Journal Homepage: -www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI:10.21474/IJAR01/10724
DOI URL: http://dx.doi.org/10.21474/IJAR01/10724

RESEARCH ARTICLE

COMPARATIVE ANALYSIS OF THE EFFICIENCY OF SIMPLE RANDOM SAMPLING AND STRATIFIED RANDOM SAMPLING TECHNIQUES USING DATA FROM 2006 POPULATION FIGURES OF THE SIX SOUTH-SOUTH STATES OF NIGERIA

Igwe N.O, Bassey U and Oyah M.P
Department of Maths/Statistics, AkanuIbiam Federal Polytechhic, Unwana, P.M.B 1007 Afikpo, Ebonyi State Nigeria.

Manuscript Info

Manuscript History
Received: 25 January 2020
Final Accepted: 27 February 2020
Published: March 2020

Key words:-
Probability Sampling Techniques

Abstract

This work compares the efficiency of the estimators in two probability sampling techniques namely: Simple Random Sampling and Stratified Random Sampling, using the 2006 population figures of the six states in the South-South geopolitical Zone of Nigeria. A table of random digit was used to select appropriate samples for each of these techniques. The selected samples were further used to estimate the population means, variance, standard error and their confidence intervals. The resulting estimates obtained from each technique were compared with the actual population figures. This comparison revealed that Stratified Random Sampling techniques proved more efficient than the simple Random Sampling using the minimum variance criterion. This interesting comparative result obtained at both 95% and 99% Confidence Intervals are shown in tables 3 and 4 in this work.

Introduction:-

Nigeria is a country with strong passion and desire for accurate census in the country for proper planning and budgeting. Data generated from census exercise is basically used to determine how to allocate resources for the determination of the demographic variables in the population (Esogbue, 2008). In view of the importance of census, plans were put in place by the Federal government to conduct the population census in the year 2006.

Wilks, (2007) stated that, to credibly and successfully conduct the 2006 Nigerian population census with a view to providing information on the population total, much resources should be put in place. Despite the huge amount spent, the 2006 population census could not be concluded to be satisfactory in having wide coverage all over the country as there were reports of non-availability of enumeration materials, absenteeism or lateness of officials of the National Population Commission, and the overwhelming number of people to be counted which scared the young enumerators.

Worthy of note is that Nigeria has been due for another census since 2016 but lack the needed fund that would cost 272 billion naira ($759 million) “according to the director of the National population commission NPC”. As a result of this Nigeria cannot conduct census regularly or efficiently as it should be.

Against all these controversies, the question one ponders on is; could the total population be obtained without wasting the huge material and human resources? The advantages of sampling over census or complete enumeration

Corresponding Author:- Igwe N.O
Address:- Department of Maths/ Statistics, AkanuIbiam Federal Polytechnhic, Unwana, P.M.B 1007 Afikpo, Ebonyi State Nigeria.

are: it reduces cost, results in greater speed, accuracy and scope of the information that could be obtained. (Gitonga, et al., 2018)

The overall population figure of Nigeria is made up of the sum of the sub-populations of the six (6) geographical Zones namely: North East, North West, North Central, South East, South West, and the South-South. Hence with sampling, it is argued that if one of the sub-populations could be isolated, and a random sample drawn there from, is used to estimate the total sub-population with high degree of success, then estimating the overall population figures of the country would be successfully attempted. In this regard, the 2006 population figures of the South-South geopolitical zone, made up of 6 States (AkwaIbom, CrossRiver, Delta, Edo, Rivers, and Bayelsa) is estimated using two (2) probability sampling techniques namely: simple random sampling and stratified random sampling.(Stuart, 1984).

The ultimate aim of this research is to compare the techniques of simple random sampling and stratified random sampling with a view to determining the one that efficiently provides good estimates and thereby serves as the best sampling plan for carrying out analysis on the population figures of the six South-South states.

Materials and Methods:-

For this research, data was collected from a secondary source precisely from the National Bureau of Statistics on the 2006 population figures of the six (South- South) states in Nigeria.

Method of determining sample size:

In the planning of a sample survey, a stage is always reached at which a decision must be made about the size of the sample. Sampling theory provides a frame work within which to think intelligently about the selection best sample size.

According to Stuart, 1984, the principal steps involved in the choice of a sample size are statement of the level of precision which shows the amount of error that we are willing to tolerate in the sample estimates, and the degree of variability in the attribute being measured.

