VOL. 3, NO.6, June 2013 ISSN 2225-7217
ARPN Journal of Science and Technology

http://www.ejournalofscience.org

New Estimation Method in Two-Stage Cluster Sampling Using Finite Population

[1] D. I. Lanlege, [2] O. M. Adetutu, [3] L.A. Nafiu
[1] Department of Mathematics and Computer Science, Ibrahim Badamasi Babangida University, Lapai, Niger State, Nigeria
[2,3]Department of Mathematics and Statistics, Federal University of Technology, Minna, Niger State, Nigeria

[1] davidlanlege@yahoo.co.uk, [2] xtutu2@yahoo.com, [3] lanafiu@futminna.edu.ng

ABSTRACT

This research investigates the use of a two-stage cluster sampling design in estimating the population total. We focus on a special design where certain number of visits is being considered for estimating the population size and a weighted factor of N_i/n_i^2 is introduced. In particular, attempt was made at deriving a new method for a three-stage sampling design. In this study, we compared the newly proposed estimator with some of the existing estimators in a two-stage sampling design. Eight (8) data sets were used to justify this paper. The first four (4) data sets were obtained from Horvitz and Thompson (1952), Raj (1972), Cochran (1977) and Okafor (2002) respectively while the second four (4) data sets represent the number of diabetic patients in Niger state, Nigeria for the years 2005, 2006, 2007 and 2008 respectively. The computation was done with software developed in Microsoft Visual C++ programming language. The population totals were obtained for illustrated data and life data. We obtained the biases of the estimated population totals for illustrated and life data respectively. For all the populations considered, the biases of our proposed estimator are the least among all estimators compared. The variance of the newly proposed method is less than the variances of those of the existing methods. All the estimated population totals are also found to fall within the computed confidence intervals for α = 5%. The coefficients of variations obtained for the estimated population totals using both illustrated and life data show that our newly proposed estimator has the least coefficient of variation. Therefore, our newly proposed estimator ($\hat{Y}_{2NPE}$) is recommended when considering a two-stage cluster sampling design.

Keywords: *Sampling, cluster, two-stage, design, finite population, estimator, bias and variance.*

1. INTRODUCTION

In a census, each unit (such as person, household or local government area) is enumerated, whereas in a sample survey, only a sample of units is enumerated and information provided by the sample is used to make estimates relating to all units [1]. In designing a study, it can be advantageous to sample units in more than one-stage. The criteria for selecting a unit at a given stage typically depend on attributes observed in the previous stages [2]. Recent work on two-stage sampling includes those of [3] and [4].

In [5], comparison between two-stage cluster sampling and simple random sampling was made; and observed that two-stage cluster sampling is better in terms of efficiency. Also, [6] gives the reason for multistage sampling as administrative convenience. The work of [7] states that multistage sampling makes fieldwork and supervision relatively easy. Multistage sampling is more efficient than single stage cluster sampling ([8]; [4] and [9]). It is concluded in [10] that sub sampling has a great variety of applications. Figure 1 shows specifically a schematic representation of a two-stage sample in which two units are selected from each cluster as given by [10].

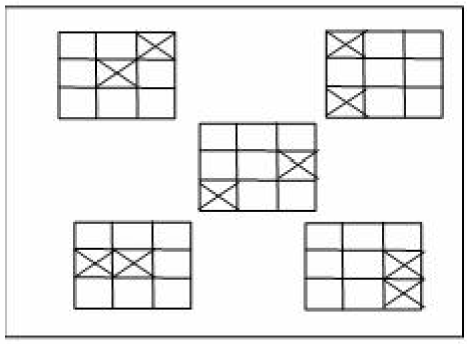

Fig 1: Schematic Representation of a Two-Stage Cluster Sampling Design

Variability in two-stage sampling includes the following:

a. In one-stage cluster sampling, the estimate varies due to one source: different samples of primary units yield different estimates.
b. In two-stage cluster sampling, the estimate varies due to two sources: different samples of primary units and then different samples of secondary units within primary units.

