ESTIMATION OF POPULATION IN A TWO-STAGE SAMPLING WITH APPLICATION TO NUTRITION SURVEYS

[1]Emmanuel Umoh, [2]Ikughur, Jonathan Atsua Ikughur, [3]Tersoo Uba, & [4]Anthony Ekpo

[1& 4]Department of Statistics, Akwa Ibom State Polytechnic, Ikot Osurua, Ikot Ekpene.
[2 & 3] Department of Statistics, Joseph SarwuanTarka University, Makurdi, Benue State, Nigeria

Authors: (1) emmanuel.umoh@akwaibompoly.edu.ng (2) Author:Ikughur.ja@uam.edu.ng (3)ekpo.anthony@uam.edu.ng (4) ubachaka@gmail.com

Abstract
In this study, the problem of estimation of population proportion in nutrition survey using two-stage cluster sampling is performed. Data used in this study were obtained using the Displacement Tracking Matrix (DTM) datasets from different locations of Internally Displaced Persons (IDPs) camps in Borno State, North-Eastern Region of Nigeria. The IDP camps formed the clusters of unequal sizes. The study population was children of ages 0-5 years that were affected by insurgency in the North East Nigeria. Efficiency of the estimator was investigated for six different set of samples using the Mean Square Error (MSE), Standard Error (SE) and Coefficient of Variation (CV). Results showed that as sample size increased, the estimator became more efficient and that the proportion of children aged 0 - 5 years with severe acute malnutrition is approximately 35%.

Keywords: Multistage sampling, Population proportion, Severe acute malnutrition, Clusters, Estimation.

Introduction

In scientific research involving human population, sample surveys play a vital role in providing required data especially in the presence of limited cost. Sampling is the process of drawing a representative group of individuals or group of individuals from a particular population clearly defined in a boundary. Both sampling and statistical inference are used in circumstances where it is impractical to obtain information from every member of the population especially when basic probability sampling design such as simple random samplingis used (Solanki*et al.*, 2012). This is because sampling procedures requires that error in selection should be controlled to barest minimum as it helps in the interpretations of results (Chambers and Clark, 2012).

In some surveys, sampling frames are readily available and this enables easy access to the population units for inclusion in the sample. There are however real life situations where information may not be accessible due to wider region involved and so, the investigators may divide the region into smaller list of groups or clusters in order to access element in such groups for data collection. In such cases, it becomes ideal to use cluster sampling technique to draw random samples from the population and when this technique is carried out in two phases, it is known as two-stage cluster sampling (Okafor, 2002).

Two-stage sampling procedure has been found to be much more acceptable for demographic surveys owing to the lower cost involved and also due to the fact that it gives rise to a representative sample (Fuller, 2011) hence, it is much employed by most governmental, non-governmental organizations and professionals to obtain estimates of population characteristics of interest.

Khan and Raza (2022) revealed that the highest number of malnourished children was found in Borno State Nigeria in 2021. The interest of this study is drawn towards this region based on the revelation by Medecins Sans Frontieres (MSF) report among others. This study therefore, implements two-stage sampling techniques in order to estimate population proportion considering the general nutritional scenario in the North Eastern part of Nigeria especially, in areas affected by humanitarian crises occasioned by insurgency.

Harkare (2021) described malnutrition as a chronic and multi-faceted concern in Borno State and is driven by the cumulative impact of displacement, insecurity, food inflation, poverty, lack of access to healthcare and health status. The problem of malnutrition in poor economies can be defined as a "syndrome of developmental impairment", which considers growth failure, delayed cognitive and behavioral development, and increased morbidity and mortality (Martorell, 1999).

The imbalance between the nutrients the body needs and the nutrients it receives is known as malnutrition; which may take the form of either under nutrition or obesity. According to the World Health Organization (WHO), under nutrition can be one of the two types namely, `protein energy malnutrition or micro nutrient deficiency. Protein energy malnutrition manifests early during the ages of 6 months to 2years, resulting from irregular or no breastfeeding, introduction to low protein food and different types of infections. It is measured by indicators such as wasting, stunting, being underweight or obesity.

Wasting is defined as the failure to receive adequate nutrition in the period immediately preceding the survey. It may be the result of inadequate food intake or a recent episode of illness causing loss of weight and the onset of malnutrition. Children whose weight-for-height (WAZ) is below minus three standard deviations (-3 SD) from the median of the reference population are considered to be severely wasted, and those below minus two standard deviations (-2 SD) are classified as wasted.

