任意標本調査法（二）

関　弥三郎

二、任意抽出法の理論

前に考察した母集団の構造を示す値（母数）を標本の構造を示す値（統計量）より推定するに当って、任意標本理論は母集団より同じ大いさの標本を可能な限り抽出構成する時、各標本より得られる統計量の分布（標本分布 sampling distribution）を求め、それと母数との関係を明らかにする事によって、現実に任意標本より得られた統計量はそのうちの一の値をとる事から、誤差の範囲を明示して母数を推定する事が出来るのである。以下母集団の平均値の推定を例にとって、標本分布を求めそれによる標本推定の方式を明らかならしめよう。

標本調査の実際に於て標本より推定すべき母数は主として平均値又は比例数であり、その他の母数が推定される場合もそれ自体の推定が目的ではなく、この平均値、比例数の推定の必要上求められる事が多い。故に以下の説明に於ては母平均及び母比例数の推定を中心に説明して行く事にする。

(1) 標本平均分布

母集団の平均値（母平均）の推定値としては、同じ性質の集団構造の簡約表章値である標本平均値を以てする事が考えられる。今母集団πはN個の単位より成り、各単位の有する量的標識をx_1 x_2……x_Nとし、その平均値を$\bar{x}$、分散をσ^2とすると、

$$\bar{x}=\frac{1}{N}\sum_{j=1}^{N}x_j,\quad \sigma^2=\frac{1}{N}\sum_{j=1}^{N}(x_j-\bar{x})^2 \tag{2.1}$$

である。この母集団よりn個の単位を任意に抽出して構成した標本Oの各単位の量的標識をx'_1 x'_2…x'_nとし、その平均値を$\bar{x}'$、分散をs^2で表すと

$$\bar{x}'=\frac{1}{n}\sum_{j=1}^{n}x'_j,\quad s^2=\frac{1}{n}\sum_{j=1}^{n}(x'_j-\bar{x}')^2 \tag{2.2}$$

である。

茲にx_jは母集団に於ける単位の有する量的標識を表し、x'_jはそれより抽出構成された標本内の単位の持つ量的標識である。又$\bar{x}$、σ^2はそれぞれ母集団の平均値、分散であり、$\bar{x}'$、s^2は標本の平均値、分散を表す。このような母集団に於ける値と標本に於ける値との記号上の区別をよく弁える事は、任意標本理論を学ぶ上に於て極めて重要である。

この標本平均値$\bar{x}'$の標本分布を求めるのであるが、その状態は標本分布の平均値、分散等を計算し、分布の形を調べる事によって規定する事が出来るのである。故に先づ標本平均分布の平均値を求めよう。

標本分布は母集団より可能な限り標本を構成する時に各標本より得られる統計量の分布であるが、茲に「可能な限りの標本構成」とは、N個の単位よりn個の単位を抽出し、次にその残り（$N-n$）単位よりn単位抽出して行ってN個の単位がなくなる迄標本を構成する事、即ちn個の単位がすべて異る単位である如き標本の可能な限りの構成ではなくして、仮令n個の単位が全く同じであっても、それの抽出の仕方が全く同じでないような標本、即ち抽出に順序を附けた時各単位の抽出される順序がすべて同じでない標本の可能な限りの構成である事は明らかであろう[1]。しからばそのような標本の理論的に構成可能な数は、N個の単位の中からn個取出す「順列」の数として求められ、故

に

$$_NP_n = N(N-1)\cdots\cdots(N-n+1)$$

通りである。

1） 若し単位の抽出に順序を附けないで考えると、標本を構成するn個の単位のすべてが同じでないような標本（従ってn個の単位が全部異る標本もあれば、只一個の単位のみを異にし残り（$n-1$）個の単位が同じであるような標本もある）の可能な限りの構成となり、その理論的な数はN個の単位の中からn個取出す「組合せ」の数として求められ

$$_NC_n = \frac{_NP_n}{n!} = \frac{N(N-1)\cdots(N-n+1)}{n(n-1)\cdots\cdots1}$$

通りである。

「可能な限りの標本の数」を$_NC_n$通りとするならば、以下の説明に於て$_NP_n$の代りに$_NC_n$と置けばよい。（但し(2.6)の式の説明の場合は若干注意を要する）

この$_NP_n$組の標本の各々に就て求めた標本平均値$\bar{x}'$の平均は

$$\frac{1}{_NP_n}\sum_{k=1}^{_NP_n}\bar{x}'_k = \frac{1}{_NP_n}\sum_{k=1}^{_NP_n}\left(\frac{1}{n}\sum_{j=1}^{n}x'_j\right) = \frac{1}{n\,_NP_n}\underbrace{\sum_{k=1}^{_NP_n}\sum_{j=1}^{n}x'_{jk}}_{=A} \qquad (2.3)$$

$$\begin{array}{c|cccc} & \multicolumn{4}{c}{n\text{（標本単位数）}} \\ \hline _NP_n\text{（標本の数）} & x'_{11} & x'_{12} & \cdots\cdots & x'_{1n} \\ & x'_{21} & x'_{22} & \cdots\cdots & x'_{2n} \\ & x'_{31} & x'_{32} & \cdots\cdots & x'_{3n} \\ & \cdots & \cdots & \cdots & \cdots \\ & \cdots & \cdots & \cdots & \cdots \end{array}$$

茲にAは$_NP_n$組の標本の構成単位の有する変量の総計であって、換言すれば$n\,_NP_n$個のx'の和である。そしてこの中には母集団に於けるN個の単位の有する変量x_1 x_2…x_Nのすべてが、同じ割合で含まれている筈であるから、このN個のx_jの和が$\frac{n\,_NP_n}{N}$組ある事になる。故に

$$A = \frac{n\,_NP_n}{N}\sum_{j=1}^{N}x_j$$

と書替えられ、従って(2.3)式は

$$\frac{1}{_NP_n}\sum_{k=1}^{_NP_n}\bar{x}'_k = \frac{1}{n\,_NP_n}\cdot\frac{n\,_NP_n}{N}\sum_{j=1}^{N}x_j$$

$$= \frac{1}{N}\sum_{j=1}^{N}x_j = \bar{x} \qquad (2.4) \qquad \text{（(2.1)より）}$$

となる。即ち理論的な標本平均分布の平均値は母平均$\bar{x}$に一致するのである。

次にその分散$\sigma^2_{\bar{x}'}$を求めよう。分散は変量（茲では$\bar{x}'$）のその平均値（(2.4)式より$\bar{x}$）よりの偏差の自乗の

平均であるから

$$\sigma^2_{\bar{x}'}=\frac{1}{{}_NP_n}\sum_{k=1}^{{}_NP_n}(\bar{x}'_k-\bar{x})^2$$

$$=\frac{1}{{}_NP_n}\sum_{k=1}^{{}_NP_n}\bar{x}'^2_k-\bar{x}^2 \qquad (2.5)\ 〔簡便計算式(6'')より〕$$

第一項：$\frac{1}{{}_NP_n}\sum_{k=1}^{{}_NP_n}\bar{x}'^2_k$

$$=\frac{1}{{}_NP_n}\sum_{k=1}^{{}_NP_n}\left\{\frac{1}{n}(x'_{1k}+x'_{2k}+\cdots\cdots+x'_{nk})\right\}^2$$

$$=\frac{1}{n^2\,{}_NP_n}\left\{\underbrace{\sum_{k=1}^{{}_NP_n}\sum_{j=1}^{n}x'^2_{jk}}_{=B}+\underbrace{\sum_{k=1}^{{}_NP_n}\sum_{j\neq j'}^{n}\sum^{n}x'_{jk}x'_{j'k}}_{=C}\right\} \qquad (2.6)$$

茲にBは(2.3)式のAに於て、x'が自乗されている場合であるから

$$B=\frac{n\,{}_NP_n}{N}\sum_{j=1}^{N}x^2_j$$

と書替えられる。Cは一の標本O_kを構成するn個の単位を、二個宛組合せて得られる値の組$(x'_{jk},\ x'_{j'k})$, $\left(\begin{smallmatrix}j,j'=1,2,\cdots,n\\ j\neq j'\end{smallmatrix}\right)$——その数は${}_nP_2$組ある——の積$x'_{jk}x'_{j'k}$の総和$\sum_{j\neq j'}^{n}\sum^{n}x'_{jk}x'_{j'k}$を、${}_NP_n$組の標本すべてに就て求め、それを合計したものであって、換言すれば${}_nP_2\,{}_NP_n$個の積$x'_jx'_{j'}$の和である。

$$
{}_NP_n\left\{\begin{array}{l}
n\left\{\begin{array}{llll}
\multicolumn{4}{c}{\overbrace{\hspace{8em}}^{n-1}}\\
x'_{11}x'_{12} & x'_{11}x'_{13} & \cdots & x'_{11}x'_{1n}\\
x'_{12}x'_{11} & x'_{12}x'_{13} & \cdots & x'_{12}x'_{1n}\\
\multicolumn{4}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots}\\
x'_{1n}x'_{11} & x'_{1n}x'_{12} & \cdots & x'_{1n}x'_{1n-1}
\end{array}\right.\\
n\left\{\begin{array}{llll}
x'_{21}x'_{22} & x'_{21}x'_{23} & \cdots & x'_{21}x'_{2n}\\
\multicolumn{4}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots}
\end{array}\right.\\
\quad\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
\quad\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots
\end{array}\right.
$$

そしてその中には母集団に於けるN個の単位の変量$x_1\ x_2\cdots x_N$の、二つ宛の可能な組合せ$(x_j, x_{j'})$, $\left(\begin{smallmatrix}j,j'=1,2,\cdots,N\\ j\neq j'\end{smallmatrix}\right)$——その数は${}_NP_2$組ある——の積$x_jx_{j'}$のすべてが、同じ割合で含まれている筈であるから、${}_NP_2$個の積$x_jx_{j'}$の和が$\frac{{}_nP_2\,{}_NP_n}{{}_NP_2}=\frac{n(n-1){}_NP_n}{N(N-1)}$組ある事になる。従って

$$C=\frac{n(n-1){}_NP_n}{N(N-1)}\sum_{j\neq j'}^{N}\sum^{N}x_jx_{j'}$$

と書く事が出来る。故に(2.6)式は

$$\frac{1}{{}_NP_n}\sum_{k=1}^{{}_NP_n}\bar{x}'^2_k=\frac{1}{n^2\,{}_NP_n}\left\{\frac{n\,{}_NP_n}{N}\sum_{j=1}^{N}x^2_j+\frac{n(n-1){}_NP_n}{N(N-1)}\sum_{j\neq j'}^{N}\sum^{N}x_jx_{j'}\right\}$$

$$=\frac{1}{nN}\sum_{j=1}^{N}x^2_j+\frac{n-1}{nN(N-1)}\sum_{j\neq j'}^{N}\sum^{N}x_jx_{j'} \qquad (2.7)$$

