Estimation of the Population Total of Nigeria Using One Unit per Stratum (Based On 2006 Census Result)

[1]T.J. Akingbade, [2]O.S. Balogun,

[1]Department of Mathematical Sciences, Kogi State University, Anyigba, Kogi State, Nigeria
[2]Department of Statistics and Operations Research, Modibbo Adama University of Technology, P.M.B 2076, Yola, Adamawa State, Nigeria

Abstract

This research has been done to show the point estimate of the population total of Nigeria based on 2006 census resultand to examine the effect of the use of auxiliary information in estimating variance in collapsing of strata with one unit per stratum. Two stage sampling method was used for the sample selection, which involves two phases of selection. In the first stage, 12 states were selected out of 36 states using random number table. The Local Government Areas (774) were taken to be the second stage and purposive sampling method was used. Based on the research work, considering one unit per stratum, that is selecting one Local Government Area was selected from each of the selected state. Based on the result obtained, using stratified random sampling with addition of auxiliary variable, the least standard error of $\hat{Y}_{str}$ was **3,501,901.105**under the collapsing of 12 strata in six into two groups with an estimated population total of **139,295,482** which was very close to the actual population total of Nigeria based on 2006 census result(**140,003,542).**
Estimation using stratified sampling with addition of an auxiliary variable gave a better result than estimation with variable of interest only

Keywords: Population, Collapse strata, Stratified sampling, Stratum

1. Introduction

The general knowledge of our day to day activities is all based to a very large extent on sample. Hence, sample survey has been very useful in almost every area of lives. Sample survey theory deals with the method and processes of sampling, data collection and estimation of the population parameters.

2. Collapsing Of Strata

A feature of many surveys Sample design is the selection of a single primary sampling unit (PSU) per stratum. The selection of a single PSU per stratum gives efficiency in design since stratification is carried out to fullest possible extent, but it does not generally permit an unbiased variance estimator to be obtained. A widely used method of variance estimation for this situation is known as collapsed strata technique (Rust and Kalton, 1987). With this techniques, strata and their corresponding sample PSU's are collapsed together in groups and then the variability among the unit within these groups is used to derive a variance estimator.If the strata can be ordered approximately in ascending order of the stratum means, the method of successive difference (Kish, 1965) is attractive. This method is an extension of collapsing of strata in pairs. Frequently, these methods have similar biases but the method of successive differences has some-what greater precision.Isaki (1983) used auxiliary information to reduce the bias of the collapse strata variance estimator. The results suggest that when auxiliary variable is highly correlated with the survey variables, there is a substantial improvement in the accuracy of variance estimation.The collapsed strata estimator (Cochran, 1977, section 5A.12) is a well-known estimator of variance estimation in one-per-stratum problem. The procedure collapses strata with one unit per stratum into groups and treats the strata in a group as independent samples from the combined stratum. In this research, collapsing can be accomplished separately among the strata containing small and medium sized districts with one district in the sample. First arrange the strata in a non-increasing sequence based on total enrolment size. Then collapse strata into pairs or groups sequentially. The variance estimator of a group is given by (5A.56) in Cochran's (1977).

2.1 Estimation of Variance with One Unit per Stratum (n_h =1)

Let the sample observation in a typical pair be y_{j1}, y_{j2}, where j goes from 1 to L/2. Let $\hat{y}_{j1} = N_{g1}y_{g1}$, $\hat{y}_{j2} = N_{g2}y_{g2}$ be the estimated stratum totals.

$$V_{CS}(\hat{Y}_{str}) = \sum_{g=1}^{L/2} \frac{L_g}{L_{g-1}} \sum_{k=1}^{L_g} (\hat{Y}_{gh} - \frac{\hat{Y}_g}{L_g})^2 \qquad \text{(Cochran, 1977)}$$

Where $\hat{Y}_g$ is the estimated total for group g for L_g =2 when $\hat{Y}_g = \hat{Y}_{g1} + \hat{Y}_{g2}$. This method of estimation is called "***collapsed strata***".

When an auxiliary variateX_h is known for each stratum that predicts the total Y_h, (Hansen, Hurwitz, and Madow, 1953) suggested the alternative variance estimator.

$$V_{CSx}(\hat{Y}_{str}) = \sum_{g=1}^{L/2} \frac{L_g}{L_g - 1} \sum_{k=1}^{L_g} \left(\hat{Y}_{gh} - \frac{x_{gh} \hat{Y}_g}{x_g} \right)^2 \qquad \text{(Hansen, Hurwitz \& Madow, 1953)}$$

Strata pairs are formed so that the strata in each pair are as similar as possible in respect to the characteristics of interest. In addition, strata that do not vary much in size as measured by an auxiliary variate are often collapsed. Pairs are not formed on the basis of selected sample units.

When to Collapse Strata

(a) When the sample contains only one unit per stratum in such a way that variance estimation within stratum (S_{yh}) is not possible to estimate.

(b) When the first stage of sampling consist of primary sampling unit such as cities or counties and the ultimate sampling unit are households.

