⑭ Samplingに於ける一問題

青　山　博次郎

§1　緒　言

サンプリングに於ては，或る一つの標識についてその母集団平均又は母集団総数を推定し，かつ所望の信頼度をもち得るように計画されるのが普通である。

例えば日本人の小学校六年生の数学の能力を測定しようという様な場合の計画がそれである。　ところが得られた結果について男女別の平均点数を比較するというような分析の場面に於てはそれほど厳密な取扱いが行われていないようである。　即ち予め男女の人数が分っていてそれに応じた標本抽出が行われておれば問題はないが，男女を考慮せずに標本抽出を行って後の分析では若干の考慮が必要である。　こゝではそのような場合について考えてみることにする。

§2．標本平均とその分散

今母集団の大きさを N とし，標識 x の母平均を推定するために，大きさ n の標本を無作為に抽出したとする。

このとき勿論標本平均

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (1)$$

は，不偏推定値で，分散は

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

で與えられる。たゞし

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2$$

こゝで n 個の中 α（例えば男子）なる属性をもつものが m 個，$\bar{\alpha}$（α でないもの，例えば女子）なる属性をもつものが $n-m$ 個あつたとして，これら α，$\bar{\alpha}$ の各々についての標識 x の母平均 $\bar{X}_\alpha$，$\bar{X}_{\bar{\alpha}}$ をどのように推定すればよいだろうかという問題が生ずる。

母集団に於ける α の大きさを Np，$\bar{\alpha}$ の大きさを Nq（従つて $p+q=1$）とすると標本として n 個のものが抽出されたとき，その中に α が m 個入つている確率は

$$\binom{Np}{m}\binom{Nq}{n-m}\Big/\binom{N}{n} \tag{3}$$

である。それ故 α の母平均推定値を $\bar{x}_\alpha$ とすれば $m \neq 0$ なるとき

$$\bar{x}_\alpha = \frac{1}{m}\sum_{i=1}^{m} x_{\alpha i} \tag{4}$$

こゝで $x_{\alpha i}$ は α なる属性をもつ i 番目のものの標識 x の値を表わす。

このとき $m=0$ の場合を除けば

$$E(\bar{x}_\alpha) = \sum_{m=1}^{n}\frac{\binom{Np}{m}\binom{Nq}{n-m}}{\binom{N}{n}-\binom{Nq}{n}}\,\underset{m}{E}\left(\frac{1}{m}\sum_{i=1}^{m} x_{\alpha i}\right)$$

$$= \sum_{m=1}^{n} \frac{\binom{Np}{m}\binom{Nq}{n-m}}{\binom{N}{n}-\binom{Nq}{n}} \frac{1}{Np} \sum_{i=1}^{Np} X_{\alpha i} = \overline{X}_{\alpha} \qquad (5)$$

となり $\overline{x}_{\alpha}$ は不偏推定値であることが分る。(5) に於て分母の $\binom{Nq}{n}$ は $m=0$ の場合を省くためでこの値のないときは 0 と考えてよい。以下の一般の場合も同様である。また $\underset{m}{E}$ は m が一定なるときの期望値を示す。

この分散は

$$V(\overline{x}_{\alpha}) = E(\overline{x}_{\alpha}^2) - \overline{X}_{\alpha}^2 \qquad (6)$$

である。然るに

$$E(\overline{x}_{\alpha}^2) = E\left\{\frac{1}{m^2}\left(\sum_{i=1}^{m} x_{\alpha i}^2 + \sum_{i \neq j}^{m}\sum^{m} x_{\alpha i} x_{\alpha j}\right)\right\}$$

$$= \sum_{m=1}^{n} \frac{\binom{Np}{m}\binom{Nq}{n-m}}{\binom{N}{n}-\binom{Nq}{n}} \frac{1}{m^2}\left(\frac{m}{Np}\sum_{i=1}^{Np} X_{\alpha i}^2 + \frac{m(m-1)}{Np(Np-1)} \sum_{\substack{i=1 \\ i \neq j}}^{Np}\sum_{j=1}^{Np} X_{\alpha i} X_{\alpha j}\right)$$

$$= \frac{1}{Np}\sum_{i=1}^{Np} X_{\alpha i}^2 \left(\sum_{m=1}^{n} \frac{\binom{Np}{m}\binom{Nq}{n-m}}{\binom{N}{n}-\binom{Nq}{n}} \frac{1}{m}\right) + \frac{1}{Np(Np-1)} \sum_{i \neq j}\sum X_{\alpha i} X_{\alpha j}$$

$$\times \left(\sum_{m=1}^{n} \frac{\binom{Np}{m}\binom{Nq}{n-m}}{\binom{N}{n}-\binom{Nq}{n}} \frac{m-1}{m}\right)$$

$$= \frac{A}{Np}\sum_{i}^{Np} X_{\alpha i}^2 + \frac{1-A}{Np(Np-1)} \sum_{\substack{i=1 \\ i \neq j}}^{Np}\sum_{j=1}^{Np} X_{\alpha i} X_{\alpha j} \qquad (7)$$