そして母集団分布に於て

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

$$\therefore\quad \frac{1}{N}\sum_{j=1}^{N}x_j{}^2=\sigma^2+\bar{x}^2 \qquad (2.8)$$

$$\bar{x}^2=\left\{\frac{1}{N}\sum_{j=1}^{N}x_j\right\}^2=\frac{1}{N^2}\left\{\sum_{j=1}^{N}x^2{}_j+\sum_{j\neq j'}^{N}\sum^{N}x_jx_{j'}\right\}$$

$$\therefore\quad \sum_{j\neq j'}^{N}\sum^{N}x_jx_{j'}=N^2\bar{x}^2-\sum_{j=1}^{N}x^2{}_j$$

$$=N^2\bar{x}^2-N(\sigma^2+\bar{x}^2) \qquad \text{〔(2.8)より〕}$$

$$=N(N-1)\bar{x}^2-N\sigma^2$$

なるため、これ等を(2.7)式に代入する事により

$$\frac{1}{{}_NP_n}\sum^{{}_NP_n}\bar{x}'^2_k=\frac{1}{n}(\sigma^2+\bar{x}^2)$$

$$+\frac{n-1}{nN(N-1)}\{N(N-1)\bar{x}^2-N\sigma^2\}$$

$$=\frac{1}{n}(\sigma^2+\bar{x}^2)+\frac{n-1}{n}\bar{x}^2-\frac{n-1}{n(N-1)}\sigma^2$$

$$=\bar{x}^2+\frac{N-n}{n(N-1)}\sigma^2$$

これを(2.5)式に代入して

$$\sigma^2{}_{\bar{x}'}=\frac{\sigma^2}{n}\,\frac{N-n}{N-1} \qquad (2.9)$$

となる。即ち理論的な標本平均分布の分散は母分散σ^2の$\dfrac{N-n}{n(N-1)}$倍である。この係数$\dfrac{N-n}{n(N-1)}$は1よりも小[1]なるため、標本平均の分布は母集団分布と平均値は同じであるが、その分布の範囲は母集団分布よりも狭いのである。

1) $n>1$なる限り $N-n<N-1<n(N-1)$

$$\therefore\quad \frac{N-n}{n(N-1)}<1$$

(2.9)式より標本平均分布の標準偏差は

$$\sigma_{\bar{x}'}=\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \qquad (2.10)$$

であり、これを平均値で除して相対化した変異係数$C_{\bar{x}'}$は、母集団分布の変異係数を$c=\dfrac{\sigma}{\bar{x}}$とすると

$$C_{\bar{x}'}=\frac{\sigma_{\bar{x}'}}{\bar{x}}=\frac{c}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \qquad (2.11)$$

である。

そして(2.9)式に於てNが無限大ならば

$$\lim_{N\to\infty}\sigma^2{}_{\bar{x}'}=\lim_{N\to\infty}\frac{1-\frac{n}{N}}{1-\frac{1}{N}}\,\frac{\sigma^2}{n}=\frac{\sigma^2}{n} \qquad (2.12)$$

となる。故に母集団の大いさが無限大即ち無限母集団

の場合は、標本平均分布の分散は母集団の大いさに無関係に、標本の大いさ並に母分散の値によって決るのである。尚母集団の大いさNが有限であっても、抽出単位数nに比して極めて大であり、$\frac{N-n}{N-1} \fallingdotseq 1$として取扱っても計算上差支えない程度の大いさであるならば(2.12)式を用いる事が出来る。従って(2.9)式の$\frac{N-n}{N-1}$を有限母集団修正 finite population correction 又は有限乗数 finite multiplier という。

(2) 確率分布、期望値計算

今母集団πよりn個の単位を抽出して標本Oを作り標本平均値$\bar{x}'$を計算する時、前項に考察した$_NP_n$個の$\bar{x}'$の何れか一の値をとる筈である。そしてn個の単位の抽出従って標本の構成が無作為 at random になされるのであるから、$_NP_n$個の$\bar{x}'$の何れがとられるかは全く偶然的に決り、従って各$\bar{x}'$の出る可能性（即ち確率[1)]）はその存在の割合（相対的度数）によって決るのであって、この場合はすべて$\frac{1}{_NP_n}$である。

このような無作為操作の結果として$x_1\ x_2 \ldots x_k$の何れかをとる変数ξがあって、ξが$x_1\ x_2 \ldots x_k$をとる確率がそれぞれ$p_1\ p_2 \ldots p_k$である時は、このような変数ξを確率変数 random variable という。この$x_1\ x_2 \ldots x_k$にそれぞれの確率$p_1\ p_2 \ldots p_k$を相対的度数として与えたものの分布を、ξの確率分布 probability distribution という。そして無作為操作の結果現実にξがとる値を確率変数の実現値という。標本抽出の場合でいえば標本平均値$\bar{x}'$は確率変数であり、それの実現値$\bar{x}'_k$（$k=1,2,\ldots,{_NP_n}$）は等確率$\frac{1}{_NP_n}$の一様確率分布をなすといえるのである。

そしてこの無作為操作を実際に多数回繰返す場合、ξの実現値の分布は予め理論的に判っているξの確率分布を以て近似される事が、実験により経験的に証明されるのである。（これを経験的大数法則 empirical

1) 「確率」の定義は次の如くである。「一の事象Eが、すべて等可能であると考えられるn個の可能な場合のうちのm個の場合に就て起るのであれば、当該事象Eの確率は$\frac{m}{n}$と定義する。」

law of large numbers という）例えば骰子を投げる場合1から6迄の各目の出る可能性は、骰子が正確な等六面体に作られて居り投げ方に特定の傾向がない限り、すべて同一であると考えられるため、投骰の結果出る目の数は1から6迄の値をとる確率変数であり、確率$\frac{1}{6}$の一様確率分布をなすといえる。実際に骰子投げを多数回繰返し出た目の度数分布を作ると、それは略〻相対的度数$\frac{1}{6}$の一様分布をなし、投骰回数が増大するに従って益〻理論的な$\frac{1}{6}$の一様確率分布に収斂する事が判るのである。従って母集団πより大いさnの標本Oを繰返して無作為に構成する時得られる標本平均値$\bar{x}'$の実現値の分布は、標本の数従って$\bar{x}'$の数が増加するに連れて、相対的度数が前項に考察した$\bar{x}'$の理論的な標本平均分布に等しくなって行くといえるのである。

確率変数ξの確率分布の状態は度数分布の場合と同様に、その平均値、分散等を求め分布の形を調べる事によって規定する事が出来る。

確率変数ξの確率分布の（算術）平均値を期望値（又は期待値）expected value といい$E(\xi)$と記す。[1] 従って期望値は平均値の定義式(2)より、確率変数ξのとり得る可能なすべての値$x_1\ x_2 \dots x_k$に、各値をとる確率$p_1\ p_2 \dots p_k$を乗じて合計する事によって求められる。

$$E(\xi)=p_1x_1+p_2x_2+\dots+p_kx_k=\sum_{i=1}^{k}p_i x_i \qquad (2.13)$$

1）確率変数ξの確率分布の平均値を「期望値」というのは、それが平均して一回の無作為操作の結果実現すると期待し得る値であるからである。蓋し無作為操作の多数回（例えばn回）反覆の結果実現する値の総和は、その実現値の分布が大数法則の作用によりその確率変数ξの理論的な確率分布を以て近似し得るのであるから

$$(np_1)x_1+(np_2)x_2+\dots+(np_k)x_k$$

である。（但しnp_iはn個の実現値中のx_iの数）従って平均一回の無作為操作に於て得られるであろう値には

$$p_1x_1+p_2x_2+\dots+p_kx_k$$

を期待してよいであろうからである。

この確率変数の期望値計算は確率分布の状態の規定に於て重要な役割を演ずるのである。故に確率変数の期望値の計算に就て成立する諸関係を説明しよう。

先づ確率変数ξの自乗ξ^2も亦確率変数であって、そ

の期望値はξのとり得るすべての値$x_i\ (i=1,2,\dots,k)$の自乗と、各値をとる確率$p_i\ (i=1,2,\dots,k)$の積の総和であるから

$$E(\xi^2)=\sum_{i=1}^{k}p_i x_i^2 \qquad (2.14.)$$

一般に確率変数ξの函数$f(\xi)$の期望値は

$$E\{f(\xi)\}=\sum_{i=1}^{k}p_i f(x_i) \qquad (2.15)$$

で定義される。

次に二の確率変数ξ、ηの和ξ＋ηも亦確率変数であって、[1] その期望値はξの期望値とηの期望値との和に等しい。即ち

$$E(\xi+\eta)=E(\xi)+E(\eta) \qquad (2.16)$$

これを証明しよう。確率変数ξは値$x_1\ x_2\dots x_k$を確率$p_1\ p_2\dots p_k$を以てとり、ηは値$y_1\ y_2\dots y_l$を確率$q_1\ q_2\dots q_l$を以てとるとする。するとξ＋ηはx_i+y_j, $\begin{pmatrix} i=1,2,\dots,k \\ j=1,2,\dots,l \end{pmatrix}$のの$kl$個の値をとり得る事となる。新確率変数ξ＋ηが値$x_i+y_j$をとる確率を$P(x_i, y_j)$で表すと、ξ＋ηの期望値は期望値の定義により

$$E(\xi+\eta)=\sum_{i=1}^{k}\sum_{j=1}^{l}P(x_i, y_j)\cdot(x_i+y_j)$$
$$=\sum_{i=1}^{k}\sum_{j=1}^{l}P(x_i, y_j)\cdot x_i+\sum_{i=1}^{k}\sum_{j=1}^{l}P(x_i, y_j)\cdot y_j$$
$$=\sum_{i=1}^{k}x_i\underbrace{\Big\{\sum_{j=1}^{l}P(x_i, y_j)\Big\}}_{=p_i}+\sum_{j=1}^{l}y_j\underbrace{\Big\{\sum_{i=1}^{k}P(x_i, y_j)\Big\}}_{=q_j}$$

