⑨ZIGZAG抽出法の抽出誤差について

兼所員　増　山　元　三　郎

前に本誌上に，ある量Yが大体大きさの順に並んでいる時は、系統抽出法を用いて抽出間隔kで抽くと、母平均の推定量の母分散は，n箇に層別して各層から1づつの無作為抽出した場合より大低大きいから好ましくないので、系統的に並べるよりZIGZAGに並べる方法が良いことを指摘した。茲に$N=nk$は母集団の大きさとする。ZIGZAG抽出法ではnが偶数になるように選ぶことが好ましいが、利用できるための必要條件ではない。

茲にはこの場合の母分散を標本から推定する方法を考えよう。

以下Cochranと異る定義を用いよう。

層＼集落	C_1　C_2 - - - - - - - - - C_k	平均
S_1	y_1　y_2 - - - - - - - - - - y_k	m_1
S_2	y_{2k}　y_{2k-1} - - - - - - - - y_{k+1}	m_2
⋮	- - - - - - - - - - - - - -	⋮
S_n		m_n
平均	ξ_1　ξ_2 - - - - - - - - - - ξ_k	m

$$k\sigma_\xi^2=\sum_{i=1}^{k}\xi_i^2-km^2$$

$$=\frac{1}{n^2}\sum_{i=1}^{nk}y_i^2+\frac{2}{n^2}\left\{\sum_{\alpha\neq\beta}y_\alpha y_\beta\right\}-km^2$$

$$=\frac{1}{n^2}\{nk\sigma_y^2+nkm^2\}+\frac{2}{n^2}\{\Sigma y_\alpha y_\beta\}-km^2$$

$$\therefore \sigma_{ZG}^2 \equiv \sigma_\xi^2=\frac{\sigma_y^2}{n}+\left(\frac{1}{n}-1\right)m^2+\frac{2}{kn^2}\{\Sigma y_\alpha y_\beta\}$$

今其変量 $C_{\alpha\beta}$ を次式で導くと、

$$\sum y_\alpha y_\beta = C_{\alpha\beta}+k m_\alpha m_\beta$$

$$\sigma_{ran}^2 \equiv \frac{\sigma_y^2}{n}=\sigma_\xi^2-\frac{2}{kn^2}\sum_{\alpha<\beta} C_{\alpha\beta}+\frac{\sigma_m^2}{n}$$

茲に σ_y^2 は Y の母分散、σ_m^2 は $m_1, m_2, \cdots\cdots, m_n$ の母分散である。

この式から、ZIGZAG法が無作為法より

$$\sigma_m^2 > \frac{2}{kn}\sum_{\alpha<\beta} C_{\alpha\beta}$$

ならよいことがわかるし、更に

$$\sum_{\alpha<\beta} C_{\alpha\beta} < 0$$

なら、ZIGZAG法は層別法よりよいことが分る。

$C_{\alpha\beta}$ の符号は、ZIGZAG法に依れば、大体次のようになり、従つて $|\beta-\alpha|$ が奇数なら負、偶数なら正、又 $|\beta-\alpha|$ が大いなると $|C_{\alpha\beta}|$ は小になるような母集団であれば大丈夫 $\sum_{\alpha<\beta} C_{\alpha\beta}$ は負になることが分る。

$\alpha \backslash \beta$	2	3	4	------------------------	n
1	−	+	−	------------------------	$(-1)^{n-1}$
2		−	+	------------------------	$(-1)^{n-2}$
3			−	------------------------	$(-1)^{n-3}$
⋮				------------------------	
⋮				------------------------	
$n-1$					−

ところで 最初（$y_1, y_2, \cdots\cdots, y_k$）の中から一つ y を乱数表で抽いたとすると、

$$nV \equiv y^2 + y_{2k-\nu+1} + \cdots - n\xi_\nu^2$$

と置いて

$$E(\nabla_\nu) = (S_y + nkm^2 - n\sum^{k}\xi_\nu^2)/(nk)$$

$$= \sigma_y^2 + m^2 - \sigma_\xi^2 - m^2$$

$$= \sigma_y^2 - \sigma_\xi^2$$

$$= (n-1)\sigma_\xi^2 + (\sigma_m^2 - \frac{2}{kn}\sum_{\alpha \neq \beta} C_{\alpha\beta})$$

従って

σ_ξ^2 の推定量として、 $\nabla_\nu/(n-1)$ を採用すると、

$$b \equiv \sqrt{(\sigma_m^2 - \frac{2}{kn}\sum_{\alpha<\beta} C_{\alpha\beta}/(n-1)}$$

だけ大きい方に偏っていることが分る。従って平均として大きく見積って差支えないなら $\nabla_n/(n-1)$ が利用できる。この時

$$(n-1)b^2 = \sigma_m^2 - \frac{2}{kn}\sum_{\alpha<\beta} C_{\alpha\beta}$$

$$= n(\sigma_{ran}^2 - \sigma_{Z.G})$$

となっていることに注意、b の推定法は別に考えたい。

1) W. G Madow and L. H. Madow : Ann. March. Stat., 15 (1944), 1

2) W. G. Cochran : Ann. Math. Stat., 17 (1946), 164.

3) W. G. Cochran : Sample survey techniques, 1948.
Chap. VI. p 66–76.