An Improvement on Some Approximate Solutions to Dalenius Equation

[1]Kareem, A.O., [2]Oshungade, I.O. and *[2]Oyeyemi, G.M.

[1]Institute for Security Studies, P.M.B. 493, Abuja, Nigeria
[2]Department of Statistics, University of Ilorin, Ilorin, Nigeria

Received: March 29, 2016; *Revised: June 22, 2016;* *Accepted: June 29, 2016*

Abstract

Dalenius researched into the problem of strata boundary determination and came up with some sets of general equations that must be satisfied to reach optimum points of stratification (OPS). These equations met a lot of criticism in terms of its difficulty and time involved in solving them as well as its practical adaptability. Thus, for easy application, sets of approximate solutions were suggested. Deficiencies in the suggested solutions include unavailability of a theory to guide the choice of class interval in their application, use of Approximate Boundary Value (ABV) and overlapping within strata. This study developed the Exact Boundary Value (EBV) approach which places the boundaries at their exact value, eliminates overlapping within strata and produces more strata formation than the ABV. In terms of the precision of the two approaches, the EBV approach was found to be much more precise than the ABV approach for both optimum and proportional allocation.

Keywords: Precision, Stratified Sampling, Approximate Boundary Value and Exact Boundary Value

1.0 Introduction

Strata boundary determination is one of the technical operations involved with the use of stratified sampling design in survey sampling. Stratified sampling is a methodology in which the elements of a heterogeneous population are classified into mutually exclusive and exhaustive homogenous subgroups called strata based on one or more characteristics of importance. The Placement of a population unit into its appropriate stratum as one of the design problems associated with stratified sampling is the focus of this study and many researchers have suggested solutions to this problem. These include; Equalization of Strata Totals (EST), the cum√f(y) method [here referred to as Dalenius and Hodge's Rule (DHR)], Ekman's Rule (EKR) Durbin's Rule (DUR), Sethis' Rule (STR), Thompson Rule (TNR), Lavalle and Hidiroglou Method (LHM) and Extended Ekman's Rule (EEKR) [1–9]. Others are Random Search Method (RSM), Geometric Stratification (GMS) and Genetic Algorithm (GA) [10–12].

Dalenius [1] was credited with the first statistical research into the problem of strata boundary determination. The researcher found the optimum stratification points for Neyman allocation to be those which satisfy the equation:

$$\frac{\sigma_h^2 + (X_h - \mu_h)^2}{\sigma_h} = \frac{\sigma_{h+1}^2 + (X_h - \mu_{h+1})^2}{\sigma_{h+1}}, h = 1,2,\ldots, L-1, L \tag{1}$$

and for proportional allocation to the number of units, the general expression for the simultaneous equation is:

$$X_h = \frac{(\mu_h + \mu_{h+1})}{2}; h = 1,2,\ldots, L-1 \tag{2}$$

*** Corresponding Author: Tel: +234(0)8068241885; Email: gmoyeyemi@gmail.com**

Cochran [13] and Sethi [6] reported that the general equation to be satisfied in order to obtain optimum stratification points for equal allocation is given as:

$$W_h[\sigma_h^2 + (X_h - \mu_h)^2] = W_{h+1}[\sigma_{h+1}^2 + (X_h - \mu_{h+1})^2], h = 1,2,...,L-1, L \quad (3)$$

These equations are solved through various steps of iteration to obtain Optimum Points of Stratification (OPS) and these equations are derived on the assumptions that the estimation variable is the stratification variable and that the frequency distribution is continuous. Detailed derivation of these sets of general equations is presented in literature [1, 14–18]. These sets of general equations had been greeted with lots of criticism in terms of their difficulty and time involved in solving the equations as well as their practical adaptability. Thus, for easy application, sets of approximate solutions have been suggested by various authors [2-7].

1.1 Some Approximate Solutions

(a.) Mahalanobis [2] suggested Equalization of Strata Totals (EST) which requires formation of stratum boundaries for equal allocation by making the product of the stratum weight and the mean equal to a constant:

$$W_h \mu_h = C, h = 1,2,...L-1, L \quad (4)$$

(b.) Ekman [4] on his part suggested that for a density over a finite range, points $\{X_h\}$ satisfying equation (2) is:

$$W_h(X_h - X_{h-1}) = C, h = 1,2,..., L-1, L \quad (5)$$

This is referred to as Ekman's Rule (EKR) in this context.

(c.) Dalenius and Hodges [3] presented a quick approximate method referred to here as Dalenius and Hodges Rule (DHR). The method requires formation of the cumulative of the square root of the frequency of the study variate i.e. cum √(f(y) and choose boundaries such that they create equal intervals on cum√(fy) scale.