けだし　$\sum_{j=1}^{l}P(x_i, y_j)=P(x_i, y_1)+P(x_i, y_2)+\dots+P(x_i, y_l)$であり、ξが或る値$x_i$をとる時ηが$y_1$から$y_l$迄可能なすべての値をとる確率の合計であるから、結局ηのとる値に無関係に、ξがx_iをとる確率p_iに等しいからである。全く同様にして$\sum_{i=1}^{k}P(x_i, y_j)=q_j$がいえる。故に

$$E(\xi+\eta)=\sum_{i=1}^{k}p_i x_i+\sum_{j=1}^{l}q_j y_j=E(\xi)+E(\eta)$$

〔証明終り〕

1) 例えば二個の骰子を投げて出る目の和を考えると、それは2―12の各値をそれぞれ特定の確率を以てとる変数であるから、これ亦確率変数である。

この関係はn個の確率変数$\xi_1\ \xi_2\dots\xi_n$の和に拡張する事が出来るのであって、その場合は次の式が成立つのである。

$$E(\xi_1+\xi_2+\dots+\xi_n)=E(\xi_1)+E(\xi_2)+\dots+E(\xi_n) \qquad (2.17)$$

これの証明も(2.16)式の証明と同様にして行う事が出来る

ので省略する。

次に二の確率変数ξ、ηの積$\xi\eta$も亦確率変数であるが、その期望値$E(\xi\eta)$は必ずしもξの期望値$E(\xi)$とηの期望値$E(\eta)$の積$E(\xi)E(\eta)$に等しくはないのであって、ただξとηとが相互に独立（無関係[1]）なる場合にのみ

$$E(\xi\eta)=E(\xi)E(\eta) \tag{2.18}$$

となるのである。これは次の様にして証明される。今ξは値$x_1\ x_2 \dots x_k$を確率$p_1\ p_2 \dots p_k$を以てとり、ηは値$y_1\ y_2 \dots y_l$を確率$q_1\ q_2 \dots q_l$を以てとるとする。すると$\xi\eta$は$x_i y_j$, $\begin{pmatrix} i=1,2,\dots,k \\ j=1,2,\dots,l \end{pmatrix}$の$kl$個の値をとり、$\xi\eta$が$x_i y_j$なる値をとる確率は、$\xi$と$\eta$とが相互に独立であるから、$p_i q_j$である。けだし$\xi$と$\eta$とが相互独立、即ち$\xi$が$x_i$をとるのと$\eta$が$y_j$をとるのとは無関係であるため、$\xi$が$x_i$をとる時に$\eta$が$y_j$をとる確率は$\xi$のとる値の如何に拘らず$q_j$であり、或は又$\eta$が$y_j$をとる時に$\xi$が$x_i$をとる確率は$\eta$がとる値に関係なく$p_i$であるから、$\xi$が$x_i$を、$\eta$が$y_j$を同時にとる確率は、確率の乗法定理[2]により$p_i q_j$であるからである。故に$\xi\eta$の期望値は

$$E(\xi\eta)=\sum_{i=1}^{k}\sum_{j=1}^{l} p_i q_j x_i y_j=\sum_{i=1}^{k} p_i x_i \Big(\sum_{j=1}^{l} q_j y_j\Big)^{3)}=E(\xi)E(\eta)$$

〔証明終り〕

1) 例えば1から10迄の数字を記入した十個の球が入っている壺の中から、引続いて二個の球を抽出する時は、最初の球の数字が例えば3であるならば、二回目の球の数字は決して3ではあり得ず3以外の数字であるから、一回目の数字と二回目の数字とは相互に無関係（独立）ではないのである。しかるに最初に抽出した球を元へ戻してから二回目を抽出するならば、二回目の数字は一回目の数字と同様1—10の何れかの値であるから、両者は相互に無関係（独立）である。

2) 確率の乗法定理は次の如くである。「事象E_1の現れる確率がp_1であり、事象E_1が現れた時に事象E_2が現れる確率がp_2であるとすれば、事象E_1、E_2が同時に現れる確率は$p_1 \times p_2$である。」けだしE_2はE_1が現れた後でなければ現れないのであるから、E_1の現れる可能性p_1の上にE_2の現れる可能性p_2が被さって始めて、E_1、E_2の併起する可能性となるからである。

3) p_i、x_iはjの計算に於ては常数となり、jに就ての和$\sum_{j=1}$の外に約し出せるからである。

一般にn個の確率変数$\xi_1\ \xi_2 \dots \xi_n$が相互に独立ならば

$$E(\xi_1\xi_2\cdots\xi_n)=E(\xi_1)E(\xi_2)\cdots E(\xi_n) \tag{2.19}$$

が成立する。何となれば$\xi_1 \cdots \xi_n$が相互に独立ならばξ_1と$\xi_2 \cdots \xi_n$とも独立であるから、(2.18)式より$E(\xi_1 \cdots \xi_n)=E(\xi_1)E(\xi_2 \cdots \xi_n)$であり、これを繰返す事により$E(\xi_1 \cdots \xi_n)=E(\xi_1)\cdots E(\xi_n)$が得られるからであ。

又確率変数ξに常数aを乗じた値$a\xi$も確率変数であって、その期望値はξの期望値のa倍である。

$$E(a\xi)=aE(\xi), \qquad a=\text{const.} \quad (2.20)$$

なんとなれば

$$E(a\xi)=p_1ax_1+p_2ax_2+\cdots+p_kax_k$$
$$=a(p_1x_1+p_2x_2+\cdots+p_kx_k)=aE(\xi)$$

〔証明終り〕

常数aの期望値はaに等しい。即ち

$$E(a)=a \qquad (2.21)$$

なんとなれば(2.13)式に於て$\xi=a$従って$x_i=a$,（$i=1,2,\cdots,k$）と置くと

$$E(a)=p_1a+p_2a+\cdots+p_ka=a\underbrace{(p_1+p_2+\cdots+p_k)}_{=1}=a$$

〔$\because$ 確率は一種の相対的度数であるから（前講P.112参照）〕　〔証明終り〕

さて次に確率分布の分散並にその計算に於て成立する諸関係を説明しよう。確率変数ξの確率分布の分散を$V(\xi)$、標準偏差をσ_ξで表すならば、標準偏差の定義式(5′)より

$$V(\xi)=\sigma_\xi^2=\sum_{i=1}^{k}p_i\{x_i-E(\xi)\}^2$$

これはξの函数$\{\xi-E(\xi)\}^2$の期望値であるため、(2.15)式より

$$V(\xi)=\sigma_\xi^2=E\{\xi-E(\xi)\}^2 \qquad (2.22)$$

と書ける。(2.22)式は次のように変形する事が出来る。

$$V(\xi)=E\{\xi-E(\xi)\}^2=E\{\xi^2-2\xi E(\xi)+(E\xi)^2\}$$
$$=E(\xi^2)-2E(\xi)E(\xi)+(E\xi)^2 \quad \left[\because E(\xi)\text{は常数であるため(2.20)(2.21)により}\right]$$
$$=E(\xi^2)-(E\xi)^2 \qquad (2.23)$$

即ち$V(\xi)$はξの自乗の期望値よりξの期望値の自乗を引いたものに等しいのである。（これは(6″)式に相当する）

又(2.23)式より

$$E(\xi^2)=V(\xi)+(E\xi)^2 \qquad (2.24)$$

故にξの自乗の期望値は、ξの分散とその期望値の自乗との和に等しいのである。この関係はよく用いられる。

若し確率変数ξに常数aが掛っているならば

$$V(a\xi)=a^2V(\xi) \qquad (2.25)$$

である。なんとなれば

$$V(a\xi)=E\{a\xi-E(a\xi)\}^2=E\{a\xi-aE(\xi)\}^2$$
$$=a^2E\{\xi-E(\xi)\}^2=a^2V(\xi)$$

〔証明終り〕

ξとηとが相互に独立な確率変数ならば

$$V(\xi+\eta)=V(\xi)+V(\eta) \qquad (2.26)$$

これを証明すると

$$V(\xi+\eta)=E\{(\xi+\eta)-E(\xi+\eta)\}^2$$
$$=E\{(\xi-E\xi)+(\eta-E\eta)\}^2$$
$$=E\{(\xi-E\xi)^2+(\eta-E\eta)^2+2(\xi-E\xi)(\eta-E\eta)\}$$
$$=E(\xi-E\xi)^2+E(\eta-E\eta)^2+2E\{(\xi-E\xi)(\eta-E\eta)\}$$

しかるにξとηとは相互独立なるため、(2.18)式より

$$E\{(\xi-E\xi)(\eta-E\eta)\}=E(\xi-E\xi)E(\eta-E\eta)=0$$

〔∵ $E(\xi-E\xi)=E\xi-E\xi=0$, 同様にして $E(\eta-E\eta)=0$〕

$$\therefore\ V(\xi+\eta)=E(\xi-E\xi)^2+E(\eta-E\eta)^2=V(\xi)+V(\eta)$$

〔証明終り〕

一般にn個の確率変数$\xi_1\xi_2\cdots\xi_n$が相互に独立ならば

$$V(\xi_1+\xi_2+\cdots+\xi_n)=V(\xi_1)+V(\xi_2)+\cdots+V(\xi_n) \qquad (2.27)$$

これは(2.26)式の証明と同様にして証明されるので省略する。

尚確率変数ξの確率分布の標準偏差σ_ξをその期望値$E(\xi)$で除して相対化した値を、ξの変異係数といいC_ξで表す

$$C_\xi=\frac{\sigma_\xi}{E(\xi)} \qquad (2.28)$$

二の確率変数ξ、ηの相関確率分布に於ては、ξ、ηの間の共変量$\mathrm{Cov}(\xi,\eta)$及び相関係数$\rho_{\xi,\eta}$が求められる。それに就ては後に必要な箇処に於て述べる事にする。

註 (2・16)(2・18)(2・21)式の証明の過程に於て確率に就ての計算を直観的に説明した箇所があるが、それの正確な証明は確率書を参照されたい。

任意標本理論に於ては確率分布たる標本分布の状態を調べるのに、その期望値、分散等を求めねばならないのであるが、以上に述べた期望値計算の諸公式によ

ると、それは極めて容易に計算する事が出来る。前項に於て既に計算した標本平均分布の期望値及び分散を、今一度この方法によって求めてみよう。

その前に母集団πより一個の単位を抽出する時得られる値x'の標本分布の期望値と分散を計算して置くと便利である。πより一個の単位を抽出する時得られる値は$x_1\ x_2 \dots x_N$の何れかであり、且任意抽出であるためどれもそれの抽かれる可能性（確率）は等しく$\frac{1}{N}$である。抽出された単位の変量x'がx_jなる値をとる確率を$P_r\{x'=x_j\}$と書くならば

$$P_r\{x'=x_1\}=P_r\{x'=x_2\}=\dots=P_r\{x'=x_N\}=\frac{1}{N}$$

従って

$$E(x')=\sum_{j=1}^{N}x_jP_r\{x'=x_j\}=\frac{1}{N}\sum_{j=1}^{N}x_j=\bar{x} \qquad (2.29)$$

$$V(x')=E(x'-\bar{x})^2=\sum_{j=1}^{N}(x_j-\bar{x})^2P_r\{x'=x_j\}$$
$$=\frac{1}{N}\sum_{j=1}^{N}(x_j-\bar{x})^2=\sigma^2 \qquad (2.30)$$

