WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Application of Successive Sampling in the Assessment of Gender Enrollment in Ekiti State Secondary Schools

Akinyemi O., Faweya O., Ilesanmi A.O., Odukoya E.A O., Ajayi E. D. Aladejana A. E.,

Department of Statistics, Ekiti State University, Ado-Ekiti, Ekiti State, Nigeria.

ABSTRACT

This study examined the use of successive sampling in the assessment of students' enrollment in Ekiti State secondary schools. The results indicated that the choice of correlation ρ between successive samples and proportion μ of overlapping units significantly affect the precision of estimates. For male enrollment, lower values of ρ minimized variance in overall mean estimation, while higher values enhanced accuracy in estimating mean differences. Similarly, for female enrollment, the estimated variances followed a consistent trend, showing the importance of optimal parameter selection in ensuring data reliability. Results showed that for male enrollment, the estimated difference between means was 19.95, with a minimum variance of 440.53. The estimated overall mean was 286.28, with its variance reducing as ρ increased, confirming that higher correlation across samples improved efficiency. Similarly, for female enrollment, the minimum variance reduced from 656.72 at $\rho = 0$ to zero at $\rho = 1$, indicating that strong correlation enhances precision. The variance of $\hat{M}$ also declined progressively from 3,694,016 to 626,1043 as ρ increased.

Keywords: Two-stage sampling, successive occasions, enrollment, variance estimation, gender.

1.0 Introduction

Sampling is an essential tool in statistical research, particularly when conducting large-scale surveys or studies where complete enumeration is not feasible. Two-stage sampling is a widely used technique that improves efficiency by reducing costs while maintaining precision (Cochran, 1977). It involves selecting primary sampling units (PSUs) in the first stage and then selecting secondary sampling units (SSUs) within each PSU in the second stage. This method is particularly useful when dealing with large, heterogeneous populations where stratification is necessary to improve representativeness (Särndal et al., 2003). Successive sampling, on the other hand, refers to repeated sampling from the same population over multiple occasions to track changes over time (Jessen, 1978). The combination of two-stage sampling and successive occasions allows researchers to monitor trends while optimizing sample size and resource allocation. However, the efficiency of different procedures in two-stage successive sampling has been a subject of interest among statisticians, as alternative approaches may yield different levels of precision and cost-effectiveness (Singh & Chaudhary, 2017). The successive method of sampling consists of selecting sample units on different occasions such that some units are repeated for both occasions. If sampling on successive occasions is done according to a specific rule, with partial replacement of sampling units, it is known as successive sampling. The method of successive sampling was developed by Jessen (1942) and extended by Patterson (1950) and Tikkiwal (1967) and also Eckler (1955). Cochran (1977) conducted a foundational study on the efficiency of multi-stage sampling techniques, emphasizing two-stage sampling as a cost-effective alternative to simple random sampling. The study aimed to determine the optimal allocation of sample units in two-stage sampling to minimize variance while maintaining representativeness. Using a theoretical approach supported by empirical case studies, they applied variance estimation formulas to compare different allocation strategies. The results demonstrated that stratified two-stage sampling significantly reduces sampling errors compared to single-stage methods, particularly in heterogeneous populations. The study concluded that the efficiency of two-stage sampling depends on the proper selection of primary and secondary sampling units, as well as the correlation between them. They recommended that researchers apply optimal allocation methods such as Neyman allocation to achieve maximum efficiency in two-stage designs. Raj (1968) explored the effectiveness of ratio estimators in improving efficiency in two-stage sampling. The study compared the relative efficiency of ratio estimation with traditional mean estimation in two-stage sampling designs and showed that ratio estimators significantly enhance precision when auxiliary information is highly correlated with the variable of interest. Singh and Chaudhary (2017) conducted a comparative study of the efficiency of various alternative procedures in two-stage sampling under different population structures and concluded that stratified two-stage sampling consistently outperformed simple random sampling in terms of variance reduction and cost-effectiveness. Mukhopadhyay (2009) assessed the impact of sample size and replacement strategies on the efficiency of successive sampling techniques and posited that an optimal replacement fraction of around 50% provides the best trade-off between cost and efficiency in most survey settings. Särndal *et al.* (2003) examined the role of model-assisted estimation techniques in improving the efficiency of two-stage sampling procedures using statistical modeling techniques to national survey data from different domains and showed that model-assisted estimators significantly improve precision by effectively utilizing auxiliary data.