(d.) Durbin [5] while reviewing the DHR, proposed the equalization of the cumulative frequencies of a distribution, g(y), which is the midway between the original distribution f(y) and a uniform distribution r(y) over the range (y_0 , y_L) of y . Where $r(y) = F(y_L)/(y_L - y_0)$ and the Optimum Points of Stratification (OPS) are obtained by equalizing the cumulative of the function g(y) = ½ { r(y) + f(y) }, i.e. the stratum boundaries are best placed by ensuring equal partitioning of cum ½ {r(y) + f(y)} scale. F(y) is the cumulative frequency of Y, i.e. cum f(y), y_L is the upper limit and y_0 is the lower limit in a given interval. This is referred to as Durbin's Rule (DUR).

(e.) Sethi [6] solved the sets of general equations (1) – (3) for some standard distributions (Normal, Beta, Gamma and various Chi-squares) with a view to evolve a good stratification system and obtained corresponding OPS for the said distributions. The author also tabulated the optimum boundaries for Neyman, equal and proportional allocations for $L \leq 6$ (for Gamma distribution) and $L \leq 10$ for Normal distribution [6, 14, 19]. Sethi's procedure is referred to as Sethi's Rule (STR).

(f.) Thomson [7] based on recommendation of Singh [20] studied another approximately optimal measure of stratification with proportional allocation and recommends equal division on cum $\sqrt[3]{f(y)}$ scale as a method of stratification. This is referred to as Thomson Rule (TNR).

It is pertinent to mention that, of all the aforementioned approximate solutions, DHR is the most commonly used and forms the basis of comparison up till date. The present study provides an improvement on three (3) (DHR, DUR and TNR) of the six (6) approximate solutions above. The improvement provided is the use of Exact Boundary Value (EBV) approach which eliminates the deficiencies identified in the approximate solutions such as unavailability of a theory to guide the choice of class interval in their application, use of Approximate Boundary Value (ABV) and overlapping within strata. Previous studies made use of the ABV approach, which may not be satisfactory as the boundaries were place on nearest available points and not on the exact point thereby not satisfying the intention of the postulated theory.

2.0 Materials and Methods

2.1 Exact Boundary Value (EBV) Approach

The EBV procedure for each of the corresponding ABV methods is given as follow:

$$EBV = X_{eh} = X_g + \frac{(P_s - P_b)}{P_g} * C$$
$$h = 1,2,\ldots, L-1 \tag{6}$$

where; X_{eh} = the Exact Boundary Value (EBV);

X_g = the lower limit of group where the stratification point P_s falls into;

P_s = the point of stratification obtained by dividing cum√ f(y); cum $\sqrt[3]{f(y)}$ and cum ½ [r(y) + f(y)] scale by the number of strata L;

P_b = the cumulative frequency on cumulative scale (as the case may be) before the group where the stratification point fall into;

P_g = the percentile of cum√f(y); cum $\sqrt[3]{f(y)}$ and cum½[r(y) + f(y)] of the group where point of stratification falls; and C = the class interval.

Therefore, in this study DHR, DUR and TNR are studied using both the ABV and EBV approaches for all the sets of data.

2.2 Estimation Procedure in Stratified Sampling

Notations

The subscript h denotes the stratum and i the unit within the stratum, for h = 1, 2, . . . , L.

L	=	Number of strata.
N_h	=	Total number of population units in stratum h.
n_h	=	Number of sample units taken in stratum h.
N	=	Total number of population units in all the L strata
n	=	Sample size of the study
Y_{hi}	=	The observed value of the i^{th} unit in the h^{th} stratum
W_h	=	N_h/N = stratum weight (population units)
w_h	=	n_h/n = stratum weight (sample units)

Sample mean, $$\bar{y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi} \tag{7}$$

True mean, $$\bar{Y}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} Y_{hi} \tag{8}$$

Sample Variance, $$s_h^2 = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}(y_{hi} - \bar{y}_h)^2 \tag{9}$$

True Variance, $$S_h^2 = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}(Y_{hi} - \bar{Y}_h)^2 \tag{10}$$

Population Mean, $$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h \tag{11}$$

Variance of the population mean, $V(\bar{y}_{st}) = \sum_{h=1}^{L} W_h^2 V(\bar{y}_h)$

$$= \sum_{h=1}^{L} W_h^2 (1 - f_h) \frac{S_h^2}{n_h} \quad (12)$$

where the subscript "st" denotes stratified.

2.3 Allocation of Samples to Strata

Once a sample is selected from each stratum, the procedure of stratified sampling is satisfied. However, this study requires at least two units in each stratum for estimation purposes. After the sample size n is chosen, there are many ways of allocating *n* into individual stratum sizes $n_1, n_2, \ldots, n_L$ with the aim of using an allocation method that gives a specified amount of information at minimum cost. However, allocation scheme is affected by the total number of units in each stratum, the variability of observations within each stratum and the cost of obtaining an observation from each stratum. The two popular ones in use are employed in this study namely optimum allocation and proportional allocation.