故にx'の確率分布の期望値は母平均に等しく、分散は母分散に等しいのである。即ちx'の標本分布は母集団分布に一致するのである。

さて標本平均分布の期望値は

$$E(\bar{x}')=E\left(\frac{1}{n}\sum_{j=1}^{n}x'_j\right)=\frac{1}{n}\sum_{j=1}^{n}E(x'_j)$$

〔(2.20)（2.17）により〕

茲に$E(x')$は母集団πより一個抽出する時得られる値の期望値である。故に(2.29)式より

$$E(\bar{x}')=\frac{1}{n}\sum_{j=1}^{n}\bar{x}=\bar{x} \qquad (2.31)$$

分散は

$$V(\bar{x}')=E(\bar{x}'-\bar{x})^2=E(\bar{x}'^2)-\bar{x}^2 \quad 〔(2.23)により〕$$

$$\text{第一項：}E(\bar{x}'^2)=E\left(\frac{1}{n}\sum_{j=1}^{n}x'_j\right)^2=\frac{1}{n^2}E\left(\sum_{j=1}^{n}x'^2_j+\sum_{j\neq j'}^{n}\sum^{n}x'_jx'_{j'}\right)$$
$$=\frac{1}{n^2}\left\{\sum_{j=1}^{n}E(x'^2_j)+\sum_{j\neq j'}^{n}\sum^{n}E(x'_jx'_{j'})\right\} \qquad (2.32)$$

右辺第一項の$E(x'^2_j)$は、母集団πより只一個抽出した時得られる値の自乗の標本分布の期望値である。故に

$$E(x'^2_j)=\sum_{j=1}^{N}x^2_jP_r\{x'=x_j\}=\frac{1}{N}\sum_{j=1}^{N}x^2_j$$

又右辺第二項の$E(x'_jx'_{j'})$は、πより二個引続いて抽出する時得られる値の組(x'_1, x'_2)の積$x'_1x'_2$の標本分布の期望値である。そしてπより二個引続いて抽出する事によって得られる値の組の理論的な数は、N個の

単位中より二個の単位を取出す「順列」の数として得られるため、${}_NP_2=N(N-1)$ 通りである。実際に二個の単位の任意抽出によって得られる値の組はこの ${}_NP_2$ 組の何れかであり、且その各々は抽出される可能性（確率）はすべて等しいから

$$P_r\{x'_1=x_j,\ x'_2=x_{j'}\}=\frac{1}{N(N-1)},\quad \left(\begin{matrix}j,j'=1,2\cdots,N\\ j\neq j'\end{matrix}\right)$$

である。従って

$$E(x'_1x'_2)=\sum_{j\neq j'}^{N}\sum^{N}x_jx_{j'}P_r\{x'_1=x_j,\ x'_2=x_{j'}\}$$

$$=\frac{1}{N(N-1)}\sum_{j\neq j'}^{N}\sum^{N}x_jx_{j'}$$

故に(2.32)式は

$$E(\bar{x}'^2)=\frac{1}{n^2}\left\{\sum_{j=1}^{n}\left(\frac{1}{N}\sum_{j=1}^{N}x_j^2\right)+\underbrace{\sum_{j\neq j'}^{n}\sum^{n}}_{1)}\left(\frac{1}{N(N-1)}\sum_{j\neq j'}^{N}\sum^{N}x_jx_{j'}\right)\right\}$$

$$=\frac{1}{n^2}\left\{\frac{n}{N}\sum_{j=1}^{N}x_j^2+\frac{n(n-1)}{N(N-1)}\sum_{j\neq j'}^{N}\sum^{N}x_jx_{j'}\right\}$$

$$=\frac{1}{nN}\sum_{j=1}^{N}x_j^2+\frac{n-1}{nN(N-1)}\sum_{j\neq j'}^{N}\sum^{N}x_jx_{j'}$$

これは(2.7)式に外ならない。従って前と同様にして

$$E(\bar{x}'^2)=\bar{x}^2+\frac{N-n}{n(N-1)}\sigma^2$$

故に

$$V(\bar{x}')=\frac{\sigma^2}{n}\,\frac{N-n}{N-1}\qquad(2.33)$$

1) $\sum_{j\neq j'}^{n}\sum^{n}1=n(n-1)$、けだし抽出された n 個の単位を二個宛（順序を附けて）組合せる場合の数は ${}_nP_2=n(n-1)$ 通りであるからである。

最後に確率分布たる標本分布の形は、母集団の分布の形並に標本として抽出される単位の数及び母数を推定するための統計量の函数形によって決るのであって、それには正規分布、二項分布、Poisson 分布、t 分布、χ^2 分布、F 分布、Z 分布等がある。しかし標本平均値及び標本比例数の分布は、標本の大いさ n が充分に大であるならば（即ち大標本の場合は）、母集団の分布形の如何に拘らず正規分布に近い分布をなす事が数学的に証明されるのである。[1] 社会統計調査に於ては抽出単位数は殆ど茲にいう大標本の場合であるから、本標本分布は正規分布をなすものとして取扱えばよいのである。かくて前項に考察した標本平均分布は正規分布をなしているという事が出来るのである。

1) 標本平均分布は、母集団の大いさ N が標本の大いさ n

に比して相当大きくその分布が極度に非対称分布でない限り、$n \geqq 50$の時は正規分布と看做し得る事が経験的に知られているのである。特に母集団分布が正規分布をなす時は、標本の大いさの如何に拘らず、標本平均分布は常に完全な正規分布をなすのである。

標本比例数分布は $np \geqq 5$であれば正規分布と見てよい事が数値計算によって判るのである。

そして確率変数ξの確率分布が正規形である時は、正規分布の特徴により

区間$\left\{\begin{matrix} E(\xi)\pm\ \sigma_\xi \\ E(\xi)\pm 2\sigma_\xi \\ E(\xi)\pm 3\sigma_\xi \end{matrix}\right\}$内にある$\xi$の実現値が有する確率は$\left\{\begin{matrix} 0.68268 \\ 0.95450 \\ 0.99730 \end{matrix}\right\}$

である。[1] 従って一回の無作為操作の結果ξが値x_0をとったとすれば、それが区間$E(\xi)\pm 2\sigma_\xi$外にある確率は5%弱に過ぎず、95%強の確率を以て$E(\xi)\pm 2\sigma_\xi$内にあるという事が出来るのである。

1) 仮令確率分布が正規形でなく非対称分布をなす場合であっても、チェビシェフの定理により、区内 $E(\xi)\pm\lambda\sigma_\xi$内にある$\xi$の実現値の有する確率は $\left(1-\frac{1}{\lambda^2}\right)$ より大である。

任意標本による時は誤差の少い母数の推定値が得られ尚且誤差の範囲を規定し得るのは、その標本推定値がこのような確率変数の実現値としての性格を有するからである。次に項を改めて任意標本による標本推定の方法を説明しよう。

(3) 標本推定の原理

以上により明らかな如く確率分布としての標本平均分布の期望値は母平均に等しく、その標本平均分布は現実に同じ大いさの任意標本を多数構成する時得られる標本平均値の分布を近似的に表すものであるため、多数回任意標本調査を行って得られた標本平均値を更に平均する事によって、母平均の値を略〻正確に知る事が出来るのである。

このように同じ方法で繰返し構成された標本より得られる標本推定値の平均が母数に等しい場合、それは不偏 unbiased であるといい、その推定値を不偏推定値 unbiased estimate という。そして標本推定値の平均が母数に一致しない時はその差を偏り bias といい、そのような推定値を偏りある推定値 biased estimate という。標本推定値が不偏推定値の場合は母数に等しい値を得る事が方法的に可能であるが、標本推定値が偏りある推定値の時はそれ

は不可能である。故に標本推定値は他の条件（推定の精度、計算の難易等）が同じであるならば、偏りあるものよりも不偏であるものの方がよい推定値である事はいう迄もないであろう。

このように標本平均は母平均の不偏推定値であるが、実際上は標本調査は只一回行われるのみであって多数回繰返して行われるものではないのであるから、只一の標本平均値を以て母平均の推定値としなければならず、従って標本推定値は当然に誤差を有するのである。しかしながら任意標本より求めた標本推定値は確率変数と看做し得るから、標本平均分布の期望値即ち母平均を中心に、その標準偏差の例えば二倍の範囲外の値である事は極めて稀であり、殆どその内部の値であると見てよいのである。従ってこの標本平均分布の標準偏差の±2倍の区間を以て推定値の誤差の範囲と規定する時は、殆どその内部に母平均を含んでいる事となるのである。

今母集団πよりn個の単位を任意に抽出して一の標本Oを構成し、それより標本平均値$\bar{x}'_0$を得たとする。

この$\bar{x}'_0$は期望値$\bar{x}$、分散$\frac{\sigma^2}{n}\frac{N-n}{N-1}$の略ゝ正規分布をなす標本平均分布の何れか一の値をとるのであるが、標本の構成即ちn個の単位の抽出が全く偶然的になされたものであるならば、その値が

$$\text{区間}\begin{Bmatrix}\bar{x}\pm\sigma_{\bar{x}},\\ \bar{x}\pm 2\sigma_{\bar{x}},\\ \bar{x}\pm 3\sigma_{\bar{x}},\end{Bmatrix}\text{内に在る確率は}\begin{Bmatrix}68.27\%\\ 95.45\%\\ 99.73\%\end{Bmatrix}$$

であるという事が出来るのである。従って逆にその標本平均値$\bar{x}'_0$に$\pm\sigma_{\bar{x}}$、$\pm 2\sigma_{\bar{x}}$、$\pm 3\sigma_{\bar{x}}$の誤差の幅を附するならば、

$$\text{区間}\begin{Bmatrix}\bar{x}'_0\pm\sigma_{\bar{x}},\\ \bar{x}'_0\pm 2\sigma_{\bar{x}},\\ \bar{x}'_0\pm 3\sigma_{\bar{x}},\end{Bmatrix}\text{内に母平均}\bar{x}\text{を含む確率は}\begin{Bmatrix}68.27\%\\ 95.45\%\\ 99.73\%\end{Bmatrix}$$

