任意標本調査法（五）

関　弥三郎

(3) 重複抽出を許さない場合

（i） 二段抽出法に於ける推定式

母集団のM個の集落よりm個の集落を抽出する場合、同一集落の重複抽出を許さないとすると各抽出は相互独立ではあり得なくなる。何となれば重複して抽出してはいけない時は、抽出された集落を元へ戻さないで抽出を続ける場合と同じであり、従って先の抽出によって如何なる集落が抽出されたかによって、後の抽出集落が異るからである。

しかし(2・17)式の確率変数の和の期望値に関する公式は、その確率変数が相互独立であると否とに拘らず成立するため、母平均の推定式(6・9)はこの場合も不偏推定である。しかるに(2・27)式の確率変数の和の分散に関する公式は、確率変数が相互に独立なる場合にのみ妥当するのであるから、(6・9)式の分散を求める過程に於て(6・11)が成立せず、かくて分散は新たに求められねばならないのである。

重複抽出を許さない場合の(6・9)式の分散を求めるに当り、最初に説明した確率比例抽出法の抽出方法を反省すると、そこに於ては各集落はそれを構成する調査単位の数丈あるものと看做されて居り、従ってM個の集落の母集団はN個の集落の母集団に拡大され[1]、それよりm個の集落を重複する事なく抽出するものと考えられ、かく解する時は確率比例抽出法は等

確率抽出法の場合に帰着する事となるのである。[2)]

1）　若し任意の値P_jによって確率比例抽出が行れる時は、その数丈同一集落があるものとされ、故にP個の集落の母集団に拡大されるのである。

2）　この考えの下に(6•9)式の期望値を求めると[3)]

$$\underset{jk}{E}(\bar{x}') = \frac{1}{m}\sum_{j=1}^{m}\underset{j}{E}\left\{\frac{P}{NP_j}\underset{k}{\overset{(j)}{E}}(X_j')\right\}$$

$$=\underset{j}{E}\left(\frac{P}{NP_j}X_j\right) = \frac{1}{P}\sum_{j=1}^{P}\frac{P}{NP_j}X_j$$

このP個の値$\frac{P}{NP_j}X_j$を同じ値のもの毎にまとめると、P_j個宛（$j=1, 2, \cdots\cdots, M$）のM組になるから

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

となり、不偏推定である事が証明される

3）　この場合はE_j、$\overset{(j)}{E_k}$共に等確率抽出の期望値を表す。

そこで(6•9)式の分散を求めると

$$\underset{jk}{V}(\bar{x}') = \underset{j}{V}\left\{\underset{k}{\overset{(j)}{E}}(\bar{x}')\right\} + \underset{j}{E}\,\underset{k}{\overset{(j)}{V}}(\bar{x}')$$

第一項：$\underset{j}{V}\left\{\underset{k}{\overset{(j)}{E}}(\bar{x}')\right\} = \underset{j}{V}\left\{\frac{1}{m}\sum_{j=1}^{m}\frac{P}{NP_j}\underset{k}{\overset{(j)}{E}}(X_j')\right\}$

$$=\frac{1}{N^2}\underset{j}{V}\left\{\frac{1}{m}\sum_{j=1}^{m}\left(\frac{P}{P_j}X_j\right)\right\}$$

この$\underset{j}{V}\left\{\frac{1}{m}\sum_{j=1}^{m}\left(\frac{P}{P_j}X_j\right)\right\}$は、$P$個の単位より成る母集団から$m$個の単位を抽出して構成した標本に於ける$\left(\frac{P}{P_j}X_j\right)$の平均の誤差分散である。従って拡大された母集団に於ける$\left(\frac{P}{P_j}X_j\right)$の分散を$\bar{\sigma}_e^2$とすると、[1)] (2•33)より

$$\underset{j}{V}\left\{\frac{1}{m}\sum_{j=1}^{m}\left(\frac{P}{P_j}X_j\right)\right\} = \frac{P-m}{P-1}\frac{\bar{\sigma}_e^2}{m}$$

となり、故に

$$\underset{j}{V}\left\{\underset{k}{\overset{(j)}{E}}(\bar{x}')\right\} = \frac{1}{N^2}\frac{P-m}{P-1}\frac{\bar{\sigma}_e^2}{m}$$

又

第二項：$\underset{j}{E}\,\underset{k}{\overset{(j)}{V}}(\bar{x}') = \underset{j}{E}\left\{\frac{1}{m^2}\sum_{j=1}^{m}\left(\frac{P}{NP_j}\right)^2\underbrace{\underset{k}{\overset{(j)}{V}}(X_j')}_{=N_j^2\frac{N_j-r_j}{N_j-1}\frac{\sigma_j^2}{r_j}\equiv A_j}\right\}$

$$=\frac{1}{m}\frac{1}{N^2}\underset{j}{E}\left\{\left(\frac{P}{P_j}\right)^2A_j\right\}$$

$$=\frac{1}{mN^2}\frac{1}{P}\sum_{j=1}^{P}\left(\frac{P}{P_j}\right)^2A_j$$

このP個の$\left(\frac{P}{P_j}\right)^2A_j$を同じもの毎にまとめると、$P_j$個宛（$j=1, 2, \dots, M$）の$M$組になるから

$$=\frac{1}{mN^2}\frac{1}{P}\sum_{j=1}^{M}P_j\left(\frac{P}{P_j}\right)^2A_j$$

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

かくて

$$\underset{jk}{V}(\bar{x}') = \frac{1}{N^2}\left\{\frac{P-m}{P-1}\frac{\bar{\sigma}_e^2}{m} + \frac{1}{m}\sum_{j=1}^{M}\frac{P}{P_j}N_j^2\frac{N_j-r_j}{N_j-1}\frac{\sigma_j^2}{r_j}\right\} \tag{6.21}$$

となる。

1）拡大されたP個の集落より成る母集団に於て、集落の値$\left(\frac{P}{P_j}X_j\right)$の平均値は

$$\frac{1}{P}\sum_{j=1}^{P}\frac{P}{P_j}X_j=\frac{1}{P}\sum_{j=1}^{M}P_j\frac{P}{P_j}X_j=\sum_{j=1}^{M}X_j=X$$

即ちM組の集落より成る元の母集団の量的大いさである。従って$\left(\frac{P}{P_j}X_j\right)$の分散は

$$V\left(\frac{P}{P_j}X_j\right)=\frac{1}{P}\sum_{j=1}^{P}\left(\frac{P}{P_j}X_j-X\right)^2$$
$$=\sum_{j=1}^{M}\frac{P_j}{P}\left(\frac{P}{P_j}X_j-X\right)^2\equiv\bar{\sigma}^2_e \quad [(6.4)]$$

である。

集落の大いさに比例する場合　この場合も推定式は(6•13)に同じであり、その分散は(6•21)に於て$P_j=N_j$, $P=N$と置いて

$$\underset{jk}{V}(\bar{x}')=\frac{N-m}{N-1}\frac{\sigma_b^2}{m}+\frac{1}{m}\sum_{j=1}^{M}\frac{N_j}{N}\frac{N_j-n_j}{N_j-1}\frac{\sigma_j^2}{n_j} \quad [(6.7)] \qquad (6.22)$$

である。

(ii)　母分散の推定

(6•21)式により誤差分散を計算するに必要な母分散$\bar{\sigma}_e^2$及びσ_j^2 ($j=1, 2, \dots, M$) を標本より推定する式を求めよう。（これの求め方は等確率抽出による二段抽出法の場合と全く同じであり、唯期望値計算を異にするのみであるためそれを参照されたい。）

先づσ_j^2の推定であるが、それを個々的に推定する事は不可能であるので

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

を推定せんとする。するとそれの不偏推定は

$$\frac{1}{m}\sum_{j=1}^{m}\left(\frac{P}{P_j}\right)^2N_j^2\frac{N_j-n_j}{N_j-1}\frac{\hat{\sigma}_j^2}{n_j} \qquad (6.23)$$

但　$\hat{\sigma}_j^2=\frac{N_j-1}{N_j}\frac{n_j}{n_j-1}s_j^2$,　$\underset{k}{E}^{(j)}(\hat{\sigma}_j^2)=\sigma_j^2$