こゝで

$$A = \sum_{m=1}^{n} \frac{\binom{Np}{m}\binom{Nq}{n-m}}{\binom{N}{n}-\binom{Nq}{n}} \frac{1}{m} = \frac{1}{\binom{N}{n}-\binom{Nq}{n}} \sum_{m=1}^{n} \frac{\binom{Np}{m}\binom{Nq}{n-m}}{m} \tag{8}$$

とおいた。 この A はまた

$$(1+x)^{Nq} \int_0^x \frac{(1+t)^{Np}-1}{t} dt \tag{9}$$

の x^n の係数を $\binom{N}{n}-\binom{Nq}{n}$ で除したものに等しい。

（7）と

$$\overline{X}_\alpha^2 = \frac{1}{(Np)^2}\left(\sum_{i=1}^{Np} X_{\alpha i}^2 + \sum_{\substack{i=1 \\ i \neq j}}^{Np} \sum_{j=1}^{Np} X_{\alpha i} X_{\alpha j} \right)$$

を（6）へ代入すれば

$$V(\overline{x}_\alpha) = \frac{NpA-1}{Np-1} \sigma_\alpha^2 \tag{10}$$

たゞし

$$\sigma_\alpha^2 = \frac{1}{Np} \sum_{i=1}^{Np} X_{\alpha i}^2 - \overline{X}_\alpha^2 \tag{11}$$

が得られる。

標本抽出後の推定値としては，α が m 個抽出されたとき

$$\left.\begin{aligned} p &\fallingdotseq \frac{m}{n} \\ \sigma_\alpha^2 &\fallingdotseq \frac{Nm-n}{N(m-1)m} \sum_{i=1}^{m} (x_{\alpha i} - \overline{x}_\alpha)^2 \\ A &\fallingdotseq \frac{1}{\binom{N}{n}-\binom{N\frac{n-m}{n}}{n}} \sum_{k=1}^{n} \binom{N\frac{m}{n}}{k}\binom{N\frac{n-m}{n}}{n-k}\frac{1}{k} \end{aligned}\right\} \tag{12}$$

用いればよい。

以上は母集団中に α，$\bar{\alpha}$ という二つの異なる属性の存する場合あつたが，三つ以上の場合も同様に得られる。

例えば三つの α_1，α_2，α_3 という属性が存するときは

$$\left.\begin{array}{ll} \alpha_1 \text{ の大きさを} & Np_1 \\ \alpha_2 \text{ の大きさを} & Np_2 \\ \alpha_3 \text{ の大きさを} & Np_3 \\ \quad p_1 + p_2 + p_3 = 1 & \\ \quad m_1 + m_2 + m_3 = n & \end{array}\right\} \qquad (13)$$

おけば

$$\bar{x}_K = \frac{1}{m_K} \sum_{i=1}^{m_K} x_{Ki} \quad ; \quad m_K \neq 0 \quad (K = 1, 2, 3) \qquad (14)$$

$$E(\bar{x}_K) = \sum_{\substack{m_1+m_2+m_3=n \\ m_K \geqq 1 \\ m_{K'} \geqq 0 \\ m_{K''} \geqq 0}} \frac{\binom{Np_1}{m_1}\binom{Np_2}{m_2}\binom{Np_3}{m_3}}{\binom{N}{n} - \sum\limits_{\substack{m_{K'}+m_{K''}=n \\ m_{K'} \geqq 0 \\ m_{K''} \geqq 0}} \binom{Np_{K'}}{m_{K'}}\binom{Np_{K''}}{m_{K''}}} \underset{m_K}{E}\left(\frac{1}{m_K}\sum_{i=1}^{m_K} x_{Ki}\right) = \bar{X}_K \qquad (15)$$

$$V(\bar{x}_K) = \frac{Np_K A_K - 1}{Np_K - 1}\, \sigma_K^2 \qquad (16)$$

$$A_K = \sum_{\substack{m_1+m_2+m_3=n \\ m_K \geqq 1 \\ m_{K'} \geqq 0 \\ m_{K''} \geqq 0}} \frac{\binom{Np_1}{m_1}\binom{Np_2}{m_2}\binom{Np_3}{m_3}\frac{1}{m_K}}{\binom{N}{n} - \sum\limits_{\substack{m_{K'}+m_{K''}=n \\ m_{K'} \geqq 0 \\ m_{K''} \geqq 0}} \binom{Np_{K'}}{m_{K'}}\binom{Np_{K''}}{m_{K''}}} \qquad (17)$$

