の抽出は層毎に別々に行れるのであるから、各層毎に計算された統計量は相互に独立な確率変数である。

(5·5)より明らかな如く層化抽出法による推定値の誤差分散は、各層毎の推定値の誤差分散の総和である。

以上は各層に於て調査単位を抽出単位とする場合であるが、若し集落を抽出単位とする時は層の量的大いさの推定に於て、集落抽出法の推定式を適用すればよいのである。

比例抽出法　各層より抽出される単位の数 n_i が層の大いさ N_i に比例し、各層の抽出率が一定 $\frac{n_i}{N_i}=f=\frac{n}{N}$ $(i=1,2,\cdots,L)$ である場合を、比例抽出法 proportionate sampling という。比例抽出法に於ては推定式並に分散は簡単な形になるのであって、それは(5・3)(5・5)に於て $\frac{n_i}{N_i}=f=\frac{n}{N}$ を代入する事によって得られる。

即ち母平均の推定式は(5・3)より

$$\bar{x}'=\frac{1}{Nf}\sum_{i=1}^{L}\sum_{j=1}^{n_i}x_{ij}'=\frac{1}{n}\sum_{i=1}^{L}\sum_{j=1}^{n_i}x_{ij}' \qquad (5\cdot7)$$

であり、分散は(5・5)より

$$V(\bar{x}')\doteqdot\frac{1}{N^2}\sum_{i=1}^{L}N_i^2\frac{N_i-n_i}{N_i}\frac{\sigma_i^2}{n_i}=\frac{1}{N^2}\left(\frac{1}{f}-1\right)\sum_{i=1}^{L}N_i\sigma_i^2=\frac{1-f}{n}\frac{1}{N}\sum_{i=1}^{L}N_i\sigma_i^2$$

$$=(1-f)\frac{\sigma_w^2{}'}{n} \qquad (5\cdot8)$$

但 $\frac{1}{N}\sum_{i=1}^{L}N_i\sigma_i^2=\sigma_w^2{}'$ …層内分散（(4)精度の項参照）

任意本調査法（四）

となる。

以上は平均値の推定の場合であるが母比例数 p を推定せんとする時は、割合を知らんとする標識を1、その他の標識を0とする事によって、推定式、分散は(5・3)(5・5)又は(5・7)(5・8)と同じであり、唯

$\sigma_i^2=p_iq_i$（但 p_i は i 番層に於ける α 標識の割合）

となるのみである。

(2) 適正配分

既に述べたように層化抽出法による推定値の誤差分散は、各層に於ける推定値の誤差分散の合計であり、各層の誤差分散はその層よりの抽出単位数の如何によって変化するため、同じ大いさの標本と雖も特定の層より抽出する単位の数が異るに従って、その推定値の誤差分散は異るのである。そして又層によって調査単位の個別調査の費用が異る時は、調査費に於ても同一ではないであろう。かくて一般に層化抽出法に於ては、各層への抽出単位数の配分の如何によって標本推定の

精度並に調査費が変化するのであり、従って要求された精度の標本推定値を最小の費用で獲得し、又は所与の費用で最高の精度の標本推定を達成し得るような、各層に於ける抽出単位数の決定即ち適正配分が問題となるのである。次にそれを便宜上有限母集団修正が無視し得る場合に就て考察しよう。

適正配分を考察するためには、先づ抽出単位数の変化による調査費用の変動を与える費用函数を決定しなければならない。今i番層に於ける一調査単位当り調査費をk_iとすると、調査費用Tは

$$T=\sum_{i=1}^{L}k_i n_i \qquad (5\cdot 9)$$

によって与えられる。

先づ調査費が与えられた場合は、(5•9)式のTが一定という条件の下に(5•6)の$V(\bar{x}')$を最小ならしめるn_i $(i=1,2,\cdots,L)$の値を求めればよいのである。それはラグランジュ乗数λを用いて

$$f(n_1,n_2\cdots,n_L)=\frac{1}{N^2}\sum_{i=1}^{L}\frac{(N_i\sigma_i)^2}{n_i}+\lambda\left(\sum_{i=1}^{L}k_i n_i-T\right)$$

をn_i $(i=1,2,\cdots,L)$でそれぞれ偏微分して0と置き、(5•9)と連立せしめて解く事によって得られる。即ち

$$\frac{\partial f}{\partial n_i}=-\frac{1}{N^2}\frac{(N_i\sigma_i)^2}{n_i^2}+\lambda k_i=0,\quad (i=1,2,\cdots,L) \qquad (\mathrm{i})$$

(i)をn_iに就て解くと

$$n_i^2=\frac{(N_i\sigma_i)^2}{\lambda k_i N^2} \qquad \therefore\ n_i=\frac{N_i\sigma_i}{\sqrt{\lambda k_i}\,N} \qquad (\mathrm{ii})$$

これを(5•9)に代入して

$$T=\frac{1}{\sqrt{\lambda}\,N}\sum_{i=1}^{L}\frac{N_i\sigma_i}{\sqrt{k_i}}k_i \qquad \therefore\ \frac{1}{\sqrt{\lambda}\,N}=\frac{T}{\sum_{i=1}^{L}N_i\sigma_i\sqrt{k_i}}$$

これを(ii)に代入してλを消去すると

$$n_i=\left(\frac{T}{\sum_{i=1}^{L}N_i\sigma_i\sqrt{k_i}}\right)\frac{N_i\sigma_i}{\sqrt{k_i}},\quad (i=1,2,\cdots,L) \qquad (5\cdot 10)$$

となる。

次に精度が要求された場合は、(5•6)式の$V(\bar{x}')$が一定という条件の下に(5•9)のTを最小ならしめるn_i $(i=1,2,\cdots,L)$の値を求めればよいのである。それは

$$f(n_1,n_2,\cdots,n_L)=\sum_{i=1}^{L}k_i n_i+\lambda\left\{\frac{1}{N^2}\sum_{i=1}^{L}\frac{(N_i\sigma_i)^2}{n_i}-V(\bar{x}')\right\}$$

をn_i $(i=1,2,\cdots,L)$でそれぞれ偏微分して0と置き(5•6)と連立させて解く事によって得られる即ち

$$\frac{\partial f}{\partial n_i}=k_i-\lambda\frac{(N_i\sigma_i)^2}{N^2 n_i^2}=0,\quad (i=1,2,\ldots,L) \qquad \text{(i)}$$

(i)をn_iに就て解くと

$$n_i^2=\frac{\lambda}{N^2}\cdot\frac{(N_i\sigma_i)^2}{k_i}\qquad \therefore\ n_i=\frac{\sqrt{\lambda}}{N}\cdot\frac{N_i\sigma_i}{\sqrt{k_i}} \qquad \text{(ii)}$$

