ISSN No:-2456-2165

Consistency of Stratified Random Sampling Estimators in Repetive Sampling

Adekunle Nurudeen MASOPA[1*]; Adisa Anthony AGBONA[2];
Sabiu MUBARAK[3]; Adefunke Rukayat MASOPA[4]
[1,2,3]Department of Statistics,Federal Polytechnic Ede
[4]Ogun State Teaching Service Commission

Correspondence Author:- Adekunle Nurudeen MASOPA[1*]

Abstract:- This study focuses on consistency of stratified random sampling in repeated sampling processes within a population with heterogeneous characteristics. The data for the study is a real life data on number of students in schools where the stratification is on the basis of ownership (public or private). Proportional allocation method was used to determine the number of units (schools) to be chosen from each stratum for a given sample size and four (4) independent samples of equal sample sizes were chosen and estimates of mean ,variance as well as confidence interval obtained with the estimators of stratified random sampling. The estimates obtained for each sample sizes were subjected to a test of significance to test the null hypothesis of no significance difference between the estimates and the actual value using the t-statistic. The analysis revealed that the estimates obtained for different samples differs but the test of significance revealed that there is no significant difference in the estimates across the independent samples as the P-values are less the level of significance $\alpha = 0.05$ except for $n = 30$ which could be considered as an outlier. Also, there is no significant difference in the estimates of variances for the various sample sizes considered for this study with a Pvalue of 0. 2344.

***Keywords*:-** *Resampling, Stratified Sampling, Proportional Allocation, Estimators, P-value.*

I. INTRODUCTION

Sampling is obviously the most widely used concepts in day to day human activities as it forms the basis of operations in virtually all areas of endeavour. It has an enormous range of application in very many real life situations which makes it a concept of concern to researchers and stakeholder in the specific areas of need with the aim of maximizing benefits in its usage by ensuring appropriateness, efficiency and cost economy. Samples survey methods have wide range of applications in both physical, natural, health science and economy (Olayiwola, Apantaku, Bisira &Adewara,2013).Sampling could be done in a probabilistic context where every units in the population has a known and non-zero probability of selection(Mweshi, & Sakyi, 2020; Clark & Steel, 2022,). Non – probabilistic sampling involves selection of units from the population without any recourse to a well stated and known principle of selection (Lohr,2021; Wu,2022). Kanaki & Kalogiannakis (2023) posited that process of selection under non-random sampling usually undermine the estimation and precision due to implementation procedure. On the other hand, the sampling methods operating under the principle of known and non-zero probability of inclusion have advantage over the non-random sampling methods particularly in the areas of procedure, estimation and precision assessment (Pawar, Verma, Daniel, & Sayyad,2023). Karunarathna, De Alvis, Gunasena and Jayawardana (2024) posited that researchers have preference for random sampling methods over non random sampling methods in order to explore the advantages of estimation and efficiency. Random sampling can be based on equal probability of selection for units in a given population as it is the case in simple random sampling (Rahman, Tabash, Salamzadeh, Abduli, & Rahaman, 2022), interval selection of units as related to systematic sampling(Iliyasu & Etikan,2021), selection of units from a population where there is intra-group heterogeneity as well as inter-group homogeneity as it is applicable in cluster sampling(Lohr, 2021) and selection of units from homogeneous sub-groups obtained after stratification of a heterogeneous population as it is in stratified sampling(Sharma,2023; Shah, 2024).Stratified random sampling is a method applicable where a given heterogeneous population is partitioned into a set of non-overlapping homogenous sub-groups called strata (Verma, Verma & Abhishek, 2024). Enzo(2021) mentioned that the population is usually partitioned into sub-groups on the basis of a stratification variable capable of grouping the units in the population in a way that the groups are mutually exclusive. Pandey (2024) opined that stratified sampling specifically have enormous areas of applications. Stratified sampling is known to have a good precision when adopted in selection of units from heterogeneous populations in the presence of a characteristic which could be used for stratification (Ahamed, Gupt, & Phukon, 2021). The principal concern of samplers and researchers is to deploy an appropriate method in order to ensure accuracy of estimates obtained through the sampling process. The major concern of stratified sampling is to reduce the variability at a given cost or reduce the cost for a given variance(Lawal, Salami, Obisesan,Yusuff & Owolabi,2018; Adebola & Ajayi,2014).Some of the issues earlier raised by researchers on stratified sampling include the determination of stratification boundary. Horgan (2006) considered the issues

