# AN ADMISSIBLE ESTIMATE FOR ANY SAMPLING DESIGN

*By* V. P. GODAMBE*

*Indian Statistical Institute, Calcutta*

*SUMMARY.* In the class of all unbiassed estimates, admissibility of a well-known estimate is established.

## 1. Introduction

The author (1955) defined the general sampling design, and the corresponding class of linear unbiassed estimates, for finite populations, as follows: Let

$$1, \ldots, \lambda, \ldots, N \qquad \ldots \quad (1.1)$$

denote the different individuals in the population, the corresponding variate values being

$$X_1, \ldots, X_\lambda, \ldots, X_N. \qquad \ldots \quad (1.2)$$

The problem is to estimate

$$T = \sum_{\lambda=1}^{N} X_\lambda \qquad \ldots \quad (1.3)$$

by observing $X$ values of a few individuals $\lambda$, from (1.1). Any sequence

$$s = \lambda_1, \ldots, \lambda_{n_s} \qquad \ldots \quad (1.4)$$

of (not necessarily distinct) individuals from (1.1) is called a 'sample' and be denoted by $S$. Further, let

$$S = \{s\} \qquad \ldots \quad (1.5)$$

be an arbitrary 'finite' set of sequences $s$ in (1.4). For every $s \epsilon S$ we define a non-negative number $P_s > 0$, such that

$$\sum_{s \epsilon S} P_s = 1. \qquad \ldots \quad (1.6)$$

Now if we put

$$P = \{P_s\} \quad s \epsilon S \qquad \ldots \quad (1.7)$$

a 'sampling design' $d$ is defined as

$$d = (S, P). \qquad \ldots \quad (1.8)$$

It is easy to see that if $D$ denotes the class of all sampling designs $d$ in (1.8), then all the known 'sample survey designs' must belong to $D$. In fact, for every $d \epsilon D$ it is possible to construct a sampling mechanism of drawing individuals from (1.1), 'one after another' (Hanumantha Rao, 1960).

* Now with Science College, Nagpur, India.

For any given sample $s \epsilon S$, we define a linear estimate $e_s$ as

$$e_s = \sum_{\lambda \epsilon s} \beta_{s\lambda} \cdot X_\lambda, \qquad \dots \quad (1.9)$$

where the summation is taken over all the 'distinct' individuals $\lambda$ in $s$. It is again clear that all the known linear estimates must be particular cases of $e_s$ in (1.9). Now, for unbiassedness of $e_s$,

$$E(e_s) = T \qquad \dots \quad (1.10)$$

for all $T$ in (1.3). A necessary and sufficient condition for $e_s$ in (1.9) to be unbiassed, in a sampling design $d$, is

$$\sum_{s \ni \lambda} \beta_{s\lambda} \cdot P_s = 1, \quad (\lambda = 1, \dots, N) \qquad \dots \quad (1.11)$$

where $s \ni \lambda$, stands 'for all $s$ which include $\lambda$'. Further if $e_s$ is unbiassed, its variance is given by,

$$V(e_s) = \sum_{\lambda=1}^{N} X_\lambda^2 \sum_{s \ni \lambda} \beta_{s\lambda}^2 \, P_s + \sum_{\lambda \neq \lambda'} X_\lambda X_{\lambda'} \sum_{s \ni \lambda, \lambda'} \beta_{s\lambda}\beta_{s\lambda'} P_s - T^2. \qquad \dots \quad (1.12)$$

## 2. An admissible estimate

The probability of any particular individual $\lambda$ being included in the sample is given by

$$\sum_{s \ni \lambda} P_s = P(\lambda) \qquad \dots \quad (2.1)$$

for $\lambda = 1, \dots, N$. Following Hájek, we call

$$\bar{e}_s = \sum_{\lambda \epsilon s} X_\lambda / P(\lambda) \qquad \dots \quad (2.2)$$

a simple linear estimate (Hájek, 1959). It is easy to check from (1.11) that $\bar{e}_s$ is unbiassed. We call an unbiassed estimate $e_s$ 'admissible' if for any other unbiassed estimate $e_s'$

$$V(e_s) < V(e_s') \qquad \dots \quad (2.3)$$

for some value of $X = (X_1, \dots, X_\lambda, \dots X_N)$ in (1.2). It is proved below that in this sense the simple linear estimate $\bar{e}_s$ in (2.2) is admissible.

Let
$$e_s' = \sum_{\lambda \epsilon s} \beta_{s\lambda}' X_\lambda \qquad \dots \quad (2.4)$$

be an unbiassed estimate. If $e_s'$ is different from $\bar{e}_s$ in (2.2) it follows that

$$\beta_{s\lambda}' \neq 1/P(\lambda) \qquad \dots \quad (2.5)$$

for some $\lambda$ and $s$, $\lambda \epsilon s$. To be specific, let us suppose for $\lambda_0 \epsilon s_0$

$$\beta_{s_0\lambda_0}' \neq 1/P(\lambda_0). \qquad \dots \quad (2.6)$$

Now if $X_{\lambda_0} = 1$ and $X_\lambda = 0$ for $\lambda \neq \lambda_0$ in (1.12), we have

$$V(\bar{e}_s) = \frac{1}{P(\lambda_0)} - 1, \qquad \dots \quad (2.7)$$

and

$$V(e'_s) = \sum_{s \ni \lambda_0} \beta'^2_{s\lambda_0} P_s - 1, \qquad \dots \quad (2.8)$$

from (1.12). From (2.7), (2.8) and (1.11) we have,

$$V(e'_s) - V(\bar{e}_s) = \sum_{s \ni \lambda_0} \left( \beta'_{s\lambda_0} - \frac{1}{P(\lambda_0)} \right)^2. \qquad \dots \quad (2.9)$$