The height-for-age (HAZ) index is an indicator of linear growth retardation and cumulative deficits. Children whose HAZ Z-scores are below minus two standard deviations (-2 SD) growth from the median of the reference population are considered short for their age, or stunted, and are chronically malnourished. Das *et al.* (2010) asserted that when Z score is less than minus three of the standard deviation (-3 SD), the child is classified as severely stunted, WHO (2020). Stunting reflects a failure to receive adequate nutrition over a long period, which is affected by recurrent and chronic illness. Malnutrition is a composite index of HAZ and WAZ. It takes into account both acute and chronic malnutrition. Children whose weight-for-age is below minus two standard deviations (-2 SD) from the median of the reference population are classified as underweight. Stunting is normally considered the most important measure when evaluating malnutrition among children; hence, this study focuses on HAZ categories. The measurement of the related Z-scores is calculated based on the reference population and its median. On the other hand, excessive fat deposition in the body can lead to being overweight or obese.

For proper planning and optimal interventions especially in cases of problems of malnutrition and mortality, it is necessary to obtain estimates of population proportions of children that are affected by this social disorder which will guide decision makers to make adequate interventions in order to tackle the unexpected child mortality in the affected regions. The basic information including the proportion of the occurrence among subgroups and the resulting child mortality in the population.

According to Dilip (2015) and Singh (2014), in multistage sampling, sampling units are first drawn from the main population and another set of sampling units are drawn. This continues until the researcher arrives at his desired number of sampling units. Similarly, Lohr (2010) explained that in conventional two-stage sampling, the first sampling stage involves the selection of a predetermined number of clusters that are mutually exclusive sub populations, most frequently constructed from recognized administrative boundaries.

Rizzo *et al.* (2010) opined that one alternative approach to a population-based mortality survey is to apply cluster sampling to estimate conflict-related mortality rates while Sarndal (2010) opined that two-stage cluster sampling designs are commonly used for household and health surveys and that in household and health surveys, the population is often sparse over a large territory and there is regularly no sampling frame. For instance, a two-stage cluster sampling designs are convenient in such situations. The population are grouped into large blocks (e.g. Local Government Areas, Villages, Constituencies, Communities, Wards, municipalities or countries etc.), called Primary Sampling Units (PSUs), which are sampled at the first stage. Only a frame of these PSUs is needed at this stage, which is easier to create. At the second stage, a list of population units is obtained inside the selected PSUs, and a sample of these population units is selected. Despite its convenience, multistage sampling has the drawback as it leads to estimators with inflated variance, compared to sampling designs where the population units are directly selected.

Solanki *et al.* (2012) stated that in the second stage, starting households are selected from each cluster as complete and adequate listings of households rarely exist as such, households are selected by the survey team in the field based on a random procedure. Most commonly, starting households are selected in the field based on the “random walk” method, which involves identifying the centre of the cluster, or another easily distinguished feature such as a main street, and selecting a random direction to walk, thus drawing a transect across the cluster.

One of the primary applications of cluster sampling is the “Area sampling”, where the clusters are countries, townships, city blocks, communities and other well-defined geographical sections of the population, which has been found to be much acceptable for studying problems involving the Mid-Upper-Arm -Circumference (MUAC), Height-for-Age Z-score (HAZ), and Weight-for-Age Z-score (WAZ) which are very important measures of nutrition. Srivastava and Garg (2009) proposed a general family of estimators for population proportion using multi-auxiliary information in two-stage sampling while Chaudhari and Singh (2013) conducted a survey on child nutrition and estimated population proportion under two-stage cluster sampling with equal cluster sizes under non-response using auxiliary characteristic.

Benson (2008) determined the effect of race and ethnicity on nutrition level. It was also observed that children originally from

African backgrounds have a higher chance of being overweight than other children elsewhere. In addition, African girls have a 6% - 7% higher chance of being overweight than Mexican-American boys.

Hackett *et al.* (2009) identified determinants of child anthropometrics in a sample of poor Colombian children living in small municipalities. They examined the influence of household consumption and public infrastructure on childhood malnutrition taking into account the endogeneity of household consumption, using two sets of instruments: household assets and municipality average wage. They found that both are important determinants of a child's nutritional status. The study also found that the coverage of a community's piped water network positively influenced a child's health, provided the parents had some basic level of education.

Goisis *et al.* (2019) showed that despite the efforts in Ethiopia to reduce childhood malnutrition for a long time, the implemented polices have not succeeded and so, there is still a long way to go to achieve this objective. Hackett *et al.* (2009), showed that malnourished, stressed, and obese mothers have a high chance of having obese children. In addition, a mother's intake of energy-driven, zero-nutritious food and less physical activity are the major contributing factors for raising obese children. Children malnourishment is one of the crucial issues of any developing country. Although, in terms of percentage, the share of malnourished children continues to show an increasing trend (Smith & Haddad, 2000). Arif *et al.* (2012) conducted a survey on gender preference on nutrition in Pakistan and concluded that malnutrition is significantly higher among girls than boys, indicating a crucial child-level gender issue while Mohseni *et al.* (2017) explained that malnourishment in children is not only a problem of present-day society, but also an issue for the future.

Ekpo and Ikughur (2021) conducted a survey on nutrition and mortality SMART survey in the Northeastern Nigeria using two-stage cluster sampling to estimate the prevalence of acute malnutrition among children aged 0-59 months (i.e. $< 5years$) using HAZ, WAZ and MUAC as estimated parameters.