ISSN No:-2456-2165

of stratification boundaries and opined that coefficient of variation would be uniform if units are selected in geometric progression from a skewed distribution. Kareem, Oshungade, Oyeyemi and Adejumo (2015) proposed the method of moving average stratification algorithm for determination of stratification boundary. The choice of method of allocation is another issue of concern in stratified sampling. Equal allocation method in stratified sampling gives no attention to the strata weights which has the tendency to affect the precision (Oyoo,2021 ; Aubry,2024). The proportional allocation strategy takes care of the effect of each stratum weight in the selection of units for estimation. Lawal, Salami, Obisesan, Yusuff and Owolabi (2018) established that optimum allocation outperform other methods based on the variances of the allocation procedures when the underlined distribution is known to be skewed. On the contrary Fayose and Adebara (2018) established that proportional allocation is the best for income data though without any attention to the distribution exhibited by the data. More importantly, it is a general phenomenon for every concept to be prone to issues of concern to researchers as it usually forms the basis of investigation into the adequacy, accuracy and appropriateness of the procedure under consideration. In other words, all sampling methods when critically examined have conventional or procedural issues that require attention of researchers.

Bolarinwa (2020) posited that a common practice in sampling is to compute estimates on the basis of the estimators of chosen method either on a single sample or over varying sample sizes with no attention to the possibility of obtaining different estimates from independent samples taken from the same population even as the sample sizes vary. Resampling of units from a larger sample or with partial replacement is often adopted in some sampling methods such as double sampling and sampling on successive occasions (Udumboso, Akanbi & Afolabi, 2019). It is a rare instance for any estimator to produce same estimates over independent samples since the units in different samples a bound to have unique measurements for the characteristics of interest. The central idea of this paper is to examine the consistency of stratified random sampling over samples independently selected from a given population for a particular size. The sample size was varied to check whether the consistency behavior is uniform across sample sizes or changes as sample size varies. The rest of this paper is arranged as follow; section 2.0 deals with methodology, 3.0 results and discussions, 4.0 conclusions and 5.0 recommendations

II. METHODOLOGY

➢ *Selection and Estimation procedure*

The data used for this study is the enrolment data into both public and private schools in Ede North local government, Osun state Nigeria. The population consists of a total of 74 schools comprising 28 public schools and 46 private schools in Ede North Local Government, Osun State Nigeria. The data consists of the number of students in the various grade levels in the schools. The summary of the data is presented in Table 1. Four independent samples of the equal sample size were drawn from the population and estimates of mean, variance, standard error and confidence interval were calculated based on the estimators of stratified sampling. Let a population of N units be stratified into L strata of sizes N_h from where n_h units is chosen with respect to a method of allocation.

Suppose X_{ij} be a random variable observed on i^{th} unit selected from the j^{th} stratum in the k^{th} sample of size n $(\sum_{L=1}^{2} n_h = n)$. Suppose $\bar{\varphi}_j^k$ is the j^{th} stratum mean and $(\delta_j^{\,2})^k$ is the j^{th} stratum variance. The stratification variable is type of school (public /private) which implies $j = 1,2$ and four (k=4) independent samples of equal sample sizes n were drawn from the population with the adoption of proportional allocation to determine the number of units to be selected in each of the strata. The use of proportional allocation is to ensure that the variation in the strata population sizes is taken into consideration. Proportional allocation is a method that considers individual stratum weights in the determination of number of units to be chosen from each stratum for a given sample size (Fayose & Adebara,2018).

