㉓ ZIGZAG抽出法

兼所員　　増山元三郎

いつでも用ひ易い方法とはいえないが，抽出台帳が，調査の対象となる量又はそれと相関度の高い量について大きさの順になつている場合に利用して効率のよい方法を考えてみた。

要点は系統抽出法を二個異つた抽出単位から出発して行い，両者に一定の條件を與えるに在る。

大きさ $N = nk$ の有限母集団を，抽出間隔 k で系統抽出を行い，大きさ n の標本を求める場合，標本平均 $\bar{y}_n$ の母分散は，

$$V_{sys} \equiv V(\bar{y}_n) = \frac{N-n}{N}\frac{\sigma_w^2}{n}\left\{1 + \frac{2}{n}\sum_{d=1}^{n-1}(n-d)\rho(kd)_w\right\}$$

となる。茲に母分散の定義は Cochran 流のもので，

$$n(k-1)\sigma_w^2 \equiv \sum_{i=1}^{N}(y_i - y_{pi})^2,$$

$$(k-1)(n-d)\sigma_w^2\rho_{(kd)w} \equiv \sum_{i=1}^{N-kd}(y_i - \bar{y}_{pi})(y_{i+kd} - \bar{y}_{p,i+kd})$$

と定義した。[1)]

記号の意味は，表の通りである。

集落＼層	S_1	S_2	……	S_n	行平均
C_1	y_1	y_{k+1}	……	$y_{(n-1)k+1}$	m_1
C_2	y_2	y_{k+2}	……	$y_{(n-1)k+2}$	m_2
⋮	⋮	⋮		⋮	⋮
C_k	y_k	y_{2k}	……	y_{nk}	m_k
列平均	$\bar{y}_{p1}$	$\bar{y}_{p2}$	……	$\bar{y}_{pn}$	$\bar{y}_p$

この場合直ぐ分るように，大体大きさの順に並んでいる以上 $\rho > 0$ の傾向を持ち，従つて，系統抽出法で集落 C_1，，C_2，………，C_k の何れかを抽く場合の標本平均 $\bar{y}_n$ の母分散は，上の表で層 S_1，S_2，………，S_n の中から，夫々1つづつ無作為抽出する場合，即ち層別比例抽出した場合の標本平均の母分散

$$V_{strat} = \frac{(N-n)}{N}\frac{\sigma_w^2}{n}$$

より大となる。[2)]

これは C_1 には各層での一番大きいもの，C_k には各層での一番小さいものが含まれるせいである。

これを避けるには，大きさの順に並べる時，次の表のようにすればよい。

	S_1	S_2	S_3	……	S_n	行平均
Z_1	y_1	y_{2k}	y_{2k+1}	……		μ_1
Z_2	y_2	y_{2k-1}	y_{2k+2}	……		μ_2
⋮	⋮	⋮	⋮			⋮
Z_{k-1}	y_{k-1}	y_{k+2}	y_{3k-1}	……		⋮
Z_k	y_k	y_{k+1}	y_{3k}	……		μ_k
列平均	$\bar{y}_{p1}$	$\bar{y}_{p2}$	$\bar{y}_{p3}$	……	$\bar{y}_{pn}$	$\bar{y}_p$

即ち圃場試験法で Beaven の用ひた方法に似て，ヂグザウに折返して並べて置き，次に集落 Z_1, Z_2, ……, Z_k からちの一つを乱数表で抽くのである。 こうすると，ρ のうち位相差 d が奇数のものは負になる傾向を持つから，系統抽出法よりは確かに効率はよくなるといえる。

この方法は従来の系統抽出法で $2k$ 箇の集落を作り，乱数表で 1 から k 迄の一つを定め，之を例えば j とすると，大きさ $n/2$ の第 j 集落と大きさ $n/2$ の第 $(2k-j+1)$ 集落とを同時に抽くことに相当する。

この考えを一般化することは容易であるが，茲には実験例を報告するに止める。実験は黒岩喜代子氏に依る。

母集団は大きさ $N = 100$ の正規型に近い形のもので，$k = 5$，従って $n = 20$ とした。

母平均 $\bar{y}_p = 0.00$，　母分散 $\sigma^2 = 22.44$ である。

標本平均について，結果を大きさ n の無作為標本での母分散 $V_{ran.}$，系統抽出法での母分散 $V_{sys.}$，層別比例抽出法での母分散 $V_{strat.}$，ジグザグ法での母分散 $V_{zig.}$ で与えて置く。

$$V_{ran.} = 0.9067 ,$$

$$V_{sys.} = 0.1370 ,$$

$$V_{strat} = 0.0133 ,$$

$$V_{zig.} = 0.000 .$$

例としては"都合のよい例"であらうが，大勢は掴めよう。抽出結果から分散を推定する方法に未だ問題が残っている。

1）W. G. Cochran : Sample Survey Techniques 1948.

2）$V_{sys} < V_{strat}$ の必充條件は，次の通り，

$$2\bar{\rho} \equiv 2\sum_{d=1}^{n-1} \rho_{(kd)w} / (n-1) < \sum_{d=1}^{n-1} d\rho_{(kd)w} \Big/ \sum_{d=1}^{n-1} d.$$