2. AIM AND OBJECTIVES

The aim of this research is to model a new estimator for two-stage sampling design and the main objectives are to:

a. Investigate some of the existing estimators used in two-stage cluster sampling design and compare them in terms of efficiency and administrative convenience.
b. Develop new estimator that is more efficient and precise than already existing estimators in two-stage cluster sampling design.
c. Compare these estimators (conventional and newly proposed) using eight (8) data sets.

3. DATA USED FOR THIS STUDY

There are eight (8) categories of data used in this paper. The first four (4) data sets were obtained and used as illustration as contained in [11]. The second four (4) data sets used are of secondary type and were collected from [12] and [13]. We constructed a sampling frame from all diabetic patients with chronic eye disease (Glaucoma and Retinopathy) in the twenty-five (25) Local Government Areas of the state between years 2005 and 2008.

4. MATERIALS AND METHODS

Let N denote the number of primary units in the population and n the number of primary units in the sample. Let M_i be the number of secondary units in the primary unit. The total number of secondary units in the population is

$$M = \sum_{i=1}^{N} M_i \qquad (1)$$

Let y_{ij} denote the value of the variable of interest of the jth secondary unit in the it primary unit. The total of the y-values in the it primary unit is

$$Y_i = \sum_{j=1}^{M_i} y_{ij} \qquad (2)$$

Accordingly, the population total for overall sample in a two-stage is given as

$$Y = \sum_{i=1}^{N} \sum_{j=1}^{M_i} y_{ij} \qquad (3)$$

For any estimation $\hat{\Theta}_h$ in the hath cell based on completely arbitrary probabilities of selection, the total variance is then the sum of the variances for all strata. The symbol E is used for the operator of expectation, V for the variance, and $\hat{V}$ for the unbiased estimate of V. We may then write

$$V(\hat{\Theta}_h) = \underset{1}{V}(\underset{>1}{E}(\hat{\Theta}_h)) + \underset{1}{E}(\underset{>1}{V}(\hat{\Theta}_h)) \qquad (4)$$

where ">1" is the symbol to represent all stages of sampling after the first.

The expression (4) may be written into three components as:

$$V(\hat{\Theta}_h) = \underset{1}{V}(\underset{2}{E}(\underset{>2}{E}(\hat{\Theta}_h))) + \underset{1}{E}(\underset{2}{V}(\underset{>2}{E}(\hat{\Theta}_h))) + \underset{1}{E}(\underset{2}{E}(\underset{>2}{V}(\hat{\Theta}_h))) \qquad (5)$$

In line with [14] and [7], we now present our proposed estimator for two-stage cluster sampling design. Here, a sample of primary unit is selected and then a sample of secondary units is chosen from each of the selected primary units. For example, the total number of individuals who are involved in the treatment of diabetes at hospitals in a state can be estimated by selecting a sample of Hospitals in the Cities within the Local Government Areas in the state and then collect the number of diabetic patients in each sampled hospital.

Let N be the number of primary units (Cities) in the population. For $i = 1,2,\cdots,N$, let y_i be the size of a population of secondary units (Hospitals) in the ith primary unit engaged in the treatment of diabetes within the state. Assuming that each secondary unit engages in the variable of interest at only one primary unit, then

$$y = \sum_{i=1}^{N} y_i \qquad (6)$$

Let N be the number of primary units sampled without replacement, M_i the number of secondary units in the ith sampled primary unit and m_i the number of secondary units selected without replacement from the ith sampled primary unit for $i = 1,2,\cdots,n$. An unbiased estimator of the total population at the ith primary unit in the sample is:

$$\begin{aligned} \hat{y}_i &= \frac{1}{\gamma_i} \sum_{j=1}^{n_i} \frac{N_i}{n_i^2} y_{ij} \\ &= \frac{M_i}{m_i} \sum_{j=1}^{n_i} \frac{N_i}{n_i^2} y_{ij} \end{aligned} \qquad (7)$$

where $\gamma_i = \frac{m_i}{M_i}$ is the known sampling fraction in the ith primary unit for $i = 1,2,\cdots,n$ and y_{ij} denotes the number of secondary units in the sample from the ith primary unit who engage in the variable of interest.

