二重抽出法の或る場合

河野, 和正

https://doi.org/10.15017/12719

出版情報：統計科学研究. 1 (2), pp.7-14, 1956-05. Research Association of Statistical Sciences
バージョン：
権利関係：

寄　　書

二重抽出法の或る場合

河　野　和　正

J. Neyman [1] の二重抽出法について，次の様な質問を受けた。──「彼の方法は予備調査のための抽出と本調査のための抽出が独立に行われることに基いているのだが，もしこのとき予備調査の対象となった調査単位の中から本調査のための抽出を行った場合，精度にどのような影響があるか」というのである。本稿の目的はこの抽出操作の相違が及ぼす精度への影響を示すことである．

まず始めに，この抽出操作の構造と記号の整理から始める．既知の大きさMをもつ有限母集団$\Pi(M)$は補助変量yによって区別された未知の大きさM_i ($i=1,2,\cdots,L$)($\sum_1^L M_i=M$) をもつL個の層から成っている．この母集団は直接の調査対象となっている変量xと補助変量yを成分とするヴエクトル$Z=(x,y)$ の集合と考えられる。予備調査の目標はまず未知のM_i又は$\overline{W}_i={M_i}/{M}$ についての情報を得ることであり，次に本調査のためのリストを作ることである．後者の目標が先のJ. Neyman の方法と異っており，前者の点では一致する．今予備調査が$\Pi(M)$からの大きさNのランダムな標本O_Nを抽出することによって行われるとする。この抽出によって，抽出単位としては，$\dot{Z}_1,\dot{Z}_2,\cdots,\dot{Z}_N$ が選ばれるのであるが，データとしては，その第二の成分である$y_1,y_2,\cdots,y_N$ のみが判る。従って Z_α ($\alpha=1,2,\cdots,N$)は確定値ではあるが未知の第一成分ξ_αと既知のy_αによって$Z_\alpha=(\xi_\alpha,y_\alpha)$ で表わされよう。このZ_αは，本調査のための二次的母集団$\pi(\dot{N})$を形づくり，既知となったy_αによって，L組の層に分けられる．この二次的層を$\pi(\dot{N}_i)$で表わす．この際，$\pi(\dot{N}_i)$の定義函数C_iを導入すると，y_αによって層別することは，$C_i(Z_\alpha)=1$なるときそのZ_αをi層へということである．このようにすると，予備調査の最初の目標は

$$(1)\qquad \sum_{\alpha=1}^{N} C_i(Z_\alpha)=N_i, \quad \sum_1^L N_i=N$$

$$i=1,2,\cdots,L$$

なる n_i によって，構成比の推定

(2) $$w_i = N_i / N$$

を利用することである．云いかえると，この目標に対しては事象 $\{n_i = a_i ; (i=1, 2, \cdots, L)\}$ の起り得る場合 $\prod_{i=1}^{L} {}_{N_i}C_{a_i}$ 通りはすべて同一影響を及ぼすに過ぎないことを意味している。一方 $Z_1, Z_2, \cdots, Z_N$ が選ばれるや否や二次的母集団 $\pi(N)$ と二次的层 $\pi(N_i)$ が確定する．但しこの层の個数及び N_i は抽出に依存する確率変数である．この二次的层 $\pi(N_i)$ からある抽出比 r_i（但し r_i は N_i にのみ依存）によって，层別抽出しそれによって得られた i 层，j 番目の第一成分を x_{ij} とすると，これが本調査のデータを構成する。このようにして得られた $n = \sum_1^L n_i = \sum_1^L N_i r_i$ から，母集団平均 $\breve{x}$ の推定

(3) $$\bar{x} = \sum_1^L w_i \bar{x}_i$$

但し，$\bar{x}_i = \frac{1}{n_i} \sum_1^{n_i} x_{ij}$

且，$\{n_i = 0\}$ に対しては $\bar{x}_i = 0$ とする．

をつくる．

このようにして作った $\breve{x}$ の推定値 $\bar{x}$ についての考察を進めることにしよう．

定理 1

$$E(\bar{x}) = \breve{x}$$

但し $\breve{x} = \frac{1}{M} \sum_i^L \sum_j^{M_i} x_{ij}$

（証）

$$E(\bar{x}) = \sum_{i=1}^L E(w_i \bar{x}_i)$$

今 E' を N_i を固定したときの条件付平均，

E'' を $\pi(N_i)$ を固定した条件付平均，但し $n_i \neq 0$ とすると

$$E(\bar{x}) = \frac{1}{N} \sum_1^L E\left\{N_i E'\left[\frac{1}{n_i} \sum_1^{n_i} x_{ij}\right]\right\}$$

$$= \frac{1}{N} \sum_1^L E\left\{N_i E'\left[\frac{1}{n_i} \sum_1^{n_i} E''(x_{ij})\right]\right\}$$

$$= \frac{1}{N} \sum_1^L E\left\{N_i \frac{1}{N_i r_i} N_i r_i \bar{\xi}_i\right\}$$