An appropriate formula used in this regard is Taro Yamane formula which is given as:

$\mathbf{n= \frac{N}{1+N(e)^2}}$, Where: n=Sample size, N=Population size, and e =Level of precision (%).

Since there are 246 units in the population using the above formula our sample size becomes (at 95%)

$$n = \frac{246}{1+246(0.05)^2}$$

$$= \frac{246}{1.615} = 152$$

Therefore, we shall use as our sample size 152 units of the entire population of 246 units to carry out the necessary comparisons between the sampling techniques employed**.**

Method of random selection:

Generally, in sampling, if our conclusions are to be valid and dependable, then the sample selection must be done without any bias. Randomization technique which guarantees equal chances of being selected was used to select the required samples for this research. The use of table of random digits was adopted in this research.

Simple Random Sampling:

Let a sample of n units be drawn from a population of N units such that all the n distinct samples have an equal probability of being drawn, then the above method of selection is referred to as simple Random sampling. In practice, a simple random sample is drawn by first numbering all the units in the population from 1 to N. Then using a table of random digits or a computer program, a series of random digits from 1 to N is drawn one after the other until n digits are selected. The units' corresponding to the n selected digits gives the required sample size n.

Stratified Random Sampling:

In Stratified sampling, the population of N units is first divided into non-overlapping (mutually exclusive) sub-populations called strata, N_1, $N_2 \cdots N_h$ units respectively such that $N = N_1 + N_2 + \cdots N_n$. Then let a sample of size $n_h \ (h = 1,2 \cdots l)$ be drawn independently from the n[th] stratum so that: $n = n_1 + n_2 + \cdots n_h$.

In other to facilitate the construction of strata before sample selection, two pieces of information are important:

1. The stratum frame and 2. The stratum weight $\left(W_h = {}^{N_h}\!/_{N} \right)$

The stratification factor also must be known. Stratification factor is the basis for dividing the heterogeneous population into homogenous sub-populations called strata. For this research, the states are used as the stratification factor.

When the population have been stratified, what follows is to determine the number of sub-sample (n_h) to be drawn from each stratum.

Proportionate stratified sampling can be achieved using the following formula:

$n_h = \frac{nN_h}{N}$ ……3.0

n_h = Sub-sample from each sample, n = Sample size, N_h= Number of units in each stratum and N =Total units in the population

Selection of simple random of size (n) = 152 units:

The method of using Table of Random digits for selecting the sample was adopted. Since there are 246 units in the population, we take first three digits random numbers on the table. The corresponding units of the random numbers selected give us our simple random sample.

Selection of stratified random sample of size n = 152 units:

Using state as the stratification factor, we proceed to obtain the sample n_h (h=1,2,3,…6) for each stratum using the proportional allocation formula. Using equation 3.0 we obtain the following:

$n_h = \frac{nN_h}{N}$ N=246, n= 152

Rivers (N_6) = 46 ; $n_6 = \frac{152(46)}{246}$ =28.42≃29 ; Edo (N_5) =36 ; $n_5 = \frac{152(36)}{246}$ =22.24≃22

Delta (N_4) = 50 ; $n_4 = \frac{152(50)}{246}$ =30.89≃31; Cross-River (N_3) = 36; $n_3 = \frac{152(36)}{246}$ =22.24 ≃ 22

Bayelsa (N_2) = 16; $n_2 = \frac{152(16)}{246}$ =9.88 ≃10 ; Akwa-Ibom (N_1) =62 ; $n_1 = \frac{152(62)}{246}$ = 38.3 ≃38

Shown above are the corresponding stratified random sample of 152units randomly selected from the 6 strata.

Estimation of the population mean, variance, standard error, and confidence intervals using simple random sample of size(n) 152.

