Comparison of One-Stage, Two-Stage, and Three-Stage Estimators Using Finite Population.

Lukman Abiodun Nafiu, Ph.D.

Department of Mathematics and Statistics, Federal University of Technology, Minna, Nigeria.

E-mail: lanconserv@yahoo.com

ABSTRACT

This paper compares the three proposed estimators in multistage cluster sampling design given by Nafiu (2012) for estimating population totals. Eight (8) sets of data were used to compare the proposed alternative estimators and it was observed that the three-stage sampling design is better than the two-stage sampling design which is in turn better than the one-stage sampling design in all cases considered.

(Keywords: sampling, cluster, one-stage, two-stage, three-stage, design, estimator, bias, standard error and variance)

INTRODUCTION

In designing a study, it can be advantageous to sample units in more than one-stage (Cochran, 1977; Kalton, 1983; and Okafor, 2002). Adams *et al.* (2003) opines that if it costs little to determine the attributes that are necessary to classify the units, it can be cost efficient to sample in stage one and then in stage two to subsample the clusters at different stages. The estimation of the finite population distribution function under several sampling strategies based on a probability proportional to size (PPS) in multistage sampling is a type of strategy that gives good results (Kuk, 1988).

Multistage sampling is where the researcher divides the population into stages, samples the stages and then resample, repeating the process until the ultimate sampling units are selected at the last of the hierarchical levels (Goldstein, 1995 and Thompson, 1992). Multistage sampling according to Ma *et al.* (2006) is generally used when it is costly or impossible to form a list of all the units in the target population. Multistage sampling is often more precise than a simple random sample of the same cost, and it is for this reason that this method is being investigated in this paper. The criteria for selecting a unit at a given stage typically depend on attributes observed in the previous stage (Kish, 1967; Hansen *et al.*, 1975; Fink, 2002; and Tate and Hudgens, 2007).

Variability in multistage sampling includes the following:

(i) In one-stage cluster sampling, the estimate varies because of different samples of primary units.

(ii) In two-stage cluster sampling, the estimate varies because of different samples of primary units and samples of secondary units within primary units.

(iii) In three-stage cluster sampling, the estimate varies because of different samples of primary units, samples of secondary units within primary units and samples of tertiary units within secondary units.

(iv) In general, if there are k stages of sub sampling, there will be k sources of variability.

MATERIALS AND METHODS

One-Stage Cluster Sampling Scheme

Alternative estimator given by Nafiu (2012) in one-stage cluster sampling is:

$$\hat{Y}_{1NPE} = \frac{1}{\gamma}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\frac{N_i}{n_i^2}y_{ij} \quad \textbf{(1)}$$

where $\gamma = \frac{n}{N}$ is the known sampling fraction and y_{ij} denotes the number of individuals in the sample.

An unbiased estimator of the variance is:

$$\hat{V}(\hat{Y}_{1NPE}) = \frac{N^2}{n^2}\sum_{i=1}^{n}\frac{N_i(N_i-n_i^2)}{n_i^4})s_i^2 \quad \textbf{(2)}$$

where $s_i^2 = \frac{1}{n_i-1}\sum_{j=1}^{n_i}(y_{ij} - \hat{Y}_{1NPE})^2$

<u>**Two-Stage Cluster Sampling Scheme**</u>

Alternative estimator given by Nafiu (2012) in two-stage cluster sampling is:

$$\hat{Y}_{2NPE} = \frac{1}{\gamma}\sum_{i=1}^{n}(\frac{1}{\gamma_i}\sum_{j=1}^{n_i}\frac{N_i}{n_i^2}y_{ij}) \quad \textbf{(3)}$$

where $\gamma = \frac{n}{N}$ **(4)**

and $\gamma_i = \frac{m_i}{M_i}$ **(5)**

with an unbiased estimator of the variance as:

$$\hat{V}(\hat{Y}_{2NPE}) = \frac{N(N-n)s_1^2}{n} + \frac{N}{n}\sum_{i=1}^{n}M_i(M_i - m_i)\frac{s_i^2}{m_i} \quad \textbf{(6)}$$

where $s_1^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \frac{\hat{Y}_{2NPE}}{N})^2}{n-1}$ **(7)**