---

[1] J. Neyman, "Contribution to the theory of sampling human populations" J. A. S. A. vol. 34 (1938).

但し，$\overline{\xi}_i = \frac{1}{N_i}\sum_{1}^{N_i}\xi_{ij}$

即ち$\pi(N_i)$が確定したときの二次的層の層内平均

$$= \frac{1}{N}\sum_{1}^{L} E\left\{N_i\overline{\xi}_i\right\} = \tilde{x}$$

$\boxed{\text{定理2}}$

> γ_i；固定したとき
>
> $$V(\overline{x}) = \sum_{1}^{L}\frac{1-\gamma_i}{N\gamma_i}W_i S_i^2 + \frac{M-N}{M}\frac{S^2}{N}$$
>
> 有限係数を省略すると
>
> $$\doteqdot \frac{1}{N}\left\{\sum\frac{W_i S_i^2}{\gamma_i} + S_b^2\right\}$$
>
> 但し　$S^2 = \frac{1}{M-1}\sum_{1}^{L}\sum_{1}^{M_i}(x_{ij}-\tilde{x})^2$
>
> $$S_i^2 = \frac{1}{M_i-1}\sum_{1}^{M_i}(x_{ij}-\tilde{x}_i)^2$$
>
> $$S_b^2 = \frac{1}{M-1}\sum_{1}^{L}(\tilde{x}_i-\tilde{x})^2$$

（証）

$$V(\overline{x}) = E\left\{\left(\sum_{1}^{L}\frac{N_i}{N}\overline{x}_i - \sum_{1}^{L}\frac{N_i}{N}\overline{\xi}_i + \sum_{1}^{L}\frac{N_i}{N}\overline{\xi} - \tilde{x}\right)^2\right\}$$

$$= E\left\{\left(\sum_{1}^{L}\frac{N_i}{N}(\overline{x}_i-\overline{\xi}_i)\right)^2\right\}$$

$$+ E\left\{(\overline{\xi}_N-\tilde{x})^2\right\} + 2E\left\{(\overline{\xi}_N-\tilde{x})\sum_{1}^{L}\frac{N_i}{N}(\overline{x}_i-\overline{\xi}_i)\right\}$$

$$= I_1 + I_2 + 2I_3$$

$$I_1 = \sum_{1}^{L}E\left(\frac{N_i^2}{N^2}E'E''(\overline{x}_i-\overline{\xi})^2\right)$$

$$= \sum_{1}^{L}E\,\frac{N_i^2}{N^2}E'\frac{N_i-n_i}{N_i}\frac{S^2_{\pi(N_i)}}{n_i}$$

$$= \sum_{1}^{L}E\left(\frac{N_i}{N}\right)^2\frac{N_i-n_i}{N_i}\frac{1}{n_i}E'(S^2_{\pi(N_i)})$$

ここで

$E'(S^2_{\pi(N_i)}) = S_i^2$　なる故

$\therefore$

$$I_1 = \sum_{i=1}^{L}S_i^2\frac{1-\gamma_i}{N\gamma_i}E(w_i) = \sum_{i=1}^{L}\frac{1-\gamma_i}{N\gamma_i}W_i S_i^2 .$$

$$I_2 = \frac{M-N}{M} \quad \frac{S^2}{N}$$

$$I_3 = 0$$

$$\therefore \quad V(\bar{x}) = \frac{1}{N}\left[\sum \frac{1-r_i}{r_i} W_i S_i^2 + \frac{M-N}{M} \quad \frac{S^2}{N}\right]$$

$S^2 = S_w^2 + S_b^2$ から

$$V(\bar{x}) \doteqdot \frac{1}{N}\left[\sum \frac{W_i S_i^2}{r_i} + S_b^2\right]$$