Mean:

Let us start by recalling the formula for the sample mean:

$$\bar{y} = \frac{\sum_{i-1}^{n} y_i}{n} \quad \bar{y} = \frac{12546651}{152} = 82543.76$$

$$S^2 = \frac{\sum_{i=1}^{n} \left(y_i - \bar{y} \right)^2}{n-1} = \frac{1}{n-1} \left(\sum_{i=1}^{n} y_i{}^2 - n\bar{y}^{-2} \right)$$

$$\sum_{i-1}^{152} y_i{}^2 = 1250340000000$$

$$\left(\sum_{i-1}^{152} y_i{}^2 - n\bar{y}^{-2} \right) = 214692208100$$

$\bar{y} = 82543.76$

$\therefore S^2 = \frac{1}{n-1}\left(\sum y_i^{\,2} - n\bar{y}^{\,2}\right)$

$= \frac{214692208100}{151}$

$= 1421807598$

Variance of the mean ➔ $V(\bar{y}) = (1-f)\frac{S^2}{n}$

Where $f = \frac{n}{N} = \frac{152}{246} = 0.62$

$V(\bar{y}) = (1-0.62)\frac{1421807598}{152}$

$= \frac{540286887.2}{152}$

$= 3554518.995$

Standard error of the mean ➔ $\delta_y = \frac{S}{\sqrt{n}}\ \sqrt{1-f}$

$= \frac{37706.86}{\sqrt{152}}\ \sqrt{0.38}$

$= 1885.34$

Variance of the population total $\left(\hat{Y}\right)$ ➔ $V\left(\hat{Y}\right) = N^2\,(1-f)\frac{S^2}{n}$

$= (246)^2\ (0.38)\frac{1421807598}{152}$

$= 2151052715$

Standard error of the population total

$\hat{Y} = \frac{NS}{\sqrt{n}}\ \sqrt{1-f}$

$= \frac{246(37706.86)\sqrt{0.38}}{\sqrt{152}}$

$= 463794.38$

Confidence interval for population mean 95%➔ $\bar{y} \pm \frac{ts}{\sqrt{n}}\ \sqrt{1-f}$

Since our sample is large $(i.e\ n > 30)$, we use the Z-value corresponding to the given confidence coefficient using the standard normal table. So our confidence interval becomes:

$\bar{y}\ \frac{Zs}{\sqrt{n}}\ \sqrt{1-f} < \bar{Y} < \bar{y} + \frac{Zs}{\sqrt{n}}\ \sqrt{1-f}$

At 95%

$Z = 1.96$

$$82543.76 - \frac{1.96\,(37706.86)(0.62)}{\sqrt{152}} < \bar{Y} < 82543.76 + \frac{1.96\,(37706.86)(0.62)}{\sqrt{152}}$$

$$82543.76 - 3716.60 < \bar{Y} < 82543.76 + 3716.60$$

$$78827.16 < \bar{Y} < 86260.36$$

$$78827 < \bar{Y} < 86260$$

At 99%

$Z = 2.58$

$$82543.76\ \frac{2.58(37706.86)(0.62)}{\sqrt{152}}\ < Y < 82543.76 + \frac{2.58(37706.86)(0.62)}{\sqrt{152}}$$

$$82543.76 - 4892.26 < \bar{Y} < 82543.76 + 4892.26$$

$$77651.5 < \bar{Y} < 87436.02$$

$$77652 < \bar{Y} < 87436$$

Confidence interval for population total $\bar{Y} \rightarrow N\bar{y} \pm \frac{ZNs}{\sqrt{n}}\sqrt{1-f}$

$$N\bar{y} - \frac{ZNs}{\sqrt{n}}\sqrt{1-f} < \bar{Y} < N\bar{y} + \frac{ZNs}{\sqrt{n}}\sqrt{1-f}$$

At 95%

$$= 246(82543.76) - \frac{1.96(246)(37706.86)(0.62) \le \bar{Y} \le}{\sqrt{152}}\ 246(82543.76) + \frac{1.96(246)(37706.86)(0.62)}{\sqrt{152}}$$

$$= 20305.76 - 914284.68 \le \bar{Y} \le 20305.76 + 914284.68$$

$$= 19391480.28 \le Y \le 21220049.68$$

$$= 19391480 \le Y \le 21220050$$

At 99%

$$= 246(82543.76) - \frac{2.58\,(246)\,(37706.86)(0.62) \le Y \le 246(82543.76)}{\sqrt{152}} + \frac{1.96(246)(37706.86)(0.62)}{\sqrt{152}}$$

$$= 20305764.96 - 1203497.18 \le Y \le 20305764.96 + 1203497.18$$

$$= 19102267.78 \le \bar{Y} \le 21509262.14$$

$$= 19102268 \le \bar{Y} \le 21509262$$

Estimation of the population mean, variance, standard errors, and confidence interval using stratified random sample of size (n) 152

Mean

$$\bar{y}_{st} = \frac{\sum_{h=1}^{L} N_h\ \overline{y_h}}{N}$$

$h = 1,2,3,\cdots,6$

Where $\bar{y}_h = \frac{\sum_{i=1}^{nh} y_{hi}}{n_h}$

Table 1:- Computational table for the mean of the stratified random sample.

