全数調査に不応答群[1)]のある場合の抽出法

兼所員　増山元三郎

法人調査を行うに就て，こういうことが問題になつた。一応全数調査を行うにしても，事実上不応答群がありうる。この時ある量 X' の和を抽出法で推定する場合，どうしたらよいかという問題である。

北川博士は，不応答群を追求するのは勿論であるが，応答群は全部手許に資料が来ているから，費用の点から見て，応答群からも更に標本を追加する方が，有効ではなかろうかと示唆されたので，Hansen-Hurwitz の論文[2) 3)]に倣って計算してみた。結論を先に云うなら，大標本では，応答群を追加して推計するより，不応答群の追求をその費用だけ増す方がよいのである。

次の記号を用いる。

$N_\alpha =$ 第 α 層にある法人総数で，郵送法を用いて調査するものとする。

$n_\alpha =$ 第 α 層から抽出される法人数

$m_\alpha =$ 第 α 層から抽出された n_α 箇の法人中返事を寄した群に属する数

$n_\alpha p_\alpha =$ 第 α 層の返事を寄した群から更に追加抽出すべき数，p_α は固定したものとする。

$S_\alpha =$ 第 α 層から抽出された法人のうち，返事を寄さなかった群に属する数，即ち

$$n_\alpha = m_\alpha + S_\alpha,$$

$\bar{x}'_{1\alpha} =$ 第 α 層での応答群に属する法人 $(m_\alpha + n_\alpha p_\alpha)$ 箇での被調査量 X の標本平均，

$\bar{x}''_{2\alpha} =$ 追求された S_α 箇の法人での標本平均，

$T =$ X の和の不偏推定量

すると，層の数を l として

$$T = \sum_{\alpha=1}^{l} \frac{N_\alpha}{n_\alpha(1+p_\alpha)} \left\{ (m_\alpha + n_\alpha p_\alpha)\bar{x}'_{1\alpha} + S_\alpha \bar{x}''_{2\alpha} \right\}$$

その母分散は

$$\sigma_T^2 = \sum_{\alpha=1}^{l} \frac{N_\alpha^2}{N_\alpha - 1} \left\{ \frac{N_\alpha}{n_\alpha(1+p_\alpha)} - 1 \right\} \sigma_\alpha^2 + \sum_{\alpha=1}^{l} \frac{N_\alpha}{n_\alpha(1+p_\alpha)^2} \frac{S_2}{S_2 - 1} (k_\alpha - 1) w_\alpha^2 \}$$

茲に，

σ_α^2 ＝ 第α層内の法人での X の母分散，

w_α^2 ＝ 第α層での不応答群に属する法人での X の母分散，

S_α ＝ 第α層での不応答群の大きさ，（s_α はその一部分）

$$S_\alpha / N_\alpha = E(s_\alpha)/n_\alpha.$$

r_α ＝ 第α層で追求された法人の数

$k_\alpha \equiv s_\alpha / r_\alpha \ (>1)$，$k_\alpha$ は固定して考える．

調査の総費用を C とすると，

$$C = \sum_{\alpha=1}^{l} n_\alpha \{ C_{1\alpha}(1+p_\alpha) + C_{2\alpha} Q_\alpha / k_\alpha + C_{3\alpha}(p_\alpha + P_\alpha + Q_\alpha / k_\alpha \}$$