(ii)を(5・6)に代入すると

$$V(\bar{x}')=\frac{1}{N^2}\sum_{i=1}^{L}(N_i\sigma_i)^2\frac{N}{\sqrt{\lambda}}\cdot\frac{\sqrt{k_i}}{N_i\sigma_i}$$

$$\therefore\ \frac{\sqrt{\lambda}}{N}=\frac{\sum_{i=1}^{L}N_i\sigma_i\sqrt{k_i}}{N^2V(\bar{x}')}$$

これを(ii)に代入してλを消去すると

$$n_i=\left(\frac{\sum_{i=1}^{L}N_i\sigma_i\sqrt{k_i}}{N^2V(\bar{x}')}\right)\frac{N_i\sigma_i}{\sqrt{k_i}},\quad (i=1,2,\ldots,L) \qquad (5\cdot11)$$

となる。

(5・10)並に(5・11)に於て右辺の括弧内の値は特定の層に無関係な一定の値であるため

$$n_i\propto\frac{N_i\sigma_i}{\sqrt{k_i}}\qquad \therefore\ \frac{n_i}{N_i}\propto\frac{\sigma_i}{\sqrt{k_i}} \qquad (5\cdot12)$$

であり、かくて層化抽出法に於ける適正配分は、費用が与えられた場合であっても精度が要求された場合であっても共に、各層よりの抽出率をその層の標準偏差σ_iに正比例させ、且その層に於ける単位当り費用k_iの平方根に反比例させればよい事が判るのである。そして(5・10)(5・11)より明らかな如く費用の増加、精度の向上（即ち$V(\bar{x}')$の減少）は、各層よりの抽出単位数を一様に増加せしめるのである。[2]

1) ∝は比例する事を表す記号である。

2) 最適抽出数（又は率）の決定に見られるこの関係は、有限母集団修正を無視し得ない場合でも同様にいゝ得るところである。

以上は層によって単位当り調査費が異る場合であるが、若しそれがすべての層に亘って同一である時は、$k_i=k\ (i=1,2,\ldots,L)$であるから(5・12)は

$$n_i\propto N_i\sigma_i\qquad \therefore\ \frac{n_i}{N_i}\propto\sigma_i \qquad (5\cdot13)$$

であり、かくて適正配分は各層よりの抽出率をその層の標準偏差σ_iに正比例させればよいのである。そしてこの場合尚層に於ける変異係数$c_i=\frac{\sigma_i}{\bar{x}_i}$がすべての層に亘って同じ値である時は、$\frac{\sigma_i}{\bar{x}_i}=c$ 故に $\sigma_i=c\bar{x}_i\ (i=1,2,\ldots,L)$ より(5・13)は

$$n_i \propto N_i\bar{x}_i \qquad \therefore \quad \frac{n_i}{N_i} \propto \bar{x}_i \tag{5・14}$$

であり、かくて各層よりの抽出率をその層の平均値$\bar{x}_i$に比例させればよいのである。

(5・13)の配分方法をネイマン(Neyman)抽出法、(5・12)の配分方法をデミング(Deming)抽出法ともいう。

(5・13)及び(5・14)より明らかな如く比例抽出法は、σ_i乃至は$\bar{x}_i$が層に無関係な場合の適正配分である。故にσ_i又は$\bar{x}_i$が層によって異る時は、(5・13)(5・14)の配分方法即ちネイマン抽出法の方が比例抽出法よりも精度は高いのである。しかしながらσ_iの値に余り差異がない時は、両者の精度の差は殆ど問題にならない位小である事が証明され得るのである（但し証明は略する）。

(3) 母分散の推定

(5・5)式により誤差分散を計算するに必要な母分散は各層に於ける分散σ_i^2の値である。層化抽出法に於てはすべての層より単位の抽出が行れるため、σ_i^2を個々的に推定する事が出来る。標本に於けるi番層の分散を

$$s_i^2=\frac{1}{n_i}\sum_{j=1}^{n_i}(x_{ij}'-\bar{x}_i')^2, \qquad \bar{x}_i'=\frac{1}{n_i}\sum_{j=1}^{n_i}x_{ij}'$$

とすると、σ_i^2の不偏推定値$\hat{\sigma}_i^2$は(2・47)より

$$\hat{\sigma}_i^2=\frac{N_i-1}{N_i}\,\frac{n_i}{n_i-1}\,s_i^2 \tag{5・15}$$

である。

次に比例抽出法による時は(5・8)式に於て必要な母分散は$\sigma_w^{2\prime}$であるが、それは$\sigma_w^{2\prime}=\frac{1}{N}\sum_{i=1}^{L}N_i\sigma_i^2$であるため結局$\sigma_i^2$の値を推定すればよく、従って$\sigma_w^{2\prime}$の不偏推定値$\hat{\sigma}_w^{2\prime}$は

$$\begin{aligned}\hat{\sigma}_w^{2\prime}&=\frac{1}{N}\sum_{i=1}^{L}N_i\,\frac{N_i-1}{N_i}\,\frac{n_i}{n_i-1}\,s_i^2\\ &\doteqdot\frac{1}{N}\sum_{i=1}^{L}N_i\,\frac{n_i}{n_i-1}\,s_i^2 && [\because\ N_i-1\doteqdot N_i]\\ &=\frac{1}{Nf}\sum_{i=1}^{L}n_i\,\frac{n_i}{n_i-1}\,s_i^2 && \left[\because\ \frac{n_i}{N_i}=f=\frac{n}{N}\right]\\ &=\frac{1}{n}\sum_{i=1}^{L}n_i\,\frac{n_i}{n_i-1}\,s_i^2 && (5・16)\end{aligned}$$

である。

(4) 層化抽出法の精度

母平均の推定の場合に就て層化抽出法の精度を単純抽出法の精度と比較して、層化抽出法の特徴を明らかならしめよう。但し比較の便宜上層化抽出法は比例抽

出法によるものとする。

その前に母集団に於ける全分散σ^2と層内分散$\sigma_w^{2\prime}$との関係を知って置かねばならない。母集団を層化した場合全分散σ^2は次のように変形する事が出来る。

$$\sigma^2=\frac{1}{N}\sum_{i=1}^{L}\sum_{j=1}^{N_i}(x_{ij}-\bar{x})^2$$

$$=\frac{1}{N}\sum_{i=1}^{L}\sum_{j=1}^{N_i}\{(x_{ij}-\bar{x}_i)+(\bar{x}_i-\bar{x})\}^2$$

$$=\frac{1}{N}\left\{\sum_{i=1}^{L}\sum_{j=1}^{N_i}(x_{ij}-\bar{x}_i)^2+2\sum_{i=1}^{L}(\bar{x}_i-\bar{x})\underbrace{\sum_{j=1}^{N_i}(x_{ij}-\bar{x}_i)}_{=0}+\sum_{i=1}^{L}N_i(\bar{x}_i-\bar{x})^2\right\}$$