であるといえるであろう。

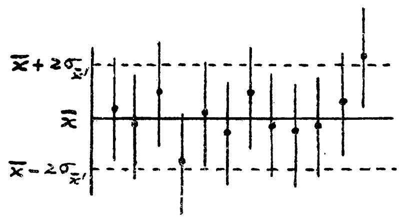
従って若し大いさnの任意標本を実際に100回構成し100個の標本平均値を得るならば、略ゝそのうち95個は区間（$\bar{x}-2\sigma_{\bar{x}}$, $\bar{x}+2\sigma_{\bar{x}}$）内の値であり、5個がその外部の値であるに過ぎないであろう。そして各標本より得られた標本平均値に$\pm 2\sigma_{\bar{x}}$の幅を見込むならば、その区間の中に母平均$\bar{x}$を含む場合が95回あり、$\bar{x}$がその区間外にある場合が5回あるであろう。（第一図参照）

かくて一の任意標本より求めた標本平均値に普通 $\pm 2\sigma_{\bar{x}'}$ の誤差を附記して、信頼度95%の母平均の推定値となすのである。例えば世帯数26万の或る都市に於て、一世帯当り平均家計費を調査するために130世帯を抽出して1/2,000の標本を構成し、家計調査の結果平均家計費18,500円を得たとする。母集団に於ける家計費分布の標準偏差の推定値を10,300円とすると、[1] $\sigma_{\bar{x}'}$ は次のようになる。この場合は N が n に比して極めて大であるため有限母集団修正は略ゝ一に等しい。故に

$$\sigma_{\bar{x}'} \fallingdotseq \frac{10{,}300}{\sqrt{130}} = 903.37\text{円}$$

従ってこの都市の平均家計費は18,500円 $\pm 2 \times 903.37$ 円即ち16,693円～20,307円の間にあるという事が出来、それの真なる確率は95%である。

第一図　信頼区間



このような標本推定値が一定の確率に於て出現する区間を信頼区間 confidence interval といい、その限界を信頼限界 confidence limit その区間内に推定値がある確率を信頼係数 confidence coefficient という。

以上のようにして任意抽出法による時は、一定の信頼度に於て信頼区間（誤差の範囲）を明示して、母数の推定をなす事が出来るのである。

1) この値は(2.47)式によって推定される。

(4) 母集団の量的大いさの推定

母集団の量的大いさ即ち母集団を構成する単位の変量の総和は、平均値に母集団の単位総数を乗ずる事によって求める事が出来る。けだし平均値はすべての単位の変量の総和を単位総数で除した値であるからである。（(1)式参照）従って母平均の推定値に母集団の単位総数 N を乗ずる事によって、母集団の量的大いさ $X=\sum_{i=1}^{N} x_i$ を推定する事が出来る。X の推定値を X' で表すと

$$X' = N\bar{x}' = \frac{N}{n}\sum_{j=1}^{n} x'_j \qquad (2.34)$$

この推定値は不偏である。

$$E(X') = NE(\bar{x}') = N\bar{x} \qquad \text{〔(2.31)より〕}$$
$$= X$$

それの分散及び変異係数は次の如くである。

$$V(X') = N^2V(\bar{x}') = N^2\frac{N-n}{N-1}\frac{\sigma^2}{n} \qquad (2.35)$$

$$C_{X'} = N\sqrt{\frac{N-n}{N-1}}\frac{\sigma}{\sqrt{n}}\frac{1}{X} = N\sqrt{\frac{N-n}{N-1}}\frac{\sigma}{\sqrt{n}}\frac{1}{N\bar{x}} \qquad \text{〔(2.33)より〕}$$
$$= \sqrt{\frac{N-n}{N-1}}\frac{c}{\sqrt{n}} \qquad (2.36)$$

(2.36)式は(2.11)式と同じである。(2.36)式を導く過程より明らかな如く、一般に$C_{X'} = C_{\bar{x}'}$である。

(5)　標本比例数分布

以上は統計集団の量的構造の簡約表章値である平均値の標本推定であった。次に統計集団の質的構造の簡約表章値である比例数の標本推定を考察しよう。標本推定の方式は第三項に考察した平均値の推定の場合と全く同じであるため、不偏な標本推定値を求めそれの標本分布の分散式を導けばよいのである。

推定すべき比例数は母集団に於ける問題たる標識αを有する単位の存在の割合であるから、それを標本より推定せんとする時は、標本に於ける標識αを有する単位の割合即ち標本比例数を以てすべきであろう。従って標本比例数の分布を調べねばならないのであるが、その場合前講「母集団分布(3)」で述べたように統計集団の質的構造に就ても、その存在割合を知らんとする標識に1、その他の標識に0なる値を与る事によって量的構造化し得るのであるから、この変換によって標本比例数分布は前述の標本平均分布の特別の場合として容易に求める事が出来るのである。

今母集団πの単位総数NのうちN_1が標識αを有し、残りN_2がα以外の標識を持って居りそれぞれの割合を$p\left(=\frac{N_1}{N}\right)$, $q\left(=\frac{N_2}{N}\right)$とする。この母集団より$n$個の単位を任意に抽出して構成した標本Oのうち$n_1$個が標識$\alpha$を、残り$n_2$個が$\alpha$以外の標識を有し、それぞれの割合を$p'\left(=\frac{n_1}{n}\right)$, $q'\left(=\frac{n_2}{n}\right)$とする。ここで$\alpha$標識を有する単位に変量$x = 1$　α以外の標識を有する単位

に変量$x=0$を与えると、母集団πは0、1のみの変量を有する単位の集合となり、その平均値はα標識の割合pに等しく、分散はpとαの以外の標識の割合qとの積pqである[1)]。

$$\bar{x}=p,\quad \sigma^2=pq$$

標本Оも変量0、1のみを有する単位の集合となるため、その平均値及び分散は同様に

$$\bar{x}'=p',\quad s^2=p'q'$$

となる。

1）これは次のようにして証明される。

$$\bar{x}=\frac{1}{N}\sum_{j=1}^{N}x_j=\frac{1}{N}\left\{\sum_{j=1}^{N_1}1+\sum_{j=1}^{N_2}0\right\}=\frac{N_1}{N}=p$$

$$\sigma^2=\frac{1}{N}\sum_{j=1}^{N}(x_j-\bar{x})^2=\frac{1}{N}\left\{\sum_{j=1}^{N_1}(1-p)^2+\sum_{j=1}^{N_2}(0-p)^2\right\}$$

$$=\frac{1}{N}(N_1q^2+N_2p^2)=pq^2+qp^2=pq(q+p)=pq$$

〔証明終り〕

従って標本比例数分布は標本平均分布に於て、$\bar{x}$、σ^2、$\bar{x}'$の代りにp、pq、p'を代入すればよい。先づ標本比例数分布の期望値は(2.31)式より

$$E(p')=p \qquad (2.37)$$

であり、故に標本比例数は母比例数の不偏推定値である。次に分散、標準偏差は(2.33)式より

$$V(p')=\frac{pq}{n}\frac{N-n}{N-1} \qquad (2.38)$$

$$\sigma_{p'}=\sqrt{\frac{pq}{n}}\sqrt{\frac{N-n}{N-1}} \qquad (2.39)$$

変異係数は、母変異係数が$c=\dfrac{\sigma}{\bar{x}}=\dfrac{\sqrt{pq}}{p}=\sqrt{\dfrac{q}{p}}$なるため、(2.11)式より

$$C_{p'}=\sqrt{\frac{q}{np}}\sqrt{\frac{N-n}{N-1}} \qquad (2.40)$$

である。

母比例数の推定値p'に母集団単位総数Nを乗ずる事によって、母集団のα標識の単位数N_1を推定する事が出来る。それは母集団に於ける標識αの比例数は、αを有する単位の数を総単位数で除した値であるからである。N_1の推定値をN_1'とすると

$$N_1'=Np' \qquad (2.41)$$

これは不偏推定値である。

$$E(N_1')=NE(p')=Np=N_1$$

その分散及び変異係数は

$$V(N_1')=N^2V(p')=N^2\frac{N-n}{N-1}\frac{pq}{n} \qquad (2.42)$$

$$C_{N_1'}=N\sqrt{\frac{N-n}{N-1}}\sqrt{\frac{pq}{n}}\frac{1}{N_1}=\sqrt{\frac{N-n}{N-1}}\sqrt{\frac{q}{np}}$$

〔$\because N_1=Np$ なるため〕 (2.43)

(2.43)式は(2.40)式と同じであるため $C_{p'}=C_{N_1'}$ である。

例えば某市の市長選挙に際して投票日三日前に60万の有権者より600人を抽出して、A、B、C何れの候補者に投票するかを調査し上の結果を得たとする。A候補の推定得票数 N_1' は

A候補	26.7%($=p_1'$)
B候補	17.9%($=p_2'$)
C候補	13.6%
未定	41.8%

$$N_1'=600{,}000\times0.267=160{,}200$$

その誤差の範囲は、母標準偏差の推定値を0.442とすると[1)]

$$\sigma_{p_1'}=\sqrt{\frac{pq}{n}}\sqrt{\frac{N-n}{N-1}}\doteqdot\sqrt{\frac{pq}{n}}=\frac{0.442}{\sqrt{600}}=0.0180$$

〔$\because N$ が n に比して極めて大なるため〕

$$\therefore\ N_1'\pm2\sigma_{N_1'}=N_1'\pm2N\sigma_{p_1'}$$

$$=160{,}200\pm2\times600{,}000\times0.018=138{,}600\sim181{,}800$$

次点B候補の推定得票数 N_2' の誤差の限界は、母標準偏差の推定値を0.383とすると[2)]

$$N_2'=600{,}000\times0.179=107{,}400$$

$$\sigma_{p_2'}\doteqdot\frac{0.383}{\sqrt{600}}=0.0156$$

$$\therefore\ N_2'\pm2\sigma_{N_2'}=N_2'\pm2N\sigma_{p_2'}$$

$$=107{,}400\pm2\times600{,}000\times0.0156=88{,}680\sim126{,}120$$

従ってB候補の最高得票数はA候補の最低得票数に及ばないため決戦投票は行われず、結局A候補が13.9万票——18.1万票を得て当選するものと、真なる確率95%で推定されるのである。