2.3.1 Optimum Allocation

With optimum allocation, samples may be selected to minimize the overall cost of the survey for a specified value of $V(\bar{y}_{st})$ or to minimize the $V(\bar{y}_{st})$ for a given overall cost of the survey.

$$n_h = \frac{nW_h S_h / \sqrt{C_h}}{\sum W_h S_h / \sqrt{C_h}} \quad (13)$$

On the other hand, if the costs are unknown or constant (the same in each stratum i.e. $C_h = C$ for h = 1, 2, . . . ,L, then expression (13) reduces to:

$$n_h = \frac{nN_h S_h}{\sum N_h S_h} \quad (14)$$

This method of allocation of total sample size n to strata is called optimum or minimum variance allocation as developed by Neyman [21]. Hence, it is often referred to as Neyman allocation, while the expression of its variance is given as:

$$V_{opt} = V_{\min}(\bar{y}_{st}) = \frac{(\sum W_h S_h)^2}{n} - \frac{\sum W_h S_h^2}{N} \quad (15)$$

2.3.2 Proportional Allocation

Proportional allocation calls for making the stratum sample size n_h proportional to stratum size N_h i.e. it requires allocation of large sampling units to large stratum and small sampling units to a small stratum. Thus, a representative sample of the population units is obtained with this allocation since the sample sizes $n_1, n_2, \ldots, n_L$ are proportional to the stratum sizes $N_1, N_2, \ldots, N_L$ i.e. $\frac{n_h}{n} = \frac{N_h}{N}$ or $\frac{n_h}{N_h} = \frac{n}{N}$ or $f_h = f$

Proportional allocation is given by the expression:

$$n_h = \frac{nN_h}{N} = nW_h, h = 1,2,\ldots L \quad (16)$$

The variance $V(\bar{y}_{st})$ of the mean in stratified random sampling when proportional allocation is used is given as:

$$V(\bar{y}_{st})_{prop} = \frac{(1-f)}{n} \sum W_h S_h^2 = \frac{\sum W_h S_h^2}{n} - \frac{\sum W_h S_h^2}{N} \quad (17)$$

3.0 Results and Data Analysis

This section deals with the empirical demonstration of the ABV and EBV approach in establishing strata boundaries for DHR, TNR and DUR. Practical demonstration with data 1 is reflected in this paper for guidance and emanating result for the other sets of data are reported. The following four (4) sets of data whose features are reflected in Table 1 were used for this study.

i. Overall cumulative average scores (OCAS) of 145 students that graduated from the Faculty of Engineering University of Ilorin 1989/90 set.
ii. Data of Kano State Ministry of Commerce and Industry Survey (2008) on manpower strength of companies and industries in the six (6) industrial Estates of Kano.
iii. Grants allocation to 774 Local Government's Council in Nigeria for the month of December, 2008 shared in January 2009.(See www.fmf.gov.ng)
iv. Population Census figures for the 774 Local Government Areas of Nigeria during the year 2006 census (see www.nigeriastat.gov.ng)

Table 1: Summary Statistics of the Data Used in this Study

DATA	N	N	Range	Skewness	Mean	Variance	Standard Deviation
1	145	48	44.7 - 68.8	0.712	55.48	20.05	4.48
2	171	57	3 – 3756	6.581	166	163923	405
3	774	258	72.2 - 365.0	3.239	108.96	700.61	26.47
4	774	258	11.7 - 1277.7	3.218	180	10281	101

3.1 Dalenius and Hodges Rule (DHR)

DHR requires selection of equal class interval, obtain the cumulative square root of the frequency (cum√(fy))of the study variate and determine the strata boundaries by dividing the total cumulative square root of the frequency by the required number of strata L and the boundary is placed at this division point. This same procedure is applicable to $\sqrt[3]{f(y)}$ TNR, hence the two operations in Table 2.

Table 2: Frequency Distribution Table of DATA 1 with the Application of DHR and TNR

S/N	Class Interval for OCAS	f (y)	√f(y)	cum √f(y)	$\sqrt[3]{f(y)}$	cum $\sqrt[3]{f(y)}$
1	40.0 - 45.0	1	1	1	1	1
2	45.1 - 50.0	10	3.162278	4.162278	2.154435	3.154435
3	50.1 - 55.0	68	8.246211	12.40849	4.081655	7.23609
4	55.1 - 60.0	43	6.557439	18.96593	3.503398	10.73949
5	60.1 - 65.0	16	4	22.96593	2.519842	13.25933
6	65.1 - 70.0	7	2.645751	25.61168	1.912931	15.17226

From Table 2, it can be observed that cum√f(y) is 25.61168 with the application of DHR; therefore stratification points for two strata is at 25.61168/2 = 12.80584. In the same vein, stratification points for three to six strata are obtained. The stratification points and various strata formed are presented in Table 3.