こゝで

$$\sigma_w'^2=\frac{1}{N}\sum_{i=1}^{L}\sum_{j=1}^{N_i}(x_{ij}-\bar{x}_i)^2=\frac{1}{N}\sum_{i=1}^{L}N_i\sigma_i^2 \text{……層内分散}^{1)}$$

$$\sigma_b^{2\prime}=\frac{1}{N}\sum_{i=1}^{L}N_i(\bar{x}_i-\bar{x})^2 \text{…………………層間分散}$$

と置くと

$$\sigma^2=\sigma_w^{2\prime}+\sigma_b^{2\prime} \qquad (5\cdot17)$$

即ち全分散は層内分散と層間分散とに分ける事が出来るのである。

1) $\sigma_w^{2\prime}$、$\sigma_b^{2\prime}$は層を集落と見た時の集落内分散、集落間分散に相当する測度である。故に層内分散、層間分散と呼ぶ事にする。

さてN個の単位より成る母集団πから単純抽出法によって大いさnの標本を構成し、母平均を推定する場合の分散をV_1とすると、(2·33)より

$$V_1=\frac{N-n}{N-1}\frac{\sigma^2}{n}\doteqdot(1-f)\frac{\sigma^2}{n} \qquad \left[\because N-1\doteqdot N,\ \frac{n}{N}=f\right]$$

である。そして母集団πをN_i個宛の単位より成るL個の層に分け、各層よりその大いさに比例してn_i個宛の単位を抽出して大いさnの標本を構成し、母平均を推定する時の分散をV_2とすると、(5·8)より

$$V_2\doteqdot(1-f)\frac{\sigma_w^{2\prime}}{n}$$

である。そこでV_1のV_2に対する比を求めると

$$\frac{V_1}{V_2}\doteqdot\frac{(1-f)\frac{\sigma^2}{n}}{(1-f)\frac{\sigma_w^{2\prime}}{n}}=\frac{\sigma^2}{\sigma_w^{2\prime}}=1+\frac{\sigma_b^{2\prime}}{\sigma_w^{2\prime}} \qquad [5\cdot17] \quad (5\cdot18)$$

となる。(5·18)の右辺第二項は決して負数とならないために$V_1<V_2$は成立せず、従って比例抽出法の精度は単純抽出法の精度よりも低くなる事はないのである。

そして

$$\left.\begin{array}{l}\dfrac{\sigma_b^{2'}}{\sigma_w^{2'}}>0 \\ \quad〃\quad =0\end{array}\right\}\text{ならば}\left\{\begin{array}{ll}V_1>V_2 & \text{(i)}\\ 〃=〃 & \text{(ii)}\end{array}\right.$$

となる。故に $\sigma_b^{2'}=0$ 即ち各層平均 $\bar{x}_i$ が総平均 $\bar{x}$ に等しい、換言すれば層間同質の場合は(ii)が成立し両者の精度は一致するが、しからざる限り(i)が成立し、$\sigma_b^{2'}$ の値が大きくなるに従って（その時は(5・17)より明らかな如く $\sigma_w^{2'}$ の値は小さくなる）、即ち層内が同質的であり層間が異質的となる程、比例抽出法の精度は単純抽出法の精度よりも高くなるのである。

一般に母集団の層化は層内同質、層間異質的となるように行われるのである。尤も母集団を層化する場合、母集団の構成単位を同質的なものと同質的でないものとに区別する基準を、調査目的たる性質に就て知り得る時は、層内同質をより完全に実現する事が出来るであろうが、それは調査に先立っては不可能であり精々過去の調査結果を利用し得るのみである。又調査目的たる性質と理論的乃至は統計的に密接な相互関係を有する他の性質に就て、同質的なものと異質的なものとの区別の基準が得られる時は、それによっても同質化の効果を或る程度挙げる事が出来るであろう。故に完全な層内同質は困難乃至は不可能であるが、大なり小なり或る程度の同質化は可能であり、かくて一般に層化抽出法は単純抽出法よりも精度は高いのである。そして仮令このような層化の基準に誤りがあり層内同質化に失敗したとしても、単純抽出法よりも精度が悪くなる事はなく、唯層化の効果を挙げ得なかったに過ぎないのである。

尚層内分散 $\sigma_b^{2'}$ の値は又層の大いさ N_i によっても影響を受け、層の平均 $\bar{x}_i$ が同じである時は N_i が大なる程 $\sigma_b^{2'}$ は大きくなるのである。故に層内の同質性を変えない限り層の大いさが大なる程、層化抽出法の精度は高くなるのである。

かくて母集団の層化に際しては層内同質となるように、そして同質性を害しない限り層の大いさを出来る丈大ならしめるようにすればよい事が判るのである。

(5) 層化二段抽出法に於ける推定式

層化二段抽出法は母集団をL個の層に分け、各層より二段抽出によって調査単位を抽出して標本を構成する方法である。層化一段抽出法に於て述べたように層化抽出法による母平均の推定は、先づ層の量的大いさを推定し、それの総計を平均して行れるのであるから、層化二段抽出法による時は各層の量的大いさの推定に於て、二段抽出法の推定式を適用すればよいのである。

今L個の層の各〻はM_i個$(i=1, 2, \cdots, L)$の集落より成り、各集落はそれぞれN_{ij}個$(j=1, 2, \cdots, M_i)$の調査単位より構成されている場合、各層よりm_i個$(i=1, 2, \cdots, L)$の集落を抽出し、抽出された集落の各〻より更にn_{ij}個$(j=1, 2, \cdots, m_i)$宛の調査単位を抽出して、大いさ$n\left(=\sum_{i=1}^{L}\sum_{j=1}^{m_i} n_{ij}\right)$の標本を構成したとする。すると$i$番層の量的大いさ$X_i$の不偏推定値$X_i'$は(4·16)より

$$X_i' = N_i\bar{x}_i' = \frac{M_i}{m_i}\sum_{j=1}^{m_i}\left(\frac{N_{ij}}{n_{ij}}\sum_{k=1}^{n_{ij}} x_{ijk}'\right) \qquad (5\cdot19)$$

$$\underset{jk}{E}(X_i') = N_i \underset{jk}{E}(\bar{x}_i') = N_i\bar{x}_i = X_i \qquad [(4.18)]$$