2.0Sampling Scheme

This study employs successive sampling scheme to assess gender enrollment in Ekiti State Secondary schools.

2.1 Sample structure and notations

Let U = (U_1, U_2 , ..., U_N) be the finite population of N units sampled over two occasions. The character under study is denoted by x(y) on the first (second) occasion, respectively. It is assumed that the information on two stable auxiliary variables z_1 and z_2 with known population means, which have positive and negative correlations respectively with the study variable x(y) on the first (second) occasion are available. A simple random sample (without replacement) of size n is drawn on the first occasion. A random subsample of size m= nλ is retained (matched) for its use on the second occasion, while a fresh simple random sample (without replacement) of size u = (n – m) = nμ is drawn on the second occasion from the entire population so that the sample size on the second occasion is also n. Here λ and μ are the fractions of the matched and fresh samples, respectively, at the current (second) occasion. The values of λ or μ are required to be chosen optimally such that μ+ λ =1.

The following notations will be used in this study:

X, Y : Population means of the study variables x and y respectively.

Z_1 , Z_2 : Population means of the auxiliary variables z_1 and z_2 respectively.

y_u , y_m , x_m , x_n , z_{ju} , z_{jm}, z_{jn} (j=1, 2,…): The sample means of the respective variables based on the sample sizes shown in suffices.

ρ_{yx} , ρ_{yz} , ρ_{xz} : Population correlation coefficients between the variables as shown in subscripts.

C_x , C_y , C_z : Coefficients of variation of the variable given in the subscripts.

$S_x^2, S_y^2,\ S_z^2$: Population variances of the variables x, y and z respectively.

To estimate the population mean Y on second (current) occasion, two different estimators may be formulated. One estimator is based on the fresh sample of size u drawn on the current occasion and the other estimator is based on matched sample of size m common to both occasions.

2.2 Selection of samples

Let a simple random sample of size n be selected on the first occasion from a population of size N and let its mean be $\bar{x}$. Let a simple random sample of size m be subsampled from the n units. Let the mean on the first occasion of the sample of size m be $\bar{x}_m$ and on the second occasion y_m'. In addition, let a simple random sample of size u be taken on the second occasion from the population N - m left after omitting the m units. Also let the total number of those sampled on the second occasion be m + u = n. We assume that simple random sampling is used and the finite population correction factor is ignored.

Let,

N be the number of units in the finite population

n be the number of units randomly sampled from the finite population.

$\bar{y}_1$be the estimate of the means on the first occasion based on initial sample size.

$\bar{y}_2$ be the estimate of the mean on the second occasion based on the current sample size.

$\bar{y}_1{}'$ be the estimate of the mean on the first occasion based on the matched sample size.

$\bar{y}_1{}''$ be the estimate of the mean on the first occasion based on the unmatched sample size.

$\bar{y}_2{}'$ be the estimate of the mean on the second occasion based on the matched sample size.

$\bar{y}_2{}''$ be the estimate of the mean on the second occasion based on the unmatched sample size.

2.3 Estimation of mean

The estimator of the population means on the second occasion based on the matched sample is given by $\bar{y}_{2m} = \bar{y}_2{}' - b(\bar{y}_1{}' - \bar{y}_2\ \ )$

The estimator of the population mean on the second occasion is $\bar{y}_2^* = \ \theta_1\bar{y}_{2m} +\ \theta_2\bar{y}_2{}''$

Where $\theta_1 +\ \theta_2 = 1$

Then, $\bar{y}_2^* = \ \theta_1\bar{y}_{2m} + (1 -\ \theta)y_2'')$

The variance of $\bar{y}_2^*$ is V($\bar{y}_2^*$) = $\theta_1^2 v(\bar{y}_{2m}) + (1 -\ \theta_1)^2\ V(\bar{y}_2'') + 2\theta_1\ (1 -\ \theta_1)cov(\bar{y}_{2m}, \bar{y}_2'')$

2.4 Sampling on Two Successive Occasions

It is assumed that the survey population remains unaltered from occasion to occasion. For the purpose of generality, let the sample size for the first occasion be n_1 and that for the second occasion be n_2 such that $n_2 = n_{12} + n_{22}$ where n_{12} is the number of common units between the 1st and the 2nd occasion and n_{22} units to be drawn afresh on the second occasion ,where the data obtained on current (second) occasion would be denoted by y and that on the previous occasion by x. Now the sampling procedure consists of the following steps:

Step (1a): From the given survey population choose a sample without replacement for survey on the first occasion. S_1 of size n_1 units by SRS;