である。何となれば

$$\underset{jk}{E}\left\{\frac{1}{m}\sum_{j=1}^{m}\left(\frac{P}{P_j}\right)^2N_j^2\frac{N_j-n_j}{N_j-1}\frac{\hat{\sigma}_j^2}{n_j}\right\}$$
$$=\frac{1}{m}\sum_{j=1}^{m}\underset{j}{E}\left\{\left(\frac{P}{P_j}\right)^2\underbrace{N_j^2\frac{N_j-n_j}{N_j-1}\frac{1}{n_j}\underset{k}{E}^{(j)}(\hat{\sigma}_j^2)}_{=A_j}\right\}$$
$$=\frac{1}{P}\sum_{j=1}^{P}\left(\frac{P}{P_j}\right)^2A_j$$
$$=\frac{1}{P}\sum_{j=1}^{M}P_j\left(\frac{P}{P_j}\right)^2A_j=\sum_{j=1}^{M}\frac{P}{P_j}N_j^2\frac{N_j-n_j}{N_j-1}\frac{\sigma_j^2}{n_j} \qquad (6.24)$$

〔証明終り〕

次に$\bar{\sigma}_e^2$の不偏推定を求めるために、標本に於ける同性質の分散$\bar{s}_e^2$の期望値を求めると、(6•16)より

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

そして$\frac{P}{P_j}X_j'=Nx'$, $m=1$であるため

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

しかし茲では重複抽出を許さないのであるから、(6·21)より

$$=\bar{\sigma}_e^2+\sum_{j=1}^{M}\frac{P}{P_j}\underbrace{N_j^2\frac{N_j-n_j}{N_j-1}\frac{\sigma_j^2}{n_j}}_{=A_j}$$

又 $X'=N\bar{x}'$ であるため

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

$$=\frac{P-m}{P-1}\frac{\bar{\sigma}_e^2}{m}+\frac{1}{m}\sum_{j=1}^{M}\frac{P}{P_j}\underbrace{N_j^2\frac{N_j-n_j}{N_j-1}\frac{\sigma_j^2}{n_j}}_{=A_j}\qquad[(6.21)]$$

故に

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

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

$$\therefore\ \frac{P-1}{P}\left\{\underset{jk}{E}\left(\frac{m}{m-1}\bar{s}_e^2\right)-\sum_{j=1}^{M}\frac{P}{P_j}A_j\right\}=\bar{\sigma}_e^2\qquad(6.25)$$

かくて(6·25)の左辺第二項にそれの不偏推定値(6·23)を代入した

$$\frac{P-1}{P}\left\{\frac{m}{m-1}\bar{s}_e^2-\frac{1}{m}\sum_{j=1}^{m}\left(\frac{P}{P_j}\right)^2N_j^2\frac{N_j-n_j}{N_j-1}\frac{\hat{\sigma}_j^2}{n_j}\right\}\approx\bar{\sigma}_e^2\qquad(6.26)$$

は $\bar{\sigma}_e^2$ の不偏推定である。

集落の大いさに比例する場合(6·22)式に於て必要な母分散 σ_b^2、$\sigma_j^2\ (j=1,2,\cdots M,)$ の不推偏定は、(6·25)(6·24)に於て $P_j=N_j$, $P=N$ と置く事により得られる。先づ(6·24)より

$$N^2\underset{jk}{E}\left\{\frac{1}{m}\sum_{j=1}^{m}\frac{N_j-n_j}{N_j-1}\frac{\hat{\sigma}_j^2}{n_j}\right\}=N^2\sum_{j=1}^{M}\frac{N_j}{N}\frac{N_j-n_j}{N_j-1}\frac{\sigma_j^2}{n_j}$$

かくて

$$\frac{1}{m}\sum_{j=1}^{m}\frac{N_j-n_j}{N_j-1}\frac{\hat{\sigma}_j^2}{n_j}\approx\sum_{j=1}^{M}\frac{N_j}{N}\frac{N_j-n_j}{N_j-1}\frac{\sigma_j^2}{n_j}\qquad(6.27)$$

次に(6·25)より

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

[(6.19)(6.7)]

この式の左辺第二項にそれの不偏推定値(6·27)を代入して

$$\frac{N-1}{N}\left\{\frac{m}{m-1}s_b^2-\frac{1}{m}\sum_{j=1}^{m}\frac{N_j-n_j}{N_j-1}\frac{\hat{\sigma}_j^2}{n_j}\right\}\approx\sigma_b^2\qquad(6.28)$$

となる。

(iii) 精度の比較

集落の大いさによる確率比例抽出に於て重複抽出を許す場合の分散を V_1、それを許さない場合の分散を V_2 とし、両者の差をとると

$$V_1-V_2=\left(1-\frac{N-m}{N-1}\right)\frac{\sigma_b^2}{m}=\frac{m-1}{N-1}\frac{\sigma_b^2}{m}>0,\quad m\neq1$$

若し N が m に比して極めて大であり $N-m\fallingdotseq N-1$ の時は

$$V_1-V_2\fallingdotseq0$$

である。かくて一般に重複抽出を許さない場合の方が推定の精度は高いのであり、唯母集団の調査単位数が標本の集落数に比して大なる時は、両者の差は殆ど問題とならないのであ

る。

層化抽出法による母平均の推定は先づ各層の量的大いさを推定し、それを総計して母集団の単位総数で除する事によつて行われるのであつた。故に以上の二段抽出法に於ける推定式を用いて、容易に層化抽出法に於ける推定式を求める事が出来る。

(iv)　層化二段抽出法に於ける推定式

今L個の層の各〻に於て、任意の値P_{ij}による確率比例抽出により、M_i個の集落よりm_i個の集落を重複する事なしに抽出し、抽出された集落のN_{ij}個の調査単位からn_{ij}個の単位を等確率で抽出して標本を構成し、母平均$\bar{x}$を推定するものとする。するとi番層の量的大いさX_iの不偏推定値X_i'は、(6・9)より

$$X_i' = N_i\bar{x}_i' = N_i\frac{1}{m_i}\sum_{j=1}^{m_i}\frac{P_i}{N_iP_{ij}}X_{ij}' = \frac{1}{m_i}\sum_{j=1}^{m_i}\frac{P_i}{P_{ij}}\left(\frac{N_{ij}}{n_{ij}}\sum_{k=1}^{n_{ij}}x_{ijk}'\right) \qquad (6.29)$$

であるため、

$$\underset{jk}{E}(X_i') = N_i\underset{jk}{E}(\bar{x}_i') = N_i\bar{x}_i = X_i \qquad [(6.10)]$$

$\bar{x}$の推定値$\bar{x}'$は

$$\bar{x}' = \frac{1}{N}\sum_{i=1}^{L}X_i' = \frac{1}{N}\sum_{i=1}^{L}\left(\frac{1}{m_i}\sum_{j=1}^{m_i}\frac{P_i}{P_{ij}}\frac{N_{ij}}{n_{ij}}\sum_{k=1}^{n_{ij}}x_{ijk}'\right) \qquad (6.30)$$

である。期望値を求めると

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

となるため、これは不偏推定である。分散は

$$\underset{jk}{V}(\bar{x}') = \frac{1}{N^2}\sum_{i=1}^{L}\underset{jk}{V}(X_i')$$
$$= \frac{1}{N^2}\sum_{i=1}^{L}\left\{\frac{P_i-m_i}{P_i-1}\frac{\bar{\sigma}_{ie}^2}{m_i} + \frac{1}{m_i}\sum_{j=1}^{M_i}\frac{P_i}{P_{ij}}N_{ij}^2\frac{N_{ij}-n_{ij}}{N_{ij}-1}\frac{\sigma_{ij}^2}{n_{ij}}\right\} \qquad [(6.21)] \quad (6.31)$$

である。

集落の大いさに比例した確率で抽出する場合は、(6・30)(6・31)に於て$P_j=N_j$, $P=N$と置けばよく、又重複抽出を許す場合の推定式も今と同様にして容易に求める事が出来るので省略する。

(4)　確率比例抽出法の精度

母集団を構成する集落の大いさがすべて相等しい時は、集落の大いさによる確率比例抽出は、何れの集落が抽出される確率も同一であるため、等確率抽出と同じ事になる。（集落の大いさに比例する場合の推定式(6・13)、分散(6・22)に於て$N_j=\bar{N}$と置き整理すると、集落の大いさが均一なる場合の等確率二段抽出法の推定式(4・24)、分散(4・25)に一致する。）従って確率比例抽出法の分散と等確率抽出法の分散の相異は、母集団の集落の大いさの変動によるのである。実際上確率比例抽出法が用い