Method

Let$Y_{1j}, Y_{2j}, \dots, Y_{Nj}$ be units in i^{th} cluster. Suppose that Y_{ij} is the j^{th} selected unit in the i^{th} cluster, j=1,2,…….., M.

$$\text{Then } Y_i = \begin{cases} 1, & i \in A \\ 0, & i \notin A \end{cases} \quad (1)$$

Let $\hat{\theta}$ be the estimator of population characteristic say θ, then the estimate of the population mean in two-stage sampling is:

$$E(\hat{\theta}) = E_1[E_2(\hat{\theta})] \quad (2)$$

while the estimated variance of the estimator is:

$$Var(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = E_1 E_2 (\hat{\theta} - \theta)^2 \quad (3)$$

which can be expressed as:

$$Var(\hat{\theta}) = V_1[E_2(\hat{\theta})] + E_2[V_2(\hat{\theta})] \quad (4)$$

To estimate the population proportion in (2) and the mean square error in (4) the following definitions hold:

Let P = θ,

Then, $\mathrm{n} = \text{number of fsu's selected in the sample,}$

$$\mathrm{M_i} = \text{number of second stage units (ssu's)in the psu}$$

$\mathrm{m_i} = \text{number of ssu's selected from } \mathrm{M_i} \text{ ssu's}$

$$\mathrm{M_0} = \sum_{i=1}^{N} \mathrm{M_i} = \text{total number of ssu' s in the population}$$

$$\overline{\mathrm{M}} = \frac{\mathrm{M_0}}{\mathrm{N}} = \text{average number of ssu's per fsu}$$

$$\bar{y}_i = \frac{y_{i.}}{m_i} = p_i \text{ as the sample proportion for } i^{th} \text{ fsu}$$

$$y_{i.} = \textstyle\sum_{j=1}^{m_i} y_{ij} = a_i = \text{ sample total for the } i^{th} \text{ fsu}$$

$$y_{..} = \sum_{i=1}^{n} y_{i.} = \text{total of y} - \text{values for the whole sample}$$

$$Y_{i.} = \sum_{j=1}^{M_i} Y_{ij} = M_i^*(M_i^* < M_i) \text{ as the total number of ssu's in the } i^{th} \text{ fsu belonging to A.}$$

$$\overline{Y}_{i.} = \frac{Y_{i.}}{M_i} = \frac{M_i^*}{M_i} = P_i \text{ as the population proportion of ssu' s in the } i^{th} \text{ fsu belonging to A.}$$

$$Y = \sum_{i=1}^{N} Y_{i.} = \sum_{i=1}^{n} M_i^* = M \text{ as the total number of ssu' s in the population that belongs to A.}$$

$$\overline{\overline{Y}} = \frac{Y}{M_o} = \frac{M_i^*}{M_o} = P \text{ as the population proportion of ssu's in A.}$$

Estimator Based on the Unit Proportion in the Sample.

$$\hat{\overline{\overline{Y}}} = \frac{1}{n}\left[\sum_{i=1}^{n} \frac{y_i}{m_i}\right] = \frac{1}{n}\left[\sum_{i=1}^{n} \frac{A}{m_i}\right] = \frac{1}{n}\left[\sum_{i=1}^{n} p_i\right]$$

Taking expectation operator on both sides,

$$E(p_i) = E\left[\frac{1}{n}\sum_{i=1}^{n} p_i\right] \tag{5}$$

$$= E_1\left[\frac{1}{n}\sum_{i=1}^{n} E_2(P_i)\right]$$

$$= E_1\left[\frac{1}{n}\sum_{i=1}^{n} P_i\right]$$

$$=\frac{1}{n}\sum_{i=1}^{n} E[p_i] = \overline{P}.$$

$$\therefore E(p_i) \neq \overline{P}. \tag{6}$$

So, p_i is a biased estimator of P whose bias is given by

$$\text{Bias } (p_i) = E(p_i) - \overline{P}. \tag{7}$$

$$= \frac{1}{N}\sum_{i=1}^{N} \widehat{P} - \frac{1}{N\overline{M}}\sum_{i=1}^{N} M_i P_i$$

$$= -\frac{1}{N\overline{M}}\left[\sum_{i=1}^{N} M_i P_i - \frac{1}{N}\left(\sum_{i=1}^{N} P_i\right)\left(\sum_{i=1}^{N} M_i\right)\right]$$

$$= \frac{1}{N\overline{M}}\sum_{i=1}^{N}(M_i - \overline{M})\left(P_i - \widehat{P}\right) \tag{8}$$

Thus, $\widehat{\text{Bias}}(p_i) = -\frac{N-1}{N\overline{M}(n-1)}\sum_{i=1}^{n}(M_i - \overline{m})(p_i - \hat{p})$ (9)