Step (1b): On the second occasion choose a set, S_c of n_{12} units from the sample taken at step (1a) either by SRS or PPS sampling depending on the situation at hand and supplement it to another set, S_f of n_{22} units taken independently from the unsurveyed N- n_1 units of the population by SRS without replacement, so that the total sample $S_2 = S_c + S_f$ on the second occasion comprises $n_2 = n_{12} + n_{22}$ units, now as S_1 acts as preliminary sample, the estimate say t_c based on y and x values of S_c and x values of S_1 would be a double sampling ratio or regression or PPS estimate or

$$t_c = \frac{1}{N_{12}} \sum_{j=i}^{n_{12}} \frac{y_j}{P_j}$$

where $p_i = \frac{x_i}{X}$, x $\sum_i^{n_{12}} x_i$.

So we have,

$$E(t_c) = \bar{Y} \; and \; V(t_c) = \frac{A}{n_1} + \frac{B}{n_{12}} - \frac{1}{N} s_y^2$$

where A and B are quantities based on the population x and y values,

Here, $A = S_y^2$, $\quad B = \sigma^2 = \sum_i^N P_i \left[\frac{y_i}{Np_i} - \bar{Y}\right]^2$

Also in view of the selection of S_f as noted in the step (1b), it is obvious for the mean y_f of S_f, where $y_f = \frac{1}{n_{22}} \sum_{i=1}^{n_{22}} y_1$

$$E(\bar{y}_f) = \bar{Y}, v(\bar{y}_f) = \left[\frac{1}{n_{22}} - \frac{1}{N}\right] S_y^2$$

Further t_c and y_f are correlated in that,

$$cov(t_c, \bar{y}_f) = cov\left[E\left({}^{t_c}/_{S_1}\right), E\left({}^{\bar{y}_f}/_{S_1}\right)\right]$$

$$= cov\left[y_i, \frac{N\bar{Y} - n_1\bar{y}_1}{N - N_1}\right] = -\frac{1}{N} S_y^2$$

3.0RESULTS AND DISCUSSION

The population for this study comprises secondary school enrollment data collected from the Ekiti State Ministry of Education for the 2021/2022 and 2022/2023 academic sessions. A comparative analysis was conducted using a two-stage successive sampling approach to examine school enrollment trends based on gender across different local government areas in Ekiti State. The data utilized for this research is cross-sectional, providing detailed information on student enrollment patterns over the specified academic years. These records offer insights into enrollment distributions, variations, and trends across schools within the state. The study adopts statistical techniques such as mean, difference, and variance to analyze the relative efficiency of alternative sampling procedures. These methods facilitate the evaluation of enrollment consistency and changes over time while ensuring reliability in the sampling approach. By applying two-stage successive sampling, the study seeks to determine the effectiveness of different selection procedures for primary and secondary sampling units.

Table 1: Analysis on Male Enrollment

Item	**Value**
N	200
n	150
μ	0.18
λ	0.82
$\bar{y}_1$	276.81
$\bar{y}_2$	275.33
$\bar{y}'_1$	268.33

$\bar{y}_1''$	278.29
$\bar{y}_2'$	260.25
$\bar{y}_2''$	305.25

Estimating the Difference between the means

$\hat{\Delta} = \emptyset(\bar{y}_2' - \bar{y}_1') + (1-\emptyset)(\bar{y}_2'' - \bar{y}_1'')$

Where $0 \leq \emptyset \leq 1$; $\emptyset = 0.2$

$\hat{\Delta}$ = 0.2(260.25-268.33) + 0.8(305.25-278.29)

$\hat{\Delta}$ = 19.95

The minimum variance is obtained with the formula;

$V(\hat{\Delta}) = \frac{1-\rho}{1-\mu\rho}\frac{2s^2}{n}$

$\mu = 0.18$

$\rho = 0.02388697$

$V(\hat{\Delta})$ = 440.531

Estimation of overall mean

$\widehat{M} = \frac{1}{2}[\gamma(\bar{y}_2' + \bar{y}_1') + (1-\gamma)(\bar{y}_2'' + \bar{y}_1'')]$

Where $0 \leq \gamma \leq 1$; $\gamma = 0.2$

$\widehat{M}$ = 0.5*0.2(260.25+268.33) + 0.8(305.25+278.29)

$\widehat{M}$ = 0.5*(572.55)

$\widehat{M}$= 286.28

The minimum variance is obtained with the formula;

