On Robust Estimation through the Use of Auxiliary Information by Ratio and Regression Estimators

Aliyu Usman, Sikiru O. Soladoye and Mohammed A. Baba
Department of Mathematics & Statistics, Kaduna Polytechnic, Kaduna Nigeria.

Abstract
The ratio and regression estimators that make use of auxiliary information for achieving higher efficiency is applied to education data. Education is critical to our development as individuals and as societies, and it paves the way to a successful and productive future. It provides the potentials for an individual's intellectual growth and productivity in the society. The objective of this paper is to estimate the ratio of pupils to classroom in Nigeria's public primary schools as well as to estimate the total pupil population in Nigeria's public primary schools using the ratio and regression estimators. The data of annual enrolment into public primary schools and the number of classrooms in 2014 were obtained from Universal Basic Education Commission. Furthermore, the sampling design used is stratified random sampling with equal allocation. Two states were selected from each geo-political zone; making a sample of 12 states The results of the ratio estimator revealed that the estimated national pupil-classroom ratio is approximately 54 and the confidence interval shows that the ratio may lie between the inter 43 and 65 approximately. Similarly, total pupils population is estimated at 20,298,309 and the confidence interval shows that the total population may lie between the inter 16,084,553 to 24,512,065 approximately. The ratio and regression estimators will save time and cost to give reliable estimates. Similarly, using the regression estimator total pupils population is estimated at 20,412,402 and the confidence interval shows that the total population may lie between the inter 16,210,204 to 24,614,600 approximately. Based on this analysis, it is therefore recommended that effort should be intensified to improve the pupil-classroom ratio nationwide and to increase pupils' enrolment.

Key words: *Bias, Enrolment, Coefficient of Variation, Confidence interval, Ratio estimation, Regression estimation Robust estimates, Standard error, Variance*

Abbreviations: *EFA - Education for All, UBE - Universal Basic Education, SRS – Simple Random Sample, UNESCO-United Nations Educational, Scientific and Cultural Organization*

1. Introduction

Estimates become better through the use of information on an auxiliary variable for improving the precision of the estimates of population means and totals (Levy & Lemeshow, 2012). The ratio and regression estimators that make use of auxiliary information for achieving higher efficiency are hereby applied. The statistical analysis of a country's on education in particular, as well as other social indicators is paramount to education planning and international comparison. The identification of demographic and economic constraints to the development of the education sector are the fundamental for implementing new policies. The forecast of such constraints should facilitate improved ownership of the final allocations, generally determined by the budget and finance ministries, to achieve greater control of education sector policy by the government.

Education is critical to our development as individuals and as societies, and it paves the way to a successful and productive future (UNESCO). It provides the potentials for an individual's intellectual growth and productivity in the society. It also contributes to the wider socio-economic and cultural development of the society. The right of all Nigerians to education has also featured in successive Nigerian constitutions. The Nigerian Government is obliged under Section 18 of the 1999 Constitution to strive to eradicate illiteracy. This commitment was reiterated with the re-enactment of the Universal Basic Education (UBE) which was subsequently reformulated into a policy of compulsory nine years of basic education in forms of six years of primary and three years of junior secondary. Hence, basic education is a fundamental right for every child in Nigeria.

The Education for All (EFA) movement is a global commitment to provide quality basic education for all children, youth and adults. At the World Education Forum (Dakar, 2000), 164 governments, including Nigeria, pledged to achieve EFA and identified six cardinal goals to be met by 2015. Governments and the private sector are working together to reach the EFA goals. The population of pupils in public primary school is paramount and needs some in-depth analysis.

Objectives

This paper has the following objectives.

1. To estimate the ratio of pupils to classroom in Nigeria's public primary schools.
2. To estimate the total pupil population in Nigeria's public primary schools through ratio estimator.

3. To estimate the total pupil population in Nigeria's public primary schools through linear regression estimator.

2. Theory of Ratio and Regression Estimators

The concept of a ratio $\bar{x}/\bar{y}$ of two sample means $\bar{x}$ and $\bar{y}$ can used for robust estimation of population mean or total. This ratio serves as an estimate of the ratio $\bar{X}/\bar{Y}$ of the means of two variables X and Y in a population. But, more importantly, it also serves as a device for obtaining a more accurate estimate of a population total X than can be obtained from the estimate x' determined by simple inflation of a sample total x by N/n, the inverse of the sampling fraction. This method is called ratio estimation, and the resulting estimates are called ratio estimates (Levy & Lemeshow, 2012).