➢ *Estimators of mean, Variance , Standard Error and Confidence Interval*

$$\bar{\varphi}_j^k = \frac{\sum_{i=1}^{n_h} x_{ij}}{n_j}$$

$$\bar{X}_{st} = N^{-1}\sum_{j=1}^{L} N_j \frac{\sum_{i=1}^{n_h} x_{ij}}{n_j}$$

$$= N^{-1}\sum_{j=1}^{L} N_j \bar{\varphi}_j^k \qquad (1)$$

$$V(\bar{X}_{st}) = V(\sum_{j=1}^{L} W_j \bar{\varphi}_j^k)$$

$$= (\sum_{j=1}^{L} (W_j)^2 (1 - f_j) V(\bar{\varphi}_j^k))$$

$$= (\sum_{j=1}^{L} (W_j)^2 (1 - f_j) \sum_{i=1}^{n_h} (n_j - 1)^{-1} (\varphi - \bar{\varphi}_j^k)^2)$$

$$= (\sum_{j=1}^{L}(W_j)^2(1-f_j)\delta_j^2$$

$$= N^{-2}\sum_{j=1}^{L}(N_j)^2 n_j^{-1}(N_j - n_j)\delta_j^2 \quad (2)$$

$$\text{Se} = [V(\bar{X}_{st})]^{0.5}$$

$$= [N^{-2}\sum_{j=1}^{L}(N_j)^2 n_j^{-1}(N_j - n_j)\delta_j^2]^{0.5} \quad (3)$$

The lower and upper boundaries of the confidence interval respectively can be expressed as

$$L_{\varphi_j^k} = \bar{X}_{st} - Z_{1-\alpha/2}[(\sum_{j=1}^{L}(W_j)^2(1-f_j)\delta_j^2]^{0.5} \quad (4)$$

$$U_{\varphi_j^k} = \bar{X}_{st} + Z_{1-\alpha/2}[(\sum_{j=1}^{L}(W_j)^2(1-f_j)\delta_j^2]^{0.5} \quad (5)$$

The proportional allocation method was adopted to ensure fairness in the determination of number of units to be chosen from each stratum for a given sample size (n) such that the stratum sample size depends on the stratum weight (W_h).The estimators in 1, 2, 3 ,4 and 5 are used in the computation to obtain the estimates on the k-samples shown in table 2 to 5. A test of significance was performed using the F-statistic and Pvalue to test the hypothesis in (6) to determine whether the estimated value of the statistic(s) at a particular sample size (n) is significantly different. $H_0 : \psi_1 = \psi_2 = \psi_3 = \psi_4 = \mu$ vs $H_1 : \psi_1 \neq \psi_2 \neq \psi_3 \neq \psi_4 \neq \mu$ for at least two pairs of k.

Decision rule: Reject if the P<0.05, otherwise do not reject.

Table 1 Showing Summary of the Data

Type of school	Number of schools	Stratum weight
Public	28	0.3784
Private	46	0.6216
Total	74	1.0000

Table 1 shows the distribution of schools with respect to ownership. It shows that there are 28 and 46 public and private schools respectively with corresponding weights of 0.3784 and 0.6216 respectively. The distribution of schools suggests that the two categories do not have the same weights as shown in column 3 where the number of private schools is almost twice the number of public schools. In order to take into consideration the variations in the stratum weights, the proportional allocation method was adopted such the stratum sample size is directly proportional to the stratum weight and expressed as $n_h = nW_h$. The various sample sizes leads to different values of n_h in the two strata while approximation was observed in cases where the value is a non-integer.

III. RESULTS AND DISCUSSIONS

Table 2 shows the estimates of mean, variance and confidence interval for the four independently selected samples of equal sizes. It was observed that the estimates of the statistic(s) are reasonably close. The estimates for various sample sizes were subjected to test of significance using the t-statistic with overall mean of 85.2466 and the Pvalue obtained was as presented in table IX. The results show that the estimates are not significantly different from the overall mean of 85.2466 indicating that the deviation of the estimates from the population mean is not significant except for n=30 which can be considered as an outlier.