States	h	N_h	n_h	$\sum_{i=1}^{nh} y_{hi}$	$\bar{y} = \frac{\sum_{h=1}^{nh} y_{hi}}{n_h}$	$N_h \bar{y}_h$
Delta	1	50	31	2577224	83136.26	4156813.00
Edo	2	36	22	2058128	93551.27	3367845.72
Cross River	3	36	22	1996671	90757.77	32678279.72
Bayelsa	4	16	10	1156912	115691.2	1851059.2
A/Ibom	5	62	38	2401332	63192.95	3917962.9
Rivers	6	46	29	3321370	114530.00	5268380
	Total	246	152			2182934054

Therefore,

$$\bar{y}_{st} = \frac{\sum_{h=1}^{L} N_h \bar{y}_h}{N}$$
$$= \frac{2182934054}{246}$$
$$= 88737.16$$

Variance of the mean:

The variance of the estimated mean for a stratified random sample is:

$$V(\bar{y}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} N_h (N_h - n_h) \frac{S_h^2}{n_h}$$
$$= \sum_{h=1}^{L} W_h^2 (1 - f_n) \frac{S_h^2}{n_h}$$

Where:

$$W_h = \frac{N_h}{N}$$

$f_h = \frac{n_h}{N_h}$ Using the proportional allocation formula:

$$n_h = \frac{nN_h}{N}$$

The variance of the mean becomes $V(y_{st}) = \frac{(1-f)}{n} \sum_{h=1}^{L} W_h S_h^2$

Table 2:- Computational table for the variance of the mean for the stratified random sample.

States	h	N_h	n_h	$w_h = \frac{N_h}{N}$	S_h^2 (From Appendix III)	$W_h S_h^2$
Delta	1	50	31	0.203	790172665.3	97258607.52
Edo	2	36	22	0.146	2883125574	420936333.8
Cross River	3	36	22	0.146	404016906.1	58986468.29
Bayelsa	4	16	10	0.065	2806428427	1824178478

A/Ibom	5	62	38	0.252	606381436	152808121.9
River	6	46	29	0.187	1088838292	203612760.6
	Total	246	152			1116020140

$$\left(\bar{y}_{st}\right)=\frac{(1-f)}{n}\sum_{h=1}^{L}W_h\ S_h{}^2$$

$$=\frac{1-0.62}{152}\left(1116020140\right)$$

$$=2790050.35$$

Standard error for the mean $\bar{y}_{st}\ \delta\bar{y}_{st}=\sqrt{V\left(\bar{y}_{st}\right)}$

$$=\sqrt{2790050.35}$$

$$=1670.34$$

Variance of the population total $\hat{Y}$

$$V\left(\hat{Y}_{st}\right)=V\left(N\bar{y}_{st}\right)$$

$$N^2\ V\left(\bar{y}_{st}\right)$$

$$=(246)^2\ (2790050.35)$$

$$=168842687$$

Standard error for the mean $\bar{y}_{st}$

$$\delta\hat{Y}_{st}=\sqrt{V\left(\hat{Y}_{st}\right)}$$

$$=\sqrt{168842687}$$

$$=410904.72$$

Confidence interval for the population Mean

$$\bar{y}_{st}\pm ZS\left(\bar{y}_{st}\right)$$

$$\bar{y}_{st}\pm ZS\left(\bar{y}_{st}\right)<\bar{Y}<\bar{y}_{st}+ZS\ \left(\bar{y}_{st}\right)$$

At 95%

$$Z=1.96$$

$$88737.16-1.96(1670.34)\le\bar{Y}\le 88737.16+1.96(1670.34)$$

$$88737.16-3273.87\le\bar{Y}\le 88737.16+3273.87$$

$$85463.3\le\bar{Y}\le 92011.03$$

$$85436\le\bar{Y}\le 92011$$

At 99%

$$Z=2.58$$

$$88737.16-2.58(1670.34)<\bar{Y}<88737.16+2.58(1670.34)$$

$$88737.16-4309.48\le Y\le 88737.16+4309.48$$

$$84427.68\le Y\le 93046.64$$

$$84428\le Y\le 93047$$

Confidence Interval for the Population total $\hat{Y}$.