Table 3: Stratification Boundaries for Two through Six Strata Constructed by DHR using ABV Approach for DATA 1

Number of strata	Stratum	Boundaries	Number of units in the stratum Nh	Points of stratification cum√f(y) / L	Interval on cum√f(y) / L	Percentage of population units.
2	1 2	40.0 – 55.0 55.1 – 70.0	79 66	12.8058	12.41 13.20	54.48 45.52
3	1 2 3	40.0 - 55.0 55.1 - 60.0 60.1 - 70.0	79 43 23	8.5373 17.0745	12.41 6.56 6.65	54.48 29.66 15.86
74	1 2 3 4	40.0 - 50.0 50.1 - 55.0 55.1 - 60.0 60.1 - 70.0	11 68 43 23	6.4029 12.8058 19.2088	4.1623 8.2462 6.5574 6.6458	7.59 46.9 29.66 15.86
5	1 2 3 4 5	40.0 - 50.0 50.1 - 55.0 55.1 - 60.0 60.1 - 65.0 65.1 - 70.0	11 68 43 16 7	5.1223 10.2446 15.3670 20.4892	4.1623 8.2462 6.5574 4.0 2.6458	7.59 46.90 29.66 11.03 4.83
6	1 2 3 4 5 6	40.0 - 50.0 50.1 - 55.0 50.1 - 55.0 55.1 - 60.0 60.1 - 65.0 65.1 - 70.0	11 68* 68* 43 16 7	4.2686 8.5372* 12.8058* 17.0744 21.343	4.1623 8.2462 8.2462 6.5574 4.0 2.6458	7.59 46.90 46.90 29.66 11.03 4.83

It can be observed from Table 3 that overlapping of units occurred in stratum 2 and 3 of six strata formation (see * in column 4 and 5 of Table 3). The nearest available points on the cum√f(y) scale (column 5 of Table 2) for the stratification points of stratum 2 and 3, i.e. 8.5372 and 12.8058 have their ABV falling into the same interval 50.1 - 55.0. Therefore, with the use of ABV approach, not more than five strata can be formed for data 1. Same operation was performed using data 2, 3 and 4 with the following results;

Data 2,overlapping occurred in stratum 1 and 2 of six strata formation as their points of stratifications 5.1262 and 10.2524 falls into the same interval 1 - 100 with 10.7703 as the nearest available point on the cum√f(y) scale Thus, only five strata are feasible with data 2.

Data 3,overlapping occurred in stratum 2 and 3 of seven strata formation as their points of stratification 19.0001 and 28.5002 falls into the same interval 90.1 - 110.0 with 30.81272 as the nearest available point on the cum√f(y) scale. Thus, six strata are feasible with data 3 using DHR - ABV approach.

Data 4, overlapping occurred in stratum 2 and 3 of seven strata situation as their points of stratification 23.5809 and 35.3714 falls into the same interval 110.1 - 160.0 with 31.5281 as nearest available point on the cum√f(y) scale. Thus, six strata are feasible with data 4 using DHR - ABV approach.

3.2 Thomson Rule (TNR)

TNR approach to strata construction requires selection of equal class interval, form the cumulative cube root of the frequencies of the study variate and establish the strata boundaries by ensuring equal partitioning of the total cum $\sqrt[3]{f(y)}$ by the number of strata L, required. It can be observed that same operations performed for DHR is required for TNR. Therefore, column 7 of Table 2 has cum $\sqrt[3]{f(y)}$ as 15.17226 for data 1. In the same manner as performed for DHR, points of stratifications were obtained by dividing cum $\sqrt[3]{f(y)}$ by the number of strata L and strata boundaries established at the nearest available points on cum $\sqrt[3]{f(y)}$ scale for TNR. Table 4 gives the summary of all obtainable strata using TNR for Data 1.

Table 4: Stratification Boundaries for Two through Six Strata Constructed By TNR Using ABV Approach for DATA 1

Number of strata	Stratum	Boundaries	Number of units in the stratum N_h	Points of stratification cum $\sqrt[3]{f(y)}$ / L	Interval on cum $\sqrt[3]{f(y)}$	% of population units
2	1 2	40.0 - 55.0 55.1 - 70.0	79 66	7.58613	7.2361 7.9362	54.48 45.52
3	1 2 3	40.0 - 55.0 55.1 - 60.0 60.1 - 70.0	79 43 23	5.05742 10.11484	7.2361 3.5034 4.4328	54.48 29.66 15.86
4	1 2 3 4	40.0 - 50.0 50.1 - 55.0 55.1 - 60.0 60.1 - 70.0	11 68 43 23	3.7931 7.5862 11.3793	3.1544 4.0817 3.5034 4.4328	7.59 46.90 29.66 15.86
5	1 2 3 4 5	40.0 - 50.0 50.1 - 55.0 55.1 - 60.0 60.1 - 65.0 65.1 - 70.0	11 68 43 16 7	3.0345 6.0689 9.1035 12.138	3.1544 4.0817 3.5034 2.5198 1.913	7.59 46.90 29.66 11.03 4.83
6	1 2 3 4 5 6	40.0 – 50.0 50.1 – 55.0 50.1 – 55.0 55.1 – 60.0 60.1 – 65.0 65.1 – 70.0	11 68* 68* 43 16 7	2.5287 5.0574* 7.5862* 10.1149 12.6436	3.1544 4.0817 4.0817 3.5034 2.5198 1.9130	7.59 46.90 46.90 29.66 11.03 4.83