A proposed unbiased estimator of the population total is:

$$\begin{aligned} \hat{Y}_{2NPE} &= \frac{1}{\gamma} \sum_{i=1}^{n} y_i \\ &= \frac{1}{\gamma} \sum_{i=1}^{n} \left(\frac{1}{\gamma_i} \sum_{j=1}^{n_i} \frac{N_i}{n_i^2} y_{ij} \right) \end{aligned} \qquad (8)$$

Where $\gamma = \frac{n}{N}$ (9)

and $\gamma_i = \frac{m_i}{M_i}$ (10)

Theorem 1:

$\hat{Y}_{2NPE}$ is unbiased for the population total Y

Proof:

We know that the expectation of $\hat{y}_i$ is conditional on a sample s_1 of primary units equals y_i, that is;

$$E(\hat{y}_i|s_1) = y_i \quad (11)$$

To obtain the expected value of $\hat{Y}_{2NPE}$ over all possible samples of primary units, let

$$z_i = \begin{cases} 1 & \text{if the ith primary unit is in the sample} \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

such that $E(z_i) = \frac{n}{N} = \gamma$, the inclusion probability for a primary units under simple random sampling.

Then, the expectation of the proposed estimator $\hat{Y}_{2NPE}$ is:

$$\begin{aligned} E(\hat{Y}_{2NPE}) &= E_1\{E_2(\hat{Y}_{2NPE}|s_1)\} \\ &= E\{E(\tfrac{1}{\gamma}\textstyle\sum_{i=1}^{n} \hat{Y}_{2NPE}|s_1)\} \\ &= E\{\tfrac{1}{\gamma}\textstyle\sum_{i=1}^{n} y_i\} \\ &= E\{\tfrac{1}{\gamma}\textstyle\sum_{i=1}^{N} z_i y_i\} \\ &= \tfrac{N}{n} E(z_i)\textstyle\sum_{i=1}^{N} y_i \\ &= \tfrac{N}{n}\tfrac{n}{N}\textstyle\sum_{i=1}^{N} y_i \\ &= \textstyle\sum_{i=1}^{N} y_i \\ &= Y \end{aligned} \quad (13)$$

This implies that the proposed estimator $\hat{Y}_{2NPE}$ is unbiased.

Hence, the variance of the newly proposed estimator $\hat{Y}_{2NPE}$ of the population total is derived as follows:

We use

$$V(\hat{Y}_{2NPE}) = V\{E(\hat{Y}_{2NPE}|s_1)\} + E\{V(\hat{Y}_{2NPE}|s_1)\}$$

$$= V\left(\frac{N}{n}\sum_{i=1}^{n} y_i\right) + E\left(\frac{N^2}{n^2}\sum_{i=1}^{N} M_i(M_i - m_i) z_i \frac{\sigma_i^2}{m_i}\right)$$

$$= \frac{N(N-n)}{n}\sigma_1^2 + \frac{N}{n}\sum_{i=1}^{N} M_i(M_i - m_i)\frac{\sigma_i^2}{m_i}$$

$$\left(E(z_i) = \frac{n}{N}\right)$$

Therefore;

$$V(\hat{Y}_{2NPE}) = \frac{N(N-n)\sigma_1^2}{n} + \frac{N}{n}\sum_{i=1}^{N} M_i(M_i - m_i)\frac{\sigma_i^2}{m_i} \quad (14)$$

where $\sigma_1^2 = \frac{\sum_{i=1}^{N}(y_i - \frac{Y_{2NPE}}{N})^2}{N-1}$ (15)

and for $i = 1,2,\cdots,N$

$$\sigma_i^2 = \frac{1}{\gamma_i^2}\sum_{j=1}^{N_i}\left(\frac{N_i^2}{n_i^4} - \frac{N_i}{n_i^2}\right)(y_{ij} - \bar{y}_i)^2 \quad (16)$$