茲に

$C_{1\alpha}$ ＝ 第α層法人一つ当りの調査費用

$C_{2\alpha}$ ＝ 第α層での一法人を追求する費用

$C_{3\alpha}$ ＝ 第α層での一法人当りの集計費用

$P_\alpha \equiv 1 - Q_\alpha$ ＝ 第α層での法人中応答群に含まれるもの割合で既知，従って，

$$E(m_\alpha) = P_\alpha n_\alpha.$$

今 $\sigma_T^2 = \varepsilon^2$ の下に C を最小にするような割当て方を求めると，添字αを落して，

$$C_2 / C_3 \equiv a. \quad C_1 / C_3 \equiv d$$

とおくと，

$$\therefore\ k = a+1+\sqrt{a^2+a} = \sqrt{a+1}\,(\sqrt{a}+\sqrt{a+1}),$$

$$p = \frac{(N-1)S^2\omega^2\left\{2Q(a+\sqrt{a^2+a})-(\alpha+1)(2a+1+2\sqrt{a^2+a})\right\}+N^2(S-1)Q\sigma^2}{(N-1)S^2(\alpha+1)(2a+1+2\sqrt{a^2+a})\omega^2-N^2(S-1)Q\sigma^2}$$

$$n = \sqrt{\lambda}\cdot\frac{kS\omega}{1+p}\sqrt{\frac{N}{(S-1)(C_2+C_3)Q}}$$

但し λ は Lagrange の乗数で，

$$C+\lambda(\sigma_T^2-\varepsilon^2) = \min.$$

の式から這入ったものである。

以上の結果では，追加標本 n_α p_α については，有限母集団としての修正を加えないで済むものとして計算してある。

元の公式で $p\to 0$ なら Hansen-Hurwitz の結果に一致する。

$$N \gg 1,\quad S \gg 1,\quad \omega^2/\sigma^2 \equiv \rho^2$$

と置くと，

$$p \fallingdotseq \frac{\rho^2\left\{2Q(a+\sqrt{a^2+a})-(\alpha+1)(2a+1+2\sqrt{a^2+a})\right\}+1}{(\alpha+1)(2a+1+2\sqrt{a^2+a})\rho^2-1}$$

次に

$p>0$ なるための条件を求めよう。

$$A = 2\rho^2 Q(a+\sqrt{a^2+a}),\quad D \equiv (d+1)(2a+1+2\sqrt{a^2+a})\rho^2-1$$

と置くと，

$$p = \frac{A-D}{D} = \frac{A}{D}-1$$

ところで $A>0$ だから，$p>0$ のためには，$D>0$ でなければいけない。従って先づ

$$(\alpha+1)(2a+1+2\sqrt{a^2+a})>1 \geqq 1/\rho^2$$

が必要である。次に

$$A-D = 1-(\sqrt{a}+\sqrt{a+1})\left\{d(\sqrt{a}+\sqrt{a+1})+2\sqrt{a}\,P+1\right\}\rho^2$$

と変形できるから，

$$1/\rho^2 \geqq (\sqrt{a}+\sqrt{a+1})\left\{d(\sqrt{a}+\sqrt{a+1})+2\sqrt{a}\,P+1\right\}>1.$$

故に $p>0$ ということは起り得ないことが分る。

即ち，応答群から追加標本を求めることは，費用，精度の点から見て好ましくないことが分る。

次に，$p+1<0$ ということは，p の定義から起り得ないことであるが，これが ρ^2 に対しどんな條件になるか調べてみよう。この時は，

$$\frac{A}{D}<0$$

$A>0$ だから，$D<0$ 従って

$$1 < (\alpha+1)(2a+1+2\sqrt{a^2+a}) < 1/\rho^2$$

これは $\sigma^2 \gg \omega^2$ でない限り成立しないが，このような場合が起らないという保證は與えられない。從って茲に考えたモデルは改良すべきように思われる。

1). Problems and methods of the estimation in corporation censers when the result involves some non respondenses, (translated by H. Sakamoto)

2). M. H. Hansen and W. N. Hurwitz: The problems of non-response in sample surveys, J. A. S. A. 41 (1946), 517.

3) W. G. Cochran: Sample survey techniques, 1948 § 5-20.

追記：この報告は筆者の研究ノートに依れば，1948年6月20日〜7月1日頃解かれている。

歐文でE. S. Sに報告[1]した時，同時に日本文が編輯に送られたが，途中で行方不明になっていることを最近やっと知り，ノートを見て書き改めたものである。