

On The Estimation of Households Characteristics Using Two-Stage Sampling Scheme

O. A. Ayansola¹; E. O. Ogunjobi²; J. T. Olajide³

¹Department of Mathematics and Statistics,
 The Polytechnic, Ibadan,
 Ibadan, Oyo State.
 email: eyitayo_ogunjobi@yahoo.com

¹Department of Mathematics and Statistics,
 The Ibarapa, Polytechnic,
 Eruwa, Oyo State.

Abstract— This paper is devoted to the empirical study of estimation of population totals and averages of study variables such as total number of persons, male, female and age, income per household head in a local area. For this purpose, when the population is structured as clusters, two stage sampling with unequal first and second stage units is suggested. The 95% confidence interval for both total and average was also adopted as one of the tools to achieve the set objectives of determining the totals and averages. The data used in this study was primarily collected in three Enumeration Area (E.A) out of the twenty-four Enumeration Area in Ibadan North-West.

Keywords: *Two-Stage Sampling, Population total, Enumeration area, Confidence interval, Primary data.*

1. INTRODUCTION

Estimation is one of the main purposes of statistics. The basic idea is that we take a sample of data and use it to make inferences about the population of interest. Estimation involves the calculation of confidence intervals for some statistic, such as Population Total, Population Mean etc.

For the purpose of estimating population parameters such as mean, total etc. when the population of interest is structured as clusters as in this study; two stage sampling with unequal first and second stage units is suggested.

It is evident that various authors have used two-stage sampling in different ways. Few of whom are cited briefly in this paper (see [1],[2],[3],[4],[5],[6]).

In the planning of sample surveys, variance estimation gives information on the accuracy of the estimators and allows to build confidence intervals. The confidence interval is given as

$$100(1 - \alpha)\% C.I = \hat{t} \pm Z_{1-\frac{\alpha}{2}} [V(\hat{t})]^{\frac{1}{2}} \quad (1)$$

In this paper, our aim is to use interval estimation to estimate the population totals and averages of households in Ibadan North-West using two-stage sampling design of unequal sizes.

II. MATERIALS AND METHODS

Sampling Strategy and Estimation Procedure

Let us consider a population of size divided into N first stage units (clusters) each having M_i second-stage units. A sample of size n is selected from N Clusters with the help of simple random sampling without replacement (SRSWOR) scheme. From each selected n-clusters, a random sample of size m_i is drawn from M_i with SRSWOR scheme.

According to the sampling set-up, we define symbolically the followings:

$$V(\hat{Y}) = N \left(\frac{N-n}{n} \right) S_1^2 + \frac{N}{n} \sum_{i=1}^n M_i \left(\frac{M_i - m_i}{m_i} \right) S_{wi}^2 \quad (2)$$

where

$$\bar{y}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} y_{ij}, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n M_i \bar{y}_i, \quad \hat{Y} = N\bar{y}$$

$$S_1^2 = \frac{\sum_{i=1}^n (M_i \bar{y}_i)^2 - n \left(\frac{1}{n} \sum_{i=1}^n M_i \bar{y}_i \right)^2}{n-1}, \quad S_{wi}^2 = \frac{\sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i)^2}{m_i - 1}$$

Empirical Study

To show the usefulness of suggested methodology presented in this paper, primary data collected from three enumeration areas in Ibadan North-West were used. The three enumeration areas are Oke Padre, Salvation Army and Ekotedo.

The population size is divided into twenty-four enumeration areas (N = 24) out of which three enumeration areas (n = 3) were selected at random with total number of households; containing $M_1 = 109$, $M_2 = 85$,

$M_3 = 64$ respectively. Of the selected enumeration areas, the number of randomly selected households is $m_1 = 53$, $m_2 = 40$, and $m_3 = 23$. The number of persons in each of these randomly selected households was examined.

The numerical values of the estimates of population means and population totals and 95% Confidence Intervals were computed.

III. ANALYSIS

The analysis begins with computation of relevant formulae for the mean, total and their standard errors using primary data collected from three of the twenty four enumeration areas in Ibadan North-West Local Government of Oyo-State. Thereafter, confidence intervals are computed for achieving targeted objectives.

