*Jour. Ind. Soc. Ag. Statistics*
*Vol. XLI, No. 1 (1989), pp. 85-90*

# AN EXTENSION OF THE CLASS OF REGULAR ESTIMATORS IN SAMPLING FROM A FINITE POPULATION

DILIP ROY
*Department of Business Administration, University of Burdwan, Golapbag, Burdwan*
(Received : August, 1984)

SUMMARY

An extension of the Class of Regular Estimators has been proposed in order to obtain a larger class within which the estimators are linearly invariant almost everywhere.
Existence of the optimal estimator in this proposed class has been verified alongwith the importance of balanced sampling for achieving the minimum variance.

*Keywords* : Balanced sampling, Minimum variance estimator, Linear transformation, Linearly invariant.

## Introduction

In an attempt to findout the optimum estimator in the linear class it was observed by Roy and Chakravarti [1] that though no best estimator exists in the linear class the best estimator exists in a more restricted class, called by them as the Class of Regular Estimators (CRE). They had shown that a regular estimator is necessarily linearly invariant in almost everywhere. Also, for the variance of any regular estimator, there exists a lower bound which is attainable through balanced sampling.
In this present work we shall try to extend the CRE in such a way that all members of that extended class are necessarily linearly invariants

with probability one. Within that extended class we shall also attempt to findout the minimum variance estimator and demonstrate therefrom the usefulness of balanced sampling.

## 2. Definitions

We shall use the following notations for our subsequent study.

$N$ = size of population,

$U = (u_1, u_2, \ldots, U_N)$, the population of $N$ units,

$y_i = y(u_i)$, the value of the variable $y$ on $i$th unit $u_i$.

$M = (1/N) \sum_{i=1}^{N} y_i$, the population mean,

$\sigma^2 = (1/N) \sum_{i=1}^{N} (y_i - M)^2$, the population variance,

$s$ = observed sample of units from $U$,

$n(s)$ = effective sample size, i.e. the number of distinct units in the sample $s$,

$t_y(s) = \sum_{i=1}^{N} a_i(s)$, an estimator based on $s$ with respect to the variable $y$.

*Definition 1.* An estimator $t_y(s)$ is said to be linearly invariant if for any $z = a + by$, a linear transformation of $y$,

$$t_z(s) = a + bt_y(s) \tag{2.1}$$

It may be verified that this condition is equivalent to

$$a_i(s) = 1 \quad \text{for all } s. \tag{2.2}$$

*Definition 2.* An estimator $t_y(s)$ is said to belong to the Class of Regular Estimators (CRE) if

$$E(t_y(s)) = M$$

and $\text{Var}(t_y(s)) = k\sigma^2$

for some constant $k$.

These above two definitions are from Roy and Chakravarti and they have shown that if $t_y(s) \in$ CRE then $\sum_{i=1}^{N} a_i(s) = 1$ almost everywhere, implying thereby that $t_y(s)$ is linearly invariant. We shall now present an extension of CRE as follows.

*Definition 3.* An estimator $t_y(s)$ is said to belong to the Extended Class of Regular Estimasors (ECRE) if

$$E(t_y(s)) = M$$

and $\quad \text{Var}(t_y(s)) = k f(y_1, \ldots, y_N)$

for some constant $k$ and any homogeneous function $f(\cdot)$ of order 2 satisfying the condition

$$f(y_1, \ldots, y_N) = 0 \quad \text{whenever } y_1 = y_2 = \ldots = y_N.$$

It may be noted that $\sigma^2$ is a particular case of $f(\cdot)$ where

$$f(y_1, \ldots, y_N) = \sum_{i=1}^{N} (y_i - M)^2/N.$$

Thus $\quad$ CRE $\in$ ECRE.

Similar is the case of square of Mean Deviation around mean or median. While the CRE is unable to cover stratification problems ECRE is in a position to include the same.

## 3. Properties of ECRE

With the above definition of ECRE we shall now examine a few properties of interest to indicate the importance of this extended class of estimators.

*Result 1,* An extended regular estimator is linearly invariant.

*Proof.* Let $t_y(s)$ be an extended regular estimator. Then, as because $t_y(s) \in$ ECRE

$$\text{Var}(t_y(s)) = k f(y_1, \ldots, y_N) \quad \text{for all } y \text{ and } s.$$

or, $\quad \text{Var}\left(\sum_{i=1}^{N} a_i(s)\, y_i\right) = k f(y_i, \ldots, y_N) \quad$ for all $y$ and $s$.

In particular when $y_1 = y_2 = \ldots = y_N = c$, say

$$c^2\, \text{Var}\left\{\sum_{i=1}^{N} a_i(s)\right\} = k f(c, \ldots, c) \text{ for all } s$$

implying thereby that

$$\text{Var}\left\{\sum_{i=1}^{N} a_i(s)\right\} = 0 \qquad \text{for all } s$$

or, $\sum_{i=1}^{N} a_i(s) = \text{Constant} \quad$ for all $s$, $a$, $c$.

$\qquad = p$, say.

