Communication in Physical Sciences, 2024, 12(1): 020-025

Evaluating The Performances of Estimators of Population Mean Weight of Babies in FMC, Imo State Under Simple Random Sampling Scheme

Loveline Chiamaka Okoro and Emmanuel John Ekpenyong
Received: 12 August 2024/Accepted: 04 November 2024/First Published 14 November 2024, doi: https://dx.doi.org/10.4314/cps.v12i1.3

Abstract: *Gestational age plays a vital role in obstetrics. Accurately estimating the average gestational age in pregnant women will help ascertain the growth of the fetus and it is also essential in structuring prenatal care, including decisions about timing and route of delivery. . This study compares the efficiency of some existing estimators of population mean using simple random sampling scheme. The estimators were compared using a real data on gestational age incorporating the weight of babies as auxiliary variable. Three samples of (n = 100, 150, 200) were selected from the population for the analysis. Of all the estimators compared, result showed that the classical regression estimator* t_6 *and Kadilar (2016) estimator* t_4 *which approximates to the regression estimator are equally efficient and also proved to be the most efficient estimators with a lowest mean squared errors and highest percent relative efficiencies. Thus,* t_4 *and* t_6 *can used to estimate the population mean of the auxiliary variable in practice.*

Keywords:: *Efficiency, Bias, Mean Squared Error, Estimator, Simple Random Sampling*

Okoro, Loveline Chiamaka
Department of Statistics, Michael Okpara University of Agriculture, Umudike, Abia State, Nigeria

Emmanuel John Ekpenyong
Department of Statistics, Michael Okpara University of Agriculture, Umudike, Abia State, Nigeria
Email: ej.ekpenyong@mouau.edu.ng
Orcid id: 0000-0003-4112-3712

1.0 Introduction

Gestational age (GA) estimation is essential for optimal maternity care. Knowledge of gestational age (GA) is a key in clinical management of individual obstetric patients, and critical to be able to calculate rates of preterm birth and small for gestational age at a population level, Elizabeth *et al.*(2024). Gestational age indicates how far along a pregnancy is and it will help ascertain the growth of the fetus and it is also essential in structuring prenatal care, including decisions about timing and route of delivery. An accurate estimation of population parameter is essential for healthcare professionals and researchers.

In sample theory, simple random sampling is a basic sampling procedure which forms the basis for all other sampling procedures. It is a sampling technique in which all the units in the population have equal chances of being selected for the sample. It is widely used in various fields like medicine, social sciences and economics. The simple random sampling method serves as a basis for a range of estimation methods, such as the ratio estimator proposed by Cochran (1940), the regression estimator proposed by Watson (1937), the product estimator proposed by Murthy (1964), the exponential estimator proposed by Bahl and Tuteja (1991) and some other estimators. Literature in survey sampling has shown that these estimators outperform the sample mean estimator due to the use of auxiliary variables.

In the estimation process, the use of auxiliary information on an auxiliary variable alongside with the study variable plays a key role in boosting the efficiency of the estimator of the population parameter under study. When estimating the finite population mean, various methods such as ratio, regression and exponential estimators come into play when there is a direct correlation between the study variables and the auxiliary variables Khazan *et al.* (2024).

Kuldeep *et al.* (2022) carried out a comparative study on ratio and regression estimators. They stated that the regression estimator always performs better than the classical ratio estimator except when the regression line passes through the origin. If the regression line passes through the origin then both the classical ratio and regression estimators are equally efficient. After carrying out a comparative analysis on both estimators, the result of the study showed that the linear regression estimator always works better than the ratio estimator.

Oke *et al.* (2023) carried out a comparative Study of Ratio and Regression Estimators using Double Sampling for Estimation of Population Mean. They used three methods of estimation namely, double sampling for ratio estimator, simple random sample without replacement, and double sampling for regression estimator. Their study aims to explore the preference order regarding the utilization of different estimation methods in sample surveys. They collected data on salary and expenditure for their numerical analysis. Also different samples were selected from the population both in the first and second phase. From the result of their analysis, the regression estimator showed the least variability, making it the most effective estimator in terms of efficiency.

