THE RELATIVE EFFICIENCY OF SOME TWO-PHASE SAMPLING SCHEMES

BY M. P. SINGH

Indian Statistical Institute, Calcutta

Introduction. It is sometimes convenient and economical to use two-phase sampling where data on the auxiliary variable x are collected on an initial first-phase sample and information on the study variable y is collected only for a sub-sample of that sample. The following procedures can be used at the second-phase for utilising the data on x collected in the first-phase: (a) stratification, (b) ratio or regression methods of estimation and (c) selection of the second-phase sample with probability proportional to size (pps), size being x. Procedures (a) and (b) are given in Cochran [1] and (c) has been considered by Saito [8], Des Raj [3] and Singh and Singh [9] using the pps with replacement scheme.

In the present note we consider the selection of N' units from the N population units in the first-phase by simple random sampling without replacement for collecting data on x, and sampling of n of these N' units in the second-phase for getting information on y using the three pps without replacement schemes given by (i) Des Raj [2], (ii) Hartley and Rao [4], and (iii) Rao, Hartley and Cochran [7]. These schemes, as is well-known, have been considered by the above authors for single-phase sampling only. The authors of schemes (ii) and (iii) have given the sampling variances of the estimators considered by them, while Pathak [7] has obtained an upper bound for the variance of the estimator in scheme (i). Here we obtain the variances of the estimators in the two-phase sampling scheme and compare them under the super-population model considered by Cochran [1].

2. Estimators and Variances. An unbiased estimator T of the population total $Y = \sum Y_j$ in the present scheme will be of the form $T = (N/N')t$, where t is an unbiased estimator of the first-phase sample total $Y' = \sum_1 Y_j$, $\sum$ and $\sum_1$ denote summations over all the units in the population and the first-phase sample respectively. In particular T for schemes (i), (ii) and (iii) respectively, will be of the form

$$(2.1)\qquad T_d = (N/nN') \textstyle\sum_2 [y_1 + y_2 \cdots + y_{j-1} + y_j/P_j'(1 - P_1' \cdots P_{j-1}')],$$

$$T_{hr} = (N/nN') \textstyle\sum_2 y_j/P_j' \quad \text{and} \quad T_{rhc} = (N/N') \textstyle\sum_2 Y_j\pi_j'/P_j',$$

where $\sum_2$ denotes summation over second-phase sample, $P_j' = X_j/\sum_1 X_j$ is the probability of selection of jth first-phase unit and $\pi_j' = \sum_{\text{group } j} P_j'$.

The sampling variance of the estimator T can be expressed as

$$(2.2)\qquad V(T) = V_1E_2(T) + E_1V_2(T)$$

where E_1, V_1 are unconditional and E_2, V_2 are conditional (given the first-phase sample) expectations and variances respectively.

Received 19 February 1966; revised 3 January 1967.