The first term on the right of the equality in equation (14) is the variance that would be obtained if every secondary units in a selected primary unit were observed, that is, if the y_i's were known for $i = 1,2,\cdots,n$. The second term contains variance due to estimating y_i's from a subsample of secondary units within the selected primary units. An unbiased estimator of the variance given in equation (14) is:

$$\hat{V}(\hat{Y}_{2NPE}) = \frac{N(N-n)s_1^2}{n} + \frac{N}{n}\sum_{i=1}^{n} M_i(M_i - m_i)\frac{s_i^2}{m_i} \quad (17)$$

where $s_1^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \frac{Y_{2NPE}}{N})^2}{n-1}$ (18)

and for $i = 1,2,\cdots,n$

$$s_i^2 = \frac{1}{\gamma_i^2}\sum_{j=1}^{n_i}\left(\frac{N_i^2}{n_i^4} - \frac{N_i}{n_i^2}\right)(y_{ij} - \bar{y}_i)^2 \quad (19)$$

Theorem 2:

$\hat{V}(\hat{Y}_{2NPE})$ is unbiased for $V(\hat{Y}_{2NPE})$

We note that

$$s_1^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}\hat{y}_i^2 - \frac{n\hat{Y}_{2NPE}^2}{n^2}\right) \quad (20)$$

Next we note that

$$E(\sum_{i=1}^{n}\hat{y}_i^2) = E\{E(\sum_{i=1}^{n}\hat{y}_i^2|s_1)\}$$

$$= E(\sum_{i=1}^{n}[V(\hat{y}_i|s_1) + \{E(\hat{y}_i|s_1)\}^2])$$
$$= E(\sum_{i=1}^{n}s_i^2 + \sum_{i=1}^{n}\hat{y}_i^2)$$
$$= E(\sum_{i=1}^{N}z_i s_i^2 + \sum_{i=1}^{N}z_i\hat{y}_i^2)$$
$$= \frac{n}{N}(\sum_{i=1}^{N}\sigma_i^2 + \sum_{i=1}^{N}y_i^2) \quad (21)$$

In addition, we also note that

$$= \frac{N(N-n)}{n}\sigma_1^2 + \frac{N}{n}\sum_{i=1}^{N}M_i(M_i - m_i)\frac{s_i^2}{m_i} + Y^2 \quad (22)$$

Together equations (20), (21) and (22) imply

$$E(s_1^2) = \frac{n}{N(n-1)}(\sum_{i=1}^{N}\sigma_i^2 + \sum_{i=1}^{N}y_i^2) - \frac{n}{N(n-1)}(\frac{N(N-n)}{n}\sigma_1^2 + \frac{N}{n}\sum_{i=1}^{N}M_i(M_i - m_i)\frac{s_i^2}{m_i} + Y^2)$$

$$= \frac{1}{n}\sum_{i=1}^{N}\sigma_1^2 + \frac{(N-1)n}{(n-1)}\{\frac{1}{N-1}(\sum_{i=1}^{N}y_i^2 - \frac{Y^2}{N})\} + \frac{(N-n)}{N(n-1)}\sigma_1^2$$

$$= \frac{1}{N}\sum_{i=1}^{N}M_i(M_i - m_i)\frac{s_i^2}{m_i} + \sigma_1^2 \quad (23)$$

Finally, we note that

$$E(\sum_{i=1}^{n}s_i^2) = E\{E(\sum_{i=1}^{n}s_i^2\,|s_1)\}$$

$$= E\{\sum_{i=1}^{N}z_i M_i(M_i - m_i)\frac{s_i^2}{m_i}\}$$
$$= \frac{n}{N}\sum_{i=1}^{N}M_i(M_i - m_i)\frac{\sigma_i^2}{m_i} \quad (24)$$

Together, equations (23) and (24) become

$$E\{\hat{V}(\hat{Y}_{2NPE})\} = \frac{N(N-n)}{n}\sigma_1^2 + \frac{(N-n)}{n}\sum_{i=1}^{N}M_i(M_i - m_i)\frac{\sigma_i^2}{m_i} + \sum_{i=1}^{N}M_i(M_i - m_i)\frac{\sigma_i^2}{m_i}$$

$$= \frac{N(N-n)}{n}\sigma_1^2 + \frac{N}{n}\sum_{i=1}^{N}M_i(M_i - m_i)\frac{\sigma_i^2}{m_i}$$
$$= V(\hat{Y}_{2NPE})$$

That is;

$$E\{\hat{V}(\hat{Y}_{2NPE})\} = V(\hat{Y}_{2NPE})$$

Hence, $\hat{V}(\hat{Y}_{2NPE})$ is an unbiased sample estimator of the proposed estimator ($\hat{Y}_{2NPE}$) in two-stage cluster sampling design.