Table 2 Showing Estimates of mean, Variance and 95% Confidence Interval

Sample size (n)	Sample number (k)	$\bar{X}_{st}$	$V(\bar{X}_{st})$	$L_{\varphi_j^k}$, $U_{\varphi_j^k}$
25	1	82.36	187.5105	55.5207 , 109.1993
	2	83.56	135.3285	60.7591 , 106.3609
	3	69.28	137.0947	46.3307 , 92.2293
	4	77.04	106.2560	56.8361 , 97.2439
30	1	78.63	138.3470	55.5763 , 101.6837
	2	79.43	93.9289	60.4343 , 98.4257
	3	71.00	132.0982	48.4729 , 93.5271
	4	75.37	103.3938	55.4401 , 95.2999
35	1	73.09	93.8435	54.1029 , 92.0771
	2	92.26	99.9712	72.6627 , 111.8573
	3	75.11	107.0401	54.8318 , 95.3882
	4	70.11	70.3275	53.6761 , 86.5469

ISSN No:-2456-2165

40	1	71.78	64.2678	56.0673 , 87.4927
	2	87.38	69.9323	70.9893 , 103.7707
	3	76.25	84.0253	58.2837 , 94.2163
	4	84.15	77.2908	66.9187 , 101.3813
45	1	73.89	45.8211	60.6226 , 87.1574
	2	86.45	49.9972	72.5911 , 100.3089
	3	81.62	85.5797	63.4888 , 99.7512
	4	82.18	54.1798	67.7531 , 96.6069
50	1	80.24	56.8281	65.4647 , 95.0153
	2	84.18	35.9509	72.4281 , 95.9319
	3	81.62	59.6646	66.4804 , 96.7596
	4	92.62	39.2463	80.3412 , 104.8988
55	1	83.93	37.0770	71.9954 , 95.8646
	2	87.49	30.3473	76.6927 , 98.2873
	3	80.09	39.9746	67.6977 , 92.4823
	4	92.62	39.2463	80.3412 , 104.8988

Table 2 Showing the Pvalue for the t-Statistics

PvalueSample size (n)	**25**	**30**	**35**	**40**	**45**	**50**	**55**
P($\bar{X}_{st}$)	0.1141	0.0119	0.2242	0.2309	0.2056	0.8474	0.7873

IV. CONCLUSIONS

The results of the analysis for this study as shown in table 1 revealed that different samples though yielded different estimates of the mean under stratified random sampling approach which are essentially not statistically significantly different from the population mean. The implication is that the estimators of stratified random sampling method may give different estimates on resampling samples of equal size from same population but enjoys consistency in terms of accuracy of estimates since the divergence in the estimates are not not significantly different from the overall estimates of means and variances.

RECOMMENDATION

Based on the result of this study, it is recommended that in studies where stratified sampling is to be adopted and resampling is also required, the process of resampling should involve minimum number of repeated samples as the stratified sampling procedure with its estimators have proven to be consistent and capable of yielding estimates that would be close to the true value. More so, it is necessary that samplers direct attention for further studies towards examining the consistency of stratified sampling in larger or lesser population with or without attention on the distribution of the variable of interest and other methods of allocation.

REFERENCES

[1]. Adebola, F.B & Ajayi ,O.S (2014).Comparison of allocation procedures in a stratified random sampling of skewed population under different distributions and sample sizes.

[2]. Ahamed, M. I., Gupt, B. K., & Phukon, M. (2021). Methods of stratification for a generalised auxiliary variable optimum allocation. Mathematics and Statistics, 9(5), 617-629.

[3]. Aubry, P. (2024). On the implementation of stratified two-stage simple random sampling without replacement, with possible collapsed strata. MethodsX, 13, 102928.

[4]. Bolarinwa, O. A. (2020). Sample size estimation for health and social science researchers: The principles and considerations for different study designs. Nigerian Postgraduate Medical Journal, 27(2), 67-75.