であるから、母平均$\bar{x}$の推定値$\bar{x}'$は、

$$\bar{x}' = \frac{1}{N}\sum_{i=1}^{L}\frac{M_i}{m_i}\sum_{j=1}^{m_i}\left(\frac{N_{ij}}{n_{ij}}\sum_{k=1}^{n_{ij}} x_{ijk}'\right)^{1)} \qquad (5\cdot20)$$

となる。この推定値は不偏である。即ち

$$\underset{jk}{E}(\bar{x}') = \frac{1}{N}\sum_{i=1}^{L}\underset{jk}{E}(X_i') = \frac{1}{N}\sum_{i=1}^{L} X_i = \bar{x} \qquad (5\cdot21)$$

(5·20)の分散は

$$\underset{jk}{V}(\bar{x}') = \frac{1}{N^2}\sum_{i=1}^{L}\underset{jk}{V}(X_i')$$

$$= \frac{1}{N^2}\sum_{i=1}^{L}\left\{M_i^2\frac{M_i-m_i}{M_i-1}\frac{\sigma_{ie}^2}{m_i} + \frac{M_i}{m_i}\sum_{j=1}^{M_i} N_{ij}^2\frac{N_{ij}-n_{ij}}{N_{ij}-1}\frac{\sigma_{ij}^2}{n_{ij}}\right\} \qquad [(4\cdot22)] \quad (5\cdot22)$$

である。

1) (5·20)式は母集団の単位総数Nが判っている事を必要とする。若しそれが不明の時はNをも推定しなければならない。

i番層に於て第一次に抽出されたij番集落の単位数N_{ij}は、第二次抽出のための枠の作成に於て判っているため、i番層の単位数N_iの不偏推定値N_i'は(3·9)より

$$N_i' = \frac{M_i}{m_i}\sum_{j=1}^{m_i} N_{ij}$$

であり、これを総計する事によってNの推定値N'が得られる。即ち

$$N' = \sum_{i=1}^{L} \frac{M_i}{m_i} \sum_{j=1}^{m_i} N_{ij} \qquad (5\cdot23)$$

これを(5・20)のNに代入する時母平均の推定値は比推定となり、従ってその分散は(5・22)ではないのである。

各層内で集落の大いさが均一なる場合　層内の集落の大いさが相等しい　$N_{ij} = \bar{N}_i$, $\begin{pmatrix} i=1, 2, \cdots, L \\ j=1, 2, \cdots, M_i \end{pmatrix}$ 従って二次抽出単位の抽出数も相等しい　$n_{ij} = \bar{n}_i$ $\begin{pmatrix} i=1, 2, \cdots, L \\ j=1, 2, \cdots, m_i \end{pmatrix}$ とすると、i番層の量的大いさの推定値は(4・24)より

$$X_i' = N_i \bar{x}_i' = \frac{N_i}{m_i \bar{n}_i} \sum_{j=1}^{m_i} \sum_{k=1}^{\bar{n}_i} x_{ijk}' \qquad (5\cdot24)$$

であるため、母平均の推定値は

$$\bar{x}' = \frac{1}{N} \sum_{i=1}^{L} \frac{N_i}{m_i \bar{n}_i} \sum_{j=1}^{m_i} \sum_{k=1}^{\bar{n}_i} x_{ijk}' \qquad (5\cdot25)$$

となり、分散は

$$V_{jk}(\bar{x}') = \frac{1}{N^2} \sum_{i=1}^{L} V_{jk}(X_i')$$
$$= \frac{1}{N^2} \sum_{i=1}^{L} N_i^2 \left\{ \frac{M_i - m_i}{M_i - 1} \frac{\sigma_{ib}^2}{m_i} + \frac{\bar{N}_i - \bar{n}_i}{\bar{N}_i - 1} \frac{\sigma_{iw}^2}{m_i \bar{n}_i} \right\} \text{1)} \quad [(4.25)] \qquad (5\cdot26)$$

若しM_i、$\bar{N}_i$がm_i、$\bar{n}_i$に比して相当大であり、有限母集団修正が無視し得る時は

$$V_{jk}(\bar{x}') \doteqdot \frac{1}{N^2} \sum_{i=1}^{L} N_i^2 \left\{ \frac{\sigma_{ib}^2}{m_i} + \frac{\sigma_{iw}^2}{m_i \bar{n}_i} \right\} \qquad (5\cdot27)$$

となる。

1) σ_{ib}^2、σ_{iw}^2はi番層に於ける集落間分散、集落内分散である。$\sigma_b^{2\prime}$、$\sigma_w^{2\prime}$即ち母集団に於ける層間分散、層内分散と記号がまぎらわしいため注意すべきである。

(6)　適　正　配　分

(5・22)(5・26)式より明らかな如く層化二段抽出法に於ては、各層への抽出単位数の配分のみならず、更に層内の各抽出段階に於ける抽出個数の配分の如何によって標本推定の精度が異るのであり、それは又調査の費用をも異ならしめるため、最適なる各層、各抽出段階への抽出単位数の配分が問題となるのである。次にそれを各層内の集落の大いさが均一である場合に就て、且便宜上有限母集団修正が無視し得るものとして考察しよう。

この場合費用函数は、i番層に於て抽出された集落

一個当りの費用をk_{i1}、調査単位一個当りの調査費をk_{i2}とすると

$$T=\sum_{i=1}^{L}(k_{i1}m_i+k_{i2}m_i\bar{n}_i) \qquad (5\cdot28)$$

である。

先づ調査費用が決っている時は、(5・28)式のTが一定という条件の下に、(5・27)の$V(\bar{x}')$を最小ならしめるm_i、$\bar{n}_i$ ($i=1, 2,\dots,L$) の値を求めればよいのであり、それは

$$f(m_1, m_2,\cdots, m_L;\ \bar{n}_1, \bar{n}_2,\cdots, \bar{n}_L) =\frac{1}{N^2}\sum_{i=1}^{L}N_i^2\left\{\frac{\sigma_{ib}^2}{m_i}+\frac{\sigma_{iw}^2}{m_i\bar{n}_i}\right\} +\lambda\left\{\sum_{i=1}^{L}(k_{i1}m_i+k_{i2}m_i\bar{n}_i)-T\right\}$$

をm_i、$\bar{n}_i$ ($i=1, 2,\dots,L$) でそれぞれ偏微分して0と置き、(5・28)と連立せしめて解く事によって得られる。その過程は省略して結果のみを示すと

$$\bar{n}_i=\sqrt{\frac{k_{i1}}{k_{i2}}}\ \frac{\sigma_{iw}}{\sigma_{ib}} \qquad (5\cdot29)$$

$$m_i=\left\{\frac{T}{\sum_{i=1}^{L}N_i(\sqrt{k_{i1}}\ \sigma_{ib}+\sqrt{k_{i2}}\ \sigma_{iw})}\right\}\frac{N_i\sigma_{ib}}{\sqrt{k_{i1}}} \qquad (i=1,2,\cdots,L) \quad (5\cdot30)$$