5. RESULTS

5.1 Estimated Population Totals

The estimated population totals computed with the aid of software developed are given in table 1 for the illustrated data and in table 2 for the life data respectively.

Table 1: Estimated Population Totals for Illustrated Data

Estimator	Case I	Case II	Case III	Case IV
$\hat{Y}_{HH}$	481	103,976	38	14,095
$\hat{Y}_{HHG}$	450	99,363	50	15,105
$\hat{Y}_{HT}$	392	98,867	32	14,583
$\hat{Y}_{RHC}$	384	142,745	29	13,886
$\hat{Y}_{C}$	429	121,717	40	15,210
$\hat{Y}_{T}$	405	98,951	37	14,275
$\hat{Y}_{O}$	472	137,264	45	13,989
$\hat{Y}_{2NPE}$	417	116,761	30	15,207

Table 2: Estimated Population Totals for Life Data

Estimator	POP1	POP 2	POP 3	POP 4
$\hat{Y}_{HH}$	28,009	28,610	29,420	29,426
$\hat{Y}_{HHG}$	24,163	26,501	26,605	28,123
$\hat{Y}_{HT}$	26,551	26,605	28,664	29,222
$\hat{Y}_{RHC}$	25,804	27,311	28,531	28,791
$\hat{Y}_{C}$	26,382	26,431	27,538	27,654
$\hat{Y}_{T}$	27,332	28,645	29,402	29,531
$\hat{Y}_{O}$	24,222	26,581	28,325	29,041
$\hat{Y}_{2NPE}$	25,841	26,675	27,204	29,300

5.2 Biases for the Estimated Population Totals

We consider the biases arising from estimated population totals using a two-stage cluster sampling design and the values presented in table 3 for illustrated data and in table 4 for life data respectively.

Table 3: Biases of Estimated Population Totals for Illustrated Data

Estimator	Case I	Case II	Case III	Case IV
$\hat{Y}_{HH}$	30	587	7	556
$\hat{Y}_{HHG}$	20	389	10	565
$\hat{Y}_{HT}$	38	532	4	250
$\hat{Y}_{RHC}$	45	569	3	669
$\hat{Y}_{C}$	25	548	5	308
$\hat{Y}_{T}$	41	563	11	551
$\hat{Y}_{O}$	56	486	5	666

$\hat{Y}_{2NPE}$	13	245	1	135

Table 4: Biases of Estimated Population Totals for Life Data

Estimator	POP1	POP 2	POP 3	POP 4
$\hat{Y}_{HH}$	218	153	153	193
$\hat{Y}_{HHG}$	162	192	133	209
$\hat{Y}_{HT}$	132	119	148	148
$\hat{Y}_{RHC}$	216	193	158	158
$\hat{Y}_{C}$	146	170	155	179
$\hat{Y}_{T}$	205	219	135	208
$\hat{Y}_{O}$	147	150	159	187
$\hat{Y}_{2NPE}$	124	107	105	112

5.3 Estimated Variances for the Estimated Population Totals

The estimated variances computed using the software developed are also given in table 5 for illustrated data and in table 6 for life data respectively.

Table 5: Variances Of The Estimated Population Totals For Illustrated Data

Estimator	Case I	Case II	Case III	Case IV
$\hat{V}(\hat{Y}_{HH})$	1,947.9059	2,981,018.9548	3.4169	60,987.3673
$\hat{V}(\hat{Y}_{HHG})$	1,660.2051	2,669,050.4552	3.2675	59,340.6001
$\hat{V}(\hat{Y}_{HT})$	1,482.6011	2,398,957.0023	1.8322	56,416.5692
$\hat{V}(\hat{Y}_{RHC})$	1,419.6012	1,995,716.9763	1.8023	55,905.4322
$\hat{V}(\hat{Y}_{C})$	1,394.7069	1,946,010.0053	1.8009	55,830.4230
$\hat{V}(\hat{Y}_{T})$	1,052.4058	1,760,015.4634	1.6794	55,590.1833
$\hat{V}(\hat{Y}_{O})$	989.4093	1,720,450.0545	1.2973	54,865.4042
$\hat{V}(\hat{Y}_{2NPE})$	915.6003	1,672,065.0125	1.2333	49,077.4359