So that

$$E\left[\widehat{\text{Bias}}(p_i)\right] = -\frac{N-1}{N\overline{M}} E_1\left[\frac{1}{n-1}\sum_{i=1}^{n} E_2\left\{(M_i - \overline{m})\left(\frac{p_i - \hat{p}}{n}\right)\right\}\right]$$

$$= -\frac{N-1}{N\overline{M}} E\left[\frac{1}{n-1}\sum_{i=1}^{n}(M_i - \overline{m})(P_i - \hat{p})\right]$$

$$= -\frac{1}{N\overline{M}}\sum_{i=1}^{N}(M_i - \overline{m})\left(P_i - \widehat{P}\right)$$

$$= \widehat{P} - P_i. \tag{10}$$

Where $\widehat{P} = \frac{1}{n}\sum_{i=1}^{n} p_i$.

However, an unbiased estimator of the variance of P_i is thus obtained as:

$$p_i + \frac{N-1}{N\overline{M}}\frac{1}{n-1}\sum_{i=1}^{n}(M_i - \overline{m})(\hat{p} - p_i). \tag{11}$$

Variance of the Estimator

In Okafor (2002), $\widehat{A} = \frac{N}{n}\sum_{i=1}^{n} M_i p_i = \frac{N}{n}\sum_{i=1}^{n} \widehat{A}_i$.

Where

$p_i = \bar{y}_{i.} = \frac{1}{m_i}\sum_{j=1}^{n} y_{ij} = \frac{a_i}{m_i}$is the sample proportion for the i^{th} fsu.

The sampling variance of $\hat{A}$ is obtained as follows:

$V(\hat{A}) = V_1E_2(\hat{A}) + E_1V_2(\hat{A})$(12)

Now,

$$V_1E_2(\hat{A}) = V_1\left(\frac{N}{n}\sum_{i=1}^{n} M_iP_i\right) = \frac{N^2(1-f_1)}{n}s_1^2 \quad (13)$$

Where

$$S_1^2 = \sum_{i=1}^{N}\frac{(A_i-\bar{A})^2}{(N-1)}; \qquad A_i = M_iP_i; \qquad \bar{A} = \frac{\sum_{i=1}^{N} A_i}{N}.$$

$$E_1V_2(\hat{A}) = E_1\left[\frac{N^2}{n^2}\sum_{i=1}^{n} M_i^2V_2(p_i)\right] = \frac{N}{n}\sum_{i=1}^{N} M_i^2\frac{1-f_{2i}}{m_i}\frac{M_iP_iQ_i}{M_i-1} \quad (14)$$

$Q_i = 1 - P_i$

adding (12) and (13), the variance of $\hat{A}$ becomes

$$V(\hat{A}) = \frac{N^2(1-f_1)}{n}s_1^2 + \frac{N}{n}\sum_{i=1}^{N} M_i^2\frac{1-f_{2i}}{m_i}\frac{M_iP_iQ_i}{M_i-1} \quad (15)$$

The unbiased sample estimator of the variance of $\hat{A}$, that is $\hat{V}(\hat{A})$ is

$$\hat{V}(\hat{A}) = \frac{N^2(1-f_1)}{n(n-1)}\sum_{i=1}^{n}\left(\hat{A}_i - \hat{\hat{A}}\right)^2 + \frac{N}{n}\sum_{i=1}^{n} M_i^2\frac{1-f_{2i}}{m_i-1}p_iq_i \quad (16)$$

Where

$\hat{\hat{A}} = \frac{1}{n}\sum_{i=1}^{n}\hat{A}_i$, $q_i = 1 - p_i$. If $M_0$is known, the unbiased estimator of the population proportion P is

$$p = \frac{\hat{A}}{M_0} \quad (17)$$

Its variance is

$$V(p) = \frac{V(\hat{A})}{M_0^2} \quad (18)$$

If, however, M_0 is not known, the ratio-to-size estimator of the population proportion P is obtained as:

$$p^* = \frac{\frac{N}{n}\sum_{i=1}^{n} M_ip_i}{\frac{N}{n}\sum_{i=1}^{n} M_i} = \frac{\hat{A}}{\hat{M}_0} = \frac{\sum_{i=1}^{n} M_ip_i}{\sum_{i=1}^{n} M_i} \quad (19)$$

This is similar to the ratio-to-size estimator of mean per element. That is

$$\hat{\hat{\bar{Y}}}_r = \frac{N}{n}\sum_{i=1}^{n} M_i\bar{y}_{i.} \Big/ \frac{N}{n}\sum_{i=1}^{n} M_i = \frac{\hat{Y}}{\hat{M}_0} \quad (20)$$

The large sample approximation to the MSE of p^* is

$$V(p^*) = \frac{N^2(1-f_1)}{nM_0^2}\frac{\sum_{i=1}^{N} M_i^2(P_i-P)^2}{N-1} + \frac{N}{nM_0^2}\sum_{i=1}^{N} M_i^2\frac{1-f_{2i}}{m_i}\frac{M_iP_iQ_i}{M_i-1} \quad (21)$$