[5]. Clark, R. G., & Steel, D. G. (2022). Sample design for analysis using high-influence probability sampling. Journal of the Royal Statistical Society Series A: Statistics in Society, 185(4), 1733-1756.

[6]. Enzo, V. (2021). The use of stratification in a mixed abilities class to facilitate note-taking in Italian L2 during Geography lessons: A case study conducted in an after-school laboratory for migrant students.

[7]. Fayose,T.S., &Adebara,L(2018). Efficiency of Proportional Allocation Procedure over other Allocation Procedures in Stratified Random Sampling. International Journal of Advanced Research and Publications.2 (4),pp.13-15.

[8]. Horgan,J.M.(2006).Stratification of skewed population. International statistical review.74(1),pp.67-76

[9]. Iliyasu, R., & Etikan, I. (2021). Comparison of quota sampling and stratified random sampling. Biom. *Biostat. Int. J. Rev*, *10*(1), 24-27.

[10]. Kanaki, K., & Kalogiannakis, M. (2023). Sample design challenges: an educational research paradigm. International Journal of Technology Enhanced Learning, 15(3), 266-285.

[11]. Kareem,A.O.,Oshungade, I.O.,& Oyeyemi, G.M(2015). A note on the precision of Stratified Sampling. Open Journal of Statistics,5(2)

[12]. Karunarathna, I., De Alvis, K., Gunasena, P., & Jayawardana, A. (2024). Designing and conducting clinical research: Methodological approaches.

[13]. Lawal,M.,Salami,A.,Obisesan,K.O.,Yusuff,K.M.,&Owolabi,A.A.(2018).On the comparison of some methods of allocation in stratified random sampling for skewed population. International Journal of Mathematics and Statistics studies. 6(3).pp.29-42.
[14]. Lohr, S. L. (2021). Sampling: Design and analysis. Chapman and Hall/CRC.
[15]. Mweshi, G. K., & Sakyi, K. (2020). Application of sampling methods for the research design. Archives of Business Review–*Vol*, *8*(11), 180-193.
[16]. Olayiwola,O.M.,Apantaku,F.S.,Bisira,H.O.,&Adewara,A.A.(2013).Efficiency of Neyman allocation procedure over other allocation procedures in stratified sampling.American Journal of Theoretical and Applied Statistics. 2(5).pp.122-127
[17]. Oyoo, D. O. (2021). Ratio Estimation of Finite Population Total in Stratified Random Sampling Under Non-response (Doctoral dissertation, University of Nairobi).
[18]. Pandey, G. P. (2024). Advancing English Language Teaching through Survey Research: Methodologies, Impacts and Applications. Nepal Journal of Multidisciplinary Research, *7*(2), 127-141.
[19]. Pawar, P. B., Verma, R., Daniel, C. O., & Sayyad, L. (2023). Foundation of Research Methodology: A Comprehensive Guide. Red Unicorn Publication.
[20]. Rahman, M. M., Tabash, M. I., Salamzadeh, A., Abduli, S., & Rahaman, M. S. (2022). Sampling techniques (probability) for quantitative social science researchers: a conceptual guidelines with examples. *Seeu Review*, *17*(1), 42-51.
[21]. Shah, M. Y.(2024). Some New Aspects of Post Stratification in Sample Surveys (Doctoral dissertation, Sher-e-Kashmir University of Agricultural Sciences & Technology of Jammu).
[22]. Sharma, L. R. (2023). Choosing appropriate probability sampling designs in research. Health, *2*, 4-21.
[23]. Udomboso,C.G.,Akanbi,O.B.,Afolabi,S.A.(2019).Application of regression type estimator in double sampling skills enrolment in Oyo State. Global Scientific Journal.7 (3).pp. 853-863.
[24]. Verma, R., Verma, S., & Abhishek, K. (2024). Research methodology. Booksclinic Publishing.
[25]. Wu, C. (2022). Statistical inference with non-probability survey samples. Survey Methodology, *48*(2), 283-311.