Table 6: Variances of the Estimated Population Totals for Life Data

Estimator	Population 1	Population 2	Population 3	Population 4
$\hat{V}(\hat{Y}_{HH})$	19,404.9028	18,484.2603	17,066.7829	16,829.1924
$\hat{V}(\hat{Y}_{HHG})$	17,460.6223	18,095.4760	15,710.4432	16,777.8457
$\hat{V}(\hat{Y}_{HT})$	14,910.3442	16,251.3901	12,770.1662	14,794.2816
$\hat{V}(\hat{Y}_{RHC})$	12,982.5024	13,097.7309	12,663.9011	13,772.7989
$\hat{V}(\hat{Y}_{C})$	11,095.2835	12,813.3486	12,153.6749	13,599.0038
$\hat{V}(\hat{Y}_{T})$	10,832.4035	12,712.5769	11,701.8641	13,410.0793
$\hat{V}(\hat{Y}_{O})$	10,630.8606	11,674.8805	11,477.7903	12,319.2565
$\hat{V}(\hat{Y}_{2NPE})$	10,131.3327	10,807.2087	10,981.8622	11,790.9118

5.4 Standard Errors for the Estimated Population Totals

Tables 7 and 8 give the standard errors of the estimated population totals using a two-stage cluster sampling design.

Table 7: Standard Errors for Estimated Population Totals for Illustrated Data

Estimator	Case I	Case II	Case III	Case IV
$\hat{Y}_{HH}$	44.1351	1,726.5628	1.8484	246.9562
$\hat{Y}_{HHG}$	40.7456	1,633.7229	1.8076	243.5993
$\hat{Y}_{HT}$	38.5046	1,548.8567	1.3536	237.5218
$\hat{Y}_{RHC}$	37.6776	1,412.6985	1.3425	236.4432
$\hat{Y}_{C}$	37.3458	1,394.9946	1.3420	236.2837
$\hat{Y}_{T}$	32.4408	1,326.6557	1.2959	235.7757
$\hat{Y}_{O}$	31.4588	1,311.6593	1.1320	234.2337
$\hat{Y}_{2NPE}$	30.2589	1,293.0835	1.1105	221.5343

Table 8: Standard Errors for Estimated Population Totals for Life Data

Estimator	POP1	POP 2	POP 3	POP 4
$\hat{Y}_{HH}$	139.3015	135.9568	123.9356	129.7274
$\hat{Y}_{HHG}$	132.1386	134.5194	118.9091	129.5293
$\hat{Y}_{HT}$	122.1079	127.4809	107.2058	121.6317
$\hat{Y}_{RHC}$	113.9408	114.4453	106.7591	117.3576
$\hat{Y}_{C}$	105.3342	113.1961	104.5861	116.6148
$\hat{Y}_{T}$	104.0788	112.7500	102.6237	115.8019
$\hat{Y}_{O}$	103.1061	108.0504	101.6363	110.9921
$\hat{Y}_{2NPE}$	100.6545	103.9577	99.4164	108.5860

5.5 Confidence Intervals for the Estimated Population Totals

The confidence intervals for estimated population totals are presented in table 9 for illustrated data and in table 10 for life data.