The sample estimator of this variance is

$$\hat{V}(p^*) = \frac{N^2(1-f_1)}{\hat{M}_0^2n(n-1)}\sum_{i=1}^{n} M_i^2(p_i - p^*)^2 + \frac{N}{n\hat{M}_0^2}\sum_{i=1}^{n} M_i^2\frac{1-f_{2i}}{m_i-1}p_iq_i \quad (22)$$

Even when M_0 is known, p^* may be more precise than p if p_i is positively correlated with M_i, in that case p^* is still preferred but M_0 then replaces $\widehat{M}_0$ in (22).
The estimated standard error therefore, is the positive square root of the variance, that is,

$$\mathrm{SE}(\mathrm{p}^*) = \sqrt{\widehat{\mathrm{V}}(\mathrm{p}^*)} \qquad (23)$$

While the coefficient of variation (CV) is given by

$$\mathrm{CV}(\mathrm{p}^*) = \frac{\mathrm{SE}(\mathrm{p}^*)}{\hat{p}} * 100 \qquad (24)$$

Data Collection

Since this study focused on anthropometric characteristics survey among the internally displaced persons' camps in Borno State, the North-East Nigeria, primary data were collected from children aged 0-5years to assess the proportion of children suffering from severe acute malnutrition (SAM) which was the main objective of this study. Here, the children Mid Upper Arm Circumference (MUAC) were measured in order to determine cases of acute malnutrition among the children. MUAC was used because it is thought to be simple particularly for screening children in emergency situations which does not need to be related to a reference population value. Using the measuring tape (MUAC tape), those children whose mid upper arm conference were less than 115mm($<115mm$) were classified as those with cases of severe acute malnutrition (SAM) and coded as one (1) and those whose measurements were greater than or equal to 115mm ($\geq 115mm$) were classified as low or moderate cases of malnutrition and were coded as zero (0).

Results

The estimates of the population characteristics namely population proportion ($\hat{P}$), bias, Mean Square Error (MSE), Standard Error (SE), Coefficient of Variation (CV), and the Confidence Interval (C.I) are presented for the two cluster sampling with a population of size N = 230 and sample sizes $n_1 = 12, n_2 = 23, n_3 = 35, n_4 = 46, n_5 = 58,$ and $n_6 = 69.$ M_i and m_i are respectively 30% of the children ≤5years. The basic statistical estimates of the study population are shown in Table 1.

Table1: Statistical properties of the Study Population

ESTIMATORS SAMPLE SIZE	PROPORTION	MSE	SE	Bias	CV	C.I
n=12	0.3414	0.1383	0.1048	0.0388	30.42	(0.1349, 0.5541)
n=23	0.3445	0.1288	0.0698	0.0449	21.48	(0.1853, 0.4645)
n=35	0.3469	0.1289	0.0568	0.0347	15.92	(0.2433, 0.4705)
n=46	0.3486	0.1277	0.0470	0.0127	13.47	(0.2548, 0.4428)
n=58	0.3489	0.1291	0.0417	0.0120	11.95	(0.3072, 0.3906)

n=69	0.3497	0.1252	0.0367	0.0546	10.48	(0.2764, 0.4229)

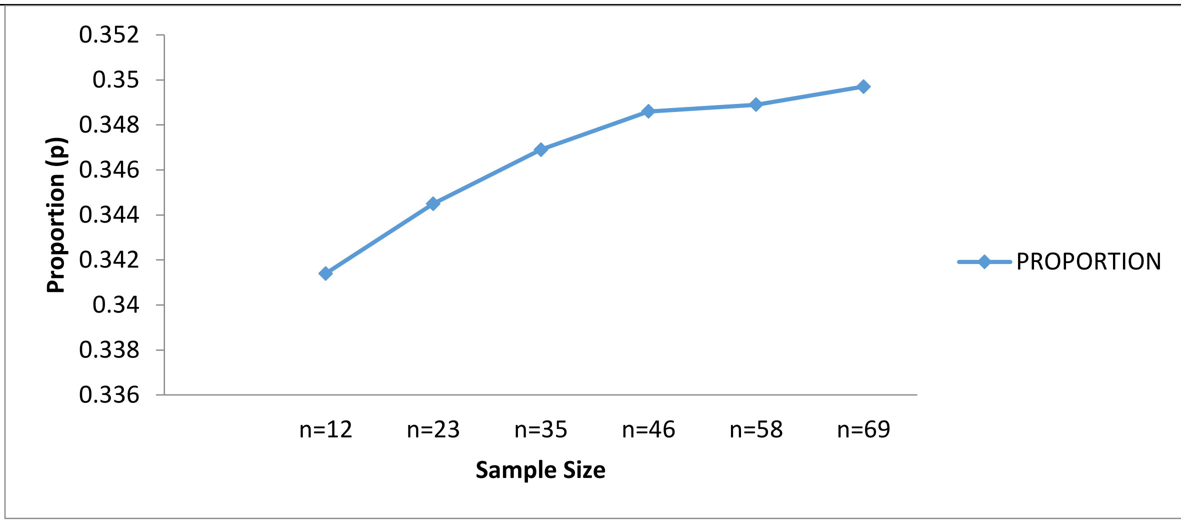

Figure 1: Plot of Estimates of the Population proportion for the Study Population for different samples.

