



On the use of auxiliary information to improve the precision of estimator for public primary pupils' enrolment in Kaduna state

^{1,*}Mudi T.A.

²Abdulrahman M.K.A.

³Arowona N.A.

^{1,2,3} Department of Mathematics & Statistics,
Kaduna Polytechnic, Nigeria

*adamu110202@gmail.com

Abstract: In this paper, a comparative estimators for public primary pupils' enrolment in Kaduna state with and without using auxiliary variable was made. The objectives are to estimate the mean enrolments and obtain a more efficient estimator for the data using the Variances of their mean enrolments. The data used is the annual enrolment into public primary schools in 2020 obtained from Kaduna State Annual School Census (ASC) report 2020. Stratified Random Sampling design was used with equal allocation in which three local governments were selected from each Senatorial zone making a sample of 9 local governments. The results of the estimator using auxiliary information (Regression) revealed that the estimated mean enrolment is approximately 95,933 with its associated variance as 159,353,239.3. For ordinary estimator (without auxiliary information) the estimated mean enrolment is approximately 96,057 with variance of 159,501,807.3. It can be infer that in estimating the public primary pupils' enrolment for the state in general, the Regression is more efficient.

Keywords: Estimator; Enrolment; Efficiency; Auxiliary Variable

1.0 INTRODUCTION

Any information that is closely related to the study variable is defined as auxiliary information. This can be used in the designing, selection and estimation stages of the survey. Use of this information usually leads to the sampling strategy with higher efficiency compared to those in which no auxiliary information is used. Higher precision can be achieved by using the auxiliary information for the dual purposes of selection and the estimation procedure. It is well established in sample surveys that auxiliary information is often used to improve the precision of estimators of population parameters [1]. It is well known that for estimating population mean \bar{Y} of the study variable y using information on auxiliary variable x (say population mean \bar{X} of x is known in advance) at the estimation stage, the ratio, product, difference and regression methods are extensively used. The properties of these methods have been dealt in detail in Cochran [2]. The regression method of estimation which has been developed by [3] is used to estimate the population mean \bar{Y} of the study variable y which is correlated with an auxiliary variable x . For example Y may be the production of wheat and x is the area cultivated. Y may denote the number of bullocks on a holding and x its area in acres. Y may be the total production and x is the number

Citation: Mudi TA., Abdulrahman MKA and Arowona NA. On the use of auxiliary information to improve the precision of estimator for public primary pupils enrolment in Kaduna state. *IJMSSC*, 2022. 1(1): p. 1-5.



of workers. y may be the pupil's enrolment and x is the number of usable classroom (as the case in this paper)

However, in estimating pupils enrolment using number of classrooms as auxiliary variable, the regression estimator is a better choice as compared to ratio estimator [4]. In the situation where the relation between the study variable y and the auxiliary variable x is a straight line and passing through the origin, the ratio and product estimators have efficiencies equals to the regression estimator and difference estimator perform equals to the regression estimator at their optimum condition. Several researchers such as [5] have proposed both stratified and post stratified estimators for regression and ratio estimation methods. In [6], a new approach was develop to use ratio estimation that produces a more precise and efficient ratio estimator that is superior to the regression estimator both in efficiency and biasedness.

School Enrolment: refers to the fact that a child's name is entered or remains on the rolls or register of a school as a pupil. The term is also used to mean the total number of pupils on the school rolls at a given time, or sometimes the average number of pupils enrolled during a given period such as a school year.

Efficiency of an Estimator: Efficiency is a term applied in the context of comparing different method of estimating the same parameter; the estimate with the lower variance being regarded as the most efficient. It is also used when comparing competing experimental designs; with one design being more efficient than another if it can achieve the same precision with fewer resources. If we have more than one consistent estimators of a parameter Y , then efficiency is the criterion which enables us to choose between them by considering the variances of the sampling distribution of the estimators. Thus, if $Y_1, Y_2, Y_3, \dots, Y_n$ are consistent estimators of parameter Y such that $Var(Y_1) < Var(Y_2) < Var(Y_3) < \dots < Var(Y_n)$; for all n , then Y_1 is said to be more efficient than Y_2, Y_3, \dots, Y_n . In other words, an estimator with lesser variability is said to be more efficient and consequently more reliable than the others. If there exist more than two consistent estimators for a parameter Y , then, considering the class of all such possible estimators, we can choose the one whose sampling variance is minimum [7]. Such an estimator is known as the most efficient estimator and provides a measure of the efficiency of the other estimators.