となる。

次に精度が与えられた場合は、(5・27)式の$V(\bar{x}')$が一定であるという条件の下に(5・28)のTを最小ならしめるm_i、$\bar{n}_i$ ($i=1, 2,\dots,L$) の値を求めるのであり、それは

$$f(m_1, m_2,\cdots, m_L;\ \bar{n}_1, \bar{n}_2,\cdots, \bar{n}_L)=\sum_{i=1}^{L}(k_{i1}m_i+k_{i2}m_i\bar{n}_i) +\lambda\left\{\frac{1}{N^2}\sum_{i=1}^{L}N_i^2\left(\frac{\sigma_{ib}^2}{m_i}+\frac{\sigma_{iw}^2}{m_i\bar{n}_i}\right)-V(\bar{x}')\right\}$$

をm_i、$\bar{n}_i$ ($i=1, 2,\dots,L$) で偏微分して0と置き、(5・27)と連立させて解く事によって得られる。その結果は

$$\bar{n}_i=\sqrt{\frac{k_{i1}}{k_{i2}}}\ \frac{\sigma_{iw}}{\sigma_{ib}}$$

$$m_i=\left\{\frac{\sum_{i=1}^{L}N_i(\sqrt{k_{i1}}\ \sigma_{ib}+\sqrt{k_{i2}}\ \sigma_{iw})}{N^2V(\bar{x}')}\right\}\frac{N_i\sigma_{ib}}{\sqrt{k_{i1}}} \qquad (i=1,3,\cdots,L) \quad (5\cdot31)$$

であり、$\bar{n}_i$は(5・29)と同じである。

以上により明らかな如く各層に於ける第二次抽出の単位数$\bar{n}_i$は、費用、精度の何れが与えられた場合も共に(5・29)式によって与えられ、それの決定には費用T又は精度$V(\bar{x}')$は関係しないのである。そして費用、精度の規定は第一次抽出の単位数m_iに影響し、費用の増額又は精度の向上（$V(\bar{x}')$の減少）は、m_iを増加せしめる事となるのである。そして(5・30)及び(5・31)に於て右辺の中括弧内の

値は特定の層に無関係な一定の値であるから

$$m_i \propto \frac{N_i\sigma_{ib}}{\sqrt{k_{i1}}} \quad \therefore \frac{m_i}{M_i} \propto \frac{\bar{N}_i\sigma_{ib}}{\sqrt{k_{i1}}} \quad [\because N_i = M_i\bar{N}_i] \qquad (5\cdot32)$$

であり、従って各層に於ける第一次抽出の率は、その層の集落の大いさN_i及び集落間標準偏差σ_{ib}に正比例し、一次抽出単位当り費用の平方根$\sqrt{k_{i1}}$に反比例させればよい事が判るのである。[1)]

要するに一次抽出単位数m_iの各層への配分に於ては、層化一段抽出法に於ける適正配分の原則が、層内に於ける二次抽出単位数n_iの決定には、二段抽出法に於ける適正配分の原則が作用しているのであるという事が出来る。

1）最適抽出数（又は率）の決定に見られるこの関係は、有限母集団修正を無視し得ない場合も同様である。

そして若し一次抽出単位当り費用及び二次抽出単位当り費用が、すべての層に於て同じであり　$k_{i1}=k_1$, $k_{i2}=k_2$, $(i=1, 2, \cdots, L)$ なる場合は、(5·32)は

$$m_i \propto N_i\sigma_{ib} \quad \therefore \frac{m_i}{M_i} \propto \bar{N}_i\sigma_{ib} \qquad (5\cdot33)$$

であり、故に第一次抽出の率は$\bar{N}_i$及びσ_{ib}に比例させればよいのである。

(5·32)は層化一段抽出法に於けるデミング抽出法に相当し、(5·33)はネイマン抽出法に相当する。

(7)　母分散の推定

(5·22)又は(5·26)より明らかな如く、達成精度の算定に必要な母分散は各層毎の分散であるから、それの不偏推定値は二段抽出法に於ける母分散の推定式によって与えられるのである。

先づ(5·22)式の誤差分散の計算に必要な母分散はσ_{ie}^2及びσ_{ij}^2 $\left(\begin{matrix} i=1, 2, \cdots, L \\ j=1, 2, \cdots, M_i \end{matrix}\right)$であり、先づ$\sigma_{ij}^2$に関する不偏推定値は(4·36)より

$$\frac{M_i}{m_i}\sum_{j=1}^{m_i} N_{ij}^2 \frac{N_{ij}-n_{ij}}{N_{ij}-1}\frac{\hat{\sigma}_{ij}^2}{n_{ij}} \approx \sum_{j=1}^{M_i} N_{ij}^2 \frac{N_{ij}-n_{ij}}{N_{ij}-1}\frac{\sigma_{ij}^2}{n_{ij}} \qquad (5\cdot34)$$

但　$\hat{\sigma}_{ij}^2 = \frac{N_{ij}-1}{N_{ij}}\frac{n_{ij}}{n_{ij}-1}s_{ij}^2$

σ_{ie}^2の不偏推定値は(4·37)より

$$\frac{M_i-1}{M_i}\left(\frac{m_i}{m_i-1}s_{ie}^2 - \frac{1}{m_i}\sum_{j=1}^{m_i} N_{ij}^2 \frac{N_{ij}-n_{ij}}{N_{ij}-1}\frac{\hat{\sigma}_{ij}^2}{n_{ij}}\right)$$

$$\approx \sigma_{ie}^2 \quad (5\cdot35)$$

である。

各層内で集落の大いさが均一なる場合　又(5·26)式の誤差分散の計算に必要な母分散は σ_{ib}^2、σ_{iw}^2 ($i=1, 2, \dots, L$) であり、σ_{iw}^2 の不偏推定値は (4·38) より

$$\frac{\bar{N}_i-1}{\bar{N}_i}\frac{\bar{n}_i}{\bar{n}_i-1}s_{iw}^2 \approx \sigma_{iw}^2 \quad (5\cdot36)$$

σ_{ib}^2 の不偏推定値は (4·39) より

$$\frac{M_i-1}{M_i}\left(\frac{m_i}{m_i-1}s_{ib}^2-\frac{\bar{N}_i-\bar{n}_i}{\bar{N}_i}\frac{s_{iw}^2}{\bar{n}_i-1}\right) \approx \sigma_{ib}^2 \quad (5\cdot37)$$