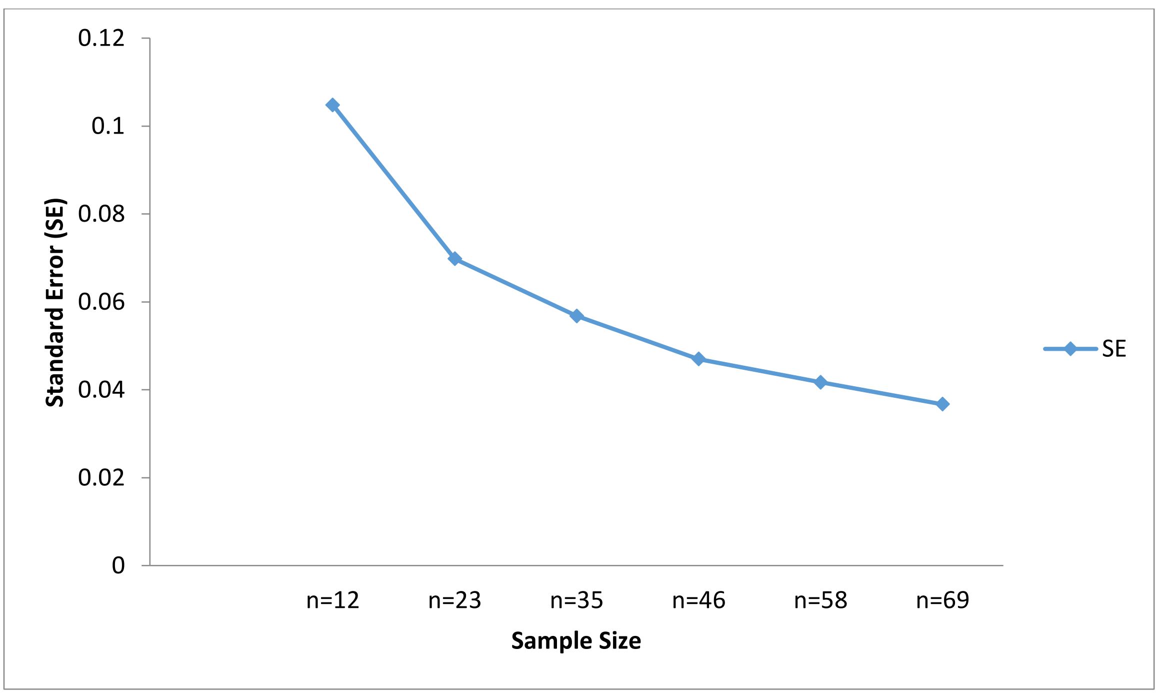

Figure 2: Plot of Estimates of the SE for the Study Population for different samples.