られるのは、層化多段抽出に於て各層より唯一個の集落を抽出する場合である。故に母集団（但し便宜上層化しないものとする）より一個の集落を抽出する場合に就て、確率比例抽出法と等確率抽出法の精度を比較して見よう。

二段抽出法による母平均の推定に於て、等確率抽出の場合の分散をV_1、集落の大いさによる確率比例抽出の場合の分散をV_2とすると、それぞれは(4・22)（但 $m=1$）及び(6・8)より

$$V_1=\frac{1}{N^2}\left\{M^2\sigma_e^{\,2}+M\sum_{j=1}^{M}N_j^{\,2}\frac{N_j-n_j}{N_j-1}\,\frac{\sigma_j^{\,2}}{n_j}\right\},$$

$$\sigma_e^{\,2}=\frac{1}{M}\sum_{j=1}^{M}(X_j-\bar{X})^2$$

$$V_2=\sigma_b^{\,2}+\sum_{j=1}^{M}\frac{N_j}{N}\,\frac{N_j-n_j}{N_j-1}\,\frac{\sigma_j^{\,2}}{n_j},$$

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

である。両者の差 V_1-V_2 をとると

$$〔第一項の差〕=\frac{1}{N^2}(M^2\sigma_e^{\,2}-N^2\sigma_b^{\,2})$$

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

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

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

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

$$\left.-\bar{N}N_j\bar{x}_j^{\,2}+2\bar{N}N_j\bar{x}_j\bar{x}-\bar{N}N_j\bar{x}^2\right\}$$

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

$$\bar{N}\bar{x}^2\sum_{j=1}^{M}(\bar{N}-N_j)=\bar{N}\bar{x}^2M(\bar{N}-\bar{N})=0$$

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

しかるに$N_j>0$, $\bar{N}>0$ なるため $N_j>N_j-\bar{N}$ 故に

$(N_j-\bar{N})N_j\bar{x}_j^{\,2}>(N_j-\bar{N})^2\bar{x}_j^{\,2}$ $(j=1,\ 2,\ \cdots,\ M)$

$$\therefore\quad >\frac{M}{N^2}\sum_{j=1}^{M}(N_j-\bar{N})^2\bar{x}_j^{\,2}>0 \qquad \text{(i)}$$

次に $\dfrac{N_j-n_j}{N_j-1}\,\dfrac{\sigma_j^{\,2}}{n_j}\equiv A_j'$ と置いて

$$〔第二項の差〕=\frac{M}{N^2}\left(\sum_{j=1}^{M}N_j^{\,2}A_j'-\frac{N}{M}\sum_{j=1}^{M}N_jA_j'\right)$$

$$=\frac{M}{N^2}\sum_{j=1}^{M}(N_j-\bar{N})N_jA_j' \qquad \left[\because\ \bar{N}=\frac{N}{M}\right]$$

前と同様にして $(N_j-\bar{N})N_jA_j'>(N_j-\bar{N})^2A_j'$ $(j=1,\ 2,\ \cdots,\ M)$ なるため

$$>\frac{M}{N^2}\sum_{j=1}^{M}(N_j-\bar{N})^2A_j'>0 \qquad \text{(ii)}$$

故に

$$V_1-V_2>0$$

かくて等確率抽出法よりも確率比例抽出法による方が推定の精度は高いのである。そして(i)(ii)の値はN_jの変動が大なる程大となるから集落の大いさの相異が激しい程確率比例抽出法の方が精度は高いのである。

以上は母集団より唯一個の集落を抽出する場合であった。

若し二個以上抽出するものとすると、V_1 は(4・22)式 V_2 は(6・22)式で与えられ、それ等の第二項の比較の結論は $m=1$ の場合と同じであるが、第一項の比較に於ては、それぞれに有限母集団修正が掛って居り

$$\frac{M-m}{M-1}<\frac{N-m}{N-1} \qquad \left[\because M<N\right]$$

であるため、第一項の比較は $m=1$ の場合の様に、V_1 の方が V_2 よりも大であるとは一般にいい得ないのであり、且 m の増加と共に $\frac{M-m}{M-1}$, $\frac{N-m}{N-1}$ 共に減少するが、そのテムポは $\frac{N-m}{N-1}$ の方が小であるため、m の数如何によっては V_2 の方が V_1 よりも大なる事があるのである。

しかし M、N が m に比して極めて大であり有限母集団修正を無視し得る時は第一項の比較の結論は $m=1$ の場合と同じであるため、一般に確率比例抽出法の方が精度は高いのである。

確率比例抽出法は集落の大いさによらなくとも、任意の値によつて同様の効果を挙げる事が出来るのである。次にその条件を考察して見よう。（但し比較の便宜上唯一個の集落を抽出する場合とする。）

任意の値による時の分散を V_1、集落の大いさによる時の分散を V_2 とすると、(6・5)及び(6・8)より

$$V_1=\frac{1}{N^2}\left\{\bar{\sigma}_e{}^2+\sum_{j=1}^{M}\frac{P}{P_j}N_j{}^2\frac{N_j-\eta_j}{N_j-1}\frac{\sigma_j{}^2}{\eta_j}\right\},$$

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

$$V_2=\sigma_b{}^2+\sum_{j=1}^{M}\frac{N_j}{N}\frac{N_j-\eta_j}{N_j-1}\frac{\sigma_j{}^2}{\eta_j},$$

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

である。茲で

$$\begin{aligned}\bar{\sigma}_e{}^2&=\sum_{j=1}^{M}\frac{P_j}{P}\left(\frac{P}{P_j}N_j\bar{x}_j-N\bar{x}\right)^2\\&=\sum_{j=1}^{M}\frac{P_j}{P}N^2\left(\frac{PN_j}{P_jN}\bar{x}_j-\bar{x}\right)^2\\&=N\sum_{j=1}^{M}N_j\left(\frac{P_jN}{PN_j}\right)\left(\frac{PN_j}{P_jN}\bar{x}_j-\bar{x}\right)^2\end{aligned}$$

と変形され得る事に注意して両者の差を求めると

$$\begin{aligned}V_1-V_2=\frac{1}{N}\sum_{j=1}^{M}N_j\left\{\frac{P_jN}{PN_j}\left(\frac{PN_j}{P_jN}\bar{x}_j-\bar{x}\right)^2-(\bar{x}_j-\bar{x})^2\right\}\\+\sum_{j=1}^{M}\left(\frac{PN_j}{P_jN}-1\right)\frac{N_j}{N}\frac{N_j-\eta_j}{N_j-1}\frac{\sigma_j{}^2}{\eta_j}\end{aligned}$$

であり、故に

$$\frac{PN_j}{P_jN}\gtreqless 1\ (j=1,2,\cdots,M) \quad ならば \quad V_1-V_2\gtreqless 0$$

即ち

$$\frac{N_j}{N}\gtreqless\frac{P_j}{P}\ (j=1,2,\cdots,M) \quad ならば \quad V_1\gtreqless V_2$$

となる。かくて集落の相対的大いさ $\frac{N_j}{N}$ に等しい相対的大いさ $\frac{P_j}{P}$ を持つ値（換言すれば N_j と正の完全相関を有する値）P_j による時は両者の精度は一致し、$\frac{N_j}{N}$ よりも大なる相対的大いさの値による時は、集落の大いさによるよりも精

度が高くなるのである。

七　比推定

今迄述べて来たのは母平均、母比例数並に母分散の推定であつた。今度は母集団に於ける比の推定を説明しよう。母集団の比を推定する場合の推定式を比推定 ratio estimate という。