である。

六　確率比例抽出法

今迄説明して来た標本構成の諸方式は、母集団の抽出単位のすべてに均等の抽出確率を与えて抽出するのであった。これに対して集落を抽出単位とする抽出方式(集落抽出法、多段抽出法)に於て、集落の大いさが著しく異る場合は、各集落にその大いさに比例した抽出確率を与えて抽出する方が、標本推定の精度を高める事が多いのである。このような抽出単位をその大いさに比例した確率で以て抽出する方式を、確率比例抽出法というのである。

しかし集落抽出法、多段抽出法に於ては、すべての集落に就てその大いさを正確に知り得ないのが普通であり、それを必要としないところにこれ等の抽出方式の長所があるのである。その時は略〻集落の大いさに比例すると考えられる値、即ち過去の大いさ又は調査せんとする性質と高度の正相関を有する他の性質に就ての量的大いさが判っている時は、それに比例する確率を与えて抽出する事によって同様の効果を挙げる事が出来るのである。

単位の確率比例抽出の実際手続は次のようにして容易に行う事が出来る。先づ集落を一番からM番迄に配列して、集落の大いさの累積和をとる。そして1より

番号	大いさ	累積和
1	N_1	N_1
2	N_2	N_1+N_2
3	N_3	$N_1+N_2+N_3$
⋮	⋮	⋮
M	N_M	$\sum_{j=1}^{M}N_j=N$
計	N	

N迄の間の数字をm個、乱数表を用いて決定し、累積和欄に於て抽出された個々の数字を含む最小の累積和に相当する番号の集落を以て抽出集落として、m個の集落を決定するのである。この抽出方法に於ては各集落の抽出される機会は同一でなく、その大いさに比例している事は明らかであろう（集落の大いさN_jが判らず、それに比例的な他の値P_jによらねばならない時も、その手続は今と全く同じである）。

この場合注意しなければならない事は、抽出する集落の数が二個以上の時は、この抽出方法では同じ集落が二度以上抽出される可能性がある事である。しかし同一集落の重複抽出を許さないとすると抽出操作が面倒であり、重複抽出を許す時は調査結果の集計に際して、重複した集落に就ての調査結果を重複回数倍して置けばよいのである（多段抽出の場合ならば、重複して抽出された集落に就ては第二段以下の抽出に於て、抽出単位数を重複回数倍丈増してもよい）。そして後に述べるように重複抽出を許さない場合の方が標本推定の精度は高いのであるが、母集団の調査単位数が抽出集落数に比して極めて大である時は両者は略〻等しいのである。

以下確率比例抽出法に於ける推定式並に分散を求めるに当り、便宜上二段抽出法に就て、先づ簡単な重複抽出を許す場合を考察し、後に重複抽出を許さない場合を取扱う事にする。そして集落の大いさN_jが判らない場合が普通であるから、最初はN_jに比例的な任意の値P_jによる確率比例抽出法を取扱い、それからN_jによる場合を誘導する事にする。その場合先づ第一次抽出の単位数が一個の場合の不偏推定並に分散を求めて置くのが便利であるから、次にそれを考察しよう。

(1)　一個の集落を抽出する場合

母集団のM個の集落から一個の集落を、任意の値P_jに比例した確率で抽出し、次にその集落のN_j個の調査単位からn_j個の単位を等確率で抽出して標本を構成し、母平均を推定するものとする（一般に多段抽出法に於

て確率比例抽出が行れるのは、第一段の抽出に於てのみであって、第二段以下の抽出は等確率抽出が普通である。)。

茲で確率比例抽出法に於ける期望値計算をよく理解して置かねばならぬ。M個の集落より値P_jに比例した確率で抽出された一個の集落の量的大いさの期望値は、j番集落の抽出される確率は$\frac{P_j}{P}$$\left(\text{但 } P=\sum_{j=1}^{M}P_j\right)$　$(j=1,2,\dots,M)$であるから、期望値の定義式(2・13)より

$$E(X_j)=\sum_{j=1}^{M}\frac{P_j}{P}X_j \tag{6.1}$$

である。若し集落の抽出がその大いさに比例した確率で行れる時は、これに於て$P_j=N_j$, $P=N$と置けばよい。(等確率抽出の場合は、j番集落の抽出される確率は$\frac{1}{M}$であるから$E(X_j)=\frac{1}{M}\sum_{j=1}^{M}X_j=\bar{X}$となる。)

次に確率比例抽出により一個の集落を抽出して、母平均を推定する場合の不偏推定を知らねばならぬ。母集団が集落より成る時は母平均$\bar{x}$は

$$\bar{x}=\frac{1}{N}\sum_{j=1}^{N}X_j=\frac{1}{N}\sum_{j=1}^{M}\frac{P_j}{P}\frac{P}{P_j}X_j$$
$$=\sum_{j=1}^{M}\frac{P_j}{P}\left(\frac{P}{NP_j}X_j\right)$$

と変形され得、これを(6・1)式と比較する事により、$\frac{P}{NP_j}X_j$の期望値であると解する事が出来る。従って集落の量的大いさX_jに$\frac{P}{NP_j}$を乗ずる事によって、母平均の不偏推定値が得られるのである。

さて二段抽出であるから、抽出された集落の量的大いさの不偏推定値

$$X_j'=\frac{N_j}{n_j}\sum_{k=1}^{n_j}x_{jk}',\quad E_k^{(j)}(X_j')=X_j$$

に$\frac{P}{NP_j}$を乗ずる事によって、母平均の不偏推定値$\bar{x}'$が得られる。即ち

$$\bar{x}'=\frac{P}{NP_j}X_j'=\frac{P}{NP_j}\left(\frac{N_j}{n_j}\sum_{k=1}^{n_j}x_{jk}'\right) \tag{6・2}$$

これの期望値を求めると

$$E_{jk}(\bar{x}')=E_j\left\{\frac{P}{NP_j}E_k^{(j)}(X_j')\right\}\overset{1)}{=}\frac{1}{N}E_j\left(\frac{P}{P_j}X_j\right)$$
$$=\frac{1}{N}\sum_{j=1}^{M}\frac{P_j}{P}\frac{P}{P_j}X_j=\frac{1}{N}\sum_{j=1}^{M}X_j=\bar{x} \tag{6・3}$$