1) この値は(2,51)式によって推定される。但しこの場合は N、n が相当に大であるので $\frac{N-1}{N}\doteqdot1$, $\frac{n}{n-1}\doteqdot1$ 従って

$$\frac{N-1}{N}\cdot\frac{n}{n-1}\sqrt{p_1'q_1'}\doteqdot\sqrt{p_1'q_1'},\quad q_1'=1-p_1'$$

$$=\sqrt{0.267\times0.733}=0.442$$

2) $\sqrt{p_2'q_2'}=\sqrt{0.179\times0.821}=0.383$, $q_2'=1-p_2'$

(6) 標本推定の精度

標本推定値の母数よりの偏差は標本誤差であるが、現実に標本調査の結果如何なる誤差が現れるかは偶然的に決るのであり、且推定値が不偏である場合は多数回の調査の繰返しによって相殺され得る誤差であるから、それは偶然誤差という事が出来る。しかしながら推定値が偏りを有するものなる場合は、標本誤差はこ

の偶然誤差のみならず、偏り（方法上、どうしても消す事の出来ない誤差）をも含んでいると見なければならぬ。

不偏推定値の標本分布の標準偏差（及び分散、変係数）は、このような標本推定値の「誤差の分布」の範囲を表す値と見る事が出来るのであって、故にそれを標準誤差 standard error（及び誤差分散、誤差変異係数）という。標準誤差が小さい時は標本推定値が期望値の附近に密に分布しているのであるから、一の任意標本より得られた推定値を以て母数の推定値となすもその誤差は余り大ではないであろう。しかし標準誤差が大きい時は標本推定値の分布が広いため、一の任意標本より得た推定値は母数より相当隔り誤差が大である可能性が大きいのである。故に標準誤差の値の大小は標本推定の精度を表すのである。若し一の母数に対して数種の標本推定値が考えられる時は、それ等が共に不偏推定値であるならば、標準誤差の小さい方が大きい方よりもよい推定値であろう。[1] 又仮令或る推定値が偏りある推定値であっても、その偏りが極めて小であって無視し得る程度であるならば、標準誤差の大きい不偏推定値よりも、標準誤差の小さい偏りある推定値の方がよりよい推定値といえるであろう。[2]

1）　或る母数 θ に対する数種の不偏推定値のうち誤差分散が最小なるものを最良不偏推定値 best unbiased estimate という。標本平均値は母平均の最良不偏推定値である。それは次のようにして証明される。

母平均 $\bar{x}$ の標本推定値を $T_n=\sum_{j=1}^{n}a_jx'_j$ とし、係数 a_j（$j=1,2,\cdots\cdots,n$）をこの推定値が不偏であり且誤差分散が最小なるように決める事によって、$\bar{x}$ の最良不偏推定値が得られる。先づ T_n が不偏であるためには次の等式が成立しなければならない。

$$E(T_n)=\bar{x} \qquad \text{(i)}$$

しかるに

$$E(T_n)=E\Big(\sum_{j=1}^{n}a_jx'_j\Big)=\sum_{j=1}^{n}a_jE(x'_j)$$

$$=\bar{x}\sum_{j=1}^{n}a_j \qquad \text{〔(2.29)により〕}$$

故にこれを(i)式と比較して

$$\sum_{j=1}^{n}a_j=1 \qquad \text{(ii)}$$

でなければならない事が判る。次に T_n の分散は

$$V(T_n)=E\{T_n-E(T_n)\}^2$$

$$=E\Big\{\sum_{j=1}^{n}a_jx'_j-\bar{x}\sum_{j=1}^{n}a_j\Big\}^2=E\Big\{\sum_{j=1}^{n}a_j(x'_j-\bar{x})\Big\}^2$$

$$=\sum_{j=1}^{n} a_j^2 E(x'_j-\bar{x})^2 + \sum_{j\neq j'}^{n}\sum^{n} a_j a_{j'} E\{(x'_j-\bar{x})(x'_{j'}-\bar{x})\}$$

しかるに　$E(x'_j-\bar{x})^2=\sigma^2$　　〔(2.30)により〕

$$E\{(x'_j-\bar{x})(x'_{j'}-\bar{x})\}$$
$$=\sum_{j\neq j'}^{N}\sum^{N}(x_j-\bar{x})(x_{j'}-\bar{x})\ P_r\{x'_j=x_j,\ x'_{j'}=x_{j'}\}$$
$$=\frac{1}{N(N-1)}\sum_{j\neq j'}^{N}\sum^{N}(x_j-\bar{x})(x_{j'}-\bar{x})\quad \left[\begin{array}{l}(2.32)\text{の}\\ \text{展開参照}\end{array}\right]$$
$$=-\frac{\sigma^2}{N-1}$$

∵ 母集団を集落に分けない場合の級内相関係数

$$\rho'=\frac{1}{\sigma^2}\left\{\frac{1}{N(N-1)}\sum_{j\neq j'}^{N}\sum^{N}(x_j-\bar{x})(x_{j'}-\bar{x})\right\}=\frac{-1}{N-1}$$

（前講 P. 127参照）より

$$\frac{1}{N(N-1)}\sum_{j\neq j'}^{N}\sum^{N}(x_j-\bar{x})(x_{j'}-\bar{x})=-\frac{\sigma^2}{N-1}$$

故に　$V(T_n)=\sigma^2\sum_{j=1}^{n}a^2-\frac{\sigma^2}{N-1}\sum_{j\neq j'}^{n}\sum^{n}a_j a_{j'}$

$$=\sigma^2\left\{\sum_{j=1}^{n}a^2_j-\frac{1}{N-1}\sum_{j\neq j'}^{n}\sum^{n}a_j a_j\right\}$$
$$=\frac{\sigma^2}{N-1}\left\{N\sum_{j=1}^{n}a^2_j-\left(\sum_{j=1}^{n}a^2_j+\sum_{j\neq j'}^{n}\sum^{n}a_j a_{j'}\right)\right\}$$
$$=\frac{\sigma^2}{N-1}\left\{N\sum_{j=1}^{n}a^2_j-\left(\sum_{j=1}^{n}a_j\right)^2\right\}$$
$$\left[\because\ \left(\sum_{j=1}^{n}a_j\right)^2=\sum_{j=1}^{n}a^2_j+\sum_{j\neq j'}^{n}\sum^{n}a_j a_{j'}\right]$$
$$=\frac{\sigma^2}{N-1}\left\{N\sum_{j=1}^{n}a^2_j-1\right\}\qquad \text{〔(ii)により〕}$$

従って(ii)の条件の下にT_nの分散を最小ならしめるには$\sum_{j=1}^{n}a^2_j$を最小ならしめればよい。そのようなa_jはラグランジュ乗数の法により

$$\frac{\partial}{\partial a_j}\left(\sum_{j=1}^{n}a^2_j+\lambda\sum_{j=1}^{n}a_j\right)=0\quad \begin{pmatrix}\lambda\text{は未定常数}\\ j=1,2,\ \cdots,\ n\end{pmatrix}\qquad \text{(iii)}$$

をa_jに関して解く事によって得られる。(iii)より

$$2a_j+\lambda=0\ \text{即ち}\ a_j=-\frac{\lambda}{2}\quad (j=1,2,\cdots\cdots,n)\qquad \text{(iv)}$$

これを(ii)に代入して

$$\sum_{j=1}^{n}a_j=-\frac{\lambda}{2}n=1\qquad \therefore\ \lambda=-\frac{2}{n}$$

これを(iv)に代入して

$$a_j=-\frac{1}{2}\left(-\frac{2}{n}\right)=\frac{1}{n}\qquad (j=1,2,\cdots\cdots,n)$$

故に

$$T_n=\sum_{j=1}^{n}a_j x'_j=\frac{1}{n}\sum_{j=1}^{n}x'_j=\bar{x}'$$

即ち標本平均値は母平均の最良不偏推定値である。

2)　この例に「比推定」がある。

今標準誤差を規定する三つの要素σ（比例数の推定の場合はp）n及びNの標準誤差に及ぼす影響を見ると次の如くである。（(2.10)(2.39)式参照）

(1)　標準誤差は母標準偏差σに正比例する。σが大即ち母集団を構成する単位が異質的であり、従って個々の単位の有する変量に大きな差異がある場合は、σが小即ち母集団が比較的同質的な単位より成って居り、個々の単位の変量が余り大きく異らない時に比して、

同じ大いさの標本による推定値は誤差が大きくなりその精度が落ちるのである。

(2) 比例数の推定の場合は標準誤差は、母比例数 p が $p<0.5$ の範囲に於ては p の増加と共に増大し、$p=0.5$ に於て極大に達し、$p>0.5$ の範囲に於ては p の増加と共に減少する。従って標準誤差を以て精度の表現となす時は、$p=0.5$ の場合が最も標本推定の精度が低く、p がそれより小又は大なるに従って精度が高まるかの如くであるが、実際は p が増大するに従って標準誤差は相対的に減小して行く事になるため、標本推定の精度は p に反比例的であるという事が出来る。この事は誤差変異係数を以て標本推定の精度を表す時に明らかとなるのである。即ち(2.40)式より明らかな如く誤差変異係数は $\sqrt{\frac{q}{p}}$（母変異係数）に正比例し、$\sqrt{\frac{q}{p}}$ は p が増加する時減少するからである。（但しこの場合 p が変動すれば q も亦変動するから（$\because\ q=1-p$）p の平方根に正確に反比例しない。）故に比例数が大即ち問題たる標識を有する単位が多い場合は、比例数が小即ち問題たる標識の単位が少い場合に比して、同じ大いさの標本による推定値は誤差が小でありその精度が高まるのである。

(3) 標準誤差は標本の大いさ n の平方根に反比例する。故により多くの単位を抽出して構成した標本による推定値の方が、少い単位の標本による推定値よりも誤差が少く精度が高いのである。標準誤差を規定する他の要素 σ（又は p）及び N は、客観的に定まっている大いさであって如何ともし難い要素であるが、この標本単位数 n は任意に決定し得る値であるため、これを増減する事によって標準誤差の大いさ従って標本推定の精度を必要な程度に管理する事が出来るのである。しかしながら標準誤差は標本の大いさの平方根に反比例するのであって、n が 4 9 16 … 100 倍となるに随って標準誤差は $1/2$ $1/3$ $1/4$ … $1/10$ となるに過ぎないため、少し標本の精度を高めるためには相当多くの標本単位数を増加しなければならないのである。そして標本単位数の増加は調査に要する費用を増大せしめるため、必要以

上の精度の標本推定を可能ならしめるような大いさの標本を構成する事は不経済であり、必要な限度に迄切下げるのが合理的であろう。かくて任意標本調査法は標本単位数の増減によって、所期の精度の標本推定を最小の費用で獲得する事を可能ならしめるのである。