Table 9: Confidence Intervals for Estimated Population Totals for Illustrated Data

Estimator	Case I	Case II	Case III	Case IV
$\hat{Y}_{HH}$	(394,568)	(100592, 107360)	(34,42)	(13611,14579)
$\hat{Y}_{HHG}$	(370,530)	(96161,102565)	(47,54)	(14628,15582)

VOL. 3, NO.6, June 2013 ISSN 2225-7217
ARPN Journal of Science and Technology

http://www.ejournalofscience.org

$\hat{Y}_{HT}$	(317,467)	(95831,101903)	(29,35)	(14117,15049)
$\hat{Y}_{RHC}$	(346,423)	(139976,145514)	(26,32)	(13423,14349)
$\hat{Y}_{C}$	(356,502)	(118983,124451)	(37,43)	(14747,15673)
$\hat{Y}_{T}$	(341,469)	(96351,101551)	(34,40)	(13913,14837)
$\hat{Y}_{O}$	(410,534)	(134693,139835)	(43,47)	(13530,14448)
$\hat{Y}_{2NPE}$	(358,476)	(114227,119295)	(28,32)	(14773,15641)

Table 10: Confidence Intervals of Estimated Population Totals for Life Data

Estimator	Population 1	Population 2	Population 3	Population 4
$\hat{Y}_{HH}$	(27740,28280)	(28340,28880)	(29180,29660)	(29170,29680)
$\hat{Y}_{HHG}$	(23900,24420)	(26240,26760)	(26370,26840)	(27870,28380)
$\hat{Y}_{HT}$	(26310,26790)	(26360,26850)	(28450,28870)	(28980,29460)
$\hat{Y}_{RHC}$	(25580,26030)	(27090,27540)	(28320,28740)	(28560,29020)
$\hat{Y}_{C}$	(26180,26590)	(26210,26650)	(27330,27740)	(27430,27880)
$\hat{Y}_{T}$	(27130,27540)	(28420,28870)	(29200,29600)	(29300,29760)
$\hat{Y}_{O}$	(24020,24420)	(26290,26710)	(28130,28520)	(28820,29260)
$\hat{Y}_{2NPE}$	(25640,26040)	(26400,26810)	(27010,27400)	(29090,29510)

5.6 Coefficients of Variation for the Estimated Population Totals

Coefficients of Variations for the estimated population totals are given in table 11 for illustrated data and in table 12 for life data.

Table 11: Coefficients of Variation for Illustrated Data

Estimator	Case I	Case II	Case III	Case IV
$\hat{Y}_{HH}$	9.18%	1.66%	4.86%	1.64%
$\hat{Y}_{HHG}$	9.05%	1.64%	3.62%	1.61%
$\hat{Y}_{HT}$	9.82%	1.57%	4.23%	1.63%
$\hat{Y}_{RHC}$	9.81%	0.98%	4.63%	1.90%
$\hat{Y}_{C}$	8.71%	1.51%	3.36%	1.90%
$\hat{Y}_{T}$	8.01%	1.34%	3.50%	1.64%
$\hat{Y}_{O}$	6.67%	0.96%	2.53%	1.67%
$\hat{Y}_{2NPE}$	7.26%	1.11%	3.70%	1.46%

Table 12: Coefficients of Variation for Life Data

Estimator	POP1	POP 2	POP 3	POP 4
$\hat{Y}_{HH}$	0.51%	0.48%	0.46%	0.44%
$\hat{Y}_{HHG}$	0.55%	0.51%	0.45%	0.46%
$\hat{Y}_{HT}$	0.46%	0.46%	0.39%	0.42%
$\hat{Y}_{RHC}$	0.44%	0.42%	0.41%	0.41%
$\hat{Y}_{C}$	0.41%	0.42%	0.40%	0.43%
$\hat{Y}_{T}$	0.39%	0.37%	0.39%	0.39%
$\hat{Y}_{O}$	0.42%	0.41%	0.38%	0.38%
$\hat{Y}_{2NPE}$	0.38%	0.33%	0.37%	0.37%

6. DISCUSSION

The population totals obtained for illustrated data are given in table 1 while the population totals obtained for life data are given in table 2. Table 3 give the biases of the estimated population totals for illustrated data for our own estimator as 13, 245, 1 and 135 for cases I – IV respectively while table 4 gives those of the four life data sets as 124, 107, 105 and 112 respectively. This implies that our own estimator has the least biases using both data sets.