Hence, this study aims to assess the performance of selected estimators of population mean in simple random sampling using real data on baby weight and gestational age from a hospital. From the assessment, the study will be able to know the estimator which will give a more precise estimate of the average of the gestational age in babies using information on baby weight as an auxiliary variable.

1.0 Sampling procedures and notations

Let $U = \{U_{1,}\ldots,U_N\}$ be a finite population of newborn babies of size 430 and let $(yi,\ xi)$ be the value of the study variable Y (gestational age) and the auxiliary variable X (baby weight) on ith unit $Ui, i = 1,\ldots N$. Let and $\bar{X}$ be the population means of the study variable Y and the auxiliary variable X respectively. Let a sample of size (n)be drawn by simple random sampling without replacement (SRSWOR) based on which we obtain the means $(\underline{x})$ and () for the auxiliary variable (X) and the study variable(Y). We assume that the population mean $\bar{X}$ and the population variance S_x^2 of the auxiliary variable are known. The following notations are defined:

$C_x = S_x \bar{X}^{-1}$, Coefficient of variation of the auxiliary variable

$C_y = S_y \bar{Y}^{-1}$, Coefficient of variation of the study variable

$\rho = S_{xy}\left(S_y S_x\right)^{-1}$, Correlation coefficient between the auxiliary and study variables

$f = \dfrac{n}{N}$, the sampling fraction

$S_x^2 = \sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2 \Big/ \left(N-1\right)$, population variance of the auxiliary variable

$S_y^2 = \sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2 \left(N-1\right)^{-1}$, population variance of the study variable

$S_{xy}=\sum_{i=1}^{N}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)\Big/N-1$, population covariance between the auxiliary and study variables

$\bar{X}=\left(\sum_{i=1}^{N}X_i\right)\left(N\right)^{-1}$, population mean of the auxiliary variable

$\bar{Y}=\left(\sum_{i=1}^{N}Y_i\right)\left(N\right)^{-1}$, population mean of the study variable

$\bar{x}=\left(\sum_{i=1}^{n}x_i\right)\left(n\right)^{-1}$, sample mean of the auxiliary variable

$\bar{y}=\left(\sum_{i=1}^{n}y_i\right)\left(n^{-1}\right)$, sample mean of the study variable

$\lambda=\left(1-f\right)/n$

2.1 Methodology

In this section, some estimators of population mean that will be used for the comparison purpose will be examined. These estimators utilize a single auxiliary variable in the context of simple random sampling. The estimators, their biases and mean squared errors are listed in the table below.

Table 1. Some existing estimators in literature used for the comparative study

S/N	Estimators	Bias	MSE
1	$t_y=\bar{y}$ (sample mean estimator)	0	$\text{Var}(t_y)=\lambda\bar{Y}^2C_y^2$
2	$t_1=\frac{\bar{y}}{\bar{x}}\bar{X}$ Cochran (1940)	$Bias(t_1)=\lambda\bar{Y}\left[C_x^2-\rho C_yC_x\right]$	$\text{MSE}(t_1)=\lambda\bar{Y}^2\left[C_y^2+C_x^2-2\rho C_yC_x\right]$
3	$t_2=\bar{y}\frac{\bar{x}}{\bar{X}}$ Murthy (1964)	$Bias(t_2)=\lambda\bar{Y}\rho C_yC_x$	$MSE(t_2)=\lambda\bar{Y}^2\left(C_y^2+C_x^2+2\rho C_yC_x\right)$
4	$t_3=\bar{y}\exp\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right)$ Bahl and Tuteja (1991)	$Bias(t_3)=\bar{Y}\lambda\left[\frac{3C_x^2}{8}-\frac{\rho CyC_x}{2}\right]$	$\text{MSE}(t_3)=\bar{Y}^2\lambda\left[C_y^2+\frac{C_x^2}{4}-\rho CyC_x\right]$
5	$t_4=\bar{y}\left(\frac{\bar{x}}{\bar{X}}\right)^{\alpha}\exp\left[\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right]$ Kadilar, (2016)	0	$MSE\left(t_4\right)=\lambda\bar{Y}^2\left(C_y^2+\frac{C_x^2}{4}+2\alpha\rho C_xC_y+\rho C_xC_y+\alpha^2C_x^2+\alpha C\right)$