系 2

$r_i = r$ とすると

$$V(\bar{x}) \doteqdot \frac{1}{N}\left[\frac{Sw^2}{r} + Sb^2\right]$$

$$Sw^2 = \sum_1^L W_i S_i^2$$

今 c_1 を予備調査の調査単位一個当りの費用 c_2 を本調査一個当りの費用とする。そうすると全費用 C は

(4) $$C = c_1 N + \sum_1^L c_2 N w_i r_i$$

となる，ここで r_i を固定すると C は w_i に関する確率変数であって，全調査に要する費用は予備調査の結果によって左右される．この全費用の平均を

(5) … $$E(C) = N(c_1 + c_2 \sum W_i r_i) = C_0$$

におさえると，その分散は

定理 3

$E(C) = C_0$ の条件で

$$V(C) = N c_2^2 \frac{M-N}{M-1} \quad S^2(r)$$

但し $S^2(r) = \sum_1^L W_i (r_i - \bar{r})^2$

$$\bar{r} = \sum_1^L W_i r_i$$

証）

$$E(C-C_0)^2 = N^2 c_2^2 E\left\{\left(\sum_1^L r_i (w_i - W_i)\right)^2\right\}$$

$$= N^2 c_2^2 \left[\sum_1^L r_i^2 E(w_i - W_i)^2 + \sum\sum_{i \neq j}^{L} r_i r_j E(w_i - W_i)(w_j - W_j)\right]$$

$$= N^2 C_2^2 \ \frac{M-N}{M-1} \left[\sum \gamma_i^2 W_i (1 - W_i) - \sum\sum_{i \neq j} \gamma_i \gamma_j W_i W_j \right] \frac{1}{N}$$

$$= N^2 C_2^2 \ \frac{M-N}{M-1} \ \frac{1}{N} \ \left[\sum W_i \gamma_i^2 - (\sum W_i \gamma_i)^2 \right]$$

$$= N C_2^2 \ \frac{M-N}{M-1} \ S^2(\gamma)$$

系 3

変動係数 ε は.

$$\varepsilon \fallingdotseq \sqrt{N} \ \frac{C_2}{C_0} \ S(\gamma)$$

定理 4

$E(C) = C_0$ の下で最適な N, γ_i ($i = 1, 2, \cdots L$) は

$$N = \frac{C_0}{C_1} \ \frac{S_b}{S_b + (\sum W_i S_i / \sqrt{\mu})}$$

$$\gamma_i = \frac{S_i}{S_b} \sqrt{\mu} \qquad 但し \quad \mu = \frac{C_1}{C_2}$$

且，そのときの分散は

$$V_1 = \frac{C_1}{C_0} \left[\frac{M-1}{M} \ S_b^2 + \frac{2}{\sqrt{\mu}} \ S_b (\sum W_i S_i) + \frac{(\sum W_i S_i)^2}{\mu} \right]$$

$$\fallingdotseq \frac{C_2}{C_0} \left[(\sum W_i S_i)^2 + 2 \ S_b (\sum W_i S_i) \sqrt{\mu} + \mu S_b^2 \right]$$

$$= \frac{C_2}{C_0} \left[S_b \sqrt{\mu} + \sum W_i S_i \right]^2$$

証明）　$G = V - \lambda E(C)$

$$\frac{\partial G}{\partial N} \equiv 0 \text{ から } \sum W_i \frac{S_i^2}{\gamma_i} + S_b^2 = N^2 \lambda (C_1 + C_2 \sum W_i \gamma_i) \cdots (a)$$

$$\frac{\partial G}{\partial \gamma_i} \equiv 0 \text{ から } \quad \gamma_i = \frac{S_i}{\sqrt{\lambda} \sqrt{C_2} \ N} \qquad (i = 1, 2, \cdots, L)$$

$$\gamma_i = S_i \ \frac{\sum W_i \gamma_i}{\sum W_i S_i} \ \cdots\cdots \ (b)$$

$$\because \quad \lambda = \frac{1}{C_2 N^2} \left(\frac{\sum W_i S_i}{\sum W_i \gamma_i} \right)^2 \ \cdots\cdots \ (c)$$

(a) に (b), (c) を代入して

又 $\sum W_i \gamma_i = \frac{\sum W_i S_i}{S_b} \sqrt{\mu}$ が得られる

$$\therefore \quad \gamma_i = \frac{S_i}{S_b} \sqrt{\mu} \qquad (i = 1, 2, \cdots, L)$$

一方 $\frac{\partial G}{\partial \lambda} \equiv 0$ から

$$N = \frac{C_0}{c_1} \frac{S_b}{[S_b + (\Sigma W_i S_i / \sqrt{\mu})]}$$

これを V に代入して

$$V_I \doteqdot \frac{c_2}{C_0} \left[\Sigma W_i S_i + \sqrt{\mu} \ S_b \right]^2$$

が得られる.