$V(\widehat{M}) = \frac{1+\rho}{1+\mu\rho}\frac{s^2}{2n}$

$\mu = 0.18$

$\rho = 0.0238$

$V(\widehat{M})$ = 33702.80

Table 2: Iterations of Estimated Variances for male enrollment Using Different Values of ρ and μ

$\boldsymbol{\rho}$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$\boldsymbol{V(\hat{\Delta})}$	449.37	411.85	372.92	332.52	290.54	246.91	201.51	154.25	104.99	53.62	0.00
$\boldsymbol{V(\widehat{M})}$	2527710.19	2731317.49	2927849.64	3117669.11	3301114.05	3478500.26	3650123.02	3816258.72	3977166.38	4133088.95	4284254.56
$\boldsymbol{\mu}$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$\boldsymbol{V(\hat{\Delta})}$	438.64	439.69	440.74	441.80	442.87	443.94	445.01	446.10	447.18	448.27	449.37
$\boldsymbol{V(\widehat{M})}$	2588089.53	2581922.10	2575783.99	2569675.00	2563594.92	2557543.54	2551520.67	2545526.09	2539559.62	2533621.05	2527710.19

Discussion

The analysis of male enrollment using two-stage successive sampling involves estimating differences between means, overall means, and their corresponding variances. With a k=total population size of $N = 200$ and a sample size of $n = 150$, the study estimates the difference between means ($\hat{\Delta}$) using weighted averages of the two-stage mean estimates ($\bar{y}_1$ and $\bar{y}_2$) with a weighting factor ($\phi = 0.2$). The computed difference is 19.95, indicating a variation in male enrollment trends across successive periods. The minimum variance for this estimate ($\boldsymbol{V(\hat{\Delta})}$) is derived using the given formula, incorporating parameters $\mu = 0.18$ and $\rho = 0.0238$, yielding a variance of 440.53. The overall mean estimate $(\widehat{\boldsymbol{M}})$ is computed similarly, incorporating weighted averages of primary and secondary sampling unit means, resulting in an estimated value of 286.28. The variance for the overall mean estimate ($\boldsymbol{V}(\widehat{\boldsymbol{M}})$) is found to be 33702.80 indicating variability in male enrollment figures.

The iterations of estimated variances for male enrollment using different values of ρ and μ in Table 1 reveal some key trends: as ρ increases from 0 to 1, $\boldsymbol{V}(\hat{\Delta})$ decreases steadily from 449.37 to 0, indicating greater stability in mean difference estimates. Conversely, $\boldsymbol{V}(\widehat{\boldsymbol{M}})$ increases from 2527710.19 to 4284254.56, highlighting rising variability in the overall mean as ρ increases. Similarly, as μ increases, $\boldsymbol{V}(\hat{\Delta})$ increases slightly, while $\boldsymbol{V}(\widehat{\boldsymbol{M}})$ gradually decreases. This suggests that higher values of ρ contribute to reduced variation in difference estimates, while increasing μ stabilizes the overall mean. The results emphasize the importance of selecting appropriate values of ρ and μ to optimize sampling efficiency and minimize error in successive-stage enrollment studies.

Table 3: Analysis on Female Enrollment

Item	Value
N	200
n	150
μ	0.25
λ	0.75
$\bar{y}_1$	288.45
$\bar{y}_2$	295.05
$\bar{y}_1'$	283.28
$\bar{y}_1''$	296.38
$\bar{y}_2'$	279.33
$\bar{y}_2''$	309.25

Estimating the Difference between the means

$\hat{\Delta} = \emptyset(\bar{y}_2' - \bar{y}_1') + (1 - \emptyset)(\bar{y}_2'' - \bar{y}_1'')$

Where $0 \leq \emptyset \leq 1$; $\emptyset = 0.2$

$\hat{\Delta}$ = 0.2(279.3333-283.2832) + 0.8(309.25-296.3784)

$\hat{\Delta}$ = 9.5073

The minimum variance is obtained with the formula;

$V(\hat{\Delta}) = \frac{1-\rho}{1-\mu\rho}\frac{2s^2}{n}$

μ = 0.25

ρ = 0.06484925

$V(\hat{\Delta})$ = 624.247

Estimation of overall mean

$\hat{M} = \frac{1}{2}[\gamma(\bar{y}_2' + \bar{y}_1') + (1-\gamma)(\bar{y}_2'' + \bar{y}_1'')]$

Where $0 \leq \gamma \leq 1$; γ= 0.2

$\hat{M}$ = 0.5*0.2(260.2544+268.3303) + 0.8(305.25+278.2927)