(4) 母集団の大いさNは正比例的に影響するがその力は小であるため、Nは標本の精度には余り関係しないのである。けだし(2.10)式に於てNが相当に大ならば$N-1 \fallingdotseq N$であるため $\sigma_{\bar{x}} \fallingdotseq \frac{\sigma}{\sqrt{n}}\sqrt{1-\frac{n}{N}}$ となり、nを一定とすると有限母集団修正はNの増加と共に1に近づくがそのテムポは極めて緩慢であるからである。従って同一精度の標本推定を得るためには、大きな母集団からも小さい母集団からも、大体同じ丈の単位数を抽出しなければならないのであって、若し標本の大いさを母集団の大いさに比例させるならば、小母集団の標本推定は精度が著しく低下する事となるのである。この事から任意標本調査法は母集団が大きい程、小さい抽出率でよい精度の推定値が得られるから、効率の高い方法である事が判るのである。

以上により明らかな如く標本推定の精度は、三つの要素のうち標本単位数nによって最も著しく影響されるのである。

普通標本推定の精度は標準誤差で表現されないで誤差変異係数を以て表される。例えば「或る都市の平均世帯人員の推定の精度（変異係数）は1%である」といい、「平均世帯人員の推定の精度（標準誤差）は0.04人である」とはいわないのである。それは標準偏差は平均値を中心に測った偏差の平均であるから、その大いさは平均値との関連に於て考えて始めてよくその変量の分散度を表し得る事から、標本推定の精度の表現としても変異係数による方がより適切であるからである。

誤差変異係数は母変異係数に正比例し、n、Nよりの影響は標準誤差の場合と全く同じである。

(7) 標本の大いさの決定

標本推定の精度は標準誤差（又は誤差変異係数）の大いさによって表されるため、標準誤差（又は誤差変異係数）を求める式を用いて、調査目的よりして要求される精度（これを目標精度 aimed at precision とい

う）の標本推定を実現するに必要な標本構成単位の数nを求める事が出来る。即ち母集団の大いさNが判って居り、母標準偏差の値を何等かの方法で推定し得るならば、[1] 調査目的より必要とされる標準誤差の値が与えられる時は、(2.10)式にそれ等の値を代入しnを未知数として解く事によって、目標精度の標本推定を実現するに必要な標本の大いさnを知る事が出来るのである。若し目標精度が誤差変異係数の値で与えられる時は、母変異係数の値を推定し、それ等を(2.11)式に代入してnを未知数として解けばよい。普通目標精度は誤差変異係数によって表されるため、必要標本数は(2.11)式

$$C_{\bar{x}'}=\frac{c}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\quad より$$

$$\left(\frac{c}{C_{\bar{x}'}}\right)^2=\frac{n(N-1)}{N-n}$$

$$N\left(\frac{c}{C_{\bar{x}'}}\right)^2=\left\{\left(\frac{c}{C_{\bar{x}'}}\right)^2+(N-1)\right\}n$$

$$\therefore\ n=\frac{N\left(\frac{c}{C_{\bar{x}'}}\right)^2}{\left(\frac{c}{C_{\bar{x}'}}\right)^2+(N-1)}\doteqdot\frac{\left(\frac{c}{C_{\bar{x}'}}\right)^2}{1+\left(\frac{c}{C_{\bar{x}'}}\right)^2\Big/N}\qquad(2.44)$$

によって求められ、無限母集団又は有限母集団修正が略々1に等しい時は

$$n=\left(\frac{c}{C_{\bar{x}'}}\right)^2\qquad(2.45)$$

によって得られる。例えば世帯数26万の某市の平均家計費調査の計画に際して、精度(誤差変異係数)を5%以下に止めるには、どれ丈の単位数が必要であるかを求めよう。母変異係数は他の都市の家計調査に於ける家計費分布の変異係数0.54に鑑みて0.55と仮定したとする。母集団は極めて大きいために必要標本数は(2.45)式より

$$n=\left(\frac{0.55}{0.05}\right)^2=121$$

故に121世帯以上を抽出しなければならない。

尚比例数の推定の場合は、母変異係数は $C=\sqrt{\frac{q}{p}}$

1) 母標準偏差の値は過去の調査又は類似の調査の結果より推定し得る場合が多い。特に母変異係数の値はこの可能性が大である。それは標準偏差の絶対的な値は、仮令同種の集団現象と雖も時処を異にする時は同じではないであろうが、それを相対化した変異係数は、その集団現象に影響する諸条件が著しく変化しない限り、時処的に異る同種の集団現象に於ても略々同じであると考えられるからである。しかし勿論正確な推定ではないから安全のために過大に推定するのが普通である。若し母標準偏差の推定に何等の手掛りも得られない時は、簡単な予備調査を行って推定するのである。

$$=\sqrt{\frac{1-p}{p}}=\sqrt{\frac{1}{p}-1}$$

であるから、結局その調査によって推定せんとする母比例数 p の値を予測すればよいのである。この時は安全のために p は過小に予測すればよいであろう。

かくて標本調査の計画に当って目標精度の達成に必要な標本構成単位の数を数理的に求める事によって、必要な精度の標本推定を最少の調査費で獲得し、又費用が与えられて居りそれより標本の大いさが決る時は、その標本より得られる推定値の精度を予め知る事が出来るので、利用目的に照らして標本の大いさ従って調査費を増減する事によって、正確な標本推定を合理的に行う事が出来るのである。

(8) 母分散の推定

標本推定値の誤差を規定するためには、(2.10)式（及び(2.39)式）より明らかな如く母集団の標準偏差従って分散の値を知らねばならぬ。故に次に母分散の標本推定が必要となるのである。

母分散の推定値としては標本に就ての同じ性質の箇約表章値である標本分散を以てする事が考えられる。しかしながら標本分散 s_2 は必ずしも母分散 σ^2 の不偏推定値ではないのである。s^2 の期望値を求めると

$$E(s^2)=E\left[\frac{1}{n}\sum_{j=1}^{n}(x'_j-\bar{x}')^2\right]$$

$$=\frac{1}{n}E\left[\sum_{j=1}^{n}\{(x'_j-\bar{x})-(\bar{x}'-\bar{x})\}^2\right]$$

$$=\frac{1}{n}E\left[\sum_{j=1}^{n}(x'_j-\bar{x})^2-2(\bar{x}'-\bar{x})\sum_{j=1}^{n}(x'_j-\bar{x})+n(\bar{x}'-\bar{x})^2\right]$$

$$=\frac{1}{n}E\left[\sum_{j=1}^{n}(x'_j-\bar{x})^2-n(\bar{x}'-\bar{x})^2\right]$$

$$\left[\because\ \sum_{j=1}^{n}(x'_j-\bar{x})=n(\bar{x}'-\bar{x})\right]$$

$$=\frac{1}{n}\left[\sum_{j=1}^{n}E(x'_j-\bar{x})^2-n\,E(\bar{x}'-\bar{x})^2\right]$$

この式の括弧内第一項に於ける期望値は、母集団より一の単位を抽出した場合得られる値 x' の標本分布の分散であるから、(2.30)式より

$$E(x'_j-\bar{x})^2=\sigma^2$$

であり、又第二項の期望値は標本平均分布の分散であるから、(2.9)式より

$$E(\bar{x}'-\bar{x})^2=\frac{\sigma^2}{n}\frac{N-n}{N-1}$$

である。故に

$$E(s^2)=\frac{1}{n}\left[n\,\sigma^2-n\frac{\sigma^2}{n}\frac{N-n}{N-1}\right]$$

$$=\frac{nN-n-N+n}{N-1}\frac{\sigma^2}{n}=\frac{N(n-1)}{N-1}\frac{\sigma^2}{n}$$

$$=\frac{N}{N-1}\frac{n-1}{n}\sigma^2$$

このようにs^2はσ^2の不偏推定値ではないのであるが、両辺に$\frac{N-1}{N}\frac{n}{n-1}$を乗ずる事により

$$E\left(\frac{N-1}{N}\frac{n}{n-1}s^2\right)=\sigma^2 \qquad (2.46)$$

となるから、$\frac{N-1}{N}\frac{n}{n-1}s^2$ がσ^2の不偏推定値であり

これを$\hat{\sigma}^2$（又はV）で表す。即ち

$$\hat{\sigma}^2=\frac{N-1}{N}\frac{n}{n-1}s^2 \qquad (2.47)$$

若し無限母集団又は母集団が極めて大きい時は

$$\hat{\sigma}^2=\frac{n}{n-1}s^2$$

であり、尚標本単位数が相当に大であり $n-1\fallingdotseq n$ なる時は

$$\hat{\sigma}^2\fallingdotseq\sigma^2$$

となり、かくて始めて標本分散は母分散の不偏推定値

となるのである。

σ^2の不偏推定値$\hat{\sigma}^2$の標本分布の分散は、無限母集団の場合は[1)]

$$V\left(\frac{n}{n-1}s^2\right)=\sigma^4\left\{\frac{\beta_2-1}{n-1}-\frac{\beta_2-3}{n(n-1)}\right\} \qquad (2.48)^{2)}$$

但し、β_2は母集団分布の尖峰度 $\beta_2=\dfrac{\frac{1}{N}\sum^{N}(x_j-\bar{x})^4}{\sigma^4}$

であり、nが相当に大ならば $\frac{\beta_2-3}{n(n-1)}\fallingdotseq 0$ となるため[3)]

$$V\left(\frac{n}{n-1}s^2\right)\fallingdotseq\sigma^4\frac{\beta_2-1}{n-1} \qquad (2.48')$$

となる。故に変異係数は

$$C_{\hat{\sigma}^2}=\sqrt{\frac{\beta_2-1}{n-1}-\frac{\beta_2-3}{n(n-1)}}\fallingdotseq\sqrt{\frac{\beta_2-1}{n-1}} \qquad (2.49)$$

である。

1) 無限母集団でない場合の$\hat{\sigma}^2$の分散は極めて複雑であるので省略する。畑村、奥野「標本調査法入門」一三四—一六頁に説明されているから参照されたい。