$$N\bar{y}_{st} \pm ZNS\left(\bar{y}_{st}\right)$$

$$N\bar{y}_{st} - ZNS\left(\bar{y}_{st}\right) \le \hat{Y} \le N\bar{y}_{st} + ZNS\left(\bar{y}_{st}\right)$$

At 95%

Z=1.96

$$= 246(88737.16) - 1.96(246)(1670.34) < Y < 246(88737.16) + 1.96(246)(1670.34)$$

$$=21829341.36 - 805371.13 \le Y \le 21829341.36 + 805371.13$$

$$=21023970.23 \le Y \le 22634712.49$$

$$= 21023970.23 < Y < 22634712$$

At 99%

$$Z = 2.58$$

$$= 246(88737.16) - 2.58(246)(1670.34) < Y < 246(88737.16) + 2.58(246)(1670.34)$$

$$21829341.36 - 1060131.39 < Y < 21829341.36 + 1060131.39$$

$$20769209.97 \le Y \le 22889472.75$$

$$20769210 \le Y \le 22889473$$

Table 3:- Comparison of the Estimation Population Mean and the Actual Population Mean for the 2006 population Figures of the South-South States in Nigeria.

Sampling Techniques	95% confidence interval for the estimated population mean	99% confidence interval for the estimated population mean	Actual population mean	Remark
Simple Random sampling	78827≤Y≤86260	77652≤Y≤87436	85545	Efficient
Stratified Random Sampling	85463≤Y≤92011	84428≤Y≤93047	85545	More Efficient

NOTE:Actual population Mean = $\dfrac{Actual\ \ population\ Total}{Total\ Number\ of\ \ units\ in\ the\ \ population}$

$$= \frac{21044081}{246} = 85545$$

The result obtained in Table 3 clearly shows that stratified random sampling Techniques is more is more efficient than the simple random sampling technique.

Table 4:- Comparison of the estimated population total and the actual population total for the 2006 population figure of the south – south in Nigeria.

Sampling Techniques	95% confidence interval for the estimated population total	99% confidence interval for the estimated population total	Actual population total	Remark
Simple Random sampling	19391480≤Y≤21220050	19102268≤Y≤21509262	21044081	Efficient
Stratified Random Sampling	21023970≤Y≤22634712	20769210≤Y≤22889473	21044081	More Efficient

Table 5:- Comparison of the estimated Variances obtained from each of the employed techniques.

Simple Technique	Estimated Variance of the Population Mean	Estimated Variance of the population Total
Simple Random Sampling	3554518.995	2151052715
Stratified Random Sampling	2790050.35	168842687

Using the minimum Variance criterion, it can be concluded that:
stratified Random Sampling is fare better than simple Random Sampling.

Conclusion:-

From the foregoing, one can empirically conclude that stratified random sampling technique provides the best and most efficient estimators of the 2006 population figures of the six states in the South- South geopolitical zones of Nigeria.

Recommendation:-

There is the need to carry out a further research, which would involve the use of the estimators in stratified Random Sampling technique on the entire 2006 population figures of Nigeria made up of all the six (6) geopolitical zones.

References:-

1. Esogbue, E (2018, June 09). The Perplexity of conduct of population Censuses in Nigeria retrieved from http://www.articlesbase.com/politics-articles-542567.html
2. Gitonga, C.W., Karanja, P.N., Kihara. J., Mwanje. M., Elizabeth J., Robert, W. S, Abdusalalam, M.N, &Brookers. (2018, October 07). Implementing school Malaria Surveys in Kenya towards a National Surveillances system retrieved from www.malariajournal.com/content
3. Stuart, A., (1984).The ideas of sampling. 3rd Edition New York, Macmillan.
4. Wilks, R., Norie, Y., Mullings, J., Namvar, Z., Figueroa, P., Tulloch- Reid, M., Trevor, F., Walters, C., Franklyn, B., Forester, T., Elizabeth, W., & Ashley, D., (2007).Factor Affecting Study Efficiency and item Non-response in Health Surveys in Developing countries.BMC Medical Research Methodology. 3(1), 52-64.hh.