Table 5 shows the variances obtained using illustrated data for our own estimator as 915.6003, 1672065.0125, 1.2333 and 49077.4359 for cases I – IV respectively while table 6 shows those of life data sets as 10131.3327, 10807.2087, 10981.8622 and 11790.9118 respectively meaning that our own estimator has the least variances using both data sets.

Table 7 shows the obtained standard errors for the estimated population totals using illustrated data for our own estimator as 30.2589, 1293.0835, 1.1105 and 221.5343 for cases I – IV respectively while table 8 shows those of life data sets as 100.6545, 103.9577, 99.4164 and 108.5860 respectively meaning that our own estimator has the least standard errors using both data sets.

The confidence intervals of the estimated populations in table 1 are given in 9 using α = 5 %. The confidence intervals of the estimated populations in table 2 are also given in table 10 for the same confidence intervals. Tables 9 and 10 show that all the estimated population totals fall within the computed intervals as expected.

For our own estimator, table 11 gives the coefficients of variations for the estimated population totals using illustrated data as 7.26%, 1.11%, 3.7% and 1.46% for cases I – IV respectively while table 12 gives that of life data sets as 0.38%, 0.33%, 0.37% and 0.37% respectively which means that our newly proposed two-stage cluster estimator has the least coefficient of variation.

7. CONCLUSION

When an unbiased estimator of high precision and an unbiased sample estimate of its variance is required for a two-stage sampling design, estimator $\hat{Y}_{2NPE}$ is preferred and hence recommended for estimating population totals.

REFERENCES

[1] Kish, L. (1967). Survey Sampling. New York: John Wiley and Sons.

[2] Kuk, A. (1988). Estimation of Distribution Functions and Medians under Sampling with Unequal Probabilities. Biometrika, 75: 97-103.

[3] Goldstein, H. (1995). Multilevel Statistical Models. New York: Halstead Press.

[4] Tate, J.E. and M. G. Hudgens (2007). Estimating Population Size with Two-stage and Three-stage sampling Designs. American Journal of Epidemiology, 165(11): 1314-1320.

[5] Fink, A. (2002). How To Sample In Surveys. Thousand Oaks, C.A.: Sage Publications.

[6] Kalton, G. (1983). Introduction to Survey Sampling. C.A.: Sage Publications.

[7] Okafor, F. (2002). Sample Survey Theory with Applications. Nigeria: Afro-Orbis Publications.

[8] Henry, G. T. (1990). Practical Sampling. Thousand Oaks, C.A.: Sage Publications.

[9] Nafiu, L. A., Oshungade, I. O. and Adewara, A. A., (2012): “Alternative Estimation Method for a Three-Stage Cluster Sampling in Finite Population”. American Journal of Mathematics and Statistics (AJMS). Vol. 2. No. 6: 12-17.

[10] Cochran, W.G. (1977). Sampling Techniques. Third Edition. New York: Wiley and Sons.

[11] Nafiu, L. A. (2012). An Alternative Estimation Method for Multistage Cluster Sampling in Finite Population, Ph.D. thesis. University of Ilorin, Nigeria.

[12] Niger State (2007). Niger State Statistical Year Book. Nigeria: Niger Press Printing & Publishing.

[13] National Bureau of Statistics (2007). Directory of Health Establishments in Nigeria. Nigeria: Abuja Printing Press.

[14] Thompson, S. K. (1992). Sampling. New York: John Wiley and Sons.

AUTHORS’ PROFILES

Mr. D. I. Lanlege received the degree in Mathematics from University of Ado-Ekiti, Ado-Ekiti, Ekiti State, Nigeria. He is a research student of Federal University of Technology, Minna, Niger State, Nigeria. Currently, he is an Assistant Lecturer at the Ibrahim Badamasi Babangida University, Lapai, Niger State, Nigeria.

Mr. O. M. Adetutu received the degree in Statistics from University of Ilorin, Ilorin, Kwara State, Nigeria. He is currently an Assistant Lecturer at the Federal University of Technology, Minna, Niger State, Nigeria.

Dr. L. A. Nafiu received the degree in Statistics from University of Ilorin, Ilorin, Kwara State, Nigeria. Currently, he is a Lecturer I at the Federal University of Technology, Minna, Niger State, Nigeria.