6	$t_5 = \bar{y}\left(\frac{\bar{X}}{\bar{x}}\right)^2$ Kadilar and Cingi (2003)		$\text{Bias}(t_5) = \lambda \bar{Y} C_x^2 (1-2k)$	$MSE(t_5) = \lambda \bar{Y}^2 \left[C_y^2 + 4C_x^2(1-k)\right]$
7	$t_6 = \bar{y} + b_{yx}\left(\bar{X} - \bar{x}\right)$ Watson (1937)		0	$MSE(t_6) = \lambda \bar{Y}^2 C_y^2 \left(1 - \rho_{yx}^2\right)$
8	$t_7 = 2^{-1}\bar{y}\left[\left(\frac{\bar{x}}{\bar{X}}\right)^{\alpha} + \exp\left[\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right]\right]$ Yunusa *et al.* (2021)		$Bias(t_7) = \bar{Y}\lambda\left[\left(\frac{\alpha(\alpha-1)}{4} + \frac{3}{16}\right)C_x^2 + \left(\frac{\alpha}{2} - \frac{1}{4}\right)\rho C_y C_x\right]$	$MSE(t_7) = \bar{Y}^2\lambda\left[C_y^2 + \left(\frac{\alpha}{2} - \frac{1}{4}\right)^2 C_x^2 + 2\left(\frac{\alpha}{2} - \frac{1}{4}\right)\rho C_y C_x\right]$
9	$t_8 = \bar{y}\left(\frac{\bar{X}+n}{\bar{x}+n}\right)^{\gamma}$ Muhammad *et al.* (2023)		$Bias(t_8) = \bar{Y}\left[\frac{\gamma(\gamma-1)}{2}\delta^2\lambda C_x^2 - \gamma\delta\rho C_y C_x\right]$	$MSE(t_8) = \bar{Y}^2\left[\lambda C_y^2 + \gamma^2\delta^2\lambda C_x^2 - 2\gamma\delta\lambda\rho C_y C_x\right]$

2.2 *Numerical study*

The data utilized for this study is a secondary data on gestational age and weight of babies sourced from Federal Medical Centre (FMC) Umuahia, Abia State Nigeria, from 2019 to 2024
Y (gestational age) is the study variable, and
X (weight of babies) is the auxiliary variable

N = 430, n = 100, $\bar{X}$ = 3.0935, $\bar{Y}$ = 38.0744, C_x = 0.1368, C_y = 0.0561, ρ_{xy} = 0.5121

N = 430, n = 150, $\bar{X}$ = 3.0935, $\bar{Y}$ = 38.0744, C_x = 0.1368, C_y = 0.0561, ρ_{xy} = 0.5121

N = 430, n = 200 $\bar{X}$ = 3.0935, $\bar{Y}$ = 38.0744, C_x = 0.1368, C_y = 0.0561, ρ_{xy} = 0.5121

The efficiency comparison was done by obtaining the percent relative efficiency (PRE) which is evaluated as

$$PRE = \frac{Var(t_y)}{MSE(t)} \; X \; 100$$

where, $t = t_y, t_1, t_2, t_3, t_4, t_5, t_6, t_7, t_8$

A PRE that is greater than 100 shows an increase in efficiency, while the PRE that is less than 100 shows a decrease in efficiency.

The mean squared errors (MSEs) and percentage relative efficiencies (PREs) of the different estimators of the population mean in relation to the sample mean based on three (3) samples from the given populations are given in Table 2.

Table 2: MSEs and PREs of the selected estimators used for the comparison.

S/N	Estimators	n = 100		n = 150		n = 200	
		MSE	**PRE**	**MSE**	**PRE**	**MSE**	**PRE**
1	t_y	0.0350	100	0.0198	100	0.0122	100

2	t_1	0.1558	22.4781	0.0881	22.4781	0.0543	22.4781
3	t_2	0.3307	10.5890	0.1870	10.5890	0.1152	10.5890
4	t_3	0.0433	80.7874	0.0245	80.7874	0.0151	80.7874
5	t_4	**0.0258**	**135.5466**	**0.0146**	**135.5466**	**0.0090**	**135.5466**
6	t_5	0.6929	5.0530	0.3920	5.0530	0.2415	5.0530
7	t_6	**0.0258**	**135.5466**	**0.0146**	**135.5466**	**0.0090**	**135.5466**
8	t_7	0.9122	3.8383	0.5160	3.8383	0.3179	3.8383
9	t_8	0.0448	78.0995	0.0254	78.0995	0.0156	78.0995

3.2 Discussion of result

Table 2 shows the numerical result of (MSE and PRE) of the estimators $t_y, t_1, t_2, t_3, t_4, t_5, t_6, t_7, t_8$ using three (3) samples from the population under study. In this study, among all the estimators considered the estimators t_4 and t_6 are equally efficient.