3. Methodology and Data Presentation

The data were collected based on female population, male population and total population for each of the 36 states and 774 local government areas (LGA) in Nigeria based on 2006 census result. The number of states was taken to be the first stage and there were 36 states in Nigeria.The second selection is known as second stage unit. The number of the Local Government Areas in the selected states was taken as the second stage. The method of selection used here was a non probability sampling schemes using purposive sampling.Based on this research work, considering one unit per stratum, the 36 states are called 36 strata and each of the state is called a stratum and each state constitute a number of LGA's Therefore, one LG was selected from each of the selected states. The LGA that has a close value to the average value of the state was chosen as a representative for the selected state.

4. Collapsed strata method used

Deterministic mixing method was employed.

Procedure: The probability $P_g=x_g/X$ were determined and were rearrange in ascending order with respect to P_i for each of the tables. That is:

Estimation based on sample size 12, were rearrange in ascending order and were also collapsed in pair, three, four and six w.r.t. P_i in order to form homogeneous collapsed strata.

For example, if L=12, L/2=6 groups, taking the first two strata as a group and the next two strata as another group until the sixth group is obtained.

Table 1: Random digits of the selected 12 states with the Population Total, Female Population of the selected LGA and the number of LGA in the selected states

Random No Digits	selected States-L.G.A.	Pop Total of LGA	Female Pop of LGA	No of LGA
03	Akwa-Ibom-Itu	127033	59467	31
29	Osun -Obokun	116511	60965	30
13	Ekiti -Aiyekire	148193	70980	16
02	Adamawa-Mubi North	151072	72850	21
16	Imo-Ikeredu	149316	73084	27
25	Nasarawa-Obi	148874	74462	13
21	Kebbi-KokoBesse	154605	76201	21
26	Niger-Paikoro	158086	77280	25
33	Sokoto-Wurno	162307	78964	23
12	Edo-Orhionmwon	182717	90051	18
11	Ebonyi-Ohaukwu	196337	103489	13
18	Kaduna-Je,a'a	278735	133068	23

Data analysis and Result

In this section, the analysis of this research was carried out. Estimation of population total ($\hat{Y}$), Bias percentage($\hat{Y}$) and variance of the population total (V($\hat{Y}$)) with one unit per stratum were estimated with n= 12 using stratified random sampling

Estimation of population total$\hat{Y}_{str}$with variable of interest only

$$\bar{Y}_{str}=\sum_{h=1}^{n} W_h \bar{y}_h \quad \hat{Y}_{str}=\frac{N}{n} M \sum_{h=1}^{n} W_h \bar{y}_h$$

$n=12,\ N=36,\ M=774,\quad \sum W_h \bar{y}_h = 54428.8825$

$$\hat{Y}_{str} = \frac{36}{12} \times 774 \times 54428.8825 = 126{,}383{,}865$$

Bias Percentage = Actual Pop. Total – Estimated Pop. Total * 100

Actual Pop. Total

$$(140003542 - 126383865) \Big/ 140003542 \times 100 = 9.728\%$$

Estimation of population total$\hat{Y}_{str}$with addition of auxiliary variable using combine Ratio Stratified Random Sampling

X=68,293,683

No ofgroups		Pop. Total (y_{gh})	Female Pop (x_{gh}).	$P_{gh} = x_{gh} \Big/ X$	N_{gh}	$N_{gh} y_{gh} = \hat{Y}_{gh}$	$\left(\hat{Y}_{gh} - \hat{Y}_g \Big/ 2\right)^2$	$\left(\hat{Y}_{gh} - \frac{x_{gh}}{x_g}\hat{Y}_g\right)^2$
1	Akwa-Ibom	127033	59467	0.000871	31	3938023	48994273062	71597227953
	Osun	116511	60965	0.000893	30	3495330	48994273062	71597227953
	Subtotal	**243544**	**120432**			**7433353**	**97988546125**	**1.43194E+11**
2	**Ekiti**	**148193**	**70980**	**0.00104**	**16**	**2371088**	**1.6057E+11**	**1.32988E+11**
	Adamawa	151072	72850	0.00107	21	3172512	1.6057E+11	1.32988E+11
	Subtotal	**299265**	**143830**			**5543600**	**3.2114E+11**	**2.65975E+11**
3	Imo	149316	73084	0.00107	27	4031532	1.09848E+12	1.15767E+12
	Nassarawa	148874	74462	0.00109	13	1935362	1.09848E+12	1.15767E+12
	Subtotal	**298190**	**147546**			**5966894**	**2.19696E+12**	**2.31533E+12**
4	Kebbi	154605	76201	0.00112	21	3246705	1.24413E+11	1.07202E+11
	Niger	158086	77280	0.00113	25	3952150	1.24413E+11	1.07202E+11
	Subtotal	**312691**	**153481**			**7198855**	**2.48826E+11**	**2.14405E+11**
5	Sokoto	162307	78964	0.00116	23	3733061	49318416006	2.04657E+11
	Edo	182717	90051	0.00132	18	3288906	49318416006	2.04657E+11
	Subtotal	**345024**	**169015**			**7021967**	**98636832013**	**4.09313E+11**