母集団の比には量的標識に関するものと質的標識に関するものとがある。前者は例えば売掛代金総額に対する貸倒総額の比率（貸倒率）又は家計費中に占める飲食費の割合（エンゲル係数）の如き、母集団に於ける二種の変量 x、y に関する量的大いさの比又は平均値の比

$$z=\frac{X}{Y}=\frac{N\bar{x}}{N\bar{y}}=\frac{\bar{x}}{\bar{y}}$$

或は又基準時（地）賃銀に対する比較時（地）賃銀の比率（賃銀指数）の如き、時間的（場所的）に異る同種変量に関する比である。

後者は β 標識の単位数 N_2 に対する α 標識の単位数 N_1 の比、又は前者の割合 $p_2=\dfrac{N_2}{N}$ に対する後者の割合 $p_1=\dfrac{N_1}{N}$ の比

$$z=\frac{N_1}{N_2}=\frac{\frac{N_1}{N}}{\frac{N_2}{N}}=\frac{p_1}{p_2}$$

である。そしてこれには労務者に対する職員の比率の如き β 標識と α 標識との関係が排反的な場合と、労務者に対する業務上の傷病者の比の如き包括的な場合とがある。

母集団に於ける比 $z=\dfrac{X}{Y}=\dfrac{\bar{x}}{\bar{y}}$ の推定は、分子並に分母を標本より推定しそれの比をとる事によつて行れる。この推定式 $z'=\dfrac{X'}{Y'}=\dfrac{\bar{x}'}{\bar{y}'}$ に於ては、分子、分母が共に標本によつて変動する確率変数であるという点に於て、今迄考察して来た一確率変数のみの推定式とその性質を異にするのである。従って比推定の分散を導くに当つて、先づ二確率変数の同時的分布の相互関係を測定する共変量、相関係数及びそれの計算に於て成立する諸関係を知らねばならないのである。

比推定は特定の標本構成の方法ではなく標本推定の方法である。故に今迄考察して来た標本構成方法の何れに対しても適用し得るのであり、従って各構成方法毎に比推定とその分散を説明しなければならないのであるが、茲では基本的な単純抽出法の場合のみを説明する事にする。

(1) 確率変数の共変量と相関係数

常に相対応して実現値をとる二の確率変数[1] ξ、η の同時的確率分布に於て、ξ、η が値 $(x_i,\ y_j)$ $\left(\begin{smallmatrix} i=1,\ 2,\ \cdots,\ k \\ j=1,\ 2,\ \cdots,\ l \end{smallmatrix}\right)$ をとる確率を p_{ij} とすると、ξ、η の間の共変量 $Cov(\xi,\ \eta)$ は共

変量の定義式(18)より

$$Cov(\xi,\ \eta)=\sum_{i=1}^{k}\sum_{j=1}^{l}p_{ij}(x_i-E\xi)(y_j-E\eta)$$
$$=E[(\xi-E\xi)(\eta-E\eta)]\quad[(2.15)]\ (7.1)$$

である。

1) 例えば n 人の学生を抽出して、英語の成績の平均点 $\bar{x}$ と数学の成績の平均点 $\bar{y}$ を求める場合、平均点の組（$\bar{x}$, $\bar{y}$）は標本によって変動するため、これは対応して実現値をとる二の確率変数である。

(7・1)は次のように変形するが出来る。

$$Cov(\xi,\eta)=E[(\xi-E\xi)(\eta-E\eta)]$$
$$=E[\xi\eta-\eta\cdot E\xi-\xi\cdot E\eta+E\xi\cdot E\eta]$$
$$=E(\xi\eta)-E(\xi)E(\eta)\qquad(7.2)$$

(7・2)より

$$E(\xi\eta)=Cov(\xi,\ \eta)+E(\xi)E(\eta)\qquad(7.3)$$

若し ξ、η に常数 a が掛っている時は

$$Cov(a\xi,\ a\eta)=E[(a\xi-Ea\xi)(a\eta-Ea\eta)]$$
$$=E[a(\xi-E\xi)\cdot a(\eta-E\eta)]$$
$$=a^2E[(\xi-E\xi)(\eta-E\eta)]=a^2Cov(\xi,\ \eta)\qquad(7.4)$$

となる。

そして二個の二確率変数の組 $(\xi_1,\ \eta_1)(\xi_2,\ \eta_2)$ に於いて、ξ_1 と η_2、ξ_2 と η_1 が相互に独立ならば

$$Cov(\xi_1+\xi_2,\ \eta_1+\eta_2)=Cov(\xi_1,\ \eta_1)+Cov(\xi_2,\ \eta_2)\qquad(7.5)$$

これは次のようにして証明される。

$$Cov(\xi_1+\xi_2,\ \eta_1+\eta_2)$$
$$=E\{(\xi_1+\xi_2)-E(\xi_1+\xi_2)\}\{(\eta_1+\eta_2)-E(\eta_1+\eta_2)\}$$
$$=E\{(\xi_1-E\xi_1)+(\xi_2-E\xi_2)\}\{(\eta_1-E\eta_1)+(\eta_2-E\eta_2)\}$$
$$=E[(\xi_1-E\xi_1)(\eta_1-E\eta_1)+(\xi_2-E\xi_2)(\eta_1-E\eta_1)$$
$$+(\xi_1-E\xi_1)(\eta_2-E\eta_2)+(\xi_2-E\xi_2)(\eta_2-E\eta_2)]$$

しかるに ξ_1 と η_2、ξ_2 と η_1 とは相互独立なるため

第二項：$E[(\xi_2-E\xi_2)(\eta_1-E\eta_1)]$

$$=E(\xi_2-E\xi_2)\cdot E(\eta_1-E\eta_1)=0\qquad[(2.18)]$$

同様にして　第三項：$E[(\xi_1-E\xi_1)(\eta_2-E\eta_2)]=0$

故に

$$Cov(\xi_1+\xi_2,\eta_1+\eta_2)=E[(\xi_1-E\xi_1)(\eta_1-E\eta_1)]$$
$$+E[(\xi_2-E\xi_2)(\eta_2-E\eta_2)]$$
$$=Cov(\xi_1,\ \eta_1)+Cov(\xi_2,\ \eta_2)\qquad〔証明終り〕$$

一般に n 個の二確率変数の組（ξ_i, η_i）（$i=1,\ 2,\ \cdots,\ n$）に於て、添数の異る（即ち組を異にする）ξ と η とが相互に独立であるならば

$$Cov\ (\sum_{i=1}^{n}\xi_i,\ \sum_{i=1}^{n}\eta_i)=\sum_{i=1}^{n}Cov(\xi_i,\ \eta_i)\qquad(7.6)$$

である事が同様にして証明される。（但し証明は略する。）

次に ξ、η 間の相関係数 $\rho_{\xi,\eta}$ は(19)より

$$\rho_{\xi,\eta}=\frac{Cov(\xi,\eta)}{\sigma_\xi\sigma_\eta} \qquad (7.7)$$

であり、これによって相関係数に関する計算は共変量と標準偏差とに分解して行れるのである。尚(17)より

$$|\rho_{\xi,\eta}|\leqq 1 \qquad (7.8)$$

である。

標本平均分布の共変量　比推定の分散を求める場合、母集団の量的大いさ又は平均値の標本推定の共変量が必要となるのである。そこで次に母平均推定値即ち標本平均値の共変量を求めよう。（この過程は標本平均分布の分散(2・33)を求める過程と全く同一であるためそれを参照されたい。）

N個の単位より成る母集団からn個の単位を抽出して構成した標本の、変量xに就ての平均値$\bar x'$とyに就ての平均値$\bar y'$との共変量 $Cov(\bar x',\bar y')$ は $E(\bar x')=\bar x$, $E(\bar y')=\bar y$ であるため

$$Cov(\bar x',\bar y')=E[(\bar x'-\bar x)(\bar y'-\bar y)]$$

$$=E(\bar x'\bar y')-\bar x\bar y \qquad [(7.2)]$$

しかるに

$$E(\bar x'\bar y')=E\left(\frac1n\sum_{j=1}^n x_j'\right)\left(\frac1n\sum_{j=1}^n y_j'\right)$$

$$=\frac{1}{n^2}E\left(\sum_{j=1}^n x_j'y_j'+\sum_{j\neq j'}^{n}\sum^{n} x'_j y'_{j'}\right)$$

$$=\frac{1}{n^2}\left\{\sum_{j=1}^n E(x_j'y_j')+\sum_{j\neq j'}^{n}\sum^{n}E(x_j'y_{j'}')\right\} \qquad (7.9)$$