系4・1

系3は定理4のN, r_i を用いると

$$\varepsilon_I = \sqrt{\frac{c_2}{C_0}} \sqrt{\frac{S_w^2 - \overline{S}^2}{S_b[S_b + (\Sigma W_i S_i / \sqrt{\mu})]}}; \ \overline{S} \equiv \Sigma W_i S_i$$

(証)

$$\varepsilon \doteqdot \sqrt{N} \ \frac{c_2}{C_0} \ S(\gamma)$$

$$\doteqdot \sqrt{\frac{C_0}{c_1} \ \frac{S_b}{S_b + \overline{S}\sqrt{\mu}} \ \frac{C_2^2}{C_0^2} - \mu \frac{1}{S_b} \left\{ \Sigma W_i S_i^2 - (\overline{S})^2 \right\}}$$

又ここで $\gamma_i = \gamma$ とすると, 全費用は

(5) …… $C = c_1 N + c_2 N \gamma \Sigma w_i$

$= N(c_1 + c_2 \gamma)$

となって, Cは確率変数ではない. 従って $C = C_0$ の下での最適なN, γを求める.

系4・2

$C = C_0$ の下でのN, γは

$$N = \frac{C_0}{c_1} \ \frac{S_b}{S_b + (S_w/\sqrt{\mu})}$$

$$\gamma = \frac{S_w}{S_b} \sqrt{\mu}$$

$$V_{II} \doteqdot \frac{c_2}{C_0} \left(\frac{M-1}{M} S_b^2 \mu + 2 \ S_w S_b \sqrt{\mu} + \frac{M-\mu}{M} S_w^2 \right)$$

$$\doteqdot \frac{c_2}{C_0} \ (S_b \sqrt{\mu} + S_w)^2$$

以上 γ_i が確定値である場合について, 解析してみたわけであるが, 既にみたように, $\gamma_i = \gamma$ の場合を除いては調査費用を一定におさえるわけにはゆかなくなる. 従って, 実際的見地から, 調査費用をおさえておいて考えてみよう. このことは

本調査の標本の大きさmを一定にすることと同意味である.

(5) $$C = c_1 N + c_2 m = N(c_1 + c_2 k)$$

$$k = \frac{m}{N},\quad k_i = \frac{m_i}{m} \qquad (i = 1, 2, \cdots, L)$$

(6) $\cdots\cdot$ $$\gamma_i = \frac{k_i}{w_i} k \qquad 且 \quad \sum_1^L \gamma_i w_i = k$$

$\boxed{定理5}$

(6)の下での最適なγ_iは

$\gamma_i = k \frac{S_i}{\Sigma w_i S_i}$ 且,そのときの分散は

$$V_{III} \doteqdot \frac{1}{N}\left[\frac{(\Sigma W_i S_i)^2}{k} + S_b^2 + \frac{S_w^2}{NR}\right]$$

証)

$$V = E\left\{\sum_1^L \frac{N_i}{N^2}\frac{N_i - m_i}{m_i} S_i^2\right\} + \frac{S^2}{N}\frac{M-N}{M-1}$$

$$= \frac{1}{N}E\left\{\sum_1^L w_i \frac{1-\gamma_i}{\gamma_i} S_i^2\right\} + \frac{S^2}{N}\frac{M-N}{M-1}$$

$$= \frac{1}{N}E\left\{\sum_1^L \frac{w_i S_i^2}{\gamma_i} - \sum_1^L w_i S_i^2\right\} + \frac{S^2}{N}\frac{M-N}{M-1}$$

$$= \frac{1}{N}\left\{E\left[\frac{(\Sigma w_i S_i)^2}{k} + \sum_1^L w_i \gamma_i \left(\frac{S_i}{\gamma_i} - \frac{\Sigma w_i S_i}{k}\right)^2 - \sum_1^L w_i S_i^2\right]\right\} + \frac{M-N}{M-1} S^2/N$$

$$\therefore \quad \min_{\gamma_i} V = \frac{1}{Nk}E\left\{(\Sigma w_i S_i)^2 - k\sum_1^L w_i S_i^2\right\} + \frac{M-N}{M-1}\frac{S^2}{N}$$