The r.h.s. of (2.9) is $> 0$ because of (2.6). Hence $\bar{e}_s$ in (2.2) is admissible.*

## 3. Illustrations

First consider the sampling mechanism of making a fixed number, say $n$, of draws, with replacement and with equal probabilities. Then for the resulting sampling design we have from (2.1),

$$P(\lambda) = 1 - \left(1 - \frac{1}{N}\right)^n. \qquad \dots \quad (3.1)$$

The simple linear estimate (2.2) in this case is given by

$$\bar{e}_s = \sum_{\lambda \epsilon s} X_\lambda / 1 - \left(1 - \frac{1}{N}\right)^n. \qquad \dots \quad (3.2)$$

Letting $v(s)$ = number of distinct individuals in $s$, we get another unbiassed estimate,

$$e'_s = N \sum_{\lambda \epsilon s} X_\lambda / v(s). \qquad \dots \quad (3.3)$$

It follows from the admissibility of $\bar{e}_s$ in (3.2) that

$$V(\bar{e}_s) < V(e'_s) \qquad \dots \quad (3.4)$$

for some value of $X = (X_1, \dots, X_\lambda, \dots, X_N)$ in (1.2).

It has been proved (Basu, 1958) that the estimate $e'_s$ in (3.3) has uniformly smaller variance than the conventional arithmetic mean. However, (3.4) proves that $e'_s$ cannot be a best unbiassed estimate. In fact, it has been demonstrated (Godambe, 1955) that in the whole class of linear unbiassed estimates of the population total, a uniformly minimum variance estimate does not exist.

Next, consider sampling with replacement and with equal probabilities, until $v$ distinct individuals are sampled, where $v$ is given in advance. In this case

$$P(\lambda) = v/N, \quad \lambda = 1, \dots, N, \qquad \dots \quad (3.5)$$

since, for the general sampling design, if $v(s)$ denote the number of distinct individuals in the sample $s$,

$$E(v(s)) = \sum_{1}^{N} P(\lambda). \qquad \dots \quad (3.6)$$

* By an independent argument, earlier the author (Godambe, 1955) specified the class of prior distributions, with respect to which, $e_s$ is the Bayes solution.

It follows from (2.2) and (3.5) that an admissible estimate in the present case is

$$\bar{e}_s = N \sum_{\lambda \epsilon s} X_\lambda / v. \quad \ldots \quad (3.7)$$

## 4. An admissible estimate which minimises maximum variance

From (1.12) and (2.2) we have,

$$V(\bar{e}_s) = \sum_{\lambda=1}^{N} X_\lambda^2 . \frac{1}{P(\lambda)} + \sum_{\lambda \neq \lambda'} X_\lambda X_{\lambda'} . \frac{P(\lambda, \lambda')}{P(\lambda) P(\lambda')} - T^2. \quad \ldots \quad (4.1)$$

If we assume that $X = (X_1, \ldots, X_\lambda, \ldots, X_N)$ in (1.2) is such that

$$X_\lambda \geqslant 0, \quad \lambda = 1, \ldots, N \quad \ldots \quad (4.2)$$

we have (since $P(\lambda, \lambda')/P(\lambda) \leqslant 1$)

$$V(\bar{e}_s) \leqslant T \sum_{1}^{N} \frac{X_\lambda}{P(\lambda)} - T^2 \leqslant T^2 . \left( \frac{1}{P(\lambda_M)} - 1 \right), \quad \ldots \quad (4.3)$$

where $$P(\lambda_M) = \text{minimum of } \{P(1), \ldots, P(\lambda), \ldots, P(N)\}. \quad \ldots \quad (4.4)$$

Thus, for any sampling design, (4.3) gives an upper bound for $V(\bar{e}_s)$, for all $X_s$'s satisfying (4.2). Moreover, for any given $T$, this upper bound is actually attained when $X_{\lambda_M} = T$ and $X_\lambda = 0$, $\lambda \neq \lambda_M$, in which case

$$V(\bar{e}_s) = T^2 \left( \frac{1}{P(\lambda_M)} - 1 \right). \quad \ldots \quad (4.5)$$

Now, the maximum variance in (4.5) is minimised for a sampling design for which $P(\lambda_M)$ in (4.4) is maximum. If we restrict ourselves to the sampling designs for which the expected number, $E(v(s))$, of distinct individuals in a sample $s$, is fixed, the design for which $P(\lambda_M)$ is maximum, or (4.5) is minimum, is immediately suggested from the fact that

$$E(v(s)) = \sum_{1}^{N} P(\lambda) \quad \ldots \quad (4.6)$$

whatever the sampling design may be. Thus, for the sampling design obtained by drawing a fixed number, $E(v(s)) = n$ (say) of individuals with equal probabilities, and without replacement, $P(\lambda_M)$ in (4.4) is maximised, and then the maximum variance in (4.5) is minimised by

$$\bar{e}_s = N \sum_{\lambda \epsilon s} X_\lambda / n. \quad \ldots \quad (4.7)$$

### References

AGGRAWAL, O. M. P. (1959): Bayes and minimax procedures in sampling from finite and infinite populations. *Ann. Math. Stat.*, **30**, No. 1.

BASU, D. (1958): On sampling with and without replacement. *Sankhyā*, **20**, 287-94.

GODAMBE, V. P. (1955): A unified theory of sampling from finite populations. *J. Roy. Stat. Soc.*, Series B, **17**, No. 2, 269-278.

HÁJEK, J. (unpublished): An optimum strategy and other problems in probability sampling.

RAO, HANUMANTH (1960): An existense theorem in sampling. (Read before Indian Science Congress, 1960).

*Paper received : September, 1959.*