右辺第一項の期望値は、母集団より1個の単位を抽出した時得られるxとyの値の積であるため

$$E(x_j'y_j')=\frac1N\sum_{j=1}^N x_jy_j=Cov(x,y)+\bar x\bar y \qquad [(7.3)]$$

右辺第二項の期望値は、母集団より二個の単位を抽出した時得られる、最初の単位のxの値と後の単位のyの値の積の期望値であるため

$$E(x'_jy'_{j'})=\frac{1}{N(N-1)}\sum_{j\neq j'}^{N}\sum^{N} x_jy_{j'}$$

しかるに　$$\bar x\bar y=\left(\frac1N\sum_{j=1}^N x_j\right)\left(\frac1N\sum_{j=1}^N y_j\right)$$

$$=\frac1{N^2}\left(\sum_{j=1}^N x_jy_j+\sum_{j\neq j'}^{N}\sum^{N}x_jy_{j'}\right)$$

$$\therefore\ \sum_{j\neq j'}^{N}\sum^{N} x_jy_{j'}=N^2\bar x\bar y-\sum_{j=1}^N x_jy_j$$

$$=N^2\bar x\bar y-N\{Cov(x,y)+\bar x\bar y\}$$

$$=N(N-1)\bar x\bar y-N\,Cov(x,y) \qquad [(20')]$$

故に

$$E(x'_jy'_{j'})=\bar x\bar y-\frac{Cov(x,y)}{N-1}$$

これ等を(7・9)に代入すると

$$E(\bar x'\bar y')=\frac1{n^2}\left\{n(Cov(x,y)+\bar x\bar y)+n(n-1)\left(\bar x\bar y-\frac{Cov(x,y)}{N-1}\right)\right\}$$

$$=\frac{N-n}{N-1}\frac{Cov(x,y)}{n}+\bar x\bar y$$

かくて

$$Cov(\bar{x}', \bar{y}') = \frac{N-n}{N-1}\frac{Cov(x, y)}{n} \tag{7.10}$$

となる。

(2) 比推定の期望値と分散

母集団の比 z の推定は、先ず分子及び分母を推定しそれの比をとる事によつて行れるのであり、その場合分子及び分母の推定には不偏推定を用いるべきであるから、分子の不偏推定値を ξ、分母の不偏推定値を η とすると、母集団の比は $z=\frac{E\xi}{E\eta}$ であり、それの推定値 z' は

$$z' = \frac{\xi}{\eta} \tag{7.11}$$

である。

この比推定は次に明らかにするように、今迄の推定値と異り偏りある推定値である。故に茲で標本推定値の偏りの意義並にその場合の標本誤差の分散度の測定を知らねばならぬ。

大体標本推定値の誤差とは、一標本より得られた推定値 ξ と母数 θ との差 $(\xi-\theta)$ である。そして標本推定値が不偏であるとは、それの期望値が母数に一致する事であり、従って標本誤差の期望値が0になる事である。即ち

$$E(\xi)=\theta$$

$$\therefore \quad E(\xi-\theta)=E(\xi)-\theta=\theta-\theta=0$$

そして標本推定値が偏りある推定値であるとは、それの期望値が母数に一致しない事であり、その差が「偏り bias」である。即ち

$$E(\xi) \neq \theta$$

$$\therefore \quad E(\xi-\theta)=E(\xi)-\theta=\text{bias}$$

偏りある推定値の標本誤差の分散度は、標本誤差 $(\xi-\theta)$ の自乗の期望値によって計算され、それを平均平方誤差 mean square error といい、$M.S.E.(\xi)$ で表す。即ち

$$M.S.E.(\xi)=E[\xi-\theta]^2$$

$M.S.E.$ は次のように変形される。

$$\begin{aligned} M.S.E.(\xi) &= E[\xi-\theta]^2 \\ &= E[(\xi-E\xi)+(E\xi-\theta)]^2 \\ &= E(\xi-E\xi)^2+2(E\xi-\theta)\underbrace{E(\xi-E\xi)}_{=0}+(E\xi-\theta)^2 \\ &= V(\xi)+(E\xi-\theta)^2 \end{aligned} \tag{7.12}$$

若し $E(\xi)=\theta$ ならば

$$M.S.E.(\xi)=V(\xi)$$

これより明らかな如く標本推定値 ξ の $M.S.E.$ は、ξ の分散と偏りの自乗より成っているのであり、ξ が不偏である時は $M.S.E.$ は分散と一致するのである。故に不偏推定値の場

合は標本誤差の分散度の規定は分散によればよいのであるが、偏りある推定値の場合は $M.S.E.$ によらねばならないのである。

さて比推定 z' の標本分布の状態を知るために先ず期望値を求めよう。今

$$\xi-E\xi=\varDelta\xi,\ \eta-E\eta=\varDelta\eta$$

と置くと

$$z'=\frac{\xi}{\eta}=\frac{E\xi+\varDelta\xi}{E\eta+\varDelta\eta}$$

$$=\frac{E\xi}{E\eta}\frac{\left(1+\frac{\varDelta\xi}{E\xi}\right)}{\left(1+\frac{\varDelta\eta}{E\eta}\right)}$$

$$=\frac{E\xi}{E\eta}\left(1+\frac{\varDelta\xi}{E\xi}\right)\left\{1-\frac{\varDelta\eta}{E\eta}+\left(\frac{\varDelta\eta}{E\eta}\right)^2-\cdots\right\}^{1)}$$

$$=\frac{E\xi}{E\eta}\left\{1+\frac{\varDelta\xi}{E\xi}-\frac{\varDelta\eta}{E\eta}+\left(\frac{\varDelta\eta}{E\eta}\right)^2-\frac{\varDelta\xi\cdot\varDelta\eta}{E\xi\cdot E\eta}+\cdots\right\}\tag{7.13}$$

但し右辺の省略してある部分は $\frac{\varDelta\xi}{E\xi}$、$\frac{\varDelta\eta}{E\eta}$ に関して三次以上の項である。

1) $(1+x)^{-1}=1-x+x^2-x^3+\cdots\cdots$, $(|x|<1)$

茲で z' の期望値を求めると

$$E(z')=\frac{E\xi}{E\eta}\left\{1+\frac{E(\varDelta\xi)}{E\xi}-\frac{E(\varDelta\eta)}{E\eta}+\frac{E(\varDelta\eta)^2}{(E\eta)^2}-\frac{E(\varDelta\xi\cdot\varDelta\eta)}{E\xi\cdot E\eta}+\cdots\cdots\right\}$$

しかるに $E(\varDelta\xi)=E[\xi-E\xi]=0$

$$E(\varDelta\eta)=E[\eta-E\eta]=0$$

$$\therefore\quad E(z')=\frac{E\xi}{E\eta}\left\{1+\frac{E(\varDelta\eta)^2}{(E\eta)^2}-\frac{E(\varDelta\xi\cdot\varDelta\eta)}{E\xi\cdot E\eta}+\cdots\right\}\tag{7.14}$$

かくて比推定の期望値は母数 $z=\frac{E\xi}{E\eta}$ に一致せず、従って比推定は偏りある推定である。しかし抽出単位数が十分に大なる時は、標本推定値はその期望値の近くに分布し、従って期望値よりの偏差 $\varDelta\xi$、$\varDelta\eta$ は $E\xi$、$E\eta$ に比して極めて小であるため、$\frac{\varDelta\xi}{E\xi}$、$\frac{\varDelta\eta}{E\eta}$ に関する二次以上の項の期望値は、$\frac{E\xi}{E\eta}$ に比して殆ど問題にするに足らない位小である。故に

$$E(z')\doteqdot\frac{E\xi}{E\eta}$$

と看做す事が出来る。かくて比推定の偏りは極めて小であり無視して差支えないのである。

次に比推定の分散度を知るため $M.S.E.$ を求めよう。z' の標本誤差 $\left(z'-\frac{E\xi}{E\eta}\right)$ は (7・13) より

$$z'-\frac{E\xi}{E\eta}=\frac{E\xi}{E\eta}\left\{\frac{\varDelta\xi}{E\xi}-\frac{\varDelta\eta}{E\eta}+\left(\frac{\varDelta\eta}{E\eta}\right)^2-\frac{\varDelta\xi\cdot\varDelta\eta}{E\xi\cdot E\eta}+\cdots\right\}$$