For ratio estimation to apply, two quantities x_i and y_i must be measured on each sample unit; y_i is often called an auxiliary variable. In the population of size N;

$$x = \sum_{i=1}^{n} x_i \text{ and } y = \sum_{i=1}^{n} y_i$$

And their ratio, is $R = \frac{x}{y} = \frac{\bar{x}}{\bar{y}}$

In the simplest use of ratio estimation, a simple random sample of size *n* is taken, and the information in both *x* and *y* is used to estimate the ratio, population mean and total. The regression estimator is done in a similar way where the regression coefficient *b* stands in place of the ration.

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

Ratio and regression estimation both take advantage of the correlation of *x* and *y* in the population; the higher the correlation, the better they work. (Lohr, 2014)

3. Materials and Methods

The data of annual enrolment into public primary schools and the number of classrooms in 2014 were obtained from Universal Basic Education Commission. Furthermore, the sampling design used is stratified random sampling with equal allocation. Two states were selected from each geo-political zone; making a sample of 12 states. The aim is to use information on these states to estimate for the entire country. Hence, we have the following sample.

Table 3.1: Classrooms and Enrolments in Nigerian Public Primary Schools

State	Number of classrooms (x)	Enrolments 2014		
		M	F	Total (y)
Adamawa	9,265	269,433	238,578	508,011
Anambra	12,996	419,117	473,992	893,109
Bauchi	8,889	431,772	339,771	771,543
Benue	16,475	341,035	302,156	643,191
Cross-River	8,491	115,243	109,670	224,913
Ebonyi	6,360	184,290	186,020	370,310
Edo	8,815	161,707	155,777	317,484
Ekiti	7,039	64,861	63,294	128,155
Jigawa	8,865	312,025	228,174	540,199
Kaduna	22,250	637,844	543,041	1,180,885
Kogi	9,919	533,324	535,068	1,068,392
Ogun	11,840	215,251	211,195	426,446
Total	131,204			7,072,638

Source: Universal Basic Education, Abuja, Nigeria, 2014

4. Results and Discussions

4.1 Auxiliary Information through Ratio Estimator

The ratio estimator and its variance are as follows:

$$\hat{R} = \frac{x}{y}$$

With variance

$$\hat{V}(\hat{R}) = \frac{1-f}{n\bar{X}^2}\left(s_y^2 + \hat{R}^2 s_x^2 - 2\hat{R}s_{xy}\right)$$

The total estimator and its variance are as follows:

$$\hat{Y} = \hat{R}X$$

With variance

$$\hat{V}(\hat{R}) = \frac{N^2(1-f)}{n}\left(s_y^2 + \hat{R}^2 s_x^2 - 2\hat{R}s_{xy}\right) = X^2\hat{V}(\hat{R})$$

The ratio:

$$\hat{R} = \frac{x}{y} = \frac{7{,}072{,}638}{131{,}204} = 53.9$$

The variance:

$$X = 376{,}592$$

$$\bar{X} = \frac{376{,}592}{37} = 10178.16$$

$$s_y^2 = \frac{\sum_{i=1}^{12} y_i - n\bar{y}^2}{n-1} = 110066537827.364$$

$$s_x^2 = \frac{\sum_{i=1}^{12} x_i - n\bar{x}^2}{n-1} = 20241518.061$$

$$s_{xy} = \frac{\sum_{i=1}^{12} x_i y_i - n\bar{x}\bar{y}}{n-1} = 1010380567.818$$

$$\therefore \hat{V}(\hat{R}) = \frac{1-f}{n\bar{X}^2}\left(s_y^2 + \hat{R}^2 s_x^2 - 2\hat{R}s_{xy}\right) = 32.59$$

The coefficient of variation:

$$CV(\hat{R}) = \frac{\sqrt{32.59}}{53.9} \times 100\% = 10.6\%$$

The 95% confidence interval for the ratio:

$$\hat{R} \pm 1.96\sqrt{\hat{V}(\hat{R})} = 53.9 \pm 1.96 \times \sqrt{32.59}$$

$$[42.7,\ 65.1]$$

The total population estimate:

$$\hat{Y} = \hat{R}X = 53.9 \times 376{,}592 = 20{,}298{,}309$$

The variance:

$$X = 376{,}592$$

$$\hat{V}(\hat{Y}) = X^2\hat{V}(\hat{R}) = (376{,}592)^2 \times 32.59 = 4621963808181.76$$

The coefficient of variation:

$$CV(\hat{R}) = \frac{\sqrt{4621963808181.76}}{20{,}298{,}309} \times 100\% = 10.6\%$$