$\gamma_i = k\frac{S_i}{\Sigma w_i S_i}$ のとき $V = \min_{\gamma_i}$

$$\min_{\gamma_i} V = \frac{1}{Nk}E(\Sigma w_i S_i)^2 + \frac{M-N}{M-1}\frac{S_b^2}{N} - \frac{N-1}{M-1}\frac{S_w^2}{N}$$

$$E(\Sigma w_i S_i)^2 = \sum_1^L E(w_i^2) S_i^2 + \sum_{i \neq j}\sum E(w_i w_j) S_i S_j$$

ここで

$$E(w_i^2) = E(w_i - W_i)^2 + W_i^2$$

$$= \frac{M-N}{M-1}\frac{W_i(1-W_i)}{N} + W_i^2$$

$$= \frac{M-N}{M-1}\frac{W_i}{N} + \frac{M(N-1)}{N(M-1)} W_i^2$$

$$E(w_i w_j) = E(w_i - W_i)(w_j - W_j) + W_i W_j$$

$$= -\frac{M-N}{M-1}\frac{W_i W_j}{N} + W_i W_j =$$

$$= \overline{W}_i \overline{W}_j \ \frac{M(N-1)}{N(M-1)}$$

$$\therefore \quad E(\Sigma w_i S_i)^2$$

$$= \frac{M-N}{M-1} \ \frac{1}{N} \ \Sigma \overline{W}_i S_i^2 + \frac{M(N-1)}{N(M-1)} \ (\Sigma \overline{W}_i S_i)^2$$

$$= \frac{M-N}{M-1} \ \frac{Sw^2}{N} + \frac{M(N-1)}{N(M-1)} \ (\Sigma \overline{W}_i S_i)^2 \ .$$

$$\min_{r_i} \ \overline{V} = \frac{M-N}{M-1} \ \frac{Sw^2}{N^2 k} + \frac{M(N-1)}{N(M-1)} \ \frac{(\Sigma \overline{W}_i S_i)^2}{Nk}$$

$$+ \frac{M-N}{M-1} \ \frac{Sb^2}{N} - \frac{N-1}{M-1} \ \frac{Sw^2}{N}$$

$$\fallingdotseq \frac{1}{N} \left[\frac{(\Sigma \overline{W}_i S_i)^2}{k} + S_b^2 + \frac{Sw^2}{Nk} \right]$$

系 5

(5) の下での最適なる k, N は

$$k = \sqrt{\frac{c_1 S_w^2 + C_0 (\Sigma \overline{W}_i S_i)^2}{c_2 Sw^2 + C_0 S_b^2}} \ \sqrt{\mu}$$

$$N = \frac{C_0}{c_1} \ \frac{\sqrt{c_2 S_w^2 + C_0 S_b^2}}{\sqrt{c_2 S_w^2 + C_0 S_b^2} + \sqrt{c_1 S_w^2 + C_0 (\Sigma \overline{W}_i S_i / \sqrt{\mu})}}$$

証）$\quad G = \min\limits_{r_i} \ \overline{V} - \lambda [C_0 - N(c_1 + c_2 k)]$

$\frac{\partial G}{\partial N} \equiv 0, \quad \frac{\partial G}{\partial k} \equiv 0, \quad \frac{\partial G}{\partial \lambda} \equiv 0$ から

$$\frac{Sw^2}{N} + (\Sigma \overline{W}_i S_i)^2 = \lambda c_2 N^2 k^2$$

$$\frac{Sw^2}{Nk} + S_b^2 = \frac{Sw^2}{k^2 N} \mu + \frac{(\Sigma \overline{W}_i S_i)^2}{k^2} \ \mu$$

$$k^2 \left(\frac{c_2}{C_0} \ S_w^2 + S_b^2 \right) = \frac{c_2}{C_0} \ \mu^2 S_w^2 + (\Sigma \overline{W}_i S_i)^2$$

$$\therefore \quad k = \sqrt{\frac{c_1 S_w^2 + C_0 (\Sigma \overline{W}_i S_i)^2}{c_2 S_w^2 + C_0 S_b^2}} \ \sqrt{\mu}$$

$$N = \frac{C_0}{c_1} \ \frac{\sqrt{c_2 S_w^2 + C_0 S_b^2}}{\sqrt{c_2 S_w^2 + C_0 S_b^2} + \sqrt{c_1 S_w^2 + C_0 (\Sigma \overline{W}_i S_i / \sqrt{\mu})}}$$