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

$$s_i^2 = \frac{1}{\gamma_i^2}\sum_{j=1}^{n_i}(\frac{N_i^2}{n_i^4} - \frac{N_i}{n_i^2})(y_{ij} - \bar{y}_i)^2 \quad \textbf{(8)}$$

<u>**Three-Stage Cluster Sampling Scheme**</u>

Alternative estimator given by Nafiu (2012) in three-stage cluster sampling is:

$$\hat{Y}_{3NPE} = \frac{N}{n}\sum_{i=1}^{n}\{\frac{M_i}{m_i}\sum_{j=1}^{m_i}(\frac{K_{ij}}{k_{ij}}\sum_{l=1}^{k_{ij}}\frac{N_i}{n_i^2}y_{ijl})\} \quad \textbf{(9)}$$

with an unbiased estimator of the variance as:

$$\hat{V}(\hat{Y}_{3NPE}) = N(N-n)\frac{s_1^2}{n} + \frac{N}{n}\sum_{i=1}^{n}M_i(M_i - m_i)\frac{s_i^2}{m_i} + \frac{N}{n}\sum_{i=1}^{n}\frac{M_i}{m_i}\sum_{j=1}^{m_i}K_{ij}(K_{ij} - k_{ij})\frac{s_{ij}^2}{k_{ij}} \quad \textbf{(10)}$$

where,

$$s_1^2 = \frac{\sum_{i=1}^{n}(y_i - \frac{\hat{Y}_{3NPE}}{n})^2}{n-1} \quad \textbf{(11)}$$

$$s_i^2 = \frac{\sum_{j=1}^{m_i}(y_{ij} - \frac{y_i}{m_i})^2}{m_i-1} \quad \textbf{(12)}$$

$$s_{ij}^2 = \frac{K_{ij}^2}{k_{ij}^2}\sum_{l=1}^{k_{ij}}(\frac{N_i^2}{n_i^4} - \frac{N_i}{n_i^2})(y_{ijl} - \bar{y}_{ij})^2 \quad \textbf{(13)}$$

RESULTS

<u>**Estimated Population Totals in each of the Sampling Designs**</u>

The estimated population totals are given in Tables 1 and 2 for illustrated and life data, respectively.

<u>**Biases for the Estimated Population Totals in Each Design**</u>

Tables 3 and 4 give the biases of the estimated population totals for illustrated and life data, respectively.

<u>**Variances for the Estimated Population Totals in Each Design**</u>

The estimated variances computed for illustrated data and life data are given in Tables 5 and 6, respectively.

<u>**Standard Errors for the Estimated Population Totals**</u>

Tables 7 and 8 give the standard errors of the estimated population totals for illustrated data and life data, respectively.

Table 1: Estimated Population Totals for Illustrated Data.

Estimator	Case I	Case II	Case III	Case IV
$\hat{Y}_{1NPE}$	421	98,966	39	13,855
$\hat{Y}_{2NPE}$	417	116,761	30	15,207
$\hat{Y}_{3NPE}$	492	99,136	42	15,016

Table 2: Estimated Population Totals for Life Data.

Estimator	Population 1	Population 2	Population 3	Population 4
$\hat{Y}_{1NPE}$	24,639	25,010	26,551	28,407
$\hat{Y}_{2NPE}$	25,841	26,675	27,204	29,300
$\hat{Y}_{3NPE}$	26,151	26,625	27,511	28,090

Table 3: Biases of Estimated Population Totals for Illustrated Data.

