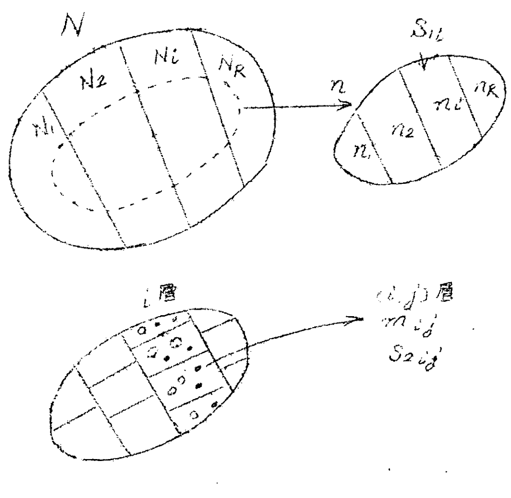
# ⑮ 二回抽出法について

青　山　　博　次　郎

標本抽出法として応用の広いものに二回抽出法がある。[1] これはNeymanが考えたものであつて，一度母集団（大きさをNとする，通常無限母集団と考える）から $n$ 個の標本を抽出しこれを調査すべき標識と関係の深い照準によつて層別して更に $m$ 個の標本を抽出し，この二通目の標本を調査する方法である。

今これを初めの抽出に際しても層別を行う場合にはどうなる[2]，きその際に必要な諸条件はどうなるかについて考えてみよう。

通常の調査では種々の標識について調査するためと，集計，分析を容易にするため比例抽出法がよく用いられる。それ故先づこの場合から考えてみることにする。




第一次層は R 個とし，第一次標本を $S_{1i}$ $(i=1,2,\cdots,R)$ とする。この $S_{1i}$ を他の照準により $L_i$ 個の第二次層に分ける。

第一次層の各々の大きさを $N_i$ とし，比例抽出法にて大きさ $n_i$ なる第一次標本 $S_{1i}$ を抽出する。次に第二次層 $(i,j)$ に属する $n_{ij}$ 個の単位より比例抽出法に

よつて $m_{ij}$ 個の標本を抽出し，この第二次標本を $S_{2ij}$ ($i=1,2,\cdots,R$; $j=1,2,\cdots,L_i$) としよう。
なおこゝで

$$m = \sum_{i=1}^{R} m_i = \sum_{i=1}^{R}\sum_{j=1}^{L_i} m_{ij} \qquad (1)$$

とする。
このとき或る標識 $x$ の母平均 $\overline{X}$ の不偏推定値 $\overline{x}$ は

$$\overline{x} = \sum_{i=1}^{R} P_i \overline{x}_i = \sum_{i=1}^{R} P_i \sum_{j=1}^{L_i} r_{ij} \overline{x}_{ij} = \sum_{i=1}^{R} P_i \sum_{j=1}^{L_i} r_{ij} \frac{1}{m_{ij}} \sum_{K=1}^{m_{ij}} x_{ijK} \qquad (2)$$

とおける。　こゝに

$$P_i = \frac{N_i}{N}, \qquad r_{ij} = \frac{n_{ij}}{n_i}$$

$N$ は母集団の大きさ，$N_i$ は第 $i$ 層の母集団の大きさ，$n_i$ は第 $i$ 層の $S_{1i}$ に属する標本の大きさ，$n_{ij}$ は第 ($i,j$) 層の $S_{2ij}$ に属する標本の大きさとする。

$r_{ij}$ と $x_{ijK}$ は独立と仮定し，有限母集団修正を無視して期望値を作れば

$$\begin{aligned} E(\overline{x}) &= \sum_i P_i \sum_j E(r_{ij}) E(\overline{x}_{ij}) \\ &= \sum_i \frac{N_i}{N} \cdot \sum_j \frac{N_{ij}}{N_i} \frac{1}{N_{ij}} \sum_{K=1}^{N_{ij}} X_{ijK} \\ &= \overline{X} \end{aligned} \qquad (3)$$