Just as recorded with DHR, overlapping occurred in 2nd and 3rd stratum of six strata situation (See * in Table 4). The nearest available points on cum $\sqrt[3]{f(y)}$ scale (column 7 of Table 2) for the stratification points of stratum 2 and 3, i.e.5.0574 and 7.5862 have their ABV falling into the same interval 50.1 - 55.0.Hence, not more than five strata formation are feasible with data 1 using TNR. In same vein, same operation was performed using cum $\sqrt[3]{f(y)}$ for data 2, 3, and 4. Data 2 recorded overlapping of units in stratum 2 and 3 of eight strata formation, hence seven strata are obtainable. Same was the experience with data 3, but points of overlapping were in 3rd and 4th stratum of eight strata formation, while data 4 recorded it's overlapping in the 3rd and 4th stratum of eleven strata formation, thus ten strata are obtainable.

3.3 Durbin's Rule (DUR)

DUR requires selection of equal class interval and establishment of strata boundaries to ensure equal interval on the cum ½[r(y) + f(y)] scale, where r (y) = F(Y) / $Y_L - Y_0$ and $Y_0, \ldots Y_L$ are strata boundaries where Y_0, Y_L are the smallest and largest value of the study variate in a given interval (Table 5). F(Y) is the cumulative of f(y).

Table 5: Frequency Distribution Table of DATA 1 with the Application of DUR

S/N	Class Interval for OCAS	f (y)	F(y)	Y_L- Y_0	r(y)	½[r(y)+ f(y)]	cum½[r(y)+f(y)]
1	40.0 - 45.0	1	1	44.7	0.022371	0.511186	0.511186
2	45.1 - 50.0	10	11	4.1	2.682927	6.341463	6.852649
3	50.1 - 55.0	68	79	4.5	17.55556	42.77778	49.63043
4	55.1 - 60.0	43	122	4.8	25.1667	34.20833	83.83876
5	60.1 - 65.0	16	138	3.96	34.84848	25.42424	109.263
6	65.1 - 70.0	7	145	3.7	39.18919	23.09459	132.3576

From column 8 in Table 5 above, total cum1/2 [r(y) + f(y)] = 132.3576. Therefore, strata boundaries are determined by dividing this estimate by the number of strata L required. For example, when two strata is required, L = 2, stratification point is at 132.3576 / 2 = 66.1788 and the nearest available point on the cum1/2 [r(y) + f(y)] scale is 49.63043. Having demonstrated the method in a two strata situation, various strata formed by DUR for Data 1 are shown in Table 6.

Table 6: Stratification Boundaries for Two through Six Strata Constructed by DUR Using ABV Approach for DATA 1

Number of strata	Stratum	Boundaries	Number of units in the stratum N_h	Points of stratification cum1/2[r(y)+f(y)] / L	Interval on cum1/2[r(y)+f(y)]	Percentage of population units.
2	1	40.0 – 55.0	79	66.1788	49.6304	54.48
	2	55.1 – 70.0	66	82.7272	82.7272	45.52
3	1	40.0 - 55.0	79	44.1192	49.6304	54.48
	2	55.1 - 60.0	43	88.2384	34.2083	29.66
	3	60.1 - 70.0	23		48.5188	15.86
4	1	40.0 - 55.0	79	33.0894	49.6304	54.48
	2	55.1 - 60.0	43	66.1788	34.2083	29.66
	3	60.1 - 65.0	16	99.2682	25.4242	11.03
	4	65.1 - 70.0	7		23.0946	4.83
5	1	40.0 - 50.0	11	26.4715	6.8527	7.59
	2	50.1 - 55.0	68	52.9430	42.7778	46.90
	3	55.1 - 60.0	43	79.4146	34.2083	29.66
	4	60.1 - 65.0	16	105.8861	25.4242	11.03
	5	65.1 - 70.0	7		23.0946	4.83
6	1	40.0 - 50.0	11	22.0596	6.8527	7.59
	2	50.1 - 55.0	68	44.1192	42.7778	46.90
	3	55.1 - 60.0	43	66.1788	34.2083	29.66
	4	60.1 - 65.0	16*	88.2384*	25.4242	11.03
	5	60.1 - 65.0	16*	110.298*	25.4242	11.03
	6	65.1 - 70.0	7		23.0946	4.83

As recorded with DHR and TNR in 2nd and 3rd stratum of six strata formation for data 1, here overlapping occurred in 4th and 5th stratum of six strata formation (See * in Column 4 & 5 in Table 6). The 4th stratum established its boundary while the nearest available point for the 5th stratum boundary also coincides with that of the 4th stratum. Therefore, only five strata formation is feasible for data 1 using DUR- ABV approach.