6	Ebonyi	196337	103489	0.00152	13	2552381	3.72205E+12	1.87383E+12
	Kaduna	278735	133068	0.00194	23	6410905	3.72205E+12	1.87383E+12
	Subtotal	**475072**	**236557**			**8963286**	**7.4441E+12**	**3.74766E+12**
	Overall Total						**1.04077E+13**	**7.09588E+12**

$$\hat{Y}_{rcstr} = \frac{\bar{y}_{str}}{\bar{x}_{str}} X$$

$$\hat{Y}_{rcstr} = \frac{\sum_{h=1}^{12} W_h \bar{y}_h}{\sum_{h=1}^{12} W_h \bar{x}_h} X = \frac{54428.88}{26685.35} * 68293683 = 139,295,482$$

Bias Percentage=

$$(140003542 - 139295482) \Big/ 140003542 \times 100 = 0.506\%$$

TABLE 2: Estimation of variance in collapsing of 12 strata pair into 6 groups

Estimation of $V(\hat{Y}str)$ in collapsing of strata in pair into six groups

a) Estimation of $V(\hat{Y}str)$ with variable of interest only

$$V_{cs}(\hat{Y}_{str}) = \sum_{g=1}^{G} \frac{L_g}{L_g - 1} \sum_{h=1}^{L_g} \left(\hat{Y}_{gh} - \frac{\hat{Y}_g}{L_g} \right)^2$$

$$L_g = 2,\ g = 1, \cdots, 6$$

$$\sum_{g=1}^{6} \sum_{h=1}^{2} \left(\hat{Y}_{gh} - \frac{\hat{Y}_g}{L_g} \right)^2 = 1.040765998 \times 10^{13}$$

$$V_{CS(2)}\left(\hat{Y}_{str}\right) = 2 \times \left(1.040765998 \times 10^{13}\right) = 2.081531996 \times 10^{13}$$

Standard error of $\hat{Y}_{str}$

$$SE_{cs(2)}\left(\hat{Y}_{str}\right) = \sqrt{V_{cs(2)}\left(\hat{Y}_{str}\right)} = 4562380.953$$

b) Estimation of $V(\hat{Y}str)$ with addition of Auxiliary Variable

$$V_{csx}(\hat{Y}_{str}) = \sum_{g=1}^{G} \frac{L_g}{L_g - 1} \sum_{h=1}^{L_g} \left(\hat{Y}_{gh} - \frac{x_{gh}}{x_g} \hat{Y}_g \right)^2$$

$$\sum_{g=1}^{6}\sum_{h=1}^{2}\left(\hat{Y}_{gh}-\frac{x_{gh}}{x_g}\hat{Y}_g\right)^2 = 7.095878997\times10^{12}$$

$$V_{csx(2)}\left(\hat{Y}_{str}\right)=2\times\left(7.095878997\times10^{12}\right)=1.419175799\times10^{13}$$

Standard error of $\hat{Y}_{str}$

$$SE_{csx(2)}\left(\hat{Y}_{str}\right)=\sqrt{V_{csx(2)}\left(\hat{Y}_{str}\right)}=3767194.977$$

Summary of the findings

Summary for the Estimated $\hat{Y}$ and $S.E\left(\hat{Y}\right)$ Based on Sample Size 12

Table 3: Summary Table for the Estimated $\hat{Y}$ and $S.E\left(\hat{Y}\right)$ Using Stratified Random Sampling with Sample Size 12

Methods	$\hat{Y}_{str}$ = **126,383,865** $Bias(\hat{Y}_{str})=9.728\%$ $S.E_{cs\left(\hat{Y}_{str}\right)}$	$\hat{Y}_{str}$ = **139,295,482** $Bias(\hat{Y}_{str})=0.506\%$ $S.E_{csx\left(Y_{str}\right)}$
Collapsing in pair	4,562,380.953	3,767,194.927
Collapsing in three	4,274,185.051	3,652,067.891
Collapsing in four	4,102,367.353	3,580,511.818
Collapsing in six	3,879,691.526	3,501,901.105

Conclusions

Stratified random sampling with addition of auxiliary variable seems to be more precise estimator, because it gives the least standard error of $\hat{Y}_{str}$ to be **3,501,901.105** under the collapsing of 12 strata in six into two groups and least bias of **0.506** percentage and the estimated population total was **139,295,482** which was very close to the actual population total of Nigeria based on 2006 census result **(140,003,542)**

References

[1] Cochran, William G., description of method of "collapsed strata" in Sampling Techniques, 1977, pp 139-140.

[2] Hansen, M.H., Hurwitz, W.N., and Madow, W.G.(1953): Sample Survey Methods and Theory. Vol.II. Methods and Applications.

[3] Kish, L. (1965): Survey sampling. John Wiley and Sons, New York.

[4] Parsons, V.L. and Eltinge, J.L. (1999). Stratum partition, collapse and mixing in construction of balanced repeated replication variance estimators. Paper presented at the 1999 Joint Statistical Meetings, Baltimore, Maryland.

[5] Rust, K. and Kalton, G. (1987). Strategies for collapsing strata for variance estimation. Journal of Official Statistics **3**, pp69-81