であるため

$$M.S.E.(z')=E\left[z'-\frac{E\xi}{E\eta}\right]^2$$

$$=\left(\frac{E\xi}{E\eta}\right)^2E\left\{\frac{\varDelta\xi}{E\xi}-\frac{\varDelta\eta}{E\eta}+\left(\frac{\varDelta\eta}{E\eta}\right)^2-\frac{\varDelta\xi\cdot\varDelta\eta}{E\xi\cdot E\eta}+\cdots\right\}^2$$

$$=\left(\frac{E\xi}{E\eta}\right)^2\left\{\frac{E(\varDelta\xi)^2}{(E\xi)^2}+\frac{E(\varDelta\eta)^2}{(E\eta)^2}-2\frac{E(\varDelta\xi\varDelta\eta)}{E\xi\cdot E\eta}-\cdots\right\}$$

この式の右辺の省略した部分は $\dfrac{\Delta\xi}{E\xi}$、$\dfrac{\Delta\eta}{E\eta}$ に関して三次以上の項の期望値であり、その値は僅少であって無視する事が出来る。そして

$$E(\Delta\xi)^2=E[\xi-E\xi]^2=V(\xi)$$

$$E(\Delta\eta)^2=E[\eta-E\eta]^2=V(\eta)$$

$$E(\Delta\xi\cdot\Delta\eta)=E[(\xi-E\xi)(\eta-E\eta)]=Cov(\xi,\eta)$$

であるため

$$M.S.E.(z')\doteqdot\left(\frac{E\xi}{E\eta}\right)^2\left\{\frac{V(\xi)}{(E\xi)^2}+\frac{V(\eta)}{(E\eta)^2}-2\frac{Cov(\xi,\eta)}{E\xi\cdot E\eta}\right\}\qquad(7.15)$$

となる。

そして比推定は偏りある推定値であるため

$$M.S.E.(z')=V(z')+(Ez'-z)^2\qquad[(7.12)]$$

しかるに右辺第二項は(7·14)より

$$(Ez'-z)^2=\left(\frac{E\xi}{E\eta}\right)^2\left\{\frac{E(\Delta\eta)^2}{(E\eta)^2}-\frac{E(\Delta\xi\cdot\Delta\eta)}{E\xi\cdot E\eta}+\cdots\right\}^2$$

であり、これはすべて $\dfrac{\Delta\xi}{E\xi}$、$\dfrac{\Delta\eta}{E\eta}$ に関して四次以上の項の期望値であるから、その値は極めて小さく

$$(Ez'-z)^2\doteqdot 0$$

と看做す事が出来る。従って

$$M.S.E.(z')\doteqdot V(z')\qquad(7.16)$$

であり、M.S.E. と分散とはその値が略々等しいのである。かくて比推定の標本誤差の分散度の規定に於ては不偏推定と同様に分散を以てすればよいのである。故に以下比推定に就ても分散を考えて行く事にする。

(3) 単純抽出法に於ける比推定

(i) 量的標識に関する比の推定

N 個の単位より成る母集団から n 個の単位を抽出して構成した標本によって、母集団に於ける二種の平均値（又は量的大いさ）の比 $z=\dfrac{\bar{x}}{\bar{y}}\left(=\dfrac{X}{Y}\right)$ を推定するものとする。$\bar{x}$ 及び $\bar{y}$ の不偏推定値は

$$\bar{x}'=\frac{1}{n}\sum_{j=1}^{n}x'_j,\quad \bar{y}'=\frac{1}{n}\sum_{j=1}^{n}y'_j$$

であるから、z の推定値 z' は

$$z'=\frac{\bar{x}'}{\bar{y}'}=\frac{\sum_{j=1}^{n}x'_j}{\sum_{j=1}^{n}y'_j}\qquad(7.17)$$

である。この比推定の分散は(7·16)(7·15)より

$$V(z')\doteqdot\left(\frac{E\bar{x}'}{E\bar{y}'}\right)^2\left\{\frac{V(\bar{x}')}{(E\bar{x}')^2}+\frac{V(\bar{y}')}{(E\bar{y}')^2}-2\frac{Cov(\bar{x}',\bar{y}')}{E\bar{x}'\cdot E\bar{y}'}\right\}$$

しかるに　$E(\bar{x}')=\bar{x},\ E(\bar{y}')=\bar{y}$　[(2.31)]

$$V(\bar{x}')=\frac{N-n}{N-1}\frac{\sigma_x^2}{n},\quad V(\bar{y}')=\frac{N-n}{N-1}\frac{\sigma_y^2}{n}\qquad[(2.33)]$$

$$Cov\,(\bar{x}',\ \bar{y}')=\frac{N-n}{N-1}\frac{Cov(x,\ y)}{n} \qquad [(7.10)]$$

$$\therefore\quad V(z')\doteqdot\left(\frac{\bar{x}}{\bar{y}}\right)^2\frac{N-n}{N-1}\frac{1}{n}\left\{\frac{\sigma^2{}_x}{\bar{x}^2}+\frac{\sigma^2{}_y}{\bar{y}^2}-2\frac{Cov(x,\ y)}{\bar{x}\bar{y}}\right\} \qquad (7.18)$$

母集団に於けるx及びyの変異係数を$\dfrac{\sigma_x}{\bar{x}}=c_x$, $\dfrac{\sigma_y}{\bar{y}}=c_y$ とすると、(19)より $Cov(x,\ y)=\rho_{xy}\sigma_x\sigma_y$ であるから

$$=z^2\frac{N-n}{N-1}\frac{1}{n}\left\{c^2{}_x+c^2{}_y-2\,\rho_{xy}c_xc_y\right\} \qquad (7.19)$$

となる。

母共変量の推定

(7・18)式によって此推定の誤差分散を計算するためには、母集団の値$\bar{x}$、$\bar{y}$、$\sigma^2{}_x$、$\sigma^2{}_y$及び $Cov(x,\ y)$ を知らねばならず、それ等を標本より推定する事が必要となるのである。しかるに母平均の不偏推定値は標本平均値であり母分散の不偏推定値は(2・47)によって与えられるため、茲では母共変量の不偏推定値を知ればよいのである。（求め方は母分散の不偏推定(2・47)の求め方と同じであるためそれを参照されたい。）

今標本に於ける共変量を $cov(x,y)$ とすると

$$cov(x,y)=\frac{1}{n}\sum_{j=1}^{n}(x'_j-\bar{x}')(y'_j-\bar{y}')$$

である。これの期望値を求めると

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

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

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

しかるに $\sum_{j=1}^{n}(y'_j-\bar{y})=n(\bar{y}'-\bar{y})$, $\sum_{j=1}^{n}(x'_j-\bar{x})=n(\bar{x}'-\bar{x})$ であるから

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

この式の括弧内第1項の期望値は、母集団より1個の単位を抽出した時得られるxとyの値の標本分布の共変量であるから、

$$E(x'_j-\bar{x})(y'_j-\bar{y})=\frac{1}{N}\sum_{j=1}^{N}(x_j-\bar{x})(y_j-\bar{y})\equiv Cov(x,y)$$

であり、又第二項の期望値は標本平均分布の共変量であるから、(7・10)より

$$E(\bar{x}'-\bar{x})(\bar{y}'-\bar{y})=\frac{N-n}{N-1}\frac{1}{n}Cov\,(x,\ y)$$

である。故に

$$E(cov(x,\ y))=\frac{1}{n}\left[n\,Cov(x,y)-n\frac{N-n}{N-1}\frac{1}{n}Cov(x,y)\right]$$

$$=\frac{N}{N-1}\frac{n-1}{n}Cov(x,\ y)$$

$$\therefore\quad E\left(\frac{N-1}{N}\frac{n}{n-1}cov(x,y)\right)=Cov(x,y) \qquad (7.20)$$

となるから、標本に於ける共変量の$\dfrac{N-1}{N}\dfrac{n}{n-1}$倍が母共変量の不推定値である。

かくて(7・18)式に必要な母数の不偏推定値

$$\bar{x}\approx\bar{x}',\quad \bar{y}\approx\bar{y}'$$

$$\sigma^2_x\approx\frac{N-1}{N}\frac{n}{n-1}s^2_x,\quad \sigma^2_y\approx\frac{N-1}{N}\frac{n}{n-1}s^2_y$$

$$Cov(x,y)\approx\frac{N-1}{N}\frac{n}{n-1}cov(x,y)$$

を(7・18)に代入して得られる

$$V=\left(\frac{\bar{x}'}{\bar{y}'}\right)^2\left(1-\frac{n}{N}\right)\frac{1}{n-1}\left\{\frac{s^2_x}{\bar{x}'^2}+\frac{s^2_y}{\bar{y}'^2}-2\frac{cov(x,y)}{\bar{x}'\bar{y}'}\right\} \tag{7.21}$$