また分散は Neyman の方法[1]に従つて簡単な計算を行えば

$$V(\bar{x}_i) = \frac{\sigma_i^2}{m_i} - \left(\frac{1}{m_i} - \frac{1}{n_i}\right)\sum_j P_{ij}\,(\bar{X}_i - \bar{X}_{ij})^2 \tag{4}$$

たゞし

$$E\,(r_{ij}) = P_{ij} = \frac{N_{ij}}{N_i} \tag{5}$$

従つて

$$V(\bar{x}) = \sum_i P_i^2 V(\bar{x}_i)$$

$$= \sum_i \frac{P_i^2\sigma_i^2}{m_i} - \sum_i P_i^2\left(\frac{1}{m_i} - \frac{1}{n_i}\right)\sum_j P_{ij}(\bar{X}_i - \bar{X}_{ij})^2$$

比例抽出法を用いたので

$$nP_i = n_i,\quad mP_i = m_i$$

となるから

$$V(\bar{x}) = \frac{\sigma^2}{m} - \frac{1}{m}\left\{\sum_i P_i(\bar{X}_i - \bar{X})^2 + \left(1 - \frac{m}{n}\right)\sum_i P_i\sigma_{bi}^2\right\} \tag{6}$$

こゝで

$$\left.\begin{aligned}
\sigma^2 &= \frac{1}{N}\sum_i\sum_j\sum_K (X_{ijK} - \bar{X})^2 \\
\sigma_{bi}^2 &= \frac{1}{N_i}\sum_j N_{ij}(\bar{X}_i - \bar{X}_{ij})^2 \\
\bar{X}_i &= \frac{1}{N_i}\sum_j X_{ij} = \frac{1}{N_i}\sum_j\sum_K X_{ijK} \\
\bar{X}_{ij} &= \frac{1}{N_{ij}}\sum_K X_{ijK}
\end{aligned}\right\} \tag{7}$$

今第一次抽出における一単位当りの費用を $C_1$，第二次抽出

における一単位当りの費用を $C_2$，総予算を $C$ とすれば

$$C = C_1 n + C_2 m \tag{8}$$

$C$ を一定として $V(\bar{x})$ が最小となる如く $m, n$ をえらぶときは

$$n^2 = \frac{\frac{C_2}{C_1} b}{\sigma^2 - a - b} m^2 \tag{9}$$

但し

$$\left.\begin{aligned} a &= \sum_i P_i (\bar{X}_i - \bar{X})^2 \\ b &= \sum_i P_i \sigma_{bi}^2 \end{aligned}\right\} \tag{10}$$

故に

$$C_1 (\sigma^2 - a - b) < C_2 b \tag{11}$$

が成立するときは，(8)，(9) を満足する如く $n$，$m$ をえらび比例抽出を行えばよい。

もし第一次抽出は比例抽出法，第二次抽出は Neyman 法に従うときは

$$V(\bar{x}_i) = \frac{1}{m_i}\Big(\sum_j P_{ij}\sigma_{ij}\Big)^2 + \sum_j m_{ij}\left(\frac{P_{ij}\sigma_{ij}}{m_{ij}} - \frac{\sum_j P_{ij}\sigma_{ij}}{m_i}\right)^2$$
$$+ \frac{1}{n_i}\sum_j P_{ij}(\bar{X}_{ij} - \bar{X}_i)^2 \tag{12}$$