2) (2.48)式の証明は少し複雑となるから省略する。斎藤、浅井「標本調査の設計」三〇四—五頁に証明されている。

3) なんとなればβ_2は普通の母集団分布に於ては8以下であり、nは普通50以上であって且分母に二次の項を含んでいるからである。

この母分散の不偏推定値を(2.9)式のσ^2に代入する事によって、標本平均分布の分散が推定されるのである。

$$V(\bar{x}') \approx^{1)} \frac{N-n}{N-1}\frac{1}{n}\left(\frac{N-1}{N}\frac{n}{n-1}s^2\right)$$

$$=\left(1-\frac{n}{N}\right)\frac{s^2}{n-1} \qquad (2.50)$$

比例数の推定の場合は母分散 $\sigma^2=pq$ 標本分散 $s^2=p'q'$ であるから、母分散の不偏推定値は

$$\hat{\sigma}^2=\frac{N-1}{N}\frac{n}{n-1}\ p'q' \qquad (2.51)$$

である。従って標本比例数分布の分散は

$$V(p') \approx \frac{N-n}{N-1}\frac{n}{n-1}\left(\frac{N-1}{N}\frac{n}{n-1}p'q'\right)$$

$$=\left(1-\frac{n}{N}\right)\frac{p'q'}{n-1} \qquad (2.52)$$

によって推定されるのである。

1) ≈は不偏推定を表す記号とする。

前項に述べたように標本調査の計画に際して、母標準偏差の値の想定に基いて目標精度の達成に必要な標本の大いさが決定されるのであるが、調査の後で母分散の値を標本より推定する事によって、果してその大いさの標本によって目標精度が達成され得たかどうかを確めて置く事が必要である。このように実際の標本によって達成された精度を達成精度 precision attained という。若し母標準偏差に就ての想定が誤って居り達成精度が目標精度以下であったならば、その調査結果は利用目的に鑑みて誤差が過大であるため正確性を欠き、又逆に達成精度が目標精度以上であったならば、標本の大いさを減少せしめる事が出来従って調査費の節減が可能であったのである。故に調査立案者は計画に当って母集団に関する智識、経験を利用して正確な母標準偏差の値を推定し、達成精度を出来る限り目標精度に一致せしめるよう努力しなければならぬ。そして標本より推定された母分散の値は、以後に於ける同種の調査の計画に際して母標準偏差想定の基礎資料となり、達成精度をより以上目標精度に近附ける事を可能ならしめるのである。

(9) 標本抽出の諸方式

以上により明らかな如く任意抽出法による標本調査

は極めて正確な且合理的な調査方法であるが、それによって実際に正確な推定値を得るためには、任意標本理論の根本的な条件である任意標本の構成、即ちN個の母集団構成単位よりのn個の標本構成単位の任意抽出が完全に実行されねばならないのである。任意抽出を行うためには先づ第一に、母集団を構成するN個の単位をすべて洩れなく個々的に明確に把握し、且それよりn個の単位を任意に抽出し得るが如き状態に置く事が必要である。このような母集団構成単位を個々的に記録したものを枠 frame といい、この枠の作成は任意標本調査の不可欠な前提条件である。このためには母集団単位の悉皆調査が行われねばならないのであるが、しかしそれは単に個々の単位を個別的に捉え他の単位との区別を明らかにする丈であって、必ずしも個々の単位に就て大量観察法的個別観察を施すものではないから、悉皆統計調査に比して遙かに容易な作業である。そしてこのようにして一度或る統計集団の構成要素の枠[1]が作成されると、それは、以後統計集団構成要素に変動がない限り、その統計集団を母集団とするすべての標本調査に於て利用し得るのであり、尚且往々にして他の目的より作成された記録を以てそのまま枠として用い得る場合があるのである。

1) 母集団は一方向の集団性より観察した時の統計集団であるから、母集団の単位の記録は実は統計集団の単位の記録である。（拙稿「統計調査法」九六頁参照）

次に枠よりn個の単位を任意に即ち無作為に抽出する手続が実行されるのであるが、それは枠上のすべての単位に一番からN番迄の一連番号を附け、乱数表を用いてn個の番号を決定しその番号の単位を抽出する事によって容易に実現されるのである。この場合n個の単位の一つ一つに就て乱数表を用いて抽出番号を決定する事は手続上煩雑であるため、先づ一番から$\frac{N}{n}$（抽出間隔 sampling interval という）番迄の間の一の数を無作為に決め、それを初項とし抽出間隔を公差とする等差級数を以て抽出番号とする事によって、n個の単位の無作為抽出は容易に行えるのである。これを系統的抽出法 systematic sampling といい実際上多

く用いられる方法である。枠に於ける一番からN番迄の母集団単位の並べ方に、問題たる標識（屬性）又はそれと密接な関連を有する他の属性に関して特定の周期的傾向があり、且それと抽出間隔とが略〻一致するならば、系統的抽出法により構成された標本は特定の標識の単位を特に多く又は少く含む傾向を有し任意標本とはならないが、枠上の配列の順序に何等の傾向もなく全く偶然的であるならば、このような便法によっても抽出行為の無作為性は決して害われないであろう。

かくして任意標本が構成されるとそれに対して大量観察の手続を施して統計を獲得し、それに統計解析的計算を加えて母数の推定値を得、又母分散を推定して信頼区間を求め標本推定の達成精度を確める事によって標本調査は完了するのである。

以上に於ては任意標本の構成は単純に、N個の母集団構成単位よりn個の単位を直接抽出する事によって行われたのであるが、任意標本の構成はこのような単純任意抽出法 simple random sampling のみならず、層化抽出法、集落抽出法（特に地域抽出法）及び多段抽出法或は又確率比例抽出法によっても行われるのであって、それぞれ推定の精度又は調査の費用に於て長所を有するのである。

前に述べたように標本推定の精度は母標準偏差の値、即ち母集団構成単位の変量の間の差異の程度に正比例するため、この母標準偏差を小くする事によって標本推定の精度を高める事が出来るのであるが、それは層化抽出法と集落抽出法とによって可能となるのである。

母集団を同質的な単位の部分集団（層 stratum）に分け、各層毎に単位を抽出して標本を構成する方法を層化抽出法 stratified sampling という。母集団を同質的な層に分つと各層に於ける単位の変量の間の差異従って標準偏差は小さくなるから、各層の値の推定の精度が高まり、故にそれを合成して母集団の値を推定する時はその精度がよくなるのである。勿論母集団の層化に際しては問題たる属性に就て同質的な単位を集めて層を作る事は不可能であるが、それに就ての過去の

統計により或は又それと密接な関連のある他の諸属性に従って同質的な層に分ける事によって、略〻その目的を達する事が出来るであろう。そして仮令このような層化の基準に誤りがあって同質化に失敗し、層内に於て問題たる標識の差異を減少せしめ得なかったとしても、各層内での抽出が正確な枠に基く任意抽出である限り推定の不偏性を害するものではなく、只推定の偶然誤差を減少させる事に成功しなかったのみである。（この事は次に述べる集落抽出法に於ても妥当するのであって、集落の構成が便宜的な基準によってなされ母集団を、たとしてもそれは標本の精度に影響するのみである。）

次に母集団構成単位を若干個集めて集落 cluster を作り母集團を集落の集合たらしめ、それより集落を抽出して標本を構成する方法[1]を集落抽出法 cluster sampling といい、特に母集団構成単位の地域的な集合を集落として用いる場合を地域抽出法 area sampling という。母集團を集落より成る集団に変換した場合、集落の大いさが略〻同じであり且集落相互間が可能な限り同質的であるように作るならば、（その事は母集団が異質的な単位より成っている限り、出来る丈異質的な単位を集めて集落を構成する事によってのみ可能であろう）集落の有する変量（集落を構成する單位の變量の總和）の間の差異は小くなるため母標準偏差は小となり、母集団の値の推定の精度は向上するであろう。しかしながら実際上集落の構成は地域的な集合によってなされる事が多いために、同質的な単位の集りであり従って集落間異質的となり易く母標準偏差の値が大きくなるため、集落抽出法による時は推定の精度はむしろ低下するのが普通である。しかし社会的集団現象は普通広大な範囲に亘って存在するため、抽出された標本単位が全域にばらばらに離れて存在するよりも出来る丈地域的に集っている方が、大量観察法的個別観察に便利でありその費用は少くて済むであろう。故に標本調査の実際に於ては精度が落ちるにも拘らず地域抽出法が多く用いられるのである。尚集落抽出法に於

ては単位の個別観察の費用のみならず枠の作成費の点からも費用を節減する事が出来るのである。それは単純抽出法による時は枠上にN個の母集団構成単位を載せねばならないが、集落抽出法による時はそれよりも小数の集落を載せればよいからである。

1) 茲に於て調査単位 unit of analysis と抽出単位 sampling unit との区別が必要となるのである。調査単位は母集団を構成する単位であって大量観察法的個別観察手続の対象たる単位であり、抽出単位は標本を構成するために母集団より抽出される単位であって、それは抽出の便宜上定められた単位に過ぎず、調査単位それ自体であるか又は調査単位の集合（集落）である。

そして抽出された集落（地域）が尚広大であり、又はそれに含まれる調査単位数が多いためにその全部観察が困難又は不可能である時は、その集落（地域）より更に単位を抽出して標本を構成する事が必要となるのである。このように母集団より一度抽出された集落の構成単位のうちより更に単位を抽出して標本を作る方法を二段抽出法 two-stage sampling（又は副次抽出法 subsampling）といい、最初の抽出単位たる集落を一次抽出単位 primary sampling unit 二回目の抽出単位（調査単位又は集落）を二次抽出単位 secondary sampling unit という。（この場合は第一次抽出のための枠は、一次抽出単位を記載し、第二次抽出のための枠は抽出された一次抽出単位に就てのみ作成すればよいのであるから、この二次抽出単位に相当する集落又は調査単位を直接抽出する一段抽出法に比して、枠の作成の費用は著しく節減される事となるのである。）このような副次抽出を更に繰返す事によって三段抽出法、四段抽出法等が考えられ、これ等を総称して多段抽出法 multi-stage sampling という。

尚集落抽出法に於て集落の大いさが著しく異る時は、それの抽出に際してすべての集落に等しい抽出の確率を与えて抽出するよりも、集落の大いさに略〻比例した確率を与えて抽出する方が推定の精度が高まる事が数理的に証明されるのである。このように抽出単位にその大いさに比例した抽出確率を与えて抽出を行う方法を確率比例抽出法 sampling with probability proportionate to size という。

以上のように諸種の標本抽出方式は、或るものは標本推定の精度を高め又或るものは調査費用を節減せしめるのである。従って標本調査の実施に当ってこれ等諸標本抽出方式を適宜組合せる事によって、与えられた費用と調査目的より要請される推定の精度との関係を最適ならしめるように、即ち必要な精度の標本推定を最小の費用を以て、或は又与えられた費用を以て最高の精度の推定値を獲得し得るように、標本の構造、規模を決定する事が出来るのである。これを標本設計 sample design という。この標本設計のためには母集団の層化、集落化に必要な資料、各種段階に於ける費用に就ての智識、標本の大いさの決定に必要な母標準偏差の値の推定、及び諸抽出方式に応ずる不偏な推定式並にその分散式を知らねばならぬ。以下各抽出方式に於ける不偏推定式及びその分散式を求め、その抽出方式の特徴を説明しよう。