となり、(6・2)は不偏推定である事が証明されるのである。

1) この場合E_jは確率比例抽出に於ける期望値を、$E_k^{(j)}$は等確率抽出に於ける期望値を意味する事に注意しなければならぬ。

次に(6・2)式の分散を求めると

$$V_{jk}(\bar{x}')=V_j\left(E_k(j)\bar{x}'\right)+E_j V_k(j)(\bar{x}')$$

第一項：$V_j\left(E_k(j)\bar{x}'\right)=V_j\left(\frac{P}{NP_j}E_k(j)X_j'\right)$

$$=\frac{1}{N^2}V_j\left(\frac{P}{P_j}X_j\right)$$

しかるに $E_j\left(\frac{P}{P_j}X_j\right)=\sum_{j=1}^{M}\frac{P_j}{P}\left(\frac{P}{P_j}X_j\right)=\sum_{j=1}^{M}X_j=X$

であるため、分散の定義式(2・22)より

$$V_j\left(\frac{P}{P_j}X_j\right)=E_j\left[\frac{P}{P_j}X_j-E_j\left(\frac{P}{P_j}X_j\right)\right]^2$$

$$=\sum_{j=1}^{M}\frac{P_j}{P}\left(\frac{P}{P_j}X_j-X\right)^2\equiv\bar{\sigma}_e^2 \qquad (6\cdot4)$$

故に

$$V_j\left(E_k(j)\bar{x}'\right)=\frac{1}{N^2}\bar{\sigma}_e^2$$

又

第二項：$E_j V_k(j)(\bar{x}')=E_j\left\{\left(\frac{P}{NP_j}\right)^2\underbrace{V_k(j)(X_j')}\right\}$

$$=\sum_{j=1}^{M}\frac{P_j}{P}\left(\frac{P}{NP_j}\right)^2A_j \qquad N_j^2\frac{N_j-n_j}{N_j-1}\frac{\sigma_j^2}{n_j}\equiv A_j$$

$$=\frac{1}{N^2}\sum_{j=1}^{M}\frac{P}{P_j}N_j^2\frac{N_j-n_j}{N_j-1}\frac{\sigma_j^2}{n_j}$$

かくて

$$V_{jk}(\bar{x}')=\frac{1}{N^2}\left\{\bar{\sigma}_e^2+\sum_{j=1}^{M}\frac{P}{P_j}N_j^2\frac{N_j-n_j}{N_j-1}\frac{\sigma_j^2}{n_j}\right\} \qquad (6\cdot5)$$

となる。

集落の大いさに比例する場合

集落の大いさによ

る確率比例抽出の場合は、上の諸式に於て $P_j=N_j$, $P=N$ と置けばよい。推定式は(6・2)より

$$\bar{x}'=\frac{1}{n_j}\sum_{k=1}^{n_j}x_{jk}'=\bar{x}_j' \qquad (6\cdot6)$$

となり、分散は

$$\bar{\sigma}_e^2=\sum_{j=1}^{M}\frac{P_j}{P}\left(\frac{P}{P_j}X_j-X\right)^2=\sum_{j=1}^{M}\frac{N_j}{N}\left(\frac{N}{N_j}N_j\bar{x}_j-N\bar{x}\right)^2$$

$$\left[\because\ \bar{x}=\frac{X}{N}\right]$$

$$=N^2\frac{1}{N}\sum_{j=1}^{M}N_j(\bar{x}_j-\bar{x})^2$$

$$=N^2\sigma_b^2,\qquad \sigma_b^2=\frac{1}{N}\sum_{j=1}^{M}N_j(\bar{x}_j-\bar{x})^2 \qquad (6.7)$$

であるため、(6・5)より

$$V_{jk}(\bar{x}')=\sigma_b^2+\sum_{j=1}^{M}\frac{N_j}{N}\frac{N_j-n_j}{N_j-1}\frac{\sigma_j^2}{n_j} \qquad (6\cdot8)$$

となる。

(2) 重複抽出を許す場合

(i) 二段抽出法に於ける推定式

M個の集落より成る母集団からm個の集落を、任意の値P_jに比例した確率で、且同一集落が二度以上重複

して抽出されても差支えないものとして抽出し、次に抽出された各集落のN_j個の調査単位よりn_j個の単位を等確率で抽出して、母平均$\bar{x}$を推定するものとする。

前節より明らかな如く一集落による母平均の不偏推定値は$\dfrac{P}{NP_j}X_j'$であり、茲ではm個の集落を抽出するのであるから、抽出された各集落より求めたこの不偏推定値を平均して、母平均の推定値$\bar{x}'$とすればよいであろう。故に

$$\bar{x}'=\frac{1}{m}\sum_{j=1}^{m}\frac{P}{NP_j}X_j'=\frac{1}{m}\sum_{j=1}^{m}\frac{P}{NP_j}\left(\frac{N_j}{n_j}\sum_{k=1}^{n_j}x_{jk}'\right) \quad (6\cdot9)$$

(但し、このm個の$\dfrac{P}{NP_j}X_j'$の中には、同一集落が重複抽出された場合は、その重複回数丈同じものがある。)

これの期望値を求めると

$$\underset{jk}{E}(\bar{x}')=\frac{1}{m}\sum_{j=1}^{m}\underset{jk}{E}\left(\frac{P}{NP_j}X_j'\right) \quad [(2\cdot17)]$$

$$=\frac{1}{m}\sum_{j=1}^{m}\bar{x}=\bar{x} \quad [(6\cdot3)] \quad (6\cdot10)$$

となり、従って(6·9)は不偏推定である。分散を求めると

$$\underset{jk}{V}(\bar{x}')=\frac{1}{m^2}\sum_{j=1}^{m}\underset{jk}{V}\left(\frac{P}{NP_j}X_j'\right)^{1)} \quad [(2\cdot27)] \quad (6\cdot11)$$

$$=\frac{1}{N^2}\left\{\frac{\bar{\sigma}_e^2}{m}+\frac{1}{m}\sum_{j=1}^{M}\frac{P}{P_j}N_j^2\frac{N_j-n_j}{N_j-1}\frac{\sigma_j^2}{n_j}\right\} \quad [(6.5)] \quad (6\cdot12)$$

となる。

1) 重複抽出が許されるとすると、m個の集落の抽出の各々は相互に独立となる。けだし同一集落を重複して抽出してもよい時は、抽出の都度それを元え戻して、母集団を常に同じ構成状態に置いて抽出を繰返す場合と同じであり、その時は先の抽出によつて如何なる集落が抽出されても、それによつて後の抽出が影響される事はないからである。

集落の大いさに比例する場合　(6·9)(6·12)に於て$P_j=N_j$, $P=N$と置けばよい。推定式は

$$\bar{x}'=\frac{1}{m}\sum_{j=1}^{m}\left(\frac{1}{n_j}\sum_{k=1}^{n_j}x_{jk}'\right)=\frac{1}{m}\sum_{j=1}^{m}\bar{x}_j' \quad (6\cdot13)$$