Table 3.: Ranks of the performance of the various estimators

S/N	Estimators	Ranks
1	t_6	1
2	t_4	1
3	t_y	2
4	t_3	3
5	t_8	4
6	t_1	5
7	t_2	6
8	t_5	7
9	t_7	8

This is because t_4 is an estimator that approximates to the regression estimator t_6 at the optimal values of the unknown constant α . Also, t_4 and t_6 were observed to have smaller mean squared errors of 0.0258, 0.0146 and 0.0090 and higher percent relative efficiencies of 135.5466%, 135.5466 % and 135.5466% respectively in the three (3) samples selected from the population for the comparative analysis, and thus they performed substantially better than the other estimators $t_y, t_1, t_2, t_3, t_5, t_7, t_8$.

2.0 Conclusion

The efficacy of nine estimators in literature has been compared using real data application on gestational age and baby weights.

From the numerical results, we can deduce that the estimators t_4 and t_6 are more efficient than all other estimators considered in this study, and hence, it is recommended for use for estimating the population mean in practice.

3.0 References

Bahl, S. & Tuteja, R. K. (1991), Ratio and product type exponential estimators. *Journal of Information and Optimization Sciences*, 12, 2, pp. 159-164.

Cochran, W. G., (1940). The estimation of the yields of the cereal experiments by sampling for the ratio of grain to total produce. *Journal of Agricultural Science,* 30, pp. 262 – 275.

Elizabeth, B., Agustin, C., Nia, W. R., Jose, V. & Aris, T. P. (2024). Accuracy of prenatal and postnatal biomarkers for estimating gestational age: A systematic review and meta-analysis. e*clinicalmedicine*, 70, pp. 1-22.

Kadilar, C. & Cingi, H. (2003). A study on the chain ratio-type estimator, Hacettepe

Journal of Mathematics and Computation, 32, pp. 105 - 108.

Kadilar, G. O. (2016). A new exponential type estimator for the population mean in simple random sampling. *Journal of Modern Applied Statistical Methods*, 15, 2, pp. 207– 214.

Khazan, S., Muhammad, A., Muhammad, M. H., Olayan, A. and Ayesha, A. (2024). Development of improved estimators of finite population mean in simple random sampling with dual auxiliaries and its application to real world problems. *Heliyon* 10(1)

Kuldeep, K. T., Sandeep, B. and Sunil K. (2022). A comparative study of ratio estimators with regression estimators. *Journal of Statistics*, 9, 2, pp. 101-104.

Murthy, M. N. (1964). Product method of estimation. *Sankhya*, 26, pp. 294-307.

Muhammad, I., Oyeyemi, G. M. and Kareem, A. O. (2023). Combined Exponential-Type Estimators for Finite Population Mean in Two-Phase Sampling. *Asian Journal of Probability and Statistics*, 2, 2, pp. 44-58.

Oke, S. A., Adesina, O. A., Oladimeji, L. A., Akinade, O. O., Oguntola, T. O., Tijani, R. A. and Adegoke, M. A. (2023). Comparative Study of Ratio and Regression Estimators using Double Sampling for Estimation of Population Mean. *International Journal of Innovative Science and Research Technology*, 8, 7, pp. 2341-2347.

Yunusa, M. A., Audu, A., Ishaq, O. O. and Beki, D. O. (2021). An efficient exponencial type estimator for estimating finite population mean under simple random sampling. *Annals computer science series* vol. XIX.

Compliance with Ethical Standards Declaration

Ethical Approval

Not Applicable

Competing interests

The authors declare that they have no known competing financial interests

Funding

The author declared no source of external funding

Availability of data and materials

Data would be made available on request**.**

Author's Contribution

Both authors contributed to the writing of the manuscript, under the supervision of EJE