であって，$m_i$，$n_i$ が定まれば $m_{ij}$ は上式の第二項を 0 とすれば $V(\bar{x}_i)$ は最小となるから

$$m_{ij} = \frac{m_i P_{ij} \sigma_{ij}}{\sum_j P_{ij} \sigma_{ij}} \tag{13}$$

また

$$n_i = n P_i \tag{14}$$

従つて

$$V(\bar{x}) = \sum_i \frac{P_i^2 a_i}{m_i} + \frac{b}{n} \tag{15}$$

こゝに

$$\sqrt{a_i} = \sum_j P_{ij} \sigma_{ij} \tag{16}$$

このときは費用一定，分散最小とする如き $m_i$，$n$ を求めると

$$C_1 a_i < C_2 b \tag{17}$$

なる條件の下に

$$m_i = \sqrt{\frac{C_1}{C_2 b}} \sqrt{a_i}\, n_i \tag{18}$$

$$n = \frac{C}{C_1 + \sqrt{\frac{C_1 C_2}{b}} \sum_i \sqrt{a_i} P_i} \tag{19}$$

となる。

またもし第一次，第二次共に比例抽出を用いず，Neyman 法に従えば

$$V(\bar{x}) = \sum_i \frac{P_i^2 a_i}{m_i} + \sum_i \frac{P_i^2 \sigma_{bi}^2}{n_i} + \sum_i P_i^2 \sum_j m_{ij} \left( \frac{P_{ij}\sigma_{ij}}{m_{ij}} - \frac{\sum_j P_{ij}\sigma_{ij}}{m_i} \right)^2 \tag{20}$$

このときも費用一定の下に分散を最小にする如き $n$, $n_i$, $m$, $m_{ij}$ は

$$C_1 a_i < C_2 \sigma_{bi}^2 \tag{21}$$

なる條件の下に

$$\left.\begin{aligned} m_{ij} &= \frac{m_i P_{ij} \sigma_{ij}}{\sum_j P_{ij}\sigma_{ij}} \\ m_i &= \sqrt{\frac{C_1}{C_2}} \frac{\sqrt{a_i}}{\sigma_{bi}} n_i \\ n_i &= \frac{n P_i \sigma_{bi}}{\sum_i P_i \sigma_{bi}} \\ n &= \frac{C \sum P_i \sigma_{bi}}{\sum_i C_1 P_i \left( \sigma_{bi} + \sqrt{\frac{C_2}{C_1}} \sqrt{a_i} \right)} \end{aligned}\right\} \tag{22}$$

となる。

以上の結果を簡単な例題に用いて相対的効果を比べてみよう。

例. $C_1 = 1$, $C_2 = 4$, $C = 500$, $P_1 = P_{i1} = \frac{1}{4}$

$P_2 = P_{i2} = \frac{2}{4}$, $P_3 = P_{i3} = \frac{1}{4}$,

$\bar{X}_1 = 1$, $\bar{X}_2 = 3$, $\bar{X}_3 = 6$, $\sigma_1 = 1$, $\sigma_2 = 2$

$\sigma_3 = 3$, $\sigma_{b1}^2 = 0.25$, $\sigma_{b2}^2 = 1$, $\sigma_{b3}^2 = 4$

$\sigma_{11}^2 = 0.5$, $\sigma_{12}^2 = 0.5$, $\sigma_{13}^2 = 1.5$, $\sigma_{21}^2 = 2$

$\sigma_{22}^2 = 3$, $\sigma_{23}^2 = 4$, $\sigma_{31}^2 = 4$, $\sigma_{32}^2 = 5$,

$\sigma_{33}^2 = 6$ とすると條件はすべて満足されていて

第一，二次抽出共に比例抽出法のときは

$$V_{rr}(\bar{x}) = 0.04358 \quad (n = 134,\ m = 92)$$

第一次抽出のみ比例抽出法のときは

$$V_{rn}(\bar{x}) = 0.04069 \quad (n = 139,\ m = 90)$$

第一，二次抽出共に Neyman 法によるときは

$$V_{nn}(\bar{x}) = 0.03833 \quad (n = 128,\ m = 93)$$

最初から random に抽出するときは

$$V(\bar{x}) = 0.0615 \quad (m = 125)$$

従って相対的効果は

$V(\bar{x})/V_{rr}(\bar{x}) = 1.41$, $V(\bar{x})/V_{rn}(\bar{x}) = 1.50$

$V(\bar{x})/V_{nn}(\bar{x}) = 1.60$ となる。

（註） 1) J. Neyman: Contribution to the Theory of Sampling Human Population J.A.S.A. 1938

2) 林知己夫： 白河に於ける言語調査，統計数理研究輯報 No. 2.