2.0 THE REGRESSION ESTIMATOR

When the regression of Y on X is linear, it is not necessary that the line should always pass through origin. Under such conditions, it is more appropriate to use the regression type estimator to estimate the population means. Suppose a random sample of size n on paired observations on (x, y) , $i=1, 2, \dots, n$ is drawn by Simple Random Sampling Without Replacement (SRSWOR). When slope of the regression line β , is unknown, it is estimated as

$$\hat{\beta} = \frac{S_{xy}}{S_x^2} \quad (1.1)$$

and then the regression estimator of \bar{Y} and Variance are given by

$$\hat{Y}_{reg} = \bar{y} + \hat{\beta}(\bar{X} - \bar{x}) \quad (1.2)$$

Citation: Mudi TA., Abdulrahman MKA and Arowona NA. On the use of auxiliary information to improve the precision of estimator for public primary pupils enrolment in Kaduna state. *IJMCS*, 2022. 1(1): p. 1-5.



$$\text{Var}(\hat{Y}_{reg}) = \frac{f}{n} [S_y^2 + \beta_0^2 S_x^2 - 2\beta_0 \rho S_x S_y] \quad (1.3)$$

and

Now, substituting $\beta = \frac{S_{xy}}{S_x^2} = \rho \frac{S_x}{S_y}$ in (1.3), the variance becomes

$$\text{Var}(\hat{Y}_{reg}) = \frac{f}{n} S_y^2 (1 - \rho^2) \quad (1.4)$$

Where, $f = \frac{N-n}{N}$, $S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2$, $S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2$ and $\rho =$ correlation coefficient

between X and Y . The regression estimator is at least as efficient as the ordinary estimator i.e.

$\text{Var}(\hat{Y}) \geq \text{Var}(\hat{Y}_{reg})$ since

$$\text{Var}(\hat{Y}) - \text{Var}(\hat{Y}_{reg}) = \left(1 - \frac{n}{N}\right) \frac{1}{n} [S_y^2 - s_{reg}^2] \geq 0 \quad (1.5)$$

Where (1.1),(1.2), (1.3), (1.4) and (1.5) are as shown in [8]

3.0 METHODOLOGY AND DATA PRESENTATION

The area of study of this research work is the primary pupils' population in Kaduna state. It covers the primary pupils' enrolment in the 23 local Government Areas (LGAs) divided into 3 senatorial zones across the state. The state Kaduna, consists of N=23 number of Local Government Areas (LGAs), using the senatorial zones as strata, simple random sample of three local governments were selected from each Senatorial zone using a table of random number making a sample of n =9 local government areas.

3.1 Data Presentation and Analysis

The data used is the annual enrolment into public primary schools in 2017 obtained from Kaduna State Annual School Census (ASC) report 2020 as presented in Table 3.1 below

Table 3.1: Public Primary Pupils' Enrolment in Kaduna State by Gender

L.G.A	No. of Classrooms (x)	Pupils Enrolment (y)		
		Primary	Pre primary	Total
Kubau	1059	127,644	25,547	153,191
Makarfi	549	63,697	7,614	71,311
Zaria	780	139,594	19,886	159,480
Birnin Gwari	794	125,121	8,406	133,527
Giwa	828	114,196	16,659	130855
Kajuru	1093	38,827	6,290	45117
Kachia	1363	80,154	4,937	85,091
Kaura	715	32,780	9,365	42,145
Sanga	931	37,317	6,480	43,797

Source: Kaduna State Annual School Census (ASC) Report 2020

Citation: Mudi TA., Abdulrahman MKA and Arowona NA. On the use of auxiliary information to improve the precision of estimator for public primary pupils enrolment in Kaduna state. *IJMCS*, 2022. 1(1): p. 1-5.



From the table above we have the following summary statistics and the summary of results

$$\sum_{i=1}^9 x_i = 8,112 \qquad \sum_{i=1}^9 x_i^2 = 7,777,706 \qquad \sum_{i=1}^9 x_i y_i = 776,342,382$$

$$\sum_{i=1}^9 y_i = 864,514 \qquad \sum_{i=1}^9 y_i^2 = 1.019095026 \times 10^{11} \qquad n = 9$$

$$\bar{x} = 901.3333 \qquad \bar{y} = 96,057.1111 \qquad \bar{X} = \frac{20,267}{23} = 881.1739$$

$$S_x^2 = \frac{\sum_{i=1}^9 x_i^2 - n\bar{x}^2}{n-1} = 58,261.25 \qquad S_y^2 = \frac{\sum_{i=1}^9 y_i^2 - n\bar{y}^2}{n-1} = 2,358,348.151 \qquad S_{xy}^2 = \frac{\sum_{i=1}^9 x_i y_i - n\bar{x}\bar{y}}{n-1} = -359,112.9167$$