Applications of DUR–ABV approach using Data 2, 3 and 4 shows that Data 2 recorded overlapping in 1st and 2nd stratum of four strata formation, therefore three strata are obtainable. Multiple overlapping occurred in nine strata formation of data 3, hence eight strata are feasible. Data 4 yielded six strata as overlapping occurred in 5th and 6th stratum of 7 strata formation.

3.4 Stratification by DHR, TNR and DUR Using Exact Boundary Value (EBV) Approach

Stratification by DHR, TNR and DUR, as discussed in this paper, used ABV approach, i.e., the strata boundary is chosen as the nearest available value on cum√f(y), cum $\sqrt[3]{f(y)}$ and cum1/2 [r(y) + f(y)] scale, respectively. When these quantities are divided by the required number of strata L, points of stratification is obtained. However, in the actual sense the boundaries are placed on nearest available point on those scales as it is rare to actually find the exact point of stratification on these scales.

With the use of the ABV approach therefore, the guiding principle of using these methods of strata construction is being violated. The principle requires selection of the strata boundaries to ensure equal partitioning of these scales by the required number of strata. It was observed that with the use of the ABV approach on data 1, besides stratum 1 of the four strata formation for DUR, distribution of the population units by these three methods of strata construction are the same. This equates the methods to mean the same operation which was not the reason for the development of each of the methods. TNR and DUR were improvement over the DHR and therefore could not have meant to equate these methods.

In view of the foregoing and the repetitive distribution of population units by the ABV approach, the Exact Boundary Value (EBV) approach establishes strata boundary at the exact point of stratification. It is as precise as the wrongly used ABV approach and can produce more number of strata with consistent gain in stratification as L increases. In this Section, the EBV approach relation (6) is therefore invoked to determine the strata boundary for Data 1.

$$EBV = X_{eh} = X_g + \frac{(P_s - P_b)}{P_g} * C \qquad \text{(From 6)}$$

$$h = 1,2,\ldots,(L-1)$$

<u>EBV for DHR using Data 1</u>

Two Strata:

Using the relation (6) on Table 2,

Point of stratification (P_s) = cum√f(y)/ L = 25.61168 / 2 = 12.80584

$$X_e = 55.1 + \frac{(12.80584 - 12.40849)}{6.557439} * 5$$

$$= 55.4$$

For three strata, stratification points are on 8.5373 and 17.0745

Using the relation (6) above, we obtain the EBV as:

$$\text{1st boundary} = X_{e1} = 50.1 + \frac{(8.5373 - 4.162278)}{8.2462} * 5 = 52.75$$

$$\text{2nd boundary} = X_{e2} = 55.1 + \frac{(17.0745 - 12.4085)}{6.5574} * 5 = 58.66$$

Table 7 shows the exact boundaries obtained using the EBV approach for DHR with the corresponding number of population units in each stratum, devoid of overlapping, whereas the ABV approach could only produce five strata formations. Similarly, exact strata boundaries were obtained for DHR, TNR and DUR using the EBV approach for other sets of data used in this study. Population units are placed in their respective stratum and simple random samples of fixed sample sizes 48, 57, 258 and 258 are selected for data 1 to 4 in order to obtain relevant statistics for the purpose of estimating the population parameters using **R-**packages (generating seed of 123).Results obtained for data 1 to 4 using optimum and proportional allocation is as presented in Tables 8–11.

4.0 Discussion and conclusion

The EBV approach to DHR, TNR and DUR in strata boundary determination eliminates the overlapping that restricts the number of strata obtainable with the previously used ABV approach to the three methods. The ABV approach tends to equate the three methods, whereas TNR and DUR were improvement over DHR. With the EBV approach, the real requirement of DHR, TNR and DUR in strata boundaries determination is accomplished as the strata boundaries are now at exact point of stratification (Tables 7 – 11 gave details of the results). In terms of the precision of the two approaches, the EBV approach is much precise than the ABV approach for both optimum and proportional allocation except with data 2 where proportional allocation performance was haphazard. Thus, the EBV approach is recommended for application when DHR, TNR and DUR are used for strata boundaries determination for its precision and large number of strata formation devoid of overlapping.

Table 7: Exact Stratification Boundaries for Two through Ten Strata Constructed by DHR Using EBV Approach for DATA 1