はz'の分散の推定値である。但しこれは不偏推定でない事は、右辺の各項が比推定である事より明らかであろう。

(ii) 比推定による量的大いさ、平均値の推定

母集団の比$z=\dfrac{X}{Y}$より　$X=zY$である。従って変量x、yの間に高い正の相関々係があり、yに関する量的大いさYの正確な値が判っている時は、それにzの推定値z'を乗ずる事によって、xに関する量的大いさXを推定する事が出来る。この比推定を用いるXの推定値をX''で表すと

$$X''=z'Y \tag{7.22}$$

である。勿論これは不偏推定ではないがその偏りは僅少である。分散は

$$V(X'')=Y^2V(z')$$

$$\doteqdot Y^2z^2\frac{N-n}{N-1}\frac{1}{n}\left\{\cdots\cdots\right\} \quad [(7.18)]$$

$$=X^2\frac{N-n}{N-1}\frac{1}{n}\left\{\frac{\sigma^2_x}{\bar{x}^2}+\frac{\sigma^2_y}{\bar{y}^2}-2\frac{Cov(x,y)}{\bar{x}\bar{y}}\right\} \tag{7.23}$$

である。

又母集団の比　$z=\dfrac{\bar{x}}{\bar{y}}$　より$\bar{x}=z\bar{y}$である。故に$\bar{y}$の正確な値が判っている時は、それにzの推定値z'を掛ける事によってxの平均値$\bar{x}$を推定する事が出来る。この比推定による$\bar{x}$の推定値を$\bar{x}''$で表すと

$$\bar{x}''=z'\bar{y} \tag{7.24}$$

であり、これも偏りある推定値である。分散は

$$V(\bar{x}'')=\bar{y}^2V(z')$$

$$\doteqdot\bar{y}^2\left(\frac{\bar{x}}{\bar{y}}\right)^2\frac{N-n}{N-1}\frac{1}{n}\left\{\cdots\cdots\right\} \quad [(7.18)]$$

$$=\bar{x}^2\frac{N-n}{N-1}\frac{1}{n}\left\{\frac{\sigma^2_x}{\bar{x}^2}+\frac{\sigma^2_y}{\bar{y}^2}-2\frac{Cov(x,y)}{\bar{x}\bar{y}}\right\} \tag{7.25}$$

である。

(iii) 質的標識に関するの比の推定

以上は量的標識に関する比の推定の場合であった。次に質的標識に関する比の推定を考察しよう。今α標識に変量xを、β標識に変量yを対応させ、α標識を有する単位には$x=1$　有しない単位には$x=0$　を与え、又β標識を持つ単位には$y=1$　持たない単位には　$y=0$　を与えると、xに就ての量的

大いさはα標識の単位数N_1を、平均値はその割合p_1を表し、yに就ての量的大いさはβ標識の単位数N_2を、平均値はその割合p_2を表す事となり、質的標識に関する比推定は量的標識の比推定に帰着するのである。

この場合 $\bar{x}=p_1,\ \bar{y}=p_2,\ \sigma^2_x=p_1(1-p_1),\ \sigma^2_y=p_2(1-p_2)$ なる事は既に明らかであるので、茲では共変量が如何になるかを調べて置こう。

母共変量

$$Cov(x,y)=\frac{1}{N}\sum_{j=1}^{N}x_jy_j-\bar{x}\bar{y} \qquad [(20')]$$

は、α標識とβ標識との関係が排反的であるか包括的であるかによって、その値が異るのである。排反的なる場合はα標識とβ標識とを併有する単位はなく、すべての単位は何れか一方のみを有し、α標識の単位の持つ変量(x_j,y_j)は$(1,0)$、β標識の単位のそれは$(0,1)$であるため$x_jy_j=0$であり、故に

$$Cov(x,y)=-p_1p_2 \qquad (7.26)$$

である。しかるにβ標識がα標識を包括する場合は、β標識のみを有する単位とα標識、β標識を併有する単位とがあり、前者は$(0,1)$後者は$(1,1)$の変量を有するため、α標識、β標識を併有する単位に於てのみ $x_jy_j=1$であり、その他の単位に於ては$x_jy_j=0$であるから、$\sum_{j=1}^{N}x_jy_j=N_1$であり、故に

$$Cov(x,y)=p_1-p_1p_2 \qquad (7.27)$$

である。

さて母集団に於けるα標識とβ標識の割合（又は単位数）の比 $z=\frac{p_1}{p_2}\left(=\frac{N_1}{N_2}\right)$の推定値$z'$は

$$\bar{x}'=\frac{1}{n}\sum_{j=1}^{n}x'_j=p_1',\quad \bar{y}'=\frac{1}{n}\sum_{j=1}^{n}y'_j=p_2'$$

であるから

$$z'=\frac{p_1'}{p_2'}=\frac{\sum_{j=1}^{n}x'_j}{\sum_{j=1}^{n}y'_j} \qquad (7.28)$$

である。この比推定の分散は、α標識とβ標識とが排反的なる場合は(7・18)(7・26)より

$$V(z')\fallingdotseq\left(\frac{p_1}{p_2}\right)^2\frac{N-n}{N-1}\frac{1}{n}\left\{\frac{p_1(1-p_1)}{p_1^2}+\frac{p_2(1-p_2)}{p_2^2}+2\frac{p_1p_2}{p_1p_2}\right\}$$
$$=z^2\frac{N-n}{N-1}\frac{1}{n}\left(\frac{1}{p_1}+\frac{1}{p_2}\right) \qquad (7.29)$$

であり、β標識がα標識を包括する時は(7・18)(7・27)より

$$V(z')\fallingdotseq\left(\frac{p_1}{p_2}\right)^2\frac{N-n}{N-1}\frac{1}{n}\left\{\frac{p_1(1-p_1)}{p_1^2}+\frac{p_2(1-p_2)}{p_2^2}-2\frac{p_1-p_1p_2}{p_1p_2}\right\}$$
$$=z^2\frac{N-n}{N-1}\frac{1}{n}\left(\frac{1}{p_1}-\frac{1}{p_2}\right) \qquad (7.30)$$

である。

これ等の分散の計算に必要な母数はp_1及びp_2であるため、標本より求めたそれ等の不偏推定値p_1'、p_2'を代入する事により、$V(z')$を推定する事が出来るのである。

尚母集団の比 $z=\dfrac{N_1}{N_2}$ より $N_1=zN_2$ であるため、β標識の単位数N_2の正確な値が判っている時は、それにzの推定値z'を乗じて

$$N_1''=z'N_2 \tag{7.31}$$

によってα標識の単位数N_1を推定する事が出来る。この推定式の分散は、(7·29)及び(7·30)式より

$$V(N_1'')=N_2^2V(z')$$

$$\doteqdot N_1^2\frac{N-n}{N-1}\,\frac{1}{n}\left(\frac{1}{p_1}+\frac{1}{p_2}\right) \tag{7.32}$$

又は $$\doteqdot N_1^2\frac{N-n}{N-1}\,\frac{1}{n}\left(\frac{1}{p_1}-\frac{1}{p_2}\right) \tag{7.33}$$

である。(7·32)式はα標識とβ標識とが排反的な場合であり、(7·33)式はβ標識がα標識を包括する場合である。

又母集団の比 $z=\dfrac{p_1}{p_2}$ より $p_1=zp_2$ であるため、β標識の割合p_2を正確に知り得る時は、それにzの推定値z'を乗じて

$$p_1''=z'p_2 \tag{7.34}$$

によってα標識の割合p_1を推定する事が出来る。その場合の分散は、両標識が排反的な場合は(7·29)式より

$$V(p_1'')=p_2^2V(z')\doteqdot p_1^2\frac{N-n}{N-1}\,\frac{1}{n}\left(\frac{1}{p_1}+\frac{1}{p_2}\right) \tag{7.35}$$

β標識がα標識を包括する時は(7·30)式より

$$\doteqdot p_1^2\frac{N-n}{N-1}\,\frac{1}{n}\left(\frac{1}{p_1}-\frac{1}{p_2}\right) \tag{7.36}$$

である。

(iv) 比推定の精度

(7·22)又は(7·34)式により比推定を用いて母平均又は母比例数を推定する事が出来る。先ず母平均$\bar{x}$を不偏推定によって推定する場合と、比推定によって推定する場合との分散を比較して、比推定の精度を調べて見よう。