$$\hat{\beta} = \frac{\sum_{i=1}^9 x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^9 x_i^2 - n\bar{x}^2} = -6.1638 \qquad \hat{\rho} = \frac{\sum_{i=1}^9 x_i y_i - n\bar{x}\bar{y}}{\sqrt{\left(\sum_{i=1}^9 x_i^2 - n\bar{x}^2\right)\left(\sum_{i=1}^9 y_i^2 - n\bar{y}^2\right)}} = -0.0306$$

$$\hat{Y}_{reg} = \bar{y} + \hat{\beta}(\bar{X} - \bar{x}) = 95,932.8516$$

$$\text{var}(\hat{Y}_{reg}) = \frac{f}{n} S_y^2 (1 - \hat{\rho}^2) = 159,353,239.3 \qquad \text{var}(\hat{Y}) = \frac{f}{n} [S_y^2] = 159,501,807.3$$

$$SE(\hat{Y}_{reg}) = \sqrt{\text{Var}(\hat{Y}_{reg})} = \sqrt{159,353,239.3} = 12,623.52 \qquad SE(\hat{Y}) = \sqrt{\text{Var}(\hat{Y})} = \sqrt{159,501,807.3} = 12,629.40$$

$$CV(\hat{Y}_{reg}) = \frac{SE(\hat{Y}_{reg})}{\hat{Y}_{reg}} \times 100 = \frac{12,623.52}{95,932.8516} \times 100 = 13.16$$

$$CV(\hat{Y}) = \frac{SE(\hat{Y})}{\hat{Y}} \times 100 = \frac{12,629.40}{96,057.1111} \times 100 = 13.15$$

Table 3.2 Summary of Results

	Estimator	Regression	Ordinary Mean
Population Mean	\bar{Y}	95,932.8516	96,057.1111
Variance	$V(\bar{Y})$	159,353,239.3	159,501,907.3
Standard Error	$SE(\bar{Y})$	12,623.52	12,629.40
Coefficient of variation	$CV(\bar{Y})$	13.16	13.15

4.0 DISCUSSION OF RESULT

In estimating pupils enrolment using number of classrooms as auxiliary variable, the regression estimator is a better choice as compared to ratio estimator both using information from auxiliary variable [4]. However, the choice of regression estimator may not lead to the sampling strategy with higher efficiency compared to those in which no auxiliary information is used. Hence the motivation for this further research. From the Table 3.2, it was observed that the regression

Citation: Mudi TA., Abdulrahman MKA and Arowona NA. On the use of auxiliary information to improve the precision of estimator for public primary pupils enrolment in Kaduna state. *IJMCS*, 2022. 1(1): p. 1-5.



estimator has the least for the variance and standard error of the population mean except for coefficient of variation of the population mean which both have approximately equal value.

5.0 CONCLUSION

From the comparison of the results obtained so far, it shows that with the same sample size, estimator using information on auxiliary variable (Regression) is more efficient than the ordinary estimator that does not use the auxiliary information (Simple Random sampling Estimator). For efficient result in the estimation of population mean of pupil enrolment, Regression estimator gives a more precise result than simple random sampling estimator with the same sample size and hence, recommended.

REFERENCES

- [1] Hansen, M. N., Hurwitz, W. N., & Madow, W. G., *Sample Survey Methods and Theory*. 1953 New York, USA.: John Wiley and Sons.
- [2] Cochran, W. G., *Sampling Techniques.*, 2 ed 1977: New York, USA: John Wiley and Sons.
- [3] Cochran, W. G., The estimation of the yields of cereal experiments by sampling for the ratio gain to total produce. 1940 *Journal of Agricultural Society.*, 1940. **30**, p 13.
- [4] Mudi, T,A., Ibrahim, K., and Abdulrahman M. K,A, Comparative estimation methods for public primary pupils enrolment in kaduna state. *Journal of remote sensing, environment and management.*, 2019. 7 p 50 – 58
- [5] Samawi, C., Yu, L., Linder, D., Cai, J., & Vogel, R., On regression estimators for different stratified sampling schemes. *Journal of Statistics and Management System.*, 2017. **20**, p18.
- [6] Etebong, P. C., & Ekaette, I. E.. On the Efficiency of Ratio Estimator over the Regression Estimator. *Journal of Communications in Statistics.*, 2016
- [7] Gupta, S.C. *Fundamental of Statistics*. 6th ed. 2004: Himalaga Publishing House, New Delhi, India
- [8] Shalabh S & Toutenburg, H. On the regression methods of estimation of population mean from incomplete survey data through importation. A discussion paper 442, 2005, University of Munich, Germany.

Citation: Mudi TA., Abdulrahman MKA and Arowona NA. *On the use of auxiliary information to improve the precision of estimator for public primary pupils enrolment in Kaduna state. IJMSSCS, 2022. 1(1): p. 1-5.*