IV. RESULTS

Table 1 and Table 2 show the results with regards to the population means and totals respectively.

TABLE 1: Descriptive statistics on averages

Estimator	Average No of persons	Average No of males	Average No of females	Average amount of house rent	Average Ages
Mean	401	221	180	₦2009	51
95% C.I	(139, 663)	(58,384)	(37, 323)		

α — Level of significance: 5%

TABLE 2: Descriptive statistics on totals

Estimator	Tot. no of pers.	Tot.no of males	Tot. no of females
Total	9620	5307	4313
95% C.I.	(3221, 15919)	(1402, 9212)	(917, 7709)

α — Level of significance: 5%

V. DISCUSSIONS

Based on the sample data, it was noticed that one would be 95% confident that the “true” population values will lie within the interval as shown in Table 1 and Table 2.

In addition, the average amount paid on house rent and average age of household heads are obtained as ₦2009 and 51years, respectively.

VI. CONCLUSION

The need for meaningful development in human life necessitates the need for information for the purpose from which government, industrialists, and scientists to mention but a few obtained facts for their respective activities.

Household survey involves fact finding about people living in an area in order to investigate their living conditions by computing some statistical estimators using the information obtained from the area. Among the information required are; number of persons in each household, type of house occupied by the households, monthly income of household, amount of money paid per month (if the household is rented occupier), age of person in the household, etc.

This paper examined the empirical analysis of households’ data which was primarily collected, by employing two-stage sampling scheme with unequal size clusters. From the empirical results, the estimated parameters will be useful in the provision of social services, satisfying consumers demand and measuring standard of living etc.

REFERENCES

[1] Hossian, M.I and Ahmed, M.S, A class of Predictive Estimators in Two-stage Sampling using Auxiliary Information, *Information and Management Sciences*, vol. 12(1), 2001, pp. 49-55.

[2] Scott. A and Smith, T.M.P., Estimation in Multistage surveys, *Journal of American Statistical Association*, vol.64, 1969,pp. 830-840.

[3] Sahoo, L.N. and Panda, P., A Class of Estimators in Two-Stage Sampling With Varying Probabilities, *South African Statistical Journal*, vol. 31, 1997, pp. 151-160.

[4] Sahoo, L.N. and Panda, P., A Class of Estimators Using Auxiliary Information in Two-Stage Sampling, *Australian & New Zealand Journal of Statistics*, vol. 41(4), 1999a, pp. 405-410.

- [5] Sahoo, L.N. and Panda, P., A Predictive Regression Type Estimator in Two-Stage Sampling, *Journal of The Indian Society of Agricultural Statistics*, Vol. 52 (3), 1999b , pp. 303-308.
- [6] Srivastava, M. and Garg N., A General Class of Estimators of A Finite Population Mean Using Multi-Auxiliary Information Under Two-Stage Sampling Scheme, *Journal Of Reliability and Statistical Studies*, vol. 1(2),2009, pp. 103-118.
- [7] Cochran, W.G. Sampling techniques, 3rd ed., John Wiley & Sons, New York, 1977.
- [8] Hansen, M.H., Hurwitz, W.N. and Madow, W.G., Sampling survey methods and theory vol. I, John Wiley & Sons, New York, 1953.
- [9] Monika S., A Class of Predictive Estimators in Two-Stage Sampling when Auxiliary Character is Estimated at SSU Level. *International Journal of Pure and Applied Mathematics*, Vol. 85(2), 2013, pp. 285-289.
- [7] C. Kadilar, H. Cingi, Ratio estimators in simple random sampling, *Applied Mathematics and Computation*, 151 (2004) 893–902.
- [8] M. Khoshnevisan, R. Singh, P. Chauhan, N. Sawan, F. Smarandache, A general family of estimators for estimating population mean using known value of some population parameter(s), *Far East Journal of Theoretical Statistics*, 22 (2) (2007) 181–191.
- [9] N. Koyuncu, C. Kadilar, Efficient estimators for the population mean, *Haceteepe Journal of Mathematics and Statistics*, 38 (2) (2009) 217–225.
- [10] M.N. Murthy, Product method of estimation, *Sankhya Series A* 26 (1964) 69–74.