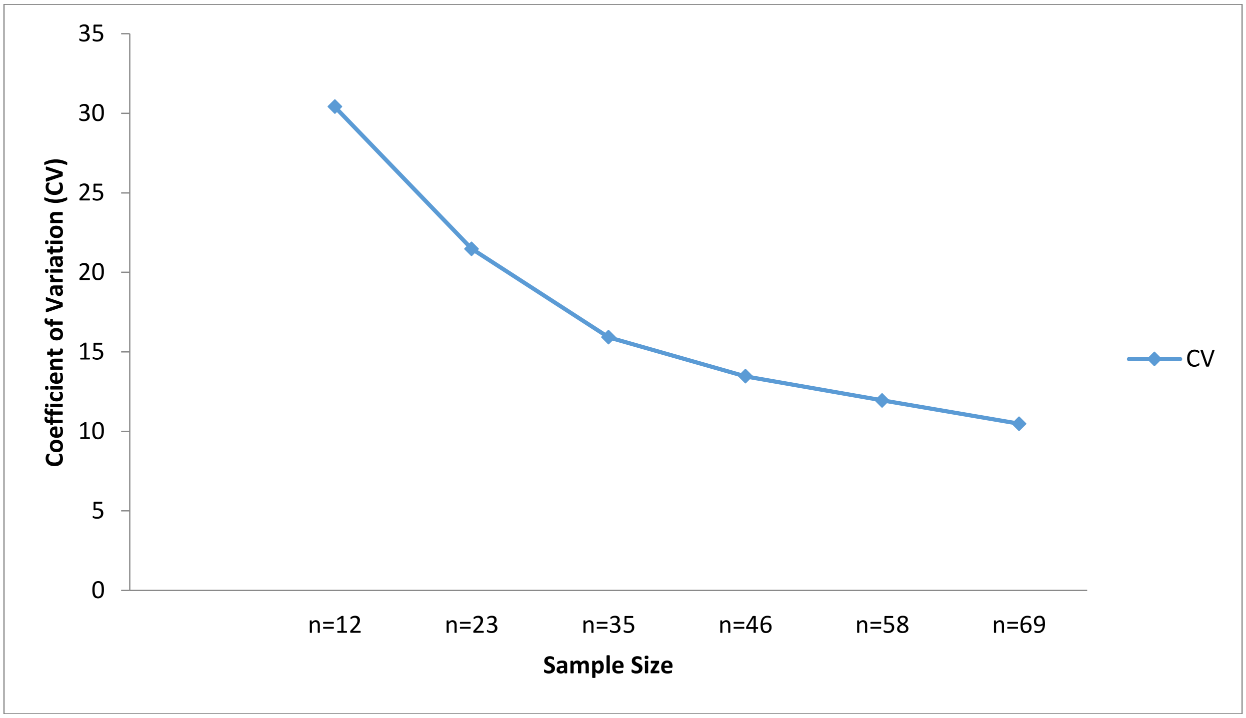

Figure 3: Plot of Estimates of the CV for the Study Population for different samples.

Discussion of Results

In this section, estimates of population proportion of malnourished children under 5 years, the bias, MSE, SE and CV computed for various sample sizes are presented in Table 1. The six sample sizes in the study consists of 5%, 10%, 15%, 20%, 25%, 30% of the population size of N=230 while the plots of population proportion, SE, and CV of the estimator are presented in figures 1-3 respectively. Estimates of the population proportion, bias, Mean Square Error, Standard Error and Coefficient of Variation are also presented on Table 1.

For the sample of size n_1=12, n_2=23,n_3=35, n_4=46, n_5=58 and n_6=69, the estimate of the population proportion of children who suffer severe acute malnutrition (SAM) are respectively, 0.3414, 0.3445, 0.3469, 0.3486, 0.3489 and 0.3497 thereby establishing the prevalence of SAM in about 35% of children of age 0 - 5 years in the IDP camps.

On the efficiency of the estimator considering different sample sizes, it is noticeable that $\text{MSE}(\hat{P}_1) = 0.1383 > MSE(\hat{P}_5) = 0.1291 > MSE(\hat{P}_3) = 0.1289 > MSE(\hat{P}_2) = 0.1288 > MSE(\hat{P}_4) = 0.1277 > MSE(\hat{P}_6) =$ 0.1252.This shows that the estimator has minimum MSE when sample size increases to 69 hence, confirming that fact that the MSE is minimized for larger samples.

Similarly, in terms of standard error and also coefficients of variation, the results showed that $\text{SE}(\hat{P}_1) = 0.1048 > SE(\hat{P}_2) = 0.0698 > SE(\hat{P}_3) = 0.0568 > SE(\hat{P}_4) = 0.0470 > SE(\hat{P}_5) = 0.0417 > SE(\hat{P}_6) = 0.0367$ while $\text{CV}(\hat{P}_1) = 30.42\ \% > CV(\hat{P}_2) = 21.48\ \% > CV(\hat{P}_3) = 15.92\ \% > CV(\hat{P}_4) = 13.47\ \% > CV(\hat{P}_5) = 11.95\ \% > CV(\hat{P}_6) =$

10.48 %. These results confirm the law of large numbers that as sample size increases, the sampling error decreases and hence, coefficient of variation is minimized. Gupta (2021) noted that the lower the value of the coefficient of variation the better or the more precise the estimate of the proportion will be. Thus, the estimate using n_6 is preferable as it has minimum SE and CV.

Looking at the 95 % confidence interval (C.I) for the six sample sizes, it is also observed that C.I. $(\hat{P}_1) = (0.1349, 0.5541) > C.I.(\hat{P}_2) = (0.1853, 0.4645) > C.I.(\hat{P}_3) = (0.2433, 0.4705) > C.I.(\hat{P}_4) = (0.2548, 0.4428) > C.I.(\hat{P}_6) = (0.2764, 0.4229) > C.I.(\hat{P}_5) = (0.3072, 0.3906)$. Here, 34.14% sample proportion of severe acute malnutrition cases will lie between 0.1349 to 0.5541 for p_1, likewise 34.45 % will lie between 0.1853 to 0.4645 for p_2; 34.69 % will lie between 0.2433 to 0.4705 for p_3; 34.86 % will lie between 0.2548 to 0.4428 for p_4; 34.89 % will lie between 0.3072 to 0.3906 for p_5; and 34.97 % will as well lie between 0.2764 to 0.4229 for p_6. Table 1 further showed that all the estimated proportions are biased whose bias values are all positive (Bolt *et al*, 2001) and tend to zero. Furthermore, larger sample sizes do not usually minimize the effect of bias (Nallasivan, 2016).