Number of Strata	Stratum	Exact Boundaries	Number of units in the Stratum N_h
	1	40.0 – 55.40	82
2	2	55.41 – 70.0	63
	1	40.0 – 52.75	35
	2	52.76 – 58.67	83
3	3	58.68 – 70.0	27
	1	40.0 – 51.47	18
	2	51.48 – 55.40	64
	3	55.41 – 60.40	47
4	4	60.41 – 70.0	16
5	1	40.0 – 50.68	13
	2	50.69 – 53.79	46
	3	53.80 – 57.36	44
	4	57.37 – 62.00	30
	5	62.01 – 70.0	12
6	1	40.0 – 50.16	11
	2	50.17 -52.75	24
	3	52.76 – 55.40	47
	4	55.41 – 58.66	36
	5	58.67 – 63.07	19
	6	63.08 – 70.0	8
7	1	40.0 – 49.30	10
	2	49.31 – 52.01	13
	3	52.02 – 54.23	45
	4	54.24 – 56.80	31
	5	56.81 – 59.59	21
	6	59.60 – 63.83	17
	7	63.84 – 70.0	8
8	1	40.0 – 48.58	7
	2	48.59 – 51.46	11
	3	51.47 - 53.40	35
	4	53.41 – 55.40	29
	5	55.41 – 57.84	29
	6	57.85 – 60.40	18
	7	60.41 – 64.41	9
	8	64.42 – 70.0	7
9	1	40.0 – 48.02	5
	2	48.03 – 51.03	10
	3	51.04 – 52.75	20
	4	52.76 – 54.48	36
	5	54.49 – 56.49	23
	6	56.50 – 58.66	24
	7	58.67 – 61.29	12
	8	61.30 – 64.85	8
	9	64.86 – 70.0	7
10	1	40.0 - 47.57	2
	2	47.58 – 50.68	11
	3	50.69 – 52.24	15
	4	52.25 – 53.79	31
	5	53.80 – 55.40	23
	6	55.41 – 57.28	21
	7	57.29 – 59.31	17
	8	59.32 – 62.00	13
	9	62.01 – 65.20	6
	10	65.21 – 70.0	6

Table 8: Variance of the Population Mean by Number of Strata and Methods of Strata Construction for Data 1 and 2 for Optimum Allocation Using ABV and EBV Approach

	Data 1						Data 2					
	DHR		TNR		DUR		DHR		TNR		DUR	
No of Strata	ABV	EBV	ABV	EBV	ABV	EBV	ABV	EBV	ABV	EBV	ABV	EBV
2	0.098690	0.089851	0.098690	0.187391	0.098690	0.104398	34.78	44.52	109.33	94.68	138.01	211.8
3	0.048783	0.064690	0.048783	0.069706	0.048783	0.056325	20.79	19.46	23.06	28.11	17.25	16.92
4	0.024145	0.029404	0.024145	0.030817	0.035503	0.035899	12.46	8.89	20.79	20.98		5.13
5	0.015230	0.017275	0.015230	0.018535	0.015230	0.018055	13.11	6.79	13.40	9.36		4.21
6		0.012371		0.011256		0.012984		4.65	11.92	7.52		2.32
7		0.007634				0.008314		3.20		5.68		1.78
8		0.006237						1.94				
9		0.005050										

Table 9: Variance of the Population Mean by Number of Strata and Methods of Strata Construction for Data 3 and 4 for Optimum Allocation Using ABV and EBV Approach

	Data 3						Data 4					
	DHR		TNR		DUR		DHR		TNR		DUR	
No of Strata	ABV	EBV	ABV	EBV	ABV	EBV	ABV	EBV	ABV	EBV	ABV	EBV
2	0.57960	0.41140	0.49185	0.49872	0.56419	0.62554	8.4273	8.3844	8.4273	7.9844	11.9348	10.6165
3	0.22953	0.19971	0.22115	0.23577	0.48165	0.39854	3.0593	3.2655	4.2899	3.9701	6.0440	5.3533
4	0.11481	0.10558	0.18010	0.16216	0.22375	0.27105	2.1646	1.8553	2.3929	2.3270	4.0490	3.7758
5	0.09489	0.07446	0.10805	0.10510	0.20396	0.16713	1.1201	1.1380	2.0549	1.6265	3.0216	2.6471
6	0.08035	0.05026	0.08453	0.07353			0.8765	0.8208	1.2320	1.0236	2.2715	1.8062
7		0.03908	0.08057	0.05001				0.6236	0.7876	0.8508		1.3174
8		0.02815		0.04183				0.4307	0.6972	0.6319		1.0841
9		0.02308		0.03424				0.3488	0.6548	0.5273		
10		0.01820		0.02739				0.3055	0.5787	0.4236		

Table 10: Variance of the Population Mean by Number of Strata and Methods of Strata Construction for DATA 1 and 2 for Proportional Allocation using ABV and EBV Approach

	Data 1						Data 2					
	DHR		TNR		DUR		DHR		TNR		DUR	
No of Strata	ABV	EBV	ABV	EBV	ABV	EBV	ABV	EBV	ABV	EBV	ABV	EBV
2	0.111925	0.099633	0.111925	0.206058	0.111925	0.113458	1495.84	256.55	1312.15	1351.38	1895.99	453.08
3	0.060491	0.071592	0.060491	0.068976	0.060491	0.06078	1260.81	1177.91	1276.79	1275.49	1484.40	355.70
4	0.040585	0.041377	0.040585	0.032724	0.036482	0.039663	1264.85	1504.10	96.10	69.02		261.36
5	0.015952	0.016288	0.015952	0.020275	0.015952	0.025168	89.33	1266.90	369.15			214.56
6		0.011637		0.012804		0.022396		83.01				1394.30
7		0.007863				0.015952		360.62				1159.15
8		0.007276										78.94
9		0.007591										101.88