となる。

更に母集団が層別されている場合にも拡張できる。 簡単のため層は二つ，属性も α，$\bar{\alpha}$ の二つのときを考えよう。

第一層，第二層の大きさをそれぞれ N_1，N_2 とし標本抽出数を n_1，n_2 としよう。

$$\left.\begin{array}{ll} \text{母集団全体の } \alpha \text{ の大きさ} & Np \\ \qquad〃 \quad \bar{\alpha} \text{ の大きさ} & Nq \\ \text{第 } i \text{ 層}(i=1,2)\text{ における } \alpha \text{ の大きさ} & N_i p_i \\ \text{第 } i \text{ 層}(i=1,2)\text{ における } \bar{\alpha} \text{ の大きさ} & N_i q_i \end{array}\right\} \qquad (18)$$

第 i 層の n_i 個の標本中 α が m_i 個含まれ $m_1+m_2=m$

（$p_i+q_i=1$，$N_1p_1+N_2p_2=Np$ であるが p_i，q_i は変り得る）

とすれば，母平均の推定値は

$$\bar{x}_\alpha = \frac{1}{m}\sum_{i=1}^{2}\sum_{j=1}^{m_i} x_{\alpha ij} \qquad (m \neq 0) \qquad (19)$$

このとき

$$E(\bar{x}_\alpha) = \sum_{N_1p_1+N_2p_2=Np} \frac{\binom{Np}{N_1p_1}\binom{Nq}{N_1q_1}}{\binom{N}{N_1}} \sum_{\substack{m=m_1+m_2\geq 1\\ n_i\geq m_i\geq 0}} \frac{\binom{N_1p_1}{m_1}\binom{N_1q_1}{n_1-m_1}\binom{N_2p_2}{m_2}\binom{N_2q_2}{n_2-m_2}}{\binom{N_1}{n_1}\binom{N_2}{n_2}-\binom{N_1q_1}{n_1}\binom{N_2q_2}{n_2}} \underset{m}{E}\left(\frac{1}{m}\sum_i\sum_j X_{\alpha ij}\right)$$

$$= \frac{1}{Np}\sum_{k=1}^{Np} X_{\alpha k} = \bar{X} \qquad (20)$$

前と同様の手続をくり返せば

$$V(\bar{x}_\alpha) = \frac{NpA-1}{Np-1}\,\sigma_\alpha^2 \qquad (21)$$

ただし

$$A = \sum_{N_1p_1+N_2p_2=Np} \frac{\binom{Np}{N_1p_1}\binom{Nq}{N_1q_1}}{\binom{N}{N_1}} \sum_{\substack{m=m_1+m_2\geq 1\\ n_i\geq m_i\geq 0}} \frac{\binom{N_1p_1}{m_1}\binom{N_1q_1}{n_1-m_1}\binom{N_2p_2}{m_2}\binom{N_2q_2}{n_2-m_2}\frac{1}{m}}{\binom{N_1}{n_1}\binom{N_2}{n_2}-\binom{N_1q_1}{n_1}\binom{N_2q_2}{n_2}} \qquad (22)$$

$$\sigma_\alpha^2 = \frac{1}{Np} \sum_{K=1}^{Np} X_{\alpha K}^2 - \overline{X}_\alpha^2 \tag{23}$$

例， $N = 5$,　$Np = 3$,　$Nq = 2$,　$n = 2$

$\alpha\ (4, 5, 6)$　$\bar{\alpha}\ (15, 20)$

の場合を考えてみよう　すべての組合せを考えれば下表の様になる。

場合	Sample	$\overline{x}_\alpha$	$\overline{x}_{\bar{\alpha}}$
(1)	4, 5	4.5	—
(2)	4, 6	5	—
(3)	5, 6	5.5	—
(4)	4, 15	4	15
(5)	4, 20	4	20
(6)	5, 15	5	15
(7)	5, 20	5	20
(8)	6, 15	6	15
(9)	6, 20	6	20
(10)	15, 20	—	17.5
平均		5	17.5

$\overline{X}_\alpha = 5$

$\sigma_\alpha^2 = 0.67$

$A_\alpha = 0.8333$

$V(\overline{x}_\alpha) = 0.458$

$\overline{X}_{\bar{\alpha}} = 17.5$

$\sigma_{\bar{\alpha}}^2 = 6.25$

$A_{\bar{\alpha}} = 0.9286$

$V(\overline{x}_{\bar{\alpha}}) = 5.375$

§ 3. あとがき

この問題は石田正次氏の呈示によるもので，ありふれたことながらまだ発表されてもおらぬようなのでこゝに記すことにした。尚後で気付いたのであるがW. G. Madowの論文（註）の定理の一つの応用例であることが分った。

（註） W. G. Madow : On the theory of systematic sampling II, Annals of Mathematical Statistics, 1949

（定理1及び定理2）