Estimator	Case I	Case II	Case III	Case IV
$\hat{Y}_{1NPE}$	16	264	2	143
$\hat{Y}_{2NPE}$	13	245	1	135
$\hat{Y}_{3NPE}$	11	219	1	112

Table 4: Biases of Estimated Population Totals for Life Data.

Estimator	Population 1	Population 2	Population 3	Population 4
$\hat{Y}_{1NPE}$	129	149	128	122
$\hat{Y}_{2NPE}$	124	107	105	112
$\hat{Y}_{3NPE}$	112	104	103	107

Table 5: Variances of the Estimated Population Totals for Illustrated Data.

Estimator	Case I	Case II	Case III	Case IV
$\hat{V}(\hat{Y}_{1NPE})$	3,052.6168	30,401,182.1107	1.3703	1,487,195.0244
$\hat{V}(\hat{Y}_{2NPE})$	915.6003	1,672,065.0125	1.2333	49,077.4359
$\hat{V}(\hat{Y}_{3NPE})$	274.6806	91,963.5764	1.1105	1,619.5559

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

Estimator	Population 1	Population 2	Population 3	Population 4
$\hat{V}(\hat{Y}_{1NPE})$	11,257.1327	12,008.3612	12,202.6286	13,101.9827
$\hat{V}(\hat{Y}_{2NPE})$	10,131.3327	10,807.2087	10,981.8622	11,790.9118
$\hat{V}(\hat{Y}_{3NPE})$	9,118.2037	9,726.4809	9,883.6215	10,611.8216

Table 7: Standard Errors of Estimated Population Total for Illustrated Data.

Estimator	Case I	Case II	Case III	Case IV
$\hat{Y}_{1NPE}$	55.2505	5,513.7267	1.1706	1,219.5061
$\hat{Y}_{2NPE}$	30.2589	1,293.0835	1.1105	221.5343
$\hat{Y}_{3NPE}$	16.5785	302.2597	1.0538	40.2437

Table 8: Standard Errors of Estimated Population Total for Life Data.

Estimator	Population 1	Population 2	Population 3	Population 4
$\hat{Y}_{1NPE}$	106.0996	109.5827	110.4655	114.4640
$\hat{Y}_{2NPE}$	100.6545	103.9577	99.4164	108.5860
$\hat{Y}_{3NPE}$	95.4893	98.6229	104.7944	103.0137

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

Estimator	Case I	Case II	Case III	Case IV
$\hat{Y}_{1NPE}$	(313,529)	(88159,109723)	(37,41)	(11465,16245)
$\hat{Y}_{2NPE}$	(358,476)	(114227,119295)	(28,32)	(14773,15641)
$\hat{Y}_{3NPE}$	(460,524)	(98542,99730)	(40,44)	(14939,15095)

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

Estimator	Population 1	Population 2	Population 3	Population 4
$\hat{Y}_{1NPE}$	(24430,24850)	(24800,25220)	(26330,26770)	(28180,28630)
$\hat{Y}_{2NPE}$	(25640,26040)	(26400,26810)	(27010,27400)	(29090,29510)
$\hat{Y}_{3NPE}$	(25960,26340)	(26410,26800)	(27310,27720)	(27890,28290)

Table 11: Coefficients of Variation for Estimated Population Totals for Illustrated Data.

Estimator	Case I	Case II	Case III	Case IV
$\hat{Y}_{1NPE}$	13.12%	5.57%	3.00%	8.80%
$\hat{Y}_{2NPE}$	7.26%	1.11%	3.70%	1.46%
$\hat{Y}_{3NPE}$	3.98%	0.31%	2.51%	0.27%

Table 12: Coefficients of Variation for Estimated Population Totals for Life Data.

Estimator	Population 1	Population 2	Population 3	Population 4
$\hat{Y}_{1NPE}$	0.40%	0.42%	0.41%	0.40%
$\hat{Y}_{2NPE}$	0.38%	0.33%	0.37%	0.37%
$\hat{Y}_{3NPE}$	0.34%	0.33%	0.35%	0.37%

Confidence Intervals of the Estimated Population Totals

Tables 9 and 10 gives the confidence intervals for illustrated data and life data, respectively.