Conclusion

This study estimated the proportion of children aged 0 to 5 years from IDP camps that suffered from severe acute malnutrition (SAM) in the North-East Nigeria. The IDP camps formed the clusters while two-stage sampling method was employed for sample selection considering six different sample sizes out of a population of size N=230 clusters. Results showed that approximately 35% (with confidence interval between 27.64% and 42.29%) of children 0-5years living in IDP camps suffered from SAM cases. The results also showed that as sample size increased, the estimator becomes stable with minimum SE and CV.

Findings in this study are very crucial in terms of the knowledge offered to assist both governmental and non-governmental agencies to tackle the menace of malnutrition among children in the North-East region of Nigeria.

References

Arif, G. M., Nazir, S., Satti, M. N. & Farooq, S. (2012). Child malnutrition in Pakistan: Trends and determinants; *Pak Institute for Development Economics*, 1-18.

Benson, T. D. (2008). Improving nutrition as a development priority: addressing under-nutrition within national policy processes in Sub-Saharan African. (156). Intl Food Policy Research Institute.

Bolt, V., & Ozturk, O., and Lee, H. (2001). Two-stage cluster samples with ranked set sampling designs. *Annals of the Institute of Statistical Mathematics*, 71: 63-91.

Chambers, R., & Clark, R. G. (2012). An introduction to model-based survey sampling with applications. Oxford: Oxford University Press.

Chaudhuri, A., & Singh, H. (2013). *Survey sampling: Theory and methods.* New York: Chapman & Hall/CRC

Das, U. N. (2010). Genes, brain, gut and environment. *Nutrition 26*(5): 459-473.

Dilip, N.C. (2015). Two-stage design for estimation of total fertility rate: with an illustration for slum dweller married women. *Electronic Journal of Applied Statistical Analysis,* 8(1): 112-121.

Ekpo, A., & Ikughur J. A (2021). *Bollori II SMART nutrition survey final report*. Available at ReliefWeb, http://reliefweb.int/report/nigeria/bollori-ii-smart-nutrition-survey-final-report-august-2021

Fuller, W.A. (2011). *Sampling statistics.* John Wiley & Sons.

Goisis, A., Martinson, M., & Sigle, M. (2019). When richer doesn't mean thinner: Ethnicity, socio-economicposition, and the risk of child obesity in the United Kingdom. *Demographic Research,* 41: 649-678.

Hackett, M., Melgar-Quinonez, H., Alvarez, M. C. (2009). Household food insecurity associated with stunting and underweight among preschool children in Antioquia, Colombia. *Rev PanamSaludPublica,* 25(6): 306-510.

Harkare, A.T. (2021). Anthropometric data quality assessment in multi-survey of child growth: A review of experience with a simplified sampling method. *Bulletin of the World Health Organization* (WHO), 60(2):253-260.

Khan, R. E., & Raza, M. A. (2014). Nutritional Status of Children in Bangladesh: measuring Composite Index of Anthropometric Failure (CIAF) and its Determinants. *Demographic Res.* 37: 134-137.

Lohr, S. L. (2010). Sampling: Design and Analysis, 2nd edition. Brooks/Code Cengage learning, Boston, MA.

Martorell, R. (1999). The nature of child malnutrition and its long-term implications. *Food Nutrition Bulletin,* 20(3): 288-292.

Mohseni, M., Aryankhesa. K., & Kalantari, N., (2017). Factors associated with nutrition among under five-year-old children in Iran: A Systematic review. *Ann. Trop Med Public Health.* 10(5): 1147l-1158.

Nallasivan, J. (2016). Contribution to the theory of sampling human population. *Journal of the American Statistical Association*, 33, 101-116.

Okafor, F.C. (2002). *Sample survey theory with application*.1st Edition, Nsukka: Afro Orbis Publication.

Rizzo, T.A., Metzger, B. E., Dooley, S.L., & Cho, N.H. (2010). Early malnutrition and child neurobehavioral development: insights from the study of children of diabetic mothers. *Child Development*, 68(1), 26-38.

Sarndal, C.E. (2010). Models in survey sampling. Statistics in Transition-new Series, 11(3), 539-554.

Singh, V.K., & Singh, R. (2014). Predictive estimation of finite population mean using generalized family of estimators. *Journal of the Turkish Statistical Association,*7(2): 43-54.

Smith, L.C., & Haddad, L.J. (2000). Overcoming child malnutrition in developing countries past achievements and future choices; vol. 30. *Intl Food Policy Res Inst.*

Solanki, R. S., Singh, H. P., & Rathour, A. (2012). An alternative estimator for estimating the finite population mean using auxiliary information in sample surveys. *ISRN Probability and Statistics,* 1- 14.

Srivastava, M., & Garg, N. (2009). A general class of estimators of a finite population mean using multi-auxiliary information under two-stage sampling scheme. *Journal of Reliability and Statistical Studies,* 1(2), 103-118.