分散は

$$\underset{jk}{V}(\bar{x}')=\frac{\sigma_b^2}{m}+\frac{1}{m}\sum_{j=1}^{M}\frac{N_j}{N}\frac{N_j-n_j}{N_j-1}\frac{\sigma_j^2}{n_j} \quad [(6\cdot7)] \quad (6\cdot14)$$

となる。

(ii)　母分散の推定

(6·12)式により誤差分散を計算するに必要な母分散は$\bar{\sigma}_e^2$及びσ_j^2 ($j=1, 2, \cdots, M$)である。しかし重複抽出可能

の場合はそれ等を別個に推定する必要はなく、(6•12)式全体の不偏推定が得られるのである。

それは$\bar{\sigma}_e^2$と同じ性質の標本に於ける分散

$$\tilde{s}_e^2=\frac{1}{m}\sum_{j=1}^{m}\left(\frac{P}{P_j}X_j'-X'\right)^2,\quad X'=\frac{1}{m}\sum_{j=1}^{m}\frac{P}{P_j}X_j'$$

の期望値を求めればよいのである。先づ$\tilde{s}_e^2$を次のように変形する。

$$\tilde{s}_e^2=\frac{1}{m}\sum_{j=1}^{m}\left\{\left(\frac{P}{P_j}X_j'-X\right)-(X'-X)\right\}^2$$

$$=\frac{1}{m}\left\{\sum_{j=1}^{m}\left(\frac{P}{P_j}X_j'-X\right)^2-2(X'-X)\underbrace{\sum_{j=1}^{m}\left(\frac{P}{P_j}X_j'-X\right)}_{=m(X'-X)}+m(X'-X)^2\right\}$$

$$=\frac{1}{m}\sum_{j=1}^{m}\left(\frac{P}{P_j}X_j'-X\right)^2-(X'-X)^2$$

茲で期望値をとると

$$\underset{jk}{E}(\tilde{s}_e^2)=\underset{jk}{E}\left(\frac{P}{P_j}X_j'-X\right)^2-\underset{jk}{E}(X'-X)^2$$

しかるに　$\underset{jk}{E}\left(\frac{P}{P_j}X_j'\right)=N\underset{jk}{E}\left(\frac{P}{NP_j}X_j'\right)=N\bar{x}=X$　[(6•3)]　(6•15)

$$\underset{jk}{E}(X')=\frac{1}{m}\sum_{j=1}^{m}\underset{jk}{E}\left(\frac{P}{P_j}X_j'\right)=X\quad[(6•15)]$$

$$\therefore\quad \underset{jk}{E}(\tilde{s}_e^2)=\underset{jk}{V}\left(\frac{P}{P_j}X_j'\right)-\underset{jk}{V}(X')\qquad(6•16)$$

ところが　$\frac{P}{P_j}X_j'$　は(6•9)式に於て　$m=1$　と置きNを乗じたものに等しい。即ち　$\frac{P}{P_j}X_j'=N\bar{x}'$,　$m=1$　故に

$$\underset{jk}{V}\left(\frac{P}{P_j}X_j'\right)=N^2\underset{jk}{V}(\bar{x}'),\quad m=1$$

$$=\bar{\sigma}_e^2+\underbrace{\sum_{j=1}^{M}\frac{P}{P_j}N_j^2\frac{N_j-n_j}{N_j-1}\cdot\frac{\sigma_j^2}{n_j}}_{=A_j}\quad[(6.12)]$$

又X'は(6•9)式に於てN倍したものに等しい。即ち　$X'=N\bar{x}'$,　故に

$$\underset{jk}{V}(X')=N^2\underset{jk}{V}(\bar{x}')$$

$$=\frac{\bar{\sigma}_e^2}{m}+\frac{1}{m}\underbrace{\sum_{j=1}^{M}\frac{P}{P_j}N_j^2\frac{N_j-n_j}{N_j-1}\cdot\frac{\sigma_j^2}{n_j}}_{=A_j}\quad[(6.12)]$$

かくて

$$\underset{jk}{E}(\tilde{s}_e^2)=\left(1-\frac{1}{m}\right)\bar{\sigma}_e^2+\left(1-\frac{1}{m}\right)\sum_{j=1}^{M}\frac{P}{P_j}A_j$$

$$=\frac{m-1}{m}\left(\bar{\sigma}_e^2+\sum_{j=1}^{M}\frac{P}{P_j}A_j\right).$$

両辺を　$(m-1)N^2$　で除して

$$\underset{jk}{E}\left\{\frac{1}{(m-1)N^2}\tilde{s}_e^2\right\}=\frac{1}{N^2}\left\{\frac{\bar{\sigma}_e^2}{m}+\frac{1}{m}\sum_{j=1}^{M}\frac{P}{P_j}N_j^2\frac{N_j-n_j}{N_j-1}\cdot\frac{\sigma_j^2}{n_j}\right\}\qquad(6•17)$$

この式の右辺は(6•12)に外ならない。故に

$$\frac{1}{(m-1)N^2}\tilde{s}_e^2\approx\underset{jk}{V}(\bar{x}')\qquad(6•18)$$

である。

集落の大いさに比例する場合(6・17)に於て$P_j=N_j$, $P=N$ と置く事により(6・14)式の不偏推定が得られる。

この場合s_e^2は

$$X'=\frac{1}{m}\sum_{j=1}^{m}\frac{P}{P_j}X_j'=N\frac{1}{m}\sum_{j=1}^{m}\bar{x}_j'=N\bar{\bar{x}}'$$

であるため

$$s_e^2=\frac{1}{m}\sum_{j=1}^{m}\left(\frac{P}{P_j}X_j'-X'\right)^2$$

$$=\frac{1}{m}\sum_{j=1}^{m}\left(\frac{N}{N_j}N_j\bar{x}_j'-N\bar{\bar{x}}'\right)^2=N^2\frac{1}{m}\sum_{j=1}^{m}(\bar{x}_j'-\bar{\bar{x}}')^2$$

$$=N^2s_b^2,\qquad s_b^2=\frac{1}{m}\sum_{j=1}^{m}(\bar{x}_j'-\bar{\bar{x}}')^2 \tag{6・19}$$

となる。従って(6・17)は

$$\underset{jk}{E}\left(\frac{1}{m-1}s_b^2\right)=\frac{\sigma_b^2}{m}+\frac{1}{m}\sum_{j=1}^{M}\frac{N_j}{N}\frac{N_j-n_j}{N_j-1}\frac{\sigma^2_j}{n_j}$$

[(6・7)]

となり、この式の右辺は(6・14)と同じであるため

$$\frac{1}{m-1}s_b^2\approx\underset{jk}{V}(\bar{\bar{x}}') \tag{6・20}$$

である。