$\widehat{M}$ = 0.5*(597.026)

$\hat{M}$ = 298.513

The minimum variance is obtained with the formula;

$V(\hat{M}) = \frac{1+\rho}{1+\mu\rho}\frac{s^2}{2n}$

μ = 0.25

ρ = 0.06484925

$V(\hat{M})$ = 3766061.5803

Table 4: Iterations of Estimated Variances for female enrollment Using Different Values of ρ and μ

$\boldsymbol{\rho}$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$\boldsymbol{V(\hat{\Delta})}$	656.72	601.88	544.99	485.94	424.60	360.83	294.49	225.42	153.44	78.37	0
$\boldsymbol{V(\hat{M})}$	3694016	3991569	4278782	4556186	4824274	5083508	5334318	5577110	5812262	6040129	626104
$\boldsymbol{\mu}$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$\boldsymbol{V(\hat{\Delta})}$	641.03	642.56	644.10	645.65	647.21	648.77	650.35	651.93	653.52	655.11	656.71
$\boldsymbol{V(\hat{M})}$	3782254	3773241	3764271	3755343	3746458	3737614	3728812	3720052	3711332	3702654	3694016

Discussion

The analysis of female enrollment, as presented in Table 2, examines the estimated variances for the difference in means ($\boldsymbol{V(\hat{\Delta})}$) and the overall mean ($\boldsymbol{V(\hat{M})}$) under different values of ρ and μ. The results indicate that as ρ increases from 0 to 1, $V(\hat{\Delta})$ steadily declines from 656.72 to 0, demonstrating that higher values of ρ contribute to reducing variability in the estimation of mean differences. Conversely, $\boldsymbol{V(\hat{M})}$ shows a progressive increase from 3,694,016 at $\rho = 0 \; to \; 6{,}261{,}043 \; at \; \rho = 1$, implying greater uncertainty in the overall mean estimation as the correlation between sampling occasions strengthens. Similarly, when μ increases from 0 to 1, $\boldsymbol{V(\hat{\Delta})}$ rises from 641.03 to 656.71, reflecting a slight increase in the variability of mean difference estimates. In contrast, $\boldsymbol{V(\hat{M})}$ decreases from 3,782,254 at $\mu = 0 \; to \; 3{,}694{,}016 \; at \; \mu = 1$, indicating improved stability in the overall mean estimate. These trends suggest that lower values of ρ are preferable for achieving precision in estimating the overall mean, while higher values help minimize fluctuations in mean differences. The results emphasize the importance of selecting optimal ρ and μ values to balance stability and efficiency in estimating female enrollment trends across successive sampling occasions.

4.0 Conclusion

This study evaluated two-stage sampling on successive occasions, focusing on data from secondary school students' enrollment in Ekiti State. The findings reveal that the choice of ρ and μ significantly impacts the precision of estimated variances for mean differences and overall mean. Lower values of ρ improve accuracy in estimating overall means, while higher values minimize fluctuations in mean differences, emphasizing the need for an optimal balance in parameter selection. The comparative analysis of the male and female enrollment highlights variations in efficiency, showing the importance of choosing appropriate sampling strategies to enhance the reliability of successive surveys. The results will contribute to improving statistical methodologies for educational data collection and policy decision-making, ensuring more accurate and consistent enrollment estimates over time.

References

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). Wiley.

Eckler, A. R. (1955). Properties of successive sampling estimators. *Journal of the American Statistical Association, 50*(271), 644–654.

Jessen, R. J. (1942). Statistical investigation of a sample survey for obtaining farm facts. *Iowa State College Journal of Science, 16*, 249–253.

Jessen, R. J. (1978). *Statistical survey techniques.* Wiley.

Mukhopadhyay, P. (2009). *Theory and methods of survey sampling*. PHI Learning.

Patterson, H. D. (1950). Sampling on successive occasions with partial replacement of units. *Journal of the Royal Statistical Society: Series B (Methodological), 12*(2), 241–255.

Raj, D. (1968). *Sampling theory*. McGraw-Hill.

Särndal, C. E., Swensson, B., and Wretman, J. (2003). *Model assisted survey sampling*. Springer.

Singh, D., and Chaudhary, F. S. (2017). *Theory and analysis of sample survey designs*. New Age International.

Tikkiwal, G. S. (1950). Successive sampling methods and their applications. *Sankhya: The Indian Journal of Statistics, 10*(3), 251–264.