$\bar{x}$の不偏推定値の分散をV_1、比推定の分散をV_2で表すと、それぞれは(2·33)(7·25)より

$$V_1=\frac{N-n}{N-1}\,\frac{\sigma^2_x}{n}$$

$$V_2\doteqdot\bar{x}^2\frac{N-n}{N-1}\,\frac{1}{n}\left\{\frac{\sigma_x^2}{\bar{x}^2}+\frac{\sigma_y^2}{\bar{y}^2}-2\frac{Cov(x,y)}{\bar{x}\bar{y}}\right\}$$

である。そこでV_2のV_1に対する比をとると

$$\frac{V_2}{V_1}\doteqdot\frac{\bar{x}^2}{\sigma_x^2}\left\{\frac{\sigma_x^2}{\bar{x}^2}+\frac{\sigma_y^2}{\bar{y}^2}-2\rho_{xy}\frac{\sigma_x}{\bar{x}}\,\frac{\sigma_y}{\bar{y}}\right\} \tag*{[(19)]}$$

$$=1+\frac{1}{c_x^2}\left(c_y^2-2\rho_{xy}c_xc_y\right)$$

$$=1+\frac{c_y}{c_x}\left(\frac{c_y}{c_x}-2\rho_{xy}\right) \tag{7.37}$$

故に

$$\frac{c_y}{c_x}\left(\frac{c_y}{c_x}-2\rho_{x'y}\right)\gtreqless 0 \quad \text{ならば} \quad V_2 \gtreqless V_1$$

即ち

$$\left.\begin{array}{ccc} \frac{1}{2}\frac{c_y}{c_x} & > & \rho_{x'y} \\ 〃 & = & 〃 \\ 〃 & < & 〃 \end{array}\right\} \text{ならば} \begin{cases} V_2>V_1 & \text{(i)} \\ 〃=〃 & \text{(ii)} \\ 〃<〃 & \text{(iii)} \end{cases}$$

となる。xとyとの間に正の高度の相関々係がある時は(iii)が成立し得、従って比推定の精度の方が不偏推定の精度よりも高くなるのであるが、xとyの相関々係が低い時は(i)が成立し不偏推定の方が精度は高いのである。かくて平均を知らんとする変量xと高い正相関を有する他の変量yがあり、yの平均値が判っている時は、不偏推定によるよりも比推定による方が標本推定の精度は高いのである。

次にa標識の母比例数p_1を推定する場合に就て、不偏推定と比推定の精度を比較しよう。p_1の不偏推定値の分散をV_1とすると、(2・38)より

$$V_1=\frac{N-n}{N-1}\frac{1}{n}p_1(1-p_1)$$

であり、比推定の分散をV_2とすると、(7・35)(7・36)より

$$V_2 \fallingdotseq p_1{}^2\frac{N-n}{N-1}\frac{1}{n}\left(\frac{1}{p_1}+\frac{1}{p_2}\right)$$

$$\text{又は} \quad \fallingdotseq p_1{}^2\frac{N-n}{N-1}\frac{1}{n}\left(\frac{1}{p_1}-\frac{1}{p_2}\right)$$

である。そこでV_2のV_1に対する比をとると、先ずa標識とβ標識とが排反的である場合は

$$\frac{V_2}{V_1} \fallingdotseq \frac{p^2{}_1\left(\frac{1}{p_1}+\frac{1}{p_2}\right)}{p_1(1-p_1)}=\frac{p_2+p_1}{p_2(1-p_1)} \qquad (7.38)$$

であり、従って

$$\frac{p_2+p_1}{p_2(1-p_1)}\gtreqless 1 \quad \text{なる時に} \quad V_2 \gtreqless V_1$$

となる。この条件式は

$$p_1(1+p_2)\gtreqless 0$$

であるが、その左辺は正であり$p_1\neq 0$なる限り0とならないために$V_2>V_1$であり、従って不偏推定の方が精度は高いのである。次にβ標識がa標識を包括する時は

$$\frac{V_2}{V_1} \fallingdotseq \frac{p_1{}^2\left(\frac{1}{p_1}-\frac{1}{p_2}\right)}{p_1(1-p_1)}=\frac{p_2-p_1}{p_2(1-p_1)} \qquad (7.39)$$

であり、故に

$$\frac{p_2-p_1}{p_2(1-p_1)}\gtreqless 1 \quad \text{なる時に} \quad V_2 \gtreqless V_1$$

となる。この条件式は

$$p_1(1-p_2)\lesseqgtr 0$$

であるが、その左辺は正であり$p_1\neq 0$なる限り0とならないために$V_2<V_1$であり、従って比推定の方が精度は高いのである。

かくてa標識の割合を推定せんとする場合、それを包括する関係にある他の標識βがありその割合が判っている時は、

比推定による方が標本推定値の精度は高いのである。しかしα標識とβ標識とが排反関係にある時は、不偏推定による方が推定値の精度は高いのである。

(v) 母比例数を用いる比の推定

尚母集団の比 $z=\dfrac{p_1}{p_2}$ を推定する場合、時には分子又は分母の正確な値が判っている事があり、その時は値の判らない何れか一方の割合のみを標本より推定して比をとり、母集団の比の推定値とする事が考えられる。

例えば分子が判っている時は、p_2を標本より推定し

$$z'=\frac{p_1}{p_2'} \tag{7.40}$$

を以てこの推定値とするのであり、（これは偏りある推定値である。しかしその偏りは小さい）その時の分散は

$$V(z')=p_1^2V\left(\frac{1}{p_2'}\right)\doteqdot p_1^2\frac{N-n}{N-1}\frac{1}{n}\frac{(1-p_2)}{p_2^3} \tag{7.41}$$

である。

又分母p_2が判っている時は、p_1を標本より推定して

$$z'=\frac{p_1'}{p_2} \tag{7.42}$$

によりzを推定するのであり、（但しこれは不偏推定である）この場合の分散は

$$V(z')=\frac{1}{p_2^2}V(p_1')=\frac{1}{p_2^2}\frac{N-n}{N-1}\frac{1}{n}p_1(1-p_1)\quad[(2.38)] \tag{7.43}$$

である。

今これ等の母比例数を利用して母集団の比zを推定する場合と、分子、分母共に標本より推定してzを推定する場合との精度を比較すると

(1) 分子p_1が判つている時は

(a) α標識とβ標識が排反的な場合はp_1の値を利用して$\dfrac{p_1}{p_2'}$による方が精度が高く、

(b) β標識がα標識を包括する時は$p_1<\dfrac{p_2}{2-p_2}$又は$p_2>\dfrac{2p_1}{1+p_1}$ならば$\dfrac{p_1}{p_2'}$によるべく、そうでない時はp_1 p_2共に標本より推定して$\dfrac{p_1'}{p_2'}$による方が精度は高い。

(2) 分母p_2が判っている時は

(c) 排反的な場合はp_2を用い$\dfrac{p_1'}{p_2}$による方が精度は高いが、

(d) 包括的な場合はp_1 p_2共に標本より推定して$\dfrac{p_1'}{p_2'}$による方が精度は高い。

のである。（但し証明は略する）

1) (7·41)式の証明

(7·40)式は分母のみが確率変数であって分子は常数である。そして一般に確率変数$1/\eta$は、比推定ξ/ηに於て$\xi=1$と置

いたものであり、従って$1/\eta$の分散は(7·15)式より

$$V\left(\frac{1}{\eta}\right) \doteqdot \left(\frac{E1}{E\eta}\right)^2 \left\{\frac{V(1)}{(E1)^2}+\frac{V(\eta)}{(E\eta)^2}-2\frac{Cov(1,\eta)}{E1\cdot E\eta}\right\}$$

$$E(1)=1,\quad V(1)=E(1-1)^2=0$$

$$Cov(1,\ \eta)=E(1-1)(\eta-E\eta)=0$$

$$\therefore\quad =\frac{1}{(E\eta)^2}\,\frac{V(\eta)}{(E\eta)^2}$$

従って

$$V\left(\frac{1}{p_2'}\right) \doteqdot \frac{1}{p_2^4}\,\frac{N-n}{N-1}\,\frac{1}{n}\,p_2(1-p_2)$$

かくて(7·41)式が得られるのである。