Then

$$p = E(p) = \sum_{i=1}^{N} E(a_i(s)) \tag{3.1}$$

Again from the fact that $t_y(s) \in$ ECRE

$$E(t_y(s)) = M \quad \text{for all } y.$$

$$\text{or, } E(a_i(s)) = \frac{1}{N} \quad \text{for all } i = 1, 2, \ldots, N. \tag{3.2}$$

Now combining (3.1) with (3.2) we get

$$p = \sum_{i=1}^{N} E(a_i(s)) = \sum_{i=1}^{N} 1/N = 1 \text{ almost everywhere.}$$

Hence from (2.2) and definition 1, $t_y(s)$ is a linearly invariant estimator. That completes the proof of the result. □

We shall next try to findout a lower bound for the variance of any extended regular estimator. For this, let us consider

$$V = E \sum_{i=1}^{N} (a_i(s) - n_i(s)/n(s))^2$$

$$\text{where } n_i(s) = \begin{cases} 1 & \text{if } u_i \in s \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$0 \leq V$$

$$= \sum_{i=1}^{N} \text{Var}(a_i(s)) + \sum_{i=1}^{N} E^2(a_i(s)) - E(1/n(s)).$$

$$\text{or, } \sum_{i=1}^{N} \text{Var}(a_i(s)) \geqslant E(1/n(s)) = 1/N \tag{3.3}$$

from (3.2).

Our next step will be to evaluate Var $(a_i(s))$ values for $i = 1, 2, \ldots, N$. For this let us introduce the following symbols.

$f_i = f(y_1, \ldots, y_N)$ evaluated at the following point

$$y_k = \begin{cases} 1 & \text{if } k = i \\ 0 & \text{otherwise,} \end{cases}$$

and

$f_{ij} = f(y_1, \ldots, y_N)$ evaluated at the following point

$$y_k = \begin{cases} 1 & \text{if } k = i \text{ and } j \\ 0 & \text{otherwise.} \end{cases}$$

Trivially, $f_{ii} = f_i$ for all $i$.

*Result 2.* under the above set up for all $s$

$$\text{Var}\,(a_i(s)) = k f_i \quad \text{for all } i$$

and $\text{Cov}\,(a_i(s), a_j(s)) = (k/2)\,(f_{ij} - f_i - f_j)$ for all $i$ and $j$, $i \neq j$.

*Proof.* Because $t_y(s) \in$ ECRE we get

$$\text{Var}\left\{ \sum_{i=1}^{N} a_i(s)\, y_i \right\} = k\, f(y_1, \ldots, y_N) \text{ for all } y \text{ and } s.$$

In particular when $y$'s are such that $y_i = 1$ and $y_k = 0$ for $k\,(\neq i) = 1, \ldots, N$.

$$\text{Var}\,(a_i(s)) = k f_i \text{ for all } s.$$

and this is true for all $i = 1, 2, \ldots, N$. Similarly considering

$$y_k = \begin{cases} 1 & \text{for } k = i \text{ and } j \\ 0 & \text{otherwise.} \end{cases}$$

$$\text{Var}\,(a_i(s) + a_j(s)) = k f_{ij} \quad \text{for all } s$$

or, $2\,\text{Cov}\,(a_i(s), a_j(s)) = k f_{ij} - k f_i - k f_j$ for all $s$

or, $\text{Cov}\,(a_i(s), a_j(s)) = (k/2)\,(f_{ij} - f_i - f_j)$ for all $s$

and this is true for all $i \neq j$.

Hence the result. □

*Result 3.* For the variance of any extended regular estimator $t_y(s)$ of $M$ there exists a lower bound given by

$$\text{Var}\,(t_y(s)) \geqslant \left\{ E\,\frac{1}{n(s)} - \frac{1}{N} \right\} \frac{f(y_1 \ldots\ldots s_N)}{\sum_{i=1}^{N} f_i}$$

*Proof.* For $t_y(s) \in$ ECRE we have from (3.3) and Result 2

$$\sum_{i=1}^{N} k f_i \geqslant E\,\frac{1}{n(s)} - \frac{1}{N}.$$

$$\text{or}\quad k \geqslant \left\{ E\,\frac{1}{n(s)} - \frac{1}{N} \right\} \frac{1}{\sum_{i=1}^{N} f_i}.$$

Then

$$\text{Var}\,(t_y(s)) = k\, f(y_1, \ldots, y_N)$$

$$\geqslant \left\{ E\,\frac{1}{n(s)} - \frac{1}{N} \right\} \frac{f(y_1, \ldots, y_N)}{\sum_{i=1}^{N} f_i}.$$ □

It may be further noted that lower bound for the variance of any extended regular estimator is attained when $V = 0$. This implies that

$$a_i(s) = n_i(s)/n(s) \quad \text{for all } i = 1, 2 \ldots\ldots N.$$

Then from (3.2)

$$E(n_i(s)/n(s)) = 1/N \qquad i = 1, 2, \ldots, N. \tag{3.4}$$

The sampling design for which (3.4) holds true constitutes a balanced sampling design. The corresponding minimum variance estimator is given by $t_y^*(s)$ where

$$t_y^*(s) = \sum_{i=1}^{N} y_i\, n_i(s)/n(s). \qquad \ldots(3.5)$$

The results (3.4) and (3.5) are independent of the choice of the function $f(\cdot)$ and hold true in general as long as $f(\cdot)$ is homogeneous of order 2 such that $f(1, 1, \ldots, 1) = 0$.

## REFERENCE

[1] Roy, J and Chakravarti, I.M. (1960) : Estimating the mean of a finite population, *AMS*, 31, 392-398.