Comparison of Horvitz and Thompson Estimator with that of Rao, Hartley and Cochran Estimator in PPS without Replacement Scheme

Anieting AE*
Department of Mathematics and Statistics, University of Uyo, P.M.B 1017, Uyo, Nigeria

Abstract
This paper was focused on the comparison of Horvitz and Thompson estimator of population total with that of Rao, Hartley and Cochran estimator in PPS without replacement scheme when a sample of size six is taken from the same finite population. The data used were from Nigerian Bureau of Statistics bulletin. The result showed that the variances of both estimators gave positive values but the variance of the estimator by Rao, Hartley and Cochran was smaller making it a better estimator.

Keywords: PPS without replacement; Rao, Hartley and Cochran estimator; Horvitz and Thompson estimator

Introduction
Unequal probability sampling scheme or proportional to size (PPS) sampling scheme ensures that the possibility of selecting each unit into the sample is in proportion to the size or measure of size of the unit provided that the sizes of the individual units in the population are known and are correlated with the variable of interest, Horvitz and Thompson (1952) were the first to give theoretical frame work of PPS without replacement [1]. The sample variance of the population total estimator given by Horvitz and Thompson (1952) has the possibility of giving negative values for some samples. This problem has made statistician to work on this scheme. They include Yates and Grundy(1953), Sen (1953), Durbin(1953), Rao, Hartley and Cochran (1962) etc. Rao, Hartley and Cochran (1962) gave a modified estimator based on PPS without replacement scheme. The variance of the estimator for population total under Rao, Hartley and Cochran is always positive. But how do the estimators given by Horvitz and Thompson, and that given by Rao, Hartley and Cochran fare when the sample size is more than two, say six, assuming that the variance of the population total estimator given by Horvitz and Thompson gives a positive value. The data for empirical verification of this study is from Nigerian Bureau of Statistics bulletin [1].

The two procedures
Horvitz and thompson estimation (procedure): First, the cumulative totals and ranges of the measure of size are formed [2]. Then to select a sample of size n without replacement, a random start between 1 and k inclusive (K=X/n) is selected using a table of random numbers. If the number selected is r then the units in the sample are those in whose ranges the numbers r,r+k,r+2k,…..,r+(n-1)k fall. The probability of selecting the unit $U_i$ in the sample of size n is

$$Pr(U_i-U_j) = \frac{X}{n}$$

If any unit in the population has its size greater than or equal to k such unit is removed before sampling and taken into the sample with probability unity. The probability that any pair of units $(U_i, U_j)$ is together in the sample is

$$\pi_{ij} = \frac{q_j}{k}$$

Where $q_j$ is the number of the random numbers between 1 and k inclusive, which will select $U_i$ and $U_j$ simultaneously in the sample!

An unbiased estimator of the population total for PPS without replacement as given by Horvitz and Thompson is

$$\hat{Y}_{HT} = \sum_{i=1}^{n} \frac{y_i}{\pi_i}$$

While the sample estimator of $V(\hat{Y}_{HT})$ is given as

$$V(\hat{Y}_{HT}) = \sum_{i=1}^{n} \frac{(1-\pi_i)(y_i - \bar{y})^2}{\pi_i \bar{y}} + 2 \sum_{i<j} \frac{\pi_i \pi_j}{\pi_i + \pi_j} \frac{y_i y_j}{\bar{y}^2}$$

provided $\pi_i > 0$.....2.1.1

Rao, hartley and cochran estimation (procedure): Divide a population of N units into n groups at random with group g containing $N_1, N_2, ..., N_n$ units (g=1,2,....,n) such that $N_1+\ldots+N_n = N$. Thereafter select one unit independently from each group [3]. This will give a total of n units selected in the sample with PPS without replacement. The probability of selecting $U_i$ in the sample in gth group is $P_{g_i} = \frac{X_g}{P_g}$ where $X_g = \sum_{i=1}^{n} X^i_g$

$X^i_g$’s are the auxiliary variables and $P_g$ is the sum of the initial probabilities in gth group.

The Rao, Hartley and Cochran estimator of the population total is

$$\hat{Y}_{RHC} = \sum_{i=1}^{n} \frac{y_i}{P_i}$$

Where $y_i$ is the value of the study variate for the ith unit in gth group and $P_i = \frac{P_{g_i}}{P_g}$

The estimator of the variance is given as

$$V(\hat{Y}_{RHC}) = \frac{N-n}{N(n-1)} \sum_{g=1}^{n} \frac{P_{g_i} y_i^2}{P_g} - \hat{Y}_{RHC}^2$$

(2.2.1)

*Corresponding author: Anieting AE, Department of Mathematics and Statistics, University of Uyo, P.M.B 1017, Uyo, Nigeria, E-mail: akananieting@yahoo.com

Received April 21, 2015; Accepted May 22, 2015; Published May 29, 2015


Copyright: © 2015 Anieting AE. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
The correlation coefficient \(r\) between the auxiliary variable \(x_i\) (the number of local government in each state) and the variable of interest \(y_i\) (the number of police stations in Nigeria in each state) is 0.60. Other results of the analysis are summarized in Tables 1-3.

### Results and discussions

Based on a sample size of six from a population of 36 units, Tables 2 and 3 show that the population total estimate using Horvitz and Thompson estimator is 1204 which is smaller than 1448, the population total estimate using Rao, Hartley and Cochran estimator. Also the variance of the estimate of the population total by Horvitz and Thompson is 690317.62 which is bigger than 102892.4, the variance of the estimate of the population total by Rao, Hartley and Cochran [4].

### Conclusion

Since the variance of the estimator given by Rao, Hartley and Cochran gives a value smaller than that given by Horvitz and Thompson with respect to sample size of six, it shows that the estimator given by Rao, Hartley and Cochran is better than that given by Horvitz and Thompson.

### References