Coefficients of Variation for the Estimated Population Totals

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

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 gives the biases of the estimated population totals for illustrated data while Table 4 gives those of the four life data. It is observed that a three-stage estimator has the least biases using both data sets.

Table 5 shows the variances obtained using illustrated data while Table 6 shows those of life data sets and it is observed that a three-stage estimator has the least variances using both data sets.

Table 7 gives the obtained standard errors for the estimated population totals using illustrated data while Table 8 gives those of life data showing that a three-stage estimator has the least standard errors using both data sets.

The confidence intervals of the estimated populations in Tables 9 and 10 for illustrated data and life data respectively show that all the estimated population totals fall within the computed intervals as expected.

Table 11 gives the coefficients of variation for the estimated population totals using illustrated data while Table 12 gives those of life data. This shows that a three-stage estimator has the least coefficient of variation.

CONCLUSION

We compared estimation methods in one-stage, two-stage and three-stage cluster sampling designs and found out that when an unbiased estimator of high precision and an unbiased sample estimate of its variance is required, the proposed three-stage sampling design estimator $(\hat{Y}_{3NPE})$ given by Nafiu (2012) is preferred and hence recommended.

REFERENCES

1. Adams, J.L., S.L. Wickstrom, M.J. Burgess, P.P. Lee, and J.J. Escarce. 2003. "Sampling Patients within Physician Practices and Health Plans: Multistage Cluster Samples in Health Services Research". *The Global Journal for Improving Health Care Delivery and Policy*. 38:1625-1640.

2. Cochran, W.G. 1977. *Sampling Techniques. Third Edition*. Wiley and Sons, New York, NY.

3. Fink, A. 2002. *How To Sample In Surveys*. Sage Publications: Thousand Oaks, CA.

4. Goldstein, H. 1995. *Multilevel Statistical Models*. Halstead Press, New York, NY.

5. Hansen, M.H., W.N. Hurwitz, and W.G. Madow. 1975. *Sample Survey Methods and Theory.* Wiley and Sons: New York, NY.

6. Kalton, G. 1983. *Introduction to Survey Sampling*. Sage Publications: Los Angeles, CA.

7. Kish, L. 1967. Survey Sampling. John Wiley and Sons, New York, NY.

8. Kuk, A. 1988. "Estimation of Distribution Functions and Medians under Sampling with Unequal Probabilities". *Biometrika*. 75: 97-103.

9. Ma, X., M. Spe, S.P. Al-Harbi, and Y. Efendiev. 2006. "A Multistage Sampling Method for Rapid Quantification of Uncertainty in History Matching Geological Models". Annual Technical Conference and Exhibition. Austin, TX. 24-27.

10. Nafiu, L.A. 2012. "An Alternative Estimation Method for Multistage Cluster Sampling in Finite Population". Unpublished Ph.D. Thesis. University of Ilorin, Nigeria.

11. Okafor, F. 2002. *Sample Survey Theory with Applications*. Afro-Orbis Publications: Lagos, Nigeria.

12. 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.

13. Thompson, S.K. 1992. *Sampling*. John Wiley and Sons: New York, NY.

ABOUT THE AUTHORS

Dr. Lukman Abiodun Nafiu, is a Lecturer I in the Department of Mathematics and Statistics, Federal University of Technology, Minna, Nigeria. He holds a Ph.D. degree in Statistics. His research interests are in the areas of sample survey designs.

SUGGESTED CITATION

Nafiu, L.A. 2012. "Comparison of One-Stage, Two-Stage, and Three Stage Estimators Using Finite Population". *Pacific Journal of Science and Technology*. 13(2):166-171.

Pacific Journal of Science and Technology