Table 11: Variance of the Population Mean by Number of Strata and Methods of Strata Construction for Data 3 and 4 for Proportional Allocation using ABV and EBV Approach

	Data 3						Data 4					
	DHR		TNR		DUR		DHR		TNR		DUR	
No of Strata	ABV	EBV	ABV	ABV	ABV	EBV	ABV	EBV	ABV	EBV	ABV	EBV
2	1.05755	0.59514	0.74587	1.03222	0.63889	0.69610	11.3638	9.0759	11.3638	15.9488	11.4214	10.3353
3	0.57920	0.53364	0.35886	0.35700	0.63832	0.56204	7.6474	7.8947	5.3349	4.8640	10.9526	9.8538
4	0.28540	0.33858	0.27413	0.25000	0.33287	0.34649	4.8101	4.7229	3.7097	3.6392	8.0210	7.2501
5	0.20067	0.20597	0.26923	0.12207	0.26235	0.32062	2.5160	2.3293	2.6602	2.3713		8.2176
6	0.17339	0.16036	0.09339	0.21786			2.2992	1.6504	1.5598	1.2487		5.6358
7		0.10176	0.19938	0.20643				1.7784	4.6530	4.5821		
8		0.06581		0.16550				1.2947	4.2920	4.2530		
9		0.04910		0.14302				0.9932	0.7137	4.0967		
10		0.03504		0.03341				0.6748	0.6177	0.5691		

References

[1] Dalenius, T. (1950). The Problem of Optimum Stratification. SkandinaviskAkturietidskrift, Vol. 33, pp. 203-213.
[2] Mahalanobis, P.C. (1952). Some Aspects of the Design of Sample Surveys. Sankhya, Vol. 12, pp. 1-7.
[3] Dalenius, T. and Hodges, J.L., Jr. (1959). Minimum Variance Stratification. JASA, Vol. 54, pp. 88–101.
[4] Ekman, G. (1959). An Approximation Useful in Univariate Stratification. Annals of Mathematical Statistics, Vol. 30, pp. 210–229.
[5] Durbin, J. (1959). Review of sampling in Sweden. Journal of Royal Statistical Societies. Vol. A.122, pp. 246–248.
[6] Sethi, V.K. (1963). A Note on Optimum Stratification for Estimating the Population Means. The Australian Journal of Statistics, Vol. 5, pp. 20–33.
[7] Thomson, J. (1976). A comparison of an approximately optimal stratification given proportional allocation with other methods of stratification and allocation. Metrika, Vol. 23, No. 1, pp.15–25.
[8] Lavalle, P. and Hidiroglou, M.A. (1988). On the Stratification of Skewed Populations. Survey Methodology, Vol. 14, No. 1, pp. 33–43.
[9] Hedlin, D. (2000). A procedure for Stratification by an Extended Ekman rule. Journal of Official Statistics, Vol. 6, No. 1, pp. 15–29.
[10] Kozak, M. (2004). Optimal Stratification Using Random Search Method in Agricultural Surveys. Statistics in Transition, Vol. 6, No. 5, pp.797–806.
[11] Gunning, P. and Horgan, J.M. (2004). A New algorithm for the construction of stratum boundaries in skewed population. Survey Methodology, Vol. 30, No. 2, pp.159-166.
[12] Keskinturk T. and Er, S. (2007). A Genetic algorithm approach to determine boundaries and Sample size of each stratum in stratified Sampling. South Pacific Journal of Natural Sciences, Vol. B.21, pp 91-95.
[13] Cochran, W.G. (1961). Comparison of Method for Determining Stratum Boundaries. Bulleting of the International Statistical Institute, Vol. 38, No. 2, pp.345-358.
[14] Murthy, M.N. (1967). Sampling Theory and methods 2nd Edition Statistical Publishing Society, Calcutta - 35, India.
[15] Sukhatme, P.V. and Sukhatme, B.V. (1970). Sampling Theory with Applications. 2nd Edition, Iowa University Press, USA.
[16] Cochran, W.G. (1977). Sampling Techniques, Third edition: John Wiley and Sons, New York.
[17] Raj, D. and Chandhok, P. (1998). Sample Survey Theory. Narosa Publishing House, 6, Community Centre, Panchsheel Park, New Delhi 100 017.
[18] Okafor, F.C. (2002). Sample Survey Theory with Applications. Afro-Orbis Publications Ltd. Nsukka, Nigeria.
[19] Hess, I., Sethi, V.K. and Balakrishnan, T.R. (1966). Stratification: A practical investigation. Journal of the American Statistical Association, Vol. 61, pp. 74-90.
[20] Singh, R. (1971). Determination of optimum boundaries. Journal of Indian society for Agricultural Statistics, Vol. 23, pp. 115-123.
[21] Neyman, J. (1934). On the Two different aspects of the representative method: The method of Stratified sampling and the method of purposive selection. Journal of Royal Statistical Society, Vol. 97, pp.558-606.