The 95% confidence interval for the total:

$$\hat{Y} \pm 1.96\sqrt{\hat{V}(\hat{Y})} = 20{,}298{,}309 \pm 1.96 \times \sqrt{4621963808181.76}$$

$$\left[16{,}084{,}553,\ 24{,}512{,}065\right]$$

4.2 Auxiliary Information through Linear Regression Estimator

The linear regression estimator for total:

$$\hat{Y}_{lr} = N\left[\bar{y} + b(\bar{x} - \bar{X})\right]$$

With variance

$$\hat{V}(\hat{Y}_{lr}) = \frac{N^2(1-f)}{n}\left(s_y^2 + b^2 s_x^2 - 2bs_{xy}\right)$$

Where;

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{s_{xy}}{s_x s_y}$$

The regression estimate of the total:

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{s_{xy}}{s_x s_y} = 49.9$$

$$N = 37$$

$$X = 376{,}592$$

$$\bar{X} = \frac{376{,}592}{37} = 10{,}178.16$$

$$\bar{x} = \frac{131{,}204}{12} = 10{,}933.67$$

$$\bar{y} = \frac{7{,}072{,}638}{12} = 589{,}386.50$$

$$\hat{Y}_{lr} = N\left[\bar{y} + b(\bar{x} - \bar{X})\right] = 20{,}412{,}402$$

The variance of the estimated total:

$$s_y^2 = \frac{\sum_{i=1}^{12} y_i - n\bar{y}^2}{n-1} = 110066537827.364$$

$$s_x^2 = \frac{\sum_{i=1}^{12} x_i - n\bar{x}^2}{n-1} = 20241518.061$$

$$s_{xy} = \frac{\sum_{i=1}^{12} x_i y_i - n\bar{x}\bar{y}}{n-1} = 1010380567.818$$

$$\hat{V}(\hat{Y}_{lr}) = \frac{N^2(1-f)}{n}\left(s_y^2 + b^2 s_x^2 - 2bs_{xy}\right) = 4596644089942.41$$

The coefficient of variation:

$$CV(\hat{Y}_{lr}) = \frac{\sqrt{4596644089942.41}}{20{,}412{,}402} \times 100\% = 10.5\%$$

The 95% confidence interval for the total:

$$\hat{Y} \pm 1.96\sqrt{\hat{V}(\hat{Y}_{lr})} = 20{,}412{,}402 \pm 1.96 \times \sqrt{4596644089942.41}$$

$$\left[16{,}210{,}204,\ 24{,}614{,}600\right]$$

5. Conclusion

The results of the ratio estimator revealed that the estimated national pupil-classroom ratio is approximately 54 and the confidence interval shows that the ratio may lie between the inter 43 and 65 approximately. Moreover, total pupils population is estimated at 20,298,309 and the confidence interval shows that the total population may fall between the intervals (16,084,553 to 24,512,065) approximately. Similarly, the result of the linear regression estimator reveals that the estimated national pupil-classroom ratio is approximately 50; it is simply the estimated regression coefficient *b*. That is number of pupils per unit classroom. Similarly, total pupils population is estimated at 20,412,402 and the confidence interval shows that the total population may fall between the intervals (16,210,204 to 24,614,600) approximately. Both methods give close results and will save time and cost to give reliable estimates. Based on this analysis, it is therefore recommended that effort should be intensified to improve the pupil-classroom ratio nationwide and to increase pupils' enrolment.

References

FME (2008). *Statistics of Education in Nigeria*: 1999-2005.
Levy, P, S. and Lemeshow, S. (2012). *Sampling of populations*. New York: John Wiley.
Lohr, S.L. (2014). *Sampling: Design and Analysis*. New York: Duxbury Press.
Thomson, S. K. (2012). *Sampling*. (3rd. ed.) New York: John Wiley.
UNESCO (2000) *Education for All (EFA 2000)* – Dakar Framework of Action

Author's Brief Biography

Dr. Aliyu Usman is a Senior Lecturer in the Department of Mathematics and Statistics, Kaduna Polytechnic, Nigeria. He is a versatile Statistician and Biostatistician with many research publications in reputable international journals. He is a member of International Statistical Institute (ISI), International Association of Survey Statisticians (IASS), Nigerian Statistical Association (NSA) and Korean International Statistical Society (KISS). He is the author of a standard reputable textbook titled "Statistical Methods for Biometric and medical Research". His research interest spans through many areas of Statistics